

UNIVERSIDAD PRIVADA ANTENOR ORREGO
ESCUELA DE POSTGRADO
SECCION DE POSTGRADO DE INGENIERÍA



**MODELO DE INTELIGENCIA DE NEGOCIO PARA LA TOMA
DE DECISIONES EN LA EMPRESA SAN ROQUE S.A.**

**TESIS PARA OPTAR EL GRADO ACADÉMICO DE
MAESTRO EN GERENCIA DE TECNOLOGÍAS DE
INFORMACIÓN Y COMUNICACIONES**

AUTOR: Br. OMAR ANTONIO SÁNCHEZ GUEVARA

ASESOR: Ms. FILIBERTO MELCHOR AZABACHE
FERNÁNDEZ

Trujillo, octubre 2014

**MODELO DE INTELIGENCIA DE NEGOCIO PARA LA TOMA DE
DECISIONES EN LA EMPRESA SAN ROQUE S.A.**

Por: Br. Omar Antonio Sánchez Guevara

Aprobado:

Dr. Luis Vladimir Urrelo Huiman

Ms. Eddy Modesto Miranda Velasquez

Ms. Ricardo Andrés Narváez Aranda

Asesor:

Ms. Filiberto Melchor Azabache Fernández

**MODELO DE INTELIGENCIA DE NEGOCIO PARA LA TOMA DE
DECISIONES EN LA EMPRESA SAN ROQUE S.A.**

Por: Br. Omar Antonio Sánchez Guevara

Aprobado:

Dr. Luis Vladimir Urrelo Huiman

Ms. Eddy Modesto Miranda Velasquez

Ms. Ricardo Andrés Narváez Aranda

Asesor:

Ms. Filiberto Melchor Azabache Fernández

ACREDITACIÓN

El **Ms. FILIBERTO MELCHOR AZABACHE FERNÁNDEZ**, que suscribe, asesor de la Tesis con Título “**Modelo de Inteligencia de Negocio para la Toma de Decisiones en la Empresa San Roque S.A.**”, desarrollado por el Bach. en Ingeniería de Sistemas: **Sánchez Guevara, Omar Antonio**, acredita haber realizado las observaciones y recomendaciones pertinentes, encontrándose expedito para su revisión por parte de los señores miembros del Jurado Evaluador.

Trujillo, Septiembre del 2014.

El Asesor:

Ms. FILIBERTO MELCHOR AZABACHE FERNÁNDEZ

El Autor:

Br. SÁNCHEZ GUEVARA OMAR ANTONIO

PRESENTACIÓN

Señores Miembros del Jurado

Cumpliendo con los requerimientos estipulados en el reglamento de Grados y Títulos de la “Universidad Privada Antenor Orrego” para optar el grado de Maestro en Gerencia de Tecnologías de Información y Comunicaciones, pongo a vuestra disposición la presente tesis titulada: **MODELO DE INTELIGENCIA DE NEGOCIOS PARA LA TOMA DE DECISIONES EN LA EMPRESA SAN ROQUE S.A.**

Gracias

Trujillo, Septiembre del 2014

Br. Omar Antonio Sánchez Guevara

DEDICATORIA

A Dios por mostrarme día a día que con humildad, paciencia y sabiduría todo es posible.

A mis padres, por la formación recibida y por su esfuerzo y ejemplo constante y porque en gran parte gracias a ustedes, hoy puedo ver alcanzada mi meta, ya que siempre estuvieron impulsándome en los momentos más difíciles.

A mi hermana, tíos, primos, abuelos y amigos, gracias por haber fomentado en mí el deseo de superación y el anhelo de triunfo en la vida.

AGRADECIMIENTOS

Un agradecimiento muy especial merece la comprensión, paciencia y el ánimo recibidos de mi familia y amigos.

A mi asesor el Ms. Filiberto Azabache Fernández quien con su paciencia y conocimiento ayudó a culminar en forma satisfactoria el presente informe, gracias a la orientación, el seguimiento y la supervisión continúa de la misma, pero sobre todo por la motivación y el apoyo recibido.

Mi más profundo y sincero agradecimiento a todas aquellas personas que con su ayuda han colaborado en la realización del presente trabajo, por sus apreciados y relevantes aportes, críticas, comentarios y sugerencias durante el desarrollo de esta investigación.

RESUMEN

MODELO DE INTELIGENCIA DE NEGOCIO PARA LA TOMA DE DECISIONES EN LA EMPRESA SAN ROQUE S.A.

Por el Br. Omar Antonio Sánchez Guevara

Las empresas comercializadoras de dulces tradicionales crecen en el mercado peruano, esto obliga a dichas empresas a volverse más competitivas y para ello toman decisiones las cuales presentan riesgos, por lo tanto para minimizar el riesgo es esencial la cantidad y calidad de información que se tiene disponible, tal es el caso de la empresa San Roque S.A. donde la información disponible para generar y analizar las propuestas de las posibles promociones de ventas, es insuficiente, generando un mayor tiempo en el proceso de toma de decisiones y produciendo escasas promociones lanzadas anualmente, por lo cual se hace importante conocer cómo mejorar la toma de decisiones en el área de ventas de dicha empresa. En tal sentido el presente trabajo muestra una propuesta de Modelo de Inteligencia de Negocios que asegure una mejora en dicho proceso y que permita transformar información clave del negocio en acciones concretas traduciéndose en beneficios tangibles que les proporcione una ventaja competitiva a los tomadores de decisiones. Para lograrlo se identificaron y compararon las características más relevantes de las principales metodologías para el desarrollo de modelos de inteligencia de negocios tanto tradicionales como de la actual tendencia de BI denominada Big Data, luego se propuso un Modelo de Inteligencia de Negocios que cuenta con las siguiente fases: planificación, modelo del negocio, análisis, diseño, construcción e implementación. Para el despliegue del modelo se utilizó la plataforma BI Pentaho EE versión 4.8 y los componentes de Big Data: Hadoop y Hive. Así se implementó el Modelo de Inteligencia de Negocio en el área de ventas de la empresa San Roque S.A. Luego se evaluaron los atributos del modelo a través de la opinión de expertos para luego clasificarlos, obteniéndose, un puntaje final de 273,66 para los atributos desfavorables y de 497,01 para los favorables, lo cual representa una mejora significativa en la mayoría de las fases del modelo propuesto. Se analizaron los datos obtenidos a través de la implementación del modelo, con lo que se demostró mejoras en los indicadores definidos para su evaluación, cuyos resultados hacen significativamente relevante, la aplicación del Modelo de Inteligencia de Negocio propuesta en el área de ventas de la empresa San Roque S.A.

Palabras clave: Modelo de Inteligencia de Negocio, Toma de Decisiones.

ABSTRACT

BUSINESS INTELLIGENCE MODEL FOR DECISION MAKING IN SAN ROQUE COMPANY S.A.

By: Br. Omar Antonio Sánchez Guevara

The traders of traditional sweets grow in the Peruvian market, this requires these companies to become more competitive and they make decisions which present risks, so to minimize the risk is essential quantity and quality of information we have available such is the case of the San Roque S.A. company where the information available to generate and analyze the proposals of possible sales promotion is insufficient, creating more time in the process of decision making and producing few promotions launched annually, so it is important to know how to improve making decisions in the area of sales of the company. In this sense, the present work shows a proposed model of Business Intelligence to ensure an improvement in the process and enabling key business transform into action information translated into tangible benefits that provide them with a competitive edge to decision makers. To achieve this we identified and compared the most relevant characteristics of the main methodologies for developing intelligence models both traditional business and the current trend of BI called Big Data, then a Model of Business Intelligence that has the following proposed phases: planning, business model, analysis, design, construction and implementation. For deployment model Pentaho BI EE platform version 4.8 and Big Data components used: Hadoop and Hive. Thus the model was implemented Business Intelligence in the area of company sales San Roque SA Model attributes were then evaluated through expert opinion and then classify , yielding a final score of 273.66 to 497.01 for unfavorable and favorable attributes, which represents a significant improvement in most phases of the proposed model. Data obtained through the implementation of the model, which improvements are demonstrated in the indicators defined for evaluation, the results make significantly relevant were analyzed, the application of the Model Business Intelligence proposal in the area of company sales San Roque S.A.

Keywords: Model of Business Intelligence, Decision Making.

ÍNDICE GENERAL

CAPITULO I: INTRODUCCION	1
1.1. SITUACION PROBLEMÁTICA	2
1.2. ANALISIS DE LA PROBLEMÁTICA	5
1.3. ANTECEDENTES DEL PROBLEMA	8
1.4. FORMULACION DEL PROBLEMA	16
1.5. HIPOTESIS	16
1.6. OBJETIVOS	16
CAPITULO II: MARCO TEORICO	18
2.1. CONCEPTOS DE NEGOCIO RELACIONADOS AL PROYECTO.	18
2.1.1 EL MARKETING EN LAS PYMES	18
2.1.2 EL PROCESO DE VENTAS	19
2.1.3 EL MODELO DE ESTRATEGIAS DE NEGOCIO	21
2.2. SISTEMAS DE SOPORTE A LA TOMA DE DECISIONES.	24
2.2.1 TOMA DE DECISIONES	24
2.2.2 INTELIGENCIA DE NEGOCIOS	25
2.3. DATA WAREHOUSE (DWH).	29
2.3.1 DEFINICIÓN	29
2.3.2 ETL	30
2.3.3 EL MODELO DIMENSIONAL	32
2.3.4 TIPOS DE MODELO DIMENSIONAL	34
2.3.5 DATA MART	37
2.3.6 DATA MART Y DATAWAREHOUSE	39
2.3.7 PROCESO ANALÍTICO EN LÍNEA (OLAP)	40
2.4. BIG DATA ANALYTICS	42
2.4.1. DEFINICIÓN	42
2.4.2. CARACTERÍSTICAS DE BIG DATA	46
2.4.3. TIPOS DE IMPLEMENTACIÓN DE BIG DATA	49
2.4.4. VENTAJAS Y DESVENTAJAS DE IMPLEMENTAR BIG DATA ANALYTICS	50
2.4.5. BIG DATA EN EL MARKETING	53
2.4.6. DIFERENCIA ENTRE DATOS ESTRUCTURADOS Y NO ESTRUCTURADOS	53
2.4.7. ARQUITECTURA BIG DATA: HADOOP Y OLAP	54
2.5. METODOLOGIAS DE DESARROLLO DE INTELIGENCIA DE NEGOCIOS.	59
2.5.1. METODOLOGÍA DE RALPH KIMBALL.	59
2.5.2. METODOLOGÍA BILL INMON.	63
2.5.3. METODOLOGÍA DE JOSEP CURTO.	66
2.5.4. METODOLOGÍA DE BIG DATA.	68
2.6. HERRAMIENTAS DE INTELIGENCIA DE NEGOCIOS.	70
2.6.1. HERRAMIENTAS DE INTELIGENCIA DE NEGOCIO DE SOFTWARE LIBRE (OPEN SOURCE)	71
2.6.2. HERRAMIENTAS OPEN SOURCE PARA BIG DATA	83
2.6.3. PENTAHO Y SU INTEGRACIÓN CON APACHE HADOOP	90
CAPITULO III: MATERIAL Y METODOS	92
3.1. MATERIAL Y MÉTODOS	92
3.1.1. MATERIAL	92
3.1.2. MÉTODO	92

CAPITULO IV: PROPUESTA DEL MODELO DE INTELIGENCIA DE NEGOCIO	100
4.1. ANÁLISIS COMPARATIVO DE METODOLOGÍAS.	100
4.1.1. METODOLOGÍAS	100
4.1.2. DEFINICIÓN DE LA METODOLOGÍA DE BI	107
4.1.3. METODOLOGÍA PROPUESTA PARA EL DESARROLLO DEL MODELO	112
4.1.3.1 PLANIFICACIÓN	114
4.1.3.2 MODELO DEL NEGOCIO	116
4.1.3.3 ANÁLISIS	118
4.1.3.4 DISEÑO	145
4.1.3.5 CONSTRUCCIÓN	153
4.1.3.6 IMPLANTACIÓN	161
CAPITULO V: APLICACIÓN DEL MODELO, CASO: EMPRESA SAN ROQUE S.A.	162
5.1 APLICACIÓN DE LA METODOLOGÍA.	162
5.1.1. FASE 1: PLANIFICACIÓN	162
5.1.1.1. PLAN DE PROYECTO	162
5.1.2. FASE 2: MODELO DEL NEGOCIO	167
5.1.2.1. COMPRESION DEL DOMINO DEL NEGOCIO	167
5.1.3. FASE 3: ANÁLISIS	185
5.1.3.1 DEFINICIÓN DE REQUERIMIENTOS Y ANÁLISIS DE LA DATA	185
5.1.3.2 ESTRUCTURACIÓN Y MODELADO DE DATOS DE FUENTES EXTERNAS	194
5.1.3.3 ANALISIS DIMENSIONAL	203
5.1.4. FASE 4: DISEÑO	208
5.1.4.1. DISEÑO DE LA ARQUITECTURA DE LA SOLUCIÓN	208
5.1.4.2. DISEÑO DIMENSIONAL FÍSICO	212
5.1.4.3. LIMPIEZA DE DATOS Y DISEÑO DE EXTRACCIÓN	221
5.1.4.4. DISEÑO DE EXPLOTACIÓN	225
5.1.5. FASE 5: CONSTRUCCIÓN	239
5.1.5.1 EVALUACIÓN Y SELECCIÓN DE LA PLATAFORMA BI DE SOFTWARE LIBRE	239
5.1.5.2 INSTALACIÓN Y CONFIGURACIÓN DE SOFTWARE	259
5.1.5.3 CONSTRUCCIÓN DE DATAMART	259
5.1.5.4 EJECUCIÓN DE ETL	260
5.1.5.5 CREACIÓN DE CUBO	263
5.1.5.6 CONSTRUCCIÓN DE INTERFACES	265
5.1.5.7 PRUEBAS	286
5.1.6. FASE 6: IMPLANTACIÓN	292
5.1.6.1. IMPLANTACIÓN DE LA SOLUCIÓN	292
CAPITULO VI: RESULTADOS	293
6.1. PROCESO DE TOMA DE DECISIONES	293
6.2. PROPUESTA DEL MODELO DE INTELIGENCIA DE NEGOCIO	294
6.3. DESPLIEGUE DEL MODELO DE INTELIGENCIA DE NEGOCIO PROPUESTO	295
6.4. VALIDACIÓN DEL MODELO DE INTELIGENCIA DE NEGOCIO PROPUESTO	295
6.4.1. CUALIDADES EVALUADAS POR EXPERTOS EN EL MODELO DE INTELIGENCIA DE NEGOCIO PROPUESTA.	296
CAPITULO VII: DISCUSIÓN DE RESULTADOS	302
7.1. ANÁLISIS DEL PROCESO DE TOMA DE DECISIONES.	302
7.2. ANÁLISIS DEL MODELO DE INTELIGENCIA DE NEGOCIO PROPUESTO.	303

7.3. ANÁLISIS DEL DESPLIEGUE DEL MODELO DE INTELIGENCIA DE NEGOCIO PROPUESTO.	305
7.4. ANÁLISIS DE LA VALIDACIÓN DEL MODELO DE INTELIGENCIA DE NEGOCIO PROPUESTO.	305
CONCLUSIONES	310
RECOMENDACIONES	313
REFERENCIAS BIBLIOGRAFICAS	315
ANEXOS	319

ÍNDICE DE TABLAS

TABLA 1: SOLUCIONES BI DE SOFTWARE LIBRE Y PROPIETARIO.	71
TABLA 2: DIFERENCIAS DE BI COMERCIAL FRENTE A OPEN SOURCE PENTAHO.	81
TABLA 3: DETALLE DE COSTOS POR TIPO DE USUARIO.....	82
TABLA 4: OPERATIVIZACIÓN DE LA VARIABLE INDEPENDIENTE.	94
TABLA 5: OPERATIVIZACIÓN DE LA VARIABLE DEPENDIENTE.....	95
TABLA 6: CUADRO COMPARATIVO KIMBALL VS INMON.	103
TABLA 7: CUADRO COMPARATIVO KIMBALL VS JOSEP CURTO.....	104
TABLA 8: CUADRO COMPARATIVO BI TRADICIONAL VS BIG DATA.	107
TABLA 9: FUENTES DE INFORMACIÓN INTERNA.	120
TABLA 10: EJEMPLO DE INDICADOR.....	120
TABLA 11: TÉCNICA MULTIDIMENSIONAL.	121
TABLA 12: FUENTES DE INFORMACIÓN EXTERNA.	125
TABLA 13: FORMATOS ANÁLISIS DIMENSIONAL.	138
TABLA 14: EJEMPLO DE HOJA DE GESTIÓN.....	138
TABLA 15: EJEMPLO DIMENSIONES Y NIVELES.....	140
TABLA 16: EJEMPLO DIMENSIONES VS JERARQUÍAS.....	141
TABLA 17: EJEMPLO DE CUADRO DE DIMENSIONES VS MEDIDAS.	142
TABLA 18: ARQUITECTURA MOLAP VS ARQUITECTURA ROLAP.....	147
TABLA 19: EJEMPLO DE DESCRIPCIÓN DE DIMENSIONES.	149
TABLA 20: FORMATO DE TABLAS FUENTES.....	151
TABLA 21: FORMATO DE FUENTE DE DATOS.....	151
TABLA 22: FORMATO DE LIMPIEZA DE DATOS.....	152
TABLA 23: FORMATO DE MAPEO DE DATOS.....	152
TABLA 24: FORMATO DE ESQUEMA DE EXTRACCIÓN.....	153
TABLA 25: STAKEHOLDERS DEL PROYECTO.	165
TABLA 26: EQUIPO GESTIÓN PARA EL PROYECTO.....	166
TABLA 27: EQUIPO DE TECNOLOGÍA DE INFORMACIÓN PARA EL PROYECTO...	166
TABLA 28: ANÁLISIS INTERNO DE SAN ROQUE S.A.....	170
TABLA 29: ANÁLISIS EXTERNO DE SAN ROQUE S.A.	171
TABLA 30: MAPA ESTRATÉGICO SAN ROQUE S.A.	172
TABLA 31: INDICADORES Y METAS DE SAN ROQUE S.A.	173
TABLA 32: PERSPECTIVA FINANCIERA.....	174
TABLA 33: PERSPECTIVA PROCESO.	175
TABLA 34: LISTADO DE ACTORES INVOLUCRADOS EN EL PROCESO DECISIONAL.	177
TABLA 35: INDICADORES DE GESTIÓN DEL PROCESO DE TOMA DE DECISIONES.	183
TABLA 36: INDICADORES DE GESTIÓN DEL PROYECTO.....	186
TABLA 37: TABLAS INVOLUCRADAS EN EL PROCESO DE VENTAS DE LA BASE DE DATOS.	190
TABLA 38: HOJA DE GESTIÓN DEL PROYECTO.	204
TABLA 39: HOJA DE ANÁLISIS VENTAS.	205
TABLA 40: CUADRO DIMENSIONES Y JERARQUÍAS DEL PROYECTO.....	206
TABLA 41: CUADRO DE MEDIDAS Y DIMENSIONES DEL PROYECTO.	206
TABLA 42: DESCRIPCIÓN DIMENSIÓN PRODUCTO.	214
TABLA 43: DESCRIPCIÓN DIMENSIÓN CLIENTE.....	216
TABLA 44: DESCRIPCIÓN DIMENSIÓN PUNTO DE VENTA.	217

TABLA 45: DESCRIPCIÓN DIMENSIÓN TIEMPO.	218
TABLA 46: DESCRIPCIÓN TABLA HECHO VENTAS.....	219
TABLA 47: TABLAS FUENTES PARA LA DIMENSIÓN PRODUCTO.	222
TABLA 48: FUENTE DE DATOS DE LA DIMENSIÓN PRODUCTO.	223
TABLA 49: LIMPIEZA DE DATOS DE LA DIMENSIÓN PRODUCTO.	223
TABLA 50: MAPEO DE DATOS DE LA DIMENSIÓN PRODUCTO.....	224
TABLA 51: ESQUEMA DE EXTRACCIÓN FINAL.....	225
TABLA 52: COLUMNAS DEL REPORTE 1.....	228
TABLA 53: MEDIDAS DEL REPORTE 1.....	228
TABLA 54: FILTROS DEL REPORTE 1.....	229
TABLA 55: FILAS DEL REPORTE 2.	230
TABLA 56: COLUMNAS DEL REPORTE 2.....	231
TABLA 57: MEDIDAS DEL REPORTE 2.....	231
TABLA 58: FILTROS DEL REPORTE 2.....	231
TABLA 59: FILAS DEL REPORTE 3.	233
TABLA 60: MEDIDAS DEL REPORTE 3.....	234
TABLA 61: FILTROS DEL REPORTE 3.....	234
TABLA 62: CUADRO COMPARATIVO – HERRAMIENTAS EXTRACCIÓN.....	242
TABLA 63: CUADRO COMPARATIVO – HERRAMIENTAS EXPLOTACIÓN.....	246
TABLA 64: RANGO DE VALORES PARA INDICADORES.	272
TABLA 65: PUNTUACIÓN DE ATRIBUTOS POR EXPERTOS, EN EL MODELO DE INTELIGENCIA DE NEGOCIOS PROPUESTO.	296
TABLA 66: RESULTADOS DE LA EVALUACIÓN DEL INDICADOR - TIEMPO DE RESPUESTA EN FORMULACIÓN DE PROPUESTAS POR REUNIÓN.....	300
TABLA 67: RESULTADOS DE LA EVALUACIÓN DEL INDICADOR – ÍNDICE DE TIEMPO DE RESPUESTA EN EL PROCESO DE ANÁLISIS DE PROPUESTA FORMULADA.....	300
TABLA 68: RESULTADOS DE LA EVALUACIÓN DEL INDICADOR – NUMERO DE PROPUESTAS FORMULADAS POR REUNIÓN.	301
TABLA 69: RELEVANCIA FINAL DE CADA ATRIBUTO DEL MODELO DE INTELIGENCIA DE NEGOCIO PROPUESTA.	306
TABLA 70: PROMEDIO FINAL DE CADA INDICADOR DE LA TOMA DE DECISIONES CON EL MODELO DE INTELIGENCIA DE NEGOCIO PROPUESTO EN COMPARACIÓN SIN EL MODELO PROPUESTO.	307
TABLA 71: PROMOCIONES DE VENTAS MENSUALES. PERIODO 2007 – 2012.....	324
TABLA 72: TABLA FUENTE DE LA DIMENSIÓN TIEMPO.....	336
TABLA 73: FUENTE DE DATOS DE LA DIMENSIÓN TIEMPO.....	336
TABLA 74: LIMPIEZA DE DATOS DE LA DIMENSIÓN TIEMPO.	337
TABLA 75: MAPEO DE DATOS DE LA DIMENSIÓN TIEMPO.	338
TABLA 76: TABLAS FUENTES PARA LA TABLA HECHOS.....	339
TABLA 77: FUENTE DE DATOS DE LA TABLA HECHO.....	341
TABLA 78: LIMPIEZA DE DATOS DE LA TABLA HECHO.....	342
TABLA 79: MAPEO DE DATOS DE LA TABLA HECHO.	344
TABLA 80: FORMATO DE ENCUESTA AL ATRIBUTO ADAPTACIÓN DEL MODELO.	345
TABLA 81: FORMATO DE ENCUESTA AL ATRIBUTO COSTO DE DESARROLLO..	346
TABLA 82: FORMATO DE ENCUESTA AL ATRIBUTO TIEMPO DE DESARROLLO..	346
TABLA 83: FORMATO DE ENCUESTA AL ATRIBUTO NIVEL DE CALIDAD DEL MODELO.....	347

TABLA 84: RESULTADO DE ENCUESTAS DE EVALUACIÓN DEL MODELO PROPUESTO.....	348
TABLA 85: RESULTADO DE APLICAR LA FORMULA AL INDICADOR TIEMPO DE RESPUESTA EN FORMULACIÓN DE PROPUESTAS POR REUNIÓN.....	352
TABLA 86: RESULTADO DE APLICAR LA FORMULA AL INDICADOR TIEMPO DE RESPUESTA EN EL PROCESO DE ANÁLISIS POR PROPUESTA FORMULADA.	353

ÍNDICE DE ILUSTRACIONES

FIGURA 1: PROCESO DE ELABORACIÓN DE PROMOCIONES DE VENTAS.	3
FIGURA 2: PROCESO DE ANÁLISIS DE PROPUESTAS DE PROMOCIÓN.	4
FIGURA 3: PLAN DE MARKETING.....	19
FIGURA 4: FASES DE LA PLANIFICACIÓN DE TOMA DE DECISIONES EN FUNCIÓN DEL ANÁLISIS DEL VALOR.	25
FIGURA 5: ARQUITECTURA DATA WAREHOUSE.....	30
FIGURA 6: FLUJO DE TRABAJO ETL.....	32
FIGURA 7: COMPONENTES DE UN DATA WAREHOUSE.....	33
FIGURA 8: MEDIDA CALCULADA.....	34
FIGURA 9: EL MODELO ESTRELLA.....	35
FIGURA 10: EL MODELO COPO DE NIEVE.....	36
FIGURA 11: COMPONENTES BÁSICOS DE UN DATA MART.....	39
FIGURA 12: CUBO OLAP.....	42
FIGURA 13: EXPLOSIÓN DE LA INFORMACIÓN.....	45
FIGURA 14: LAS CUATRO V DE BIG DATA.....	47
FIGURA 15: ARQUITECTURA DE BIG DATA.....	56
FIGURA 16: ARQUITECTURA BIG DATA PROPUESTA POR PENTAHOS.....	59
FIGURA 17: ETAPAS DE LA METODOLOGÍA DE KIMBALL, DENOMINADA BUSINESS DIMENSIONAL LIFECYCLE.....	62
FIGURA 18: BILL INMON: DATA WAREHOUSE CORPORATIVO.....	64
FIGURA 19: ROADMAP BI.....	66
FIGURA 20: FASES DE UN PROYECTO BI.....	67
FIGURA 21: FASES DE UN PROYECTO BIG DATA.....	68
FIGURA 22: ARQUITECTURA DE COMPONENTES PENTAHOS.....	73
FIGURA 23: COMPARATIVA DE PENTAHOS ENTERPRISE Y COMUNIDAD.....	78
FIGURA 24: DETALLE DE PRECIOS DE PENTAHOS BI PARA EL NIVEL SILVER.....	79
FIGURA 25: EL COSTO PROMEDIO POR USUARIO, POR EL PROVEEDOR Y EL ESCENARIO.....	82
FIGURA 26: ARQUITECTURA DE HADOOP.....	87
FIGURA 27: ARQUITECTURA DE HDFS.....	88
FIGURA 28: ARQUITECTURA DE HIVE.....	90
FIGURA 29: HADOOP Y PENTAHOS.....	91
FIGURA 30: METODOLOGÍA PROPUESTA DE LA SOLUCIÓN BI.....	114
FIGURA 31: EJEMPLO DE MAPEO Y REDUCCIÓN.....	127
FIGURA 32: DATOS NO ESTRUCTURADOS.....	133
FIGURA 33: RESULTADO DE PENTAHOS MAPPER.....	134
FIGURA 34: RESULTADO DE PENTAHOS REDUCER.....	134
FIGURA 35: RESULTADO DE JOB MAPREDUCE.....	135
FIGURA 36: RESULTADO DEL JOB HADOOP - HIVE.....	136
FIGURA 37: EJEMPLO DE DIMENSIONES Y JERARQUÍAS.....	142
FIGURA 38: EJEMPLO DE ANÁLISIS DIMENSIONAL APLICANDO LA TÉCNICA DIMENSIONAL.....	143
FIGURA 39: EJEMPLO ANÁLISIS DIMENSIONAL VENTAS. MEDIDAS Y DIMENSIONES.....	144
FIGURA 40: EJEMPLO DE ANÁLISIS DIMENSIONAL FINAL.....	144
FIGURA 41: EJEMPLO DE DISEÑO DIMENSIONAL FÍSICO CLIENTE.....	149
FIGURA 42: EJEMPLO DE DISEÑO FÍSICO DE LA TABLA HECHO.....	150
FIGURA 43: ORGANIGRAMA DE LA EMPRESA SAN ROQUE S.A.....	169
FIGURA 44: PARTICIPACIÓN EN EL MERCADO.....	176
FIGURA 45: PROCESO DE TOMA DE DECISIONES ACTUAL DE LAS PROPUESTAS DE PROMOCIONES DE PRODUCTOS DE VENTAS Y LANZAMIENTO AL MERCADO.....	182
FIGURA 46: MUESTRA DE CUADRO DE TOTAL DE VENTAS POR AÑO.....	188
FIGURA 47: MUESTRA DE CUADRO DE TOTAL DE CANTIDAD VENDIDA POR AÑO.	189

FIGURA 48: BASE DE DATOS BDNAVA02 (PROCESO VENTAS).....	191
FIGURA 49: VISTA PREVIA DE LOS DATOS PRE PROCESADOS.....	196
FIGURA 50: ESQUEMA FINAL DE PENTAHO MAPPER.....	197
FIGURA 51: ESQUEMA FINAL DE PENTAHO REDUCER.....	198
FIGURA 52: ESQUEMA FINAL DEL JOB MAP REDUCE.....	198
FIGURA 53: ESQUEMA FINAL DEL JOB HADOOP - HIVE.....	199
FIGURA 54: CONFIGURACIÓN DE LOS DATOS DE ENTRADA DESDE HIVE.....	200
FIGURA 55: CONFIGURACIÓN DE LOS DATOS DE SALIDA HACIA CADA TABLA DEL RDBMS.....	200
FIGURA 56: ESQUEMA FINAL DEL JOB HIVE - RDBMS.....	201
FIGURA 57: BASE DE DATOS RELACIONAL WEBLOG.....	202
FIGURA 58: ANÁLISIS DIMENSIONAL VENTAS.....	207
FIGURA 59: ANÁLISIS DIMENSIONAL WEBLOG.....	208
FIGURA 60: ARQUITECTURA DE LA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS.	210
FIGURA 61: DETERMINANDO EL GRANO.....	213
FIGURA 62: DIMENSIÓN PRODUCTO.....	214
FIGURA 63: DIMENSIÓN CLIENTE.....	215
FIGURA 64: DIMENSIÓN PUNTO DE VENTA.....	216
FIGURA 65: DIMENSIÓN TIEMPO.....	217
FIGURA 66: TABLA HECHO VENTAS.....	218
FIGURA 67: DISEÑO DIMENSIÓN FÍSICO FINAL DE LA FUENTE INTERNA.....	220
FIGURA 68: DISEÑO DIMENSIÓN FÍSICO FINAL DE LA FUENTE EXTERNA.....	221
FIGURA 69: DISEÑO DE REPORTE DE VENTAS REALIZADAS.....	227
FIGURA 70: DISEÑO DE REPORTE TOTAL DE VENTAS POR PRODUCTOS, SUBFAMILIA, TIPOS DE CLIENTE DE UNA DETERMINADA SUCURSAL, Y PERIODO DE TIEMPO.....	230
FIGURA 71: DISEÑO DEL REPORTE NÚMERO DE VISITAS POR PÁGINA WEB.....	233
FIGURA 72: PROTOTIPO DE PANEL DE CONTROL DE ANÁLISIS DE VENTAS.....	235
FIGURA 73: PROTOTIPO DE TABLERO DE MANDO ANÁLISIS DE VENTAS POR PRODUCTOS.....	237
FIGURA 74: PROTOTIPO DE PANEL DE CONTROL DE NÚMERO DE VISITAS A LA PÁGINA DE UN DETERMINADO PRODUCTO.....	239
FIGURA 75: HERRAMIENTAS BI PARA LA ARQUITECTURA DE LA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS.....	249
FIGURA 76: DATA MART VENTAS SMART.....	260
FIGURA 77: PANTALLA DE BIENVENIDA DE PENTAHO DATA INTEGRATION.....	261
FIGURA 78: ESQUEMA DE POBLAMIENTO (JOB).....	262
FIGURA 79: CUBO OLAP VENTAS.....	263
FIGURA 80: CUBO OLAP WEBLOG.....	264
FIGURA 81: PUBLICACIÓN EXITOSA DEL CUBO VENTAS.....	265
FIGURA 82: REPORTE DE TOTAL Y CANTIDAD DE VENTAS.....	267
FIGURA 83: VISTA FINAL DEL REPORTE PUBLICADO.....	268
FIGURA 84: VISTA FINAL DEL REPORTE TOTAL DE VENTAS POR CLIENTE.....	269
FIGURA 85: VISTA FINAL DEL REPORTE NÚMERO DE VISITAS POR PÁGINA.....	270
FIGURA 86: VISTA FINAL DEL TABLERO DE MANDO ANÁLISIS DE VENTAS PARA EL AÑO 2011.....	272
FIGURA 87: TABLERO DE MANDO TOTAL DE VENTAS POR PRODUCTO PARA LA SUCURSAL CHICLAYO.....	273
FIGURA 88: VISTA FINAL DEL TABLERO DE MANDO TOTAL DE VENTAS POR CLIENTE.....	275
FIGURA 89: VISTA FINAL DEL TABLERO DE MANDO RANKING DE LAS FRASES MÁS UTILIZADAS.....	276
FIGURA 90: VISTA FINAL DEL TABLERO DE MANDO NÚMERO DE VISITAS A LA PÁGINA DE UN DETERMINADO PRODUCTO.....	277
FIGURA 91: CANTIDAD DE VENTAS POR AÑO.....	278
FIGURA 92: CANTIDAD DE VENTAS POR SUCURSAL Y POR AÑO.....	278
FIGURA 93: ANALYZER REPORT - TOTAL DE VENTAS.....	279

FIGURA 94: ANALYZER REPORT - CANTIDAD DE VENTAS	280
FIGURA 95: ANALYZER REPORT – TOTAL DE COSTO Y VENTAS.....	280
FIGURA 96: ANALYZER REPORT – TOTAL DE VENTAS POR TIPO DE CLIENTE...	281
FIGURA 97: ANALYZER REPORT – TOTAL DE VENTAS DE UN PRODUCTO ESPECÍFICO.....	282
FIGURA 98: ANALYZER REPORT – TOTAL DE VENTAS POR AÑO.	282
FIGURA 99: ANALYZER REPORT – NRO DE VISITAS POR URL PREVIA.	283
FIGURA 100: ANALYZER REPORT – NRO DE VISITAS POR FRASES UTILIZADAS.	284
FIGURA 101: ANALYZER REPORT – NRO DE VISITAS POR PÁGINAS.	284
FIGURA 102: ANALYZER REPORT – TOP 10 DEL NRO DE VISITAS POR URL Y NOMBRE DE LA PÁGINA WEB.	285
FIGURA 103: ANALYZER REPORT – NRO DE VISITAS POR MESES DE UN DETERMINADO AÑO.	285
FIGURA 104: ANALYZER REPORT – TOP 5 DE LOS MESES CON MAYOR NÚMERO DE VISITAS.....	286
FIGURA 105: PROCESO DE CARGA SATISFACTORIA.....	287
FIGURA 106: DATA MART VENTAS POBLADO.....	288
FIGURA 107: EJECUCIÓN ANÁLISIS DE VENTAS - MUESTRA.	289
FIGURA 108: EJECUCIÓN ANÁLISIS DE VENTAS - TABLERO.....	290
FIGURA 109: EJECUCIÓN ANÁLISIS DE VENTAS POR PRODUCTOS - MUESTRA.	291
FIGURA 110: EJECUCIÓN ANÁLISIS DE VENTAS POR PRODUCTOS - TABLERO.	291
FIGURA 111: MODELO DEL PROCESO DE TOMA DE DECISIONES ACTUAL.....	293
FIGURA 112: MODELO DE INTELIGENCIA DE NEGOCIO PROPUESTA.	294
FIGURA 113: PROCESO DE TOMA DE DECISIONES APLICANDO EL MODELO BI PROPUESTO.....	295
FIGURA 114: PUNTUACIÓN PROMEDIO DE CUALIDADES EVALUADAS POR EXPERTOS EN EL MODELO DE INTELIGENCIA DE NEGOCIOS.	297
FIGURA 115: OBTENCIÓN DEL ALFA DE CROMBACH CON EL SOFTWARE SPSS.	298
FIGURA 116: PROPORCIÓN DE LOS ATRIBUTOS FAVORABLES Y DESFAVORABLES DEL MODELO DE INTELIGENCIA DE NEGOCIOS PROPUESTA.	306
FIGURA 117: PROMEDIO DE LOS INDICADORES DE LA TOMA DE DECISIONES SIN EL MODELO PROPUESTO VERSUS EL MODELO PROPUESTO.	308
FIGURA 118: VENTAS MENSUALES EN MILES DE UNIDADES. 2007 – 2012	325

CAPITULO I: INTRODUCCION

Un sistema de información es un conjunto de componentes interrelacionados que permiten reunir, procesar, almacenar y distribuir información para apoyar el control, el análisis y visión en una empresa. (Laundon & Laundon, 2012). Los sistemas de información contienen información acerca de personas, lugares y cosas importantes dentro de una organización o en su entorno. Tanto las empresas grandes como pequeñas usan sistemas de información para realizar sus actividades, con el fin de hacerlas más eficientes y competitivas. Los sistemas de información ayudan a las Compañías a:

- Extender su alcance a lugares lejanos.
- Ofrecer nuevos productos y servicios.
- Modificar los puestos de flujo de trabajo.
- Alterar la forma en que se hacen los negocios.

Tres actividades de un sistema de información producen la información que las organizaciones necesitan para controlar operaciones, analizar problemas, producir servicios nuevos. Estas actividades son: entrada, procesamiento y salida.

- La entrada captura o recolecta datos del interior de la organización o de su entorno para ser procesados en un sistema de información.
- El procesamiento convierte las entradas brutas en una forma que tiene más sentido para las personas.
- La salida transfiere la información procesada a las personas que la usarán o las actividades en las que será usada.

Los sistemas de información también requieren retroalimentación que consiste en salidas que se devuelven a los miembros apropiados de la organización para ayudarles a evaluar o corregir la etapa de entrada.

1.1. SITUACION PROBLEMÁTICA

1.1.1. Planteamiento del Problema

San Roque S.A. es una empresa dedicada al rubro de elaboración de dulces tradicionales, ubicada en la carretera Panamericana Norte Km. 780 en la ciudad de Lambayeque. Esta empresa cuenta con una gran variedad de productos entre los cuales destaca el King Kong, alfajor de gran tamaño y de diversos sabores. Según la entrevista realizada a la jefa de comercialización y ventas, mostrada en el ANEXO 01, la comercialización de estos productos se realiza a través de sus puntos de venta; a nivel nacional cuentan con seis tiendas en Chiclayo y con dos tiendas en las ciudades de Piura, Trujillo y Lambayeque, además los productos se distribuyen en supermercados y autoservicios en las ciudades de Lima, Chimbote, Tacna y Arequipa. Y de manera internacional en los países de Estados Unidos, Suiza, Italia, Canadá y Costa Rica. En septiembre del 2004 se implanto un sistema de información, el cual incluía la gestión de procesos de las áreas de Contabilidad, Compras, Almacén, Tesorería, Recursos Humanos y Comercial. Según la entrevista al responsable del área de sistemas, mostrada en el ANEXO 02, el sistema estaba funcionando de manera correcta sobre todo en la parte contable, pero el rápido crecimiento de la empresa, volvió limitado al sistema en cuanto a escalabilidad, presentando inconsistencia de datos, por lo que se vieron en la necesidad de realizar una licitación para la implantación de un Sistema ERP desde enero del año 2011.

En el área de ventas la toma decisiones es de vital importancia, por eso se realizan reuniones todos los días martes que tienen por duración de 3 horas aproximadamente, según la entrevista al Gerente General, mostrada en el

ANEXO 03. Ahí se definen uno de los puntos estratégicos como es el de establecer promociones de ventas a través de la técnica “*Brainstorming*” (Clark, 2011), con opiniones de los involucrados pero sin basarse en datos históricos de ventas, el número promedio de propuestas establecidas por reuniones es de tres. Desde que se inicia la idea hasta que se lanza el producto al mercado demora entre dos a tres meses, eso depende del tiempo que le dediquen durante las reuniones ya definidas. En la figura 1 se puede observar de manera general todo el proceso de las promociones de ventas.

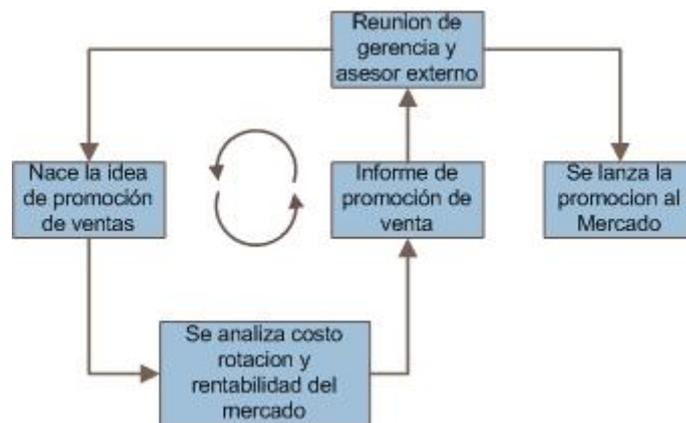


Figura 1: Proceso de Elaboración de Promociones de Ventas.

Fuente: Gerente General. Jorge Piscoya Madueño.

Según la entrevista realizada a la encargada de analizar las propuestas de promoción, mostrada en el ANEXO 04, el tiempo utilizado es de 4 días en promedio para el análisis de propuestas. En la Figura 2 se observa que esta propuesta es analizada en primer lugar según la rotación y ventas del producto o los productos de la promoción por tiendas en los dos últimos años y los meses anteriores.

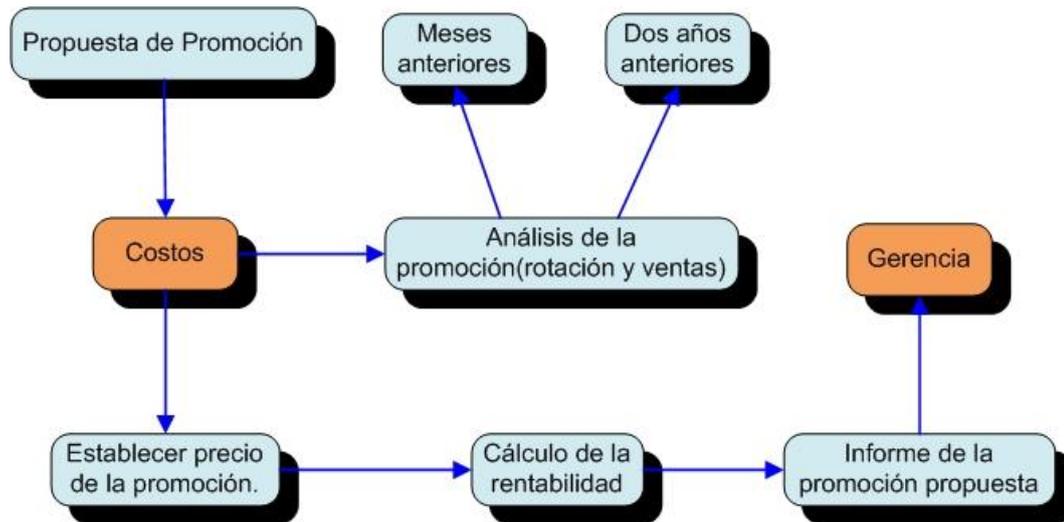


Figura 2: Proceso de Análisis de Propuestas de Promoción.

Fuente: Contadora de Costos. Ana Mendoza Saavedra.

Toda esa información la consigue utilizando los reportes del sistema los cuales se exportan a una hoja de cálculo, analizando los datos obtenidos de la rotación y venta del producto para establecer de esta manera el posible descuento de la promoción calculando si la rentabilidad satisface al indicador ya determinado.

Luego se realiza un informe general para ser presentada en la siguiente reunión de Gerencia para tomar la decisión si la promoción se lanza al mercado o si se cambia a otra propuesta de promoción, si es la primera opción se enviará el informe a la asesora externa de marketing para la publicidad respectiva, a logística para el pedido de envases e insumos y a producción para que se encargue de la elaboración de la promoción, caso contrario se repite el proceso de análisis de la contadora de costos, hasta que la promoción sea la aceptada por Gerencia.

Cada reunión se realiza una vez por semana como ya se mencionó con anterioridad, pero todas estas promociones

realizadas han dado, en algunos meses beneficios en el nivel de ventas en la empresa, según los datos obtenidos en las promociones de ventas mensuales del periodo 2007 al 2012, mostradas en el ANEXO 05, se ha detallado las promociones realizadas, que no son muy frecuentes ya que se han realizado un total de 10 promociones desde abril del 2011 hasta diciembre del año 2012. Se contrasta esto con los datos de las ventas mensuales, mostradas en el ANEXO 06, donde se observa que el nivel de ventas aumenta en ciertos meses siendo los picos más altos a partir de la segunda mitad del año.

1.1.2. Delimitación

Por lo expuesto anteriormente, el presente trabajo de investigación se delimita a estudiar la problemática del proceso de toma de decisiones del área de ventas de la empresa San Roque S.A.

1.2. ANALISIS DE LA PROBLEMÁTICA

1.2.1. Análisis

Los problemas encontrados en el área mencionada son los siguientes:

- ✓ Empleo de tiempo excesivo para la toma de decisiones.
- ✓ La información disponible es insuficiente y por lo tanto no ayuda para la toma de decisiones.
- ✓ Gestión de recurso humano para el proceso de toma de decisiones inadecuada.

De acuerdo a lo expuesto, se puede priorizar la problemática de la siguiente manera:

- De acuerdo a lo expuesto en la Figura 1, se deduce que el tiempo invertido antes del lanzamiento de las promociones, que mayormente son en meses por fechas festivas, es amplio ya que se reúnen hasta con 3 meses de anticipación para la toma de decisiones, de igual forma de acuerdo a lo indicado en la Figura 2, se deduce que existe demora al analizar los costos de la posible promoción siendo el tiempo de análisis de 4 días promedio, porque su única herramienta de ayuda son las hojas de cálculo que contienen los reportes del sistema.
- De acuerdo a lo expuesto en la Figura 1, se deduce que en cada reunión de Gerencia al realizar el “*Brainstorming*” para la toma de decisiones de los procesos de ventas, cada una de las ideas no tiene un sustento sólido en base a hechos, lo que origina más duración en la reunión al no ser analizados correctamente los datos históricos de las ventas.
- De acuerdo a lo expuesto en la Figura 2, existe sobreutilización del recurso humano convirtiendo este proceso en un ciclo casi cerrado al momento de realizar los análisis de los costos de la posible promoción de venta hasta que sea aprobado por Gerencia, lo que genera según el ANEXO 05, escasas promociones de ventas lanzadas anualmente.

Por lo tanto se puede concluir que la empresa San Roque S.A usa un estilo de toma de decisiones tradicional.

1.2.2. Alcance

El alcance del proyecto está delimitado al área de ventas de la empresa San Roque S.A, al uso de la plataforma de Business

Intelligence Pentaho y a la implementación de reportes, tablas de análisis y tableros de mando.

1.2.3. Justificación

Esta investigación se justifica académicamente porque permite recrear la manera de brindar herramientas que soporten un nuevo estilo de toma de decisiones el cual se caracteriza por el uso compartido de una solución de inteligencia de negocios con plataforma Pentaho, la cual es una herramienta Open Source, de la que no se tiene mucha documentación y cruzada con información externa.

Esta investigación se justifica socialmente porque permite reducir el tiempo y los recursos en el proceso de toma de decisiones en la elaboración de promociones en la empresa San Roque S.A.

1.2.4. Aportes

El proyecto realiza los siguientes aportes:

a) Empresa San Roque:

- Permitirá mejorar el proceso de toma de decisiones en el departamento de ventas.
- Permitirá que el departamento de marketing genere un mayor número de estrategias de ventas a través de la información consolidada proporcionada por la solución BI desarrollada.
- Permitirá mejorar las relaciones de los clientes internos (empleados) con el nuevo modelo de inteligencia de negocio propuesto ya que se producirá una herramienta de inteligencia de negocios innovadora y de calidad.

b) Sociedad:

- El Modelo a desarrollar podrá ser utilizado por otras empresas del sector comercializador de dulces tradicionales, para que les ayude a mejorar el proceso de toma de decisiones en el área de ventas.

c) Tesista:

- Plasmar los conocimientos adquiridos a lo largo de la maestría a fin de lograr el grado de Maestro en Ingeniería de Sistemas.
- Obtener los conocimientos más profundos sobre las diferentes metodologías y herramientas para el desarrollo de proyectos de Inteligencia de Negocios.
- Sentar las bases para trabajos futuros en la presente área.

1.3. ANTECEDENTES DEL PROBLEMA

Las siguientes investigaciones forman parte de los antecedentes:

1.3.1. Título de Investigación: Análisis, diseño e implementación de una solución de inteligencia de negocios para el área de finanzas del Supermercado el Super S.A.C

Autores: Torres Meléndez Manuel, Rojas Santisteban Miguel.

Fuente: Biblioteca de la Escuela de Post Grado. Programa de Maestría en Administración de Tecnologías de Información de la Universidad Privada César Vallejo.

País: Perú.

Año: 2010.

Problemática:

Este trabajo de investigación surgió por la necesidad que tenían los usuarios de acceder de manera rápida a información confiable relacionada con su trabajo. El procedimiento a realizar por los usuarios para acceder a esta información, era solicitar al área de sistemas la exportación de los datos de la Base de Datos, vaciar estos datos en un archivo Excel, depurar, ordenar, organizar y clasificar dichos datos para obtener la información que requieren para realizar sus labores. Este procedimiento manual genera diversos problemas tales como: dependencia del área de Sistemas, información generada con una alta posibilidad de error (proceso manual), inversión de tiempo en procedimientos mecánicos e información dispersa en archivos de Excel, dificultando la consulta de datos históricos.

Conclusiones:

Para dar solución a esta problemática se realizó el análisis, diseño e implementación de una Solución de Inteligencia de Negocios para el Área de Finanzas del Supermercado El Super S.A.C. de la ciudad de Chiclayo, utilizaron la metodología de Ralph Kimball y la herramienta Microsoft BI para su desarrollo, como resultado obtuvieron una aplicación que logro automatizar el procedimiento que ha sido explicado anteriormente y que en un entorno amigable, permita a los usuarios acceder a información de mejor calidad, más confiable, en menor tiempo y en un repositorio que permita acceder a información histórica, así mismo la solución permitió eliminar la dependencia con el área de Sistemas para realizar el requerimiento de los datos.

1.3.2. Título de Investigación: Desarrollo e Implementación de un Sistema de Soporte de Decisiones para efectivizar la Toma de

Decisiones en el Proceso de Ventas de la Empresa Productos Razzeto & Nestorovic S.A.C.

Autor: Castillo Alfaro, Vladimir Sandino.

Fuente: Biblioteca de la Escuela de Post Grado. Programa de Maestría en Administración de Tecnologías de Información de la Universidad Privada César Vallejo.

País: Perú.

Año: 2010.

Problemática:

En esta investigación se trató el problema de que la información con la que contaba la empresa para el proceso de ventas no era utilizada de manera apropiada al momento de tomar decisiones, el cual limitaba el aumentar su nivel de calidad, buscar mayor participación del mercado, lograr mayor satisfacción de sus clientes, mejorar su competitividad. Se necesitaba reducir enormemente el tiempo que tomaba el solicitar información necesaria para tomar decisiones esto por motivo que la información tenía que solicitarse de diferentes departamentos o a través de los administradores de TI, quienes asignaban a varias personas a integrar distintos reportes. El tiempo de respuesta de las consultas realizadas por el usuario era de varios días incluso semanas, por lo tanto la organización era rica en datos pero pobres en información. El reto fue transformar los datos en información útil, así los empleados podrían utilizar esa información para incrementar la rentabilidad de la compañía.

Conclusiones:

En este contexto para dar solución a la problemática se desarrolló un modelo de Inteligencia de Negocios basado en la metodología Ralph Kimball y utilizando la Suite Microsoft BI para su implementación.

Como resultado se logró obtener información más adecuada y sólida para la toma de decisiones en el proceso de ventas y contribuir al logro de los objetivos empresariales, así mismo le permitió obtener mejores resultados analizando sus tendencias de ventas, con cifras de años anteriores o con los resultados de otros equipos de ventas y de esta manera sugerir formas de mejorar el desempeño. La fuerza de ventas también se benefició al utilizar BI para analizar datos sobre marcas, clientes y distribuidores es decir enfocarse en clientes y productos altamente rentables. Este trabajo de investigación presenta una solución que las empresas pueden implementar para satisfacer sus necesidades de gestión, análisis y toma de decisiones y otorga un panorama de lo que está sucediendo en la organización.

1.3.3. Título de Investigación: Formulación de un modelo de implementación de un CRM analítico para Supermercados del Norte S.A.C.

Autor: Peña Solís Cesar, Primo Bonilla Roberto.

Fuente: Biblioteca de la Escuela de Post Grado. Programa de Maestría en Administración de Tecnologías de Información de la Universidad Privada César Vallejo.

País: Perú.

Año: 2010.

Problemática:

La problemática de esta investigación era que no se estaba analizando la información con respecto al conocimiento de sus clientes, detectando sus necesidades, aumentando su grado de satisfacción, incrementando su fidelidad a la empresa, mediante el análisis de la información extraída desde los diferentes orígenes de datos.

Conclusiones:

Por lo tanto como solución a esta problemática este trabajo de investigación, pretendió identificar y segmentar los clientes más rentables que posee la empresa, diseñando en primera instancia el paquete DTS (Data Transformation Services) para la extracción, transformación y carga de datos, que ejecutara una tarea determinada dentro del conjunto de operaciones empaquetadas. Después se crearon los Data Marts y los respectivos cubos OLAP. Para la realización de la solución se utilizó la metodología de Bill Inmon, para el desarrollo de esta aplicación utilizaron la suit Microsoft BI.

Como resultado se creó una solución de inteligencia de negocios el cual ayudo a que la organización tenga toda su información consolidada y ordenada en un solo lugar, lo cual es muy importante en este tipo de organizaciones debido a la sensibilidad e importancia de la información. Los reportes generados por la solución de BI satisficieron las necesidades de los usuarios para una adecuada toma de decisiones. Además de ello, les ayudo a reducir tiempos de respuesta en el procesamiento y análisis de información, lo cual genero una ventaja competitiva para la empresa.

1.3.4. Título de Investigación: Sistema de soporte de decisiones para la gestión operativa en el área comercial de la empresa Ferretería Boca E.I.R.L.

Autor: Delgado Gonzales José, Tarrillo Mondragón Walter.

Fuente: Biblioteca de la Escuela de Post Grado. Programa de Maestría en Administración de Tecnologías de Información de la Universidad Privada César Vallejo.

País: Perú.

Año: 2011.

Problemática:

La realidad problemática de la presente investigación era la dificultad que existía al momento de tomar decisiones en la gerencia del área comercial de la empresa, al no contar con una herramienta que le permita procesar datos según la información que requiera, los principales problemas con los que se enfrentaba esta empresa eran:

- Dificultad para acceder a la información.
- Toma de decisiones fuera de tiempo.
- Falta de datos consolidados.
- "Islas" de información Excel.
- Falta de una herramienta ágil para visualización.
- Sobreutilización del recurso humano para elaboración de informes.

Todo ello creaba incoherencias en la toma de decisiones y necesidades muy altas del uso de los recursos del departamento de Tecnologías de la Información, el cual, debido a la necesidad de información de los usuarios, aumentaba en personal, máquinas, y horas extra. Esto provocaba toma de decisiones tardías, debido a la imposibilidad de obtener la información en el momento preciso.

Conclusiones:

Por lo tanto como solución a esta problemática se desarrolló un modelo de inteligencia de negocios, utilizando la herramienta Pentaho BI. Se elaboró el cubo necesario para que los analistas del área comercial de la empresa, pudieran elaborar informes a partir de la información brindada por la herramienta BI. Se implementó un cuadro de mando, que permitiera el control por indicadores de dicha área. Se

integraron todas las hojas excel dispersas. La metodología utilizada fue la de Ralph Kimball.

Los resultados obtenidos son los siguientes:

- Se logró tener información para la toma de decisiones en un menor tiempo.
- Reducción al mínimo del uso de plantillas Excel susceptibles a manipulación o integración de las mismas en los cubos OLAP.
- Reducción de horas hombre en la preparación de informes.
- Liberación del personal de sistemas a tareas más acordes con el área.

1.3.5. Título de Investigación: SIG utilizando Data Marts para mejorar el proceso de toma de decisiones estratégicas en las empresas de hidrocarburos en el departamento de Lambayeque. Caso: Grifo Nor Oriente S.A.C.

Autor: Barbadillo Piscoya Juan, Carbajal Cornejo Carlos.

Fuente: Biblioteca de la Escuela de Post Grado del Programa de Maestría en Administración de Tecnologías de Información de la Universidad Privada César Vallejo.

País: Perú.

Año: 2011.

Problemática:

La realidad problemática de esta investigación se centra en el proceso de toma de decisiones, el cual está limitado al análisis de informes generados mensualmente, por el personal de las áreas de ventas y contabilidad, necesitando invertir tiempo considerable, ya que dependen del área de Tecnologías de Información para obtener la información necesaria para tomar decisiones estratégicas.

Conclusiones:

Teniendo en cuenta el problema planteado, se determinó como solución al problema el desarrollar un SIG utilizando Data Marts, mediante la metodología de Ralph Kimball. Como herramienta BI se utilizó la Suit Pentaho BI. Los resultados obtenidos fueron los siguientes:

- El uso de una interfaz BI de reportes permitió un manejo intuitivo y sencillo a los usuarios finales para generar sus propios reportes y análisis acorde a las necesidades del negocio en comparación del uso de hojas de cálculo.
- Mejoro el proceso de toma de decisiones estratégicas, realizando decisiones más confiables y acertadas aportando un mayor grado de satisfacción y confiabilidad a la gerencia.
- Mejoraron los tiempos en la obtención de la información relevante y aminoraron los costos que involucra la obtención de dicha información.
- El software libre utilizado en este proyecto permitió implementar todo lo necesario para crear una plataforma de Inteligencia de negocios y el desarrollo de los reportes y procesos ETL, se usó el software libre debido a su principal ventaja que es su bajo costo económico.

En los antecedentes presentados para el presente proyecto destaca el uso de metodologías de BI tradicional como kimball e Inmon para el desarrollo de un modelo de inteligencia de negocios como solución a cada uno de los problemas presentados en cada antecedente, lo cual aplica de manera adecuada al tema de la presente investigación como un buen aporte, sin embargo los

trabajos realizados en estos antecedentes se pueden considerar un tanto insuficientes ya que no se aplica la metodología Big Data, es decir no se llegó a trabajar con data no estructurada lo cual permitiría un análisis más profundo y robusto, al permitir analizar información externa no estructurada y de gran cantidad de datos.

1.4. FORMULACION DEL PROBLEMA

1.4.1. Enunciado

¿Cómo mejorar la toma de decisiones en el área de ventas de la empresa San Roque?

1.5. HIPOTESIS

1.5.1 General

El desarrollo de un Modelo de Inteligencia de Negocios, permitirá una mejora en la toma de decisiones en el área de ventas en la empresa SAN ROQUE S.A.

1.6. OBJETIVOS

1.6.1 General

Desarrollar un Modelo de Inteligencia de Negocio para el área de Ventas de la empresa San Roque S.A.

1.6.2 Específicos

- Describir el proceso de toma de decisiones en el área de ventas de la empresa mediante revisión de documentos, encuestas y entrevistas, realizados a los responsables de la toma de decisiones.
- Diseñar un modelo de inteligencia de negocios basándose en metodologías para el desarrollo de modelos de

inteligencia de negocio, como la metodología de Josep Curto, Ralph Kimball y Big Data.

- Desplegar el modelo desarrollando los artefactos de inteligencia de negocios aplicando la Suit Pentaho para el despliegue.
- Validar el modelo de inteligencia de negocios a través de juicio de expertos.

CAPITULO II: MARCO TEORICO

En el presente capítulo, se han recogido puntos básicos a revisar en el contexto teórico en el que se desarrolla el proyecto de inteligencia de negocios. Lo que se busca es revisar algunos conceptos de negocio relacionados al proyecto, también acerca de los sistemas de soporte a la toma de decisiones, definición de un Data Warehouse (DWH), Big Data Analytics, y por ultimo revisar las metodologías y herramientas de desarrollo de inteligencia de negocios más representativas.

2.1. CONCEPTOS DE NEGOCIO RELACIONADOS AL PROYECTO.

2.1.1 El Marketing en las Pymes

El marketing de hoy se centra en ofrecer productos de alta calidad para el comprador lo que permitirá tener precios más estables como el caso de éxito de Starbucks, Wong y El Comercio que son empresas que generan expectativas de beneficio en sus clientes. En el Perú ya se ha vuelto algo común que cuando a una empresa le va bien inmediatamente surgen los imitadores de menor precio, pero en la mayoría de los casos los imitadores terminan cediendo ante la fuerte imagen de la marca de alta calidad.

Según (Sainz de Vicuña, 2010), en su Libro “El Plan de Marketing en la PYME” menciona que una vez establecidos los objetivos que se desean alcanzar, se requiere definir cuál es la estrategia a seguir. En la siguiente figura, se muestra ejemplos de estrategias de marketing donde normalmente se empieza con la estrategia de cartera.

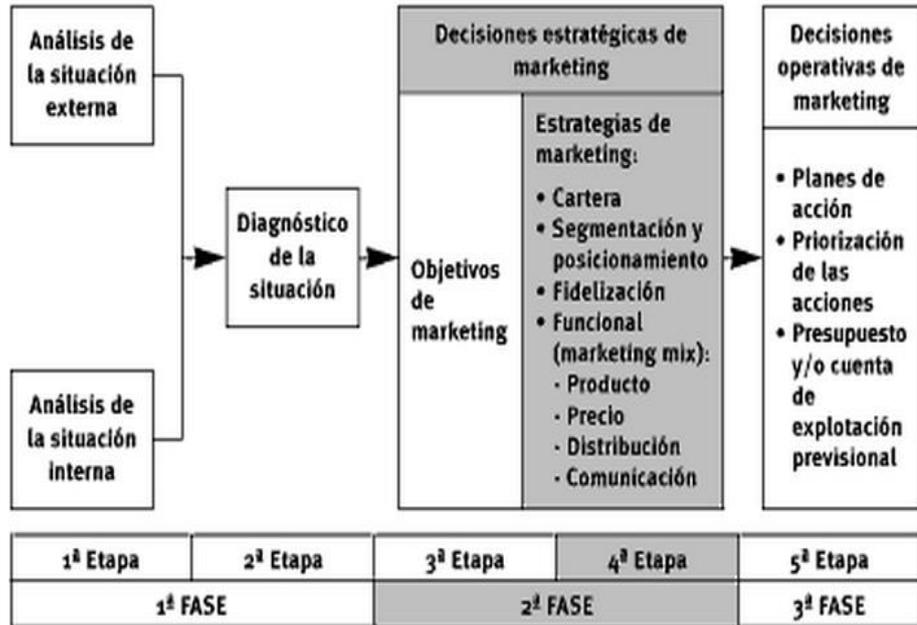


Figura 3: Plan de Marketing.

Fuente: (Sainz de Vicuña, 2010).

Hoy en día existen determinadas actividades de marketing para las Pymes, no son las más recomendables de acuerdo al análisis de la situación actual del mercado, sin embargo muchas veces son también las únicas posibles, en función de cómo ha quedado el presupuesto anual de marketing o por no basarse en datos históricos de ventas.

2.1.2 El proceso de ventas

Según (Stanton, Etzel, & Walker, 2009) autores del libro "Fundamentos de Marketing", indican que el proceso de ventas es una secuencia lógica de cuatro pasos que emprende el vendedor con un comprador potencial y que tiene por objeto producir alguna reacción deseada en el cliente (usualmente la compra).

A continuación, se detallan los cuatro pasos o fases del proceso de venta indicados por el autor:

1. Prospección:

Consiste en la búsqueda de clientes en perspectiva; es decir, aquellos que aún no son clientes de la empresa pero que tienen grandes posibilidades de serlo.

La prospección involucra un proceso de tres etapas:

- Identificar a los clientes en perspectiva
- Calificar a los candidatos en función a su potencial de compra
- Elaborar una lista de clientes en perspectiva

2. El acercamiento previo o "pre entrada":

Consiste en la obtención de información más detallada de cada cliente en perspectiva y la preparación de la presentación de ventas adaptada a las particularidades de cada cliente.

Esta fase involucra el siguiente proceso:

- Investigación de las particularidades de cada cliente en perspectiva
- Preparación de la presentación de ventas enfocada en el posible cliente
- Obtención de la cita o planificación de las visitas en frío: por ejemplo tocando las puertas de cada domicilio en una zona determinada (muy útil para abordar amas de casa con decisión de compra).

3. La presentación del mensaje de ventas:

Consiste en contarle la historia del producto al consumidor, siguiendo la fórmula AIDA de captar la Atención, conservar el Interés, provocar un Deseo y obtener la Acción (compra).

La presentación del mensaje de ventas se basa en una estructura basada en 3 pilares:

- **Las características del producto:** Lo que es el producto en sí, sus atributos
- **Las ventajas:** Aquello que lo hace superior a los productos de la competencia
- **Los beneficios que obtiene el cliente:** Aquello que busca el cliente de forma consciente o inconsciente

4. Servicios posventa:

La etapa final del proceso de venta es una serie de actividades posventa que fomentan la buena voluntad del cliente y echan los cimientos para negocios futuros.

Los servicios de posventa tienen el objetivo de asegurar la satisfacción e incluso la complacencia del cliente. Es en esta etapa donde la empresa puede dar un valor agregado que no espera el cliente pero que puede ocasionar su lealtad hacia la marca o la empresa.

Los servicios de posventa, pueden incluir todas o algunas de las siguientes actividades:

- Verificación de que se cumplan los tiempos y condiciones de envío.
- Verificación de una entrega correcta.
- Instalación.
- Asesoramiento para un uso apropiado.
- Garantías en caso de fallas de fábrica.
- Servicio y soporte técnico.
- Posibilidad de cambio o devolución en caso de no satisfacer las expectativas del cliente.
- Descuentos especiales para compras futuras.

2.1.3 El modelo de estrategias de Negocio

2.1.3.1 El Modelo de Negocio

Según el autor (Harvard Business Essentials, 2010) en su libro: “Como Crear Una Empresa Exitosa”, mencionan que el término “Modelo de Negocio”, se hizo popular por primera vez a finales de los ochenta, después de que mucha gente hubiera adquirido experiencia con los ordenadores personales y con las hojas de cálculo. Por otro lado hace referencia que un modelo de negocio es el conjunto de decisiones y negociaciones realizadas por una compañía para conseguir un beneficio, es decir define al modelo de negocio como la arquitectura fundacional de un negocio.

Por su parte los autores (Anthony, Sinfield, Johnson, & Altman, 2010) en su libro “Guía del Innovador para crecer” hace mención de una definición más amplia y diversa, indican que el modelo de negocio debe ser visto como parte de una estrategia de negocio general, pero es también una categoría singular de disciplina de management, relacionada, pero diferenciada de la estrategia competitiva, la innovación de producto y proceso, las operaciones y la organización.

Un modelo de negocio no es nada más que un conjunto de deducciones que se plantean para que la empresa cree valor para todos los involucrados (clientes, inversionistas y socios).

2.1.3.2 Estrategias de Negocio

Las estrategias de negocio, según el autor (Aaker, 2010) en el libro “Estrategia de la Cartera de Marcas”, describe los siguientes aspectos:

- a) **En el ámbito del producto-mercado**, se refiere al lugar donde competirá el negocio, en que productos y mercados hará énfasis y en cuales desenfatará o serán abandonados;
- b) **La proposición de valor**, indica la oferta para el cliente, proporcionar un beneficio al producto para satisfacer la necesidad del cliente, consiguiendo su fidelización, y en que se puede diferenciar de la competencia.
- c) **Activos estratégicos**: identificar los activos que permitirán que el negocio tenga éxito en el tiempo en cada producto.

El autor menciona que para obtener mejor rentabilidad en una empresa, se debe basar en proporcionarle mayor valor al cliente. De esa manera seleccionar a tipos de clientes, y realizar una segmentación adecuada. También menciona que un buen diseño de negocio necesita de un análisis interno de todo lo que involucra darle un valor mayor e importancia a las necesidades y satisfacción del cliente, en este caso en el producto o servicio a brindar por parte de la empresa y de qué manera este le ofrece más ganancias al empresario.

2.2. SISTEMAS DE SOPORTE A LA TOMA DE DECISIONES.

2.2.1 Toma de Decisiones

Según el autor (Amaya Amaya, 2010) en el libro “Toma de decisiones gerenciales: métodos cuantitativos para la administración” menciona que la toma de decisiones empresariales constituye un proceso de selección entre las diferentes alternativas posibles. A partir de la elección de una alternativa se podrá pasar al proceso de la planificación de actividades, desarrollo de las estrategias y formulación de programas y presupuestos. Para dirigir eficazmente la pequeña y mediana empresa se necesita utilizar los recursos disponibles de una manera eficiente, al ser los recursos limitados los gerentes deberán tratar de aprovechar todas las oportunidades.

La toma de decisiones se debe realizar en los diferentes niveles de la empresa, esto trae consigo que las decisiones individuales, correspondientes a cada nivel de responsabilidad, afectan de una manera directa al desarrollo de todo el sistema organizativo y condicionan tanto la consecución de los objetivos particulares como la de los objetivos generales o globales. En la siguiente figura se observa las fases de la planificación de toma de decisiones en función del análisis del valor.

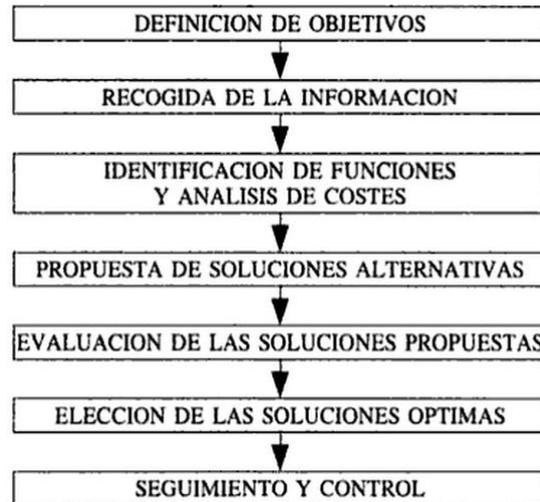


Figura 4: Fases de la Planificación de Toma de Decisiones en Función del Análisis del Valor.

Fuente: (Amaya Amaya, 2010).

Un gerente debe tomar muchas decisiones todos los días, algunas de ellas son decisiones de rutina o intrascendentes mientras que otras tienen una repercusión drástica en las operaciones de la empresa donde trabaja. Con frecuencia, las decisiones de rutina se toman rápidamente, quizás inconscientemente, sin necesidad de elaborar un proceso detallado de consideración. Sin embargo, cuando las decisiones son complejas, críticas o importantes, es necesario tomarse el tiempo para decidir sistemáticamente. El decisor (la persona que tiene el problema) debe responder en forma rápida ante situaciones que parecen sucederse a un paso cada vez más veloz.

2.2.2 Inteligencia de Negocios

Según el autor (Howson, 2009) en el libro "Business intelligence: estrategias para una implementación exitosa", indica que la inteligencia de negocios es un conjunto de

tecnologías y procesos que permiten a personas de todos los niveles en una organización, tener acceso a datos y su análisis. Sin personas para interpretar información y actuar con base en ella, la inteligencia de negocios nada logra. Son las personas quienes harán de los esfuerzos del BI un enorme éxito o un rotundo fracaso.

Un depósito de información no es sinónimo de inteligencia de negocios, aunque exista un almacén de información, solamente se puede decir que la compañía utiliza inteligencia de negocios una vez que ponga en manos de los usuarios herramientas para llegar a la información y hacerla útil. Una clave de inteligencia de negocios exitosa, es el grado en que impacta en el desempeño comercial. La BI puede confirmar decisiones que los individuos toman en cada paso de un proceso, para ayudar a modernizar un proceso, al medir cuanto tardan los subprocesos e identificar áreas de mejoría.

Las ventajas de utilizar inteligencia de negocios son:

- **Reducción de costes y mayor eficiencia:** La mayoría de las empresas no son conscientes de la cantidad de trabajo y recursos necesarios para generar toda esa información y distribuirla entre las personas que la necesitan. Una parte muy importante de las horas de trabajo de oficina consiste en la recopilación de datos y generación de informes.
- **Más capacidad para tomar decisiones:** Al reducir los costes de obtener información, se abre un nuevo horizonte para la empresa con muchas más posibilidades.

La empresa puede comenzar a plantearse cuestiones o proyectos que antes eran imposibles de valorar por falta

de tiempo o recursos necesarios para hacer una valoración fiable

- **Mejor capacidad de respuesta:** Se puede programar al sistema para que genere notificaciones cuando un valor concreto se encuentre fuera de los parámetros que hemos establecido.
- **Mejor visibilidad, mayor comprensión del negocio:** Los sistemas de inteligencia de negocios proporcionan información analítica y fiable de manera rápida y lo hacen mediante formatos visuales atractivos que facilitan su comprensión y visualización. Los sistemas de inteligencia de negocios nos ofrecen múltiples vistas, permitiéndonos mejorar nuestra perspectiva.

Según los autores (Curto Díaz & Conesa Caralt, 2010) en el libro “Introducción al Business Intelligence” mencionan que prácticamente todas las empresas de la actualidad disponen de bases de datos que almacenan datos sobre sus actividades y sus colaboradores mediante distintos programas informáticos. Por lo tanto podemos decir que las empresas disponen, por norma general, de multitud de datos históricos, fiables y rigurosos de todas las actividades realizadas. Es lógico pensar que dichos datos podrían ser refinados, agrupados, tratados y analizados para intentar extraer información que permitiera ayudar en la toma de decisiones de la empresa. Encontrar patrones de conducta en la compra de nuestros clientes, presentar información en tiempo real sobre el rendimiento de las distintas sucursales de una empresa a su dirección, o identificar los clientes que no nos son rentables, son ejemplos que muestran que se podría obtener información a partir de los datos de la empresa. Este hecho, la conversión de los datos operativos de las empresas

en información que de soporte a la toma de decisiones, es lo que se conoce como inteligencia de negocio o Business Intelligence.

Inteligencia de negocios nos permite responder algunas preguntas como:

- ¿Quiénes son mis mejores clientes?
- ¿Dónde se ubican?
- ¿Qué compran?
- ¿Qué los caracteriza?
- ¿Cuáles líneas de productos son las más rentables?
- ¿Qué productos son mis top 10? ¿Cuáles son mis peores 10?
- ¿Qué productos son los más vendidos este mes?
- ¿El mes anterior?
- ¿Dónde se concentran el 80% de mis ventas, mis gastos, mi rentabilidad?

Según Conesa y Curto, se entiende por Business Intelligence al conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que facilita la tarea de tomar decisiones a los usuarios de una organización.

La inteligencia de negocios implica beneficios a las empresas traducidos en una capacidad inteligente de poder actuar, reducción del costo oportunidad, una mejor proyección de los resultados del negocio y obtener información mucho más consolidada.

2.3. DATA WAREHOUSE (DWH).

2.3.1 Definición

Según los autores (Kommineni Sivaganesh & Suresh Chandra, 2012) en su artículo "Optimization of ETL Work Flow in Data Warehouse" de la revista "*International Journal on Computer Science and Engineering.*" afirman que un Data Warehouse (almacén de datos) es una base de datos relacional que está diseñado para análisis y consultas en lugar de para procesar transacciones. Contiene datos históricos anteriores derivados de múltiples fuentes heterogéneas. Se separa la carga de trabajo de análisis de la carga de trabajo de transacción y permite a las organizaciones consolidar datos de varias fuentes. En la siguiente figura se muestra la arquitectura de un Data Warehouse. Las principales características de un almacén de datos son:

- Orientado al tema
- Integrado
- De tiempo variante
- No volátil

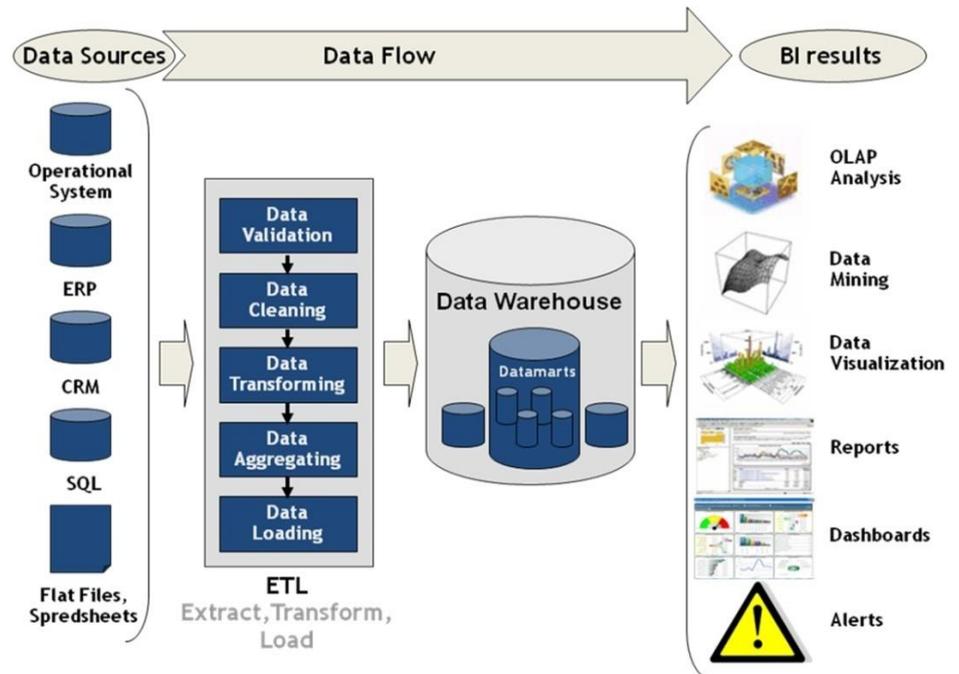


Figura 5: Arquitectura Data Warehouse.

Fuente: (Kommineni Sivaganesh & Suresh Chandra, 2012).

2.3.2 ETL (*Extract, Transform and Load*)

Según los autores (Kommineni Sivaganesh & Suresh Chandra, 2012) en su artículo "Optimization of ETL Work Flow in Data Warehouse" de la revista *"International Journal on Computer Science and Engineering."* afirman que el ETL es el responsable de extraer la información o los datos de diferentes áreas, la personalización, la función de transformación de los datos y, finalmente, la carga en el almacén de datos. El implementar el escenario de un ETL es un proceso por etapas, que contiene sus relaciones, atributos y las transformaciones entre ellos. Los procesos ETL manejan el gran volumen de datos y gestión de la carga de trabajo.

Extraer

Consiste en identificar la información a partir de los sistemas fuentes:

- Bases de Datos Transaccionales
- Hojas de Calculo
- Archivos Texto
- XML

Transformar

A partir de la información identificada para extraer, se le tiene que dar forma a los datos, de tal manera que estén preparados para ser poblados en los Data Marts respectivos. En algunos casos tanto el origen como el destino coinciden directamente, pero en otros es necesario realizar las transformaciones propias.

Cargar

La fase de carga es el momento en el cual los datos de la fase anterior son cargados en el destino. Dependiendo de los requerimientos de la organización, este proceso puede abarcar una amplia variedad de procesos diferentes. Algunos almacenes de datos sobrescriben información antigua con nuevos datos. Los sistemas más complejos pueden mantener un historial de los registros de manera que se pueda hacer una auditoría de los mismos y disponer de un rastro de toda la historia de un dato. En la siguiente figura se puede observar todo el Flujo de Trabajo ETL.

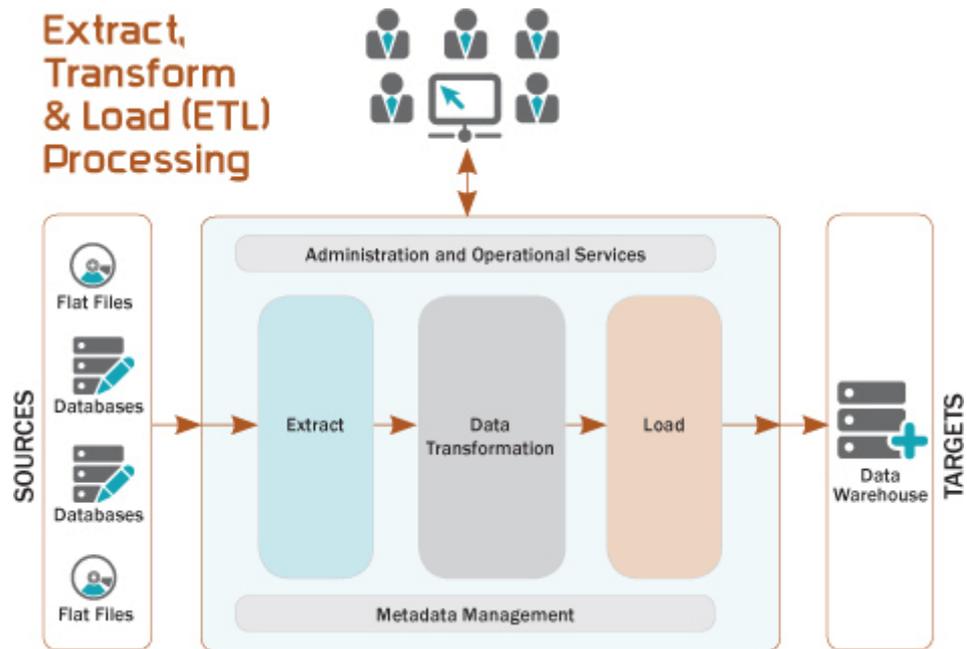


Figura 6: Flujo de Trabajo ETL.

Fuente: (Kommineni Sivaganesh & Suresh Chandra, 2012).

2.3.3 EI MODELO DIMENSIONAL

Según los autores (Salcedo Parra, Milena Galeano, & Rodriguez B, 2010) en el artículo "Modelamiento Dimensional de Datos", de la revista "Red de Revistas Científicas de América Latina, el Caribe, España y Portugal", indican que el Modelo Dimensional es una técnica de diseño lógico enfocada a presentar la data en una arquitectura estándar que es altamente intuitiva y busca ejecutar rápidos accesos. Surge en los años 60 pero ha tomado mayor presencia desde el uso del Data Warehouse en las empresas.

El modelo dimensional se encuentra conformado por tablas hecho y tablas dimensionales, en la siguiente figura podemos observar los componentes de un Dara Warehouse.

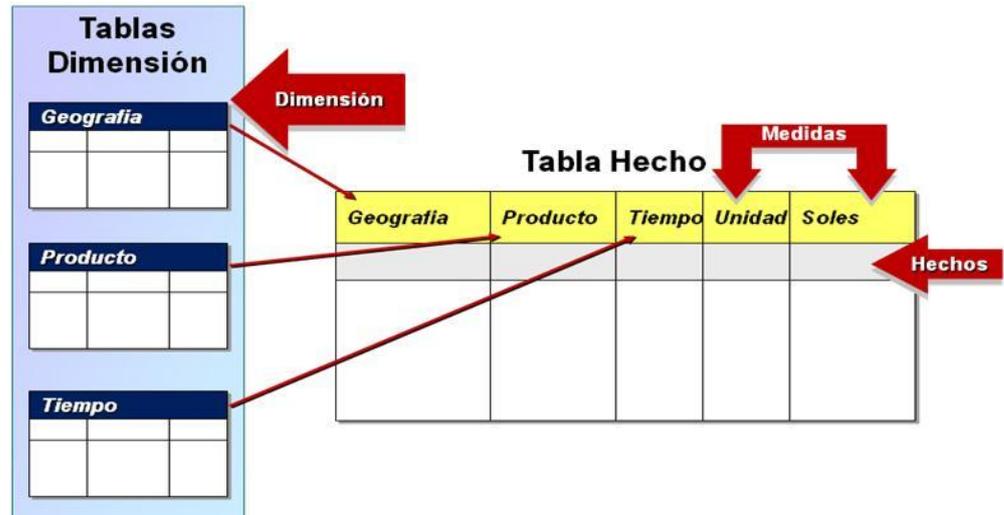


Figura 7: Componentes de un Data Warehouse.

Fuente: (Salcedo Parra, Milena Galeano, & Rodriguez B, 2010).

Tabla Hecho: Incluye las medidas como parte de sus atributos, es lo que se desea analizar, además en ella se ubican las claves foráneas de las dimensiones.

Medidas: Representan el valor a ser analizado. Estas medidas deben ser numéricas y permitirán realizar agregados de la información y servirán de base para ejecutar cálculos en un futuro. Por ejemplo podemos citar: los montos vendidos, el peso de materia prima, entre otros.

Es posible tener medidas pre-calculadas siempre que mejoren el tiempo de respuesta de las consultas a realizar, esto es cuando la formula tenga algo de complejidad.

En la siguiente figura podemos observar que el Precio Oferta representa una formula algo compleja que en tiempo de ejecución podría restarle rendimiento a las consultas. Esta puede ser una medida natural.

<i>TiempoKey</i>	<i>Product Key</i>	<i>Precio</i>	<i>Descuento</i>	<i>Rebatir</i>	<i>PrecioOferta</i>
7	2	20.00	.10	5.00	13.00
13	...				
25		$((\text{Precio} - (\text{Precio} * \text{Descuento})) - \text{Rebatir}) = \text{PrecioOferta}$			
7	5	10.00	.10	5.00	4.00
...

Figura 8: Medida Calculada.

Fuente: (Salcedo Parra, Milena Galeano, & Rodriguez B, 2010).

2.3.4 TIPOS DE MODELO DIMENSIONAL

Según los autores (Salcedo Parra, Milena Galeano, & Rodriguez B, 2010) en el artículo “Modelamiento Dimensional de Datos”, de la revista “Red de Revistas Científicas de América Latina, el Caribe, España y Portugal”, indican que en un Data Warehouse relacional encontramos modelos en estrella y en copo de nieve que premian por encima de todo las consultas.

2.3.4.1. El modelo Estrella

Es un modelo que presenta a la tabla hecho como eje central y a su alrededor se ubican las dimensiones, es más sencillo de comprender, siendo un modelo desnormalizado. Tal y como se muestra en la siguiente figura, cada dimensión posee una Clave Primaria la cual, una vez establecida la relación con la tabla hecho, permitirá analizar una determinada medida en la dimensión respectiva.

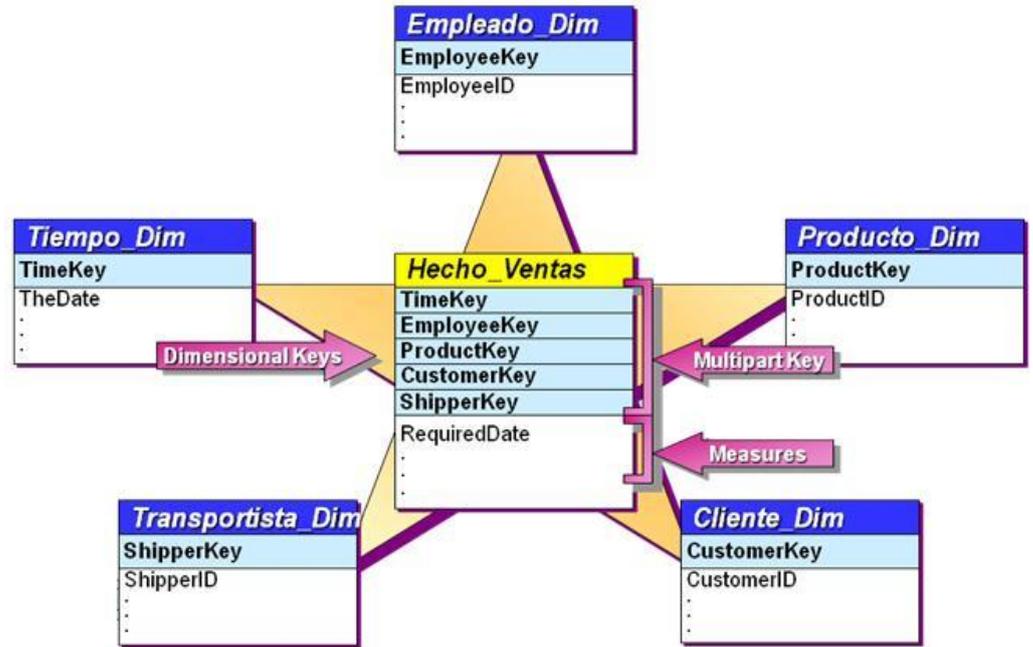


Figura 9: El modelo Estrella.

Fuente: (Salcedo Parra, Milena Galeano, & Rodriguez B, 2010).

2.3.4.2. El modelo copo de Nieve

Consiste en descomponer una dimensión, lo cual podría incrementar la complejidad del modelo. En la siguiente figura se observa que la Dimensión Producto es la conformación adicional de Línea y Categoría.

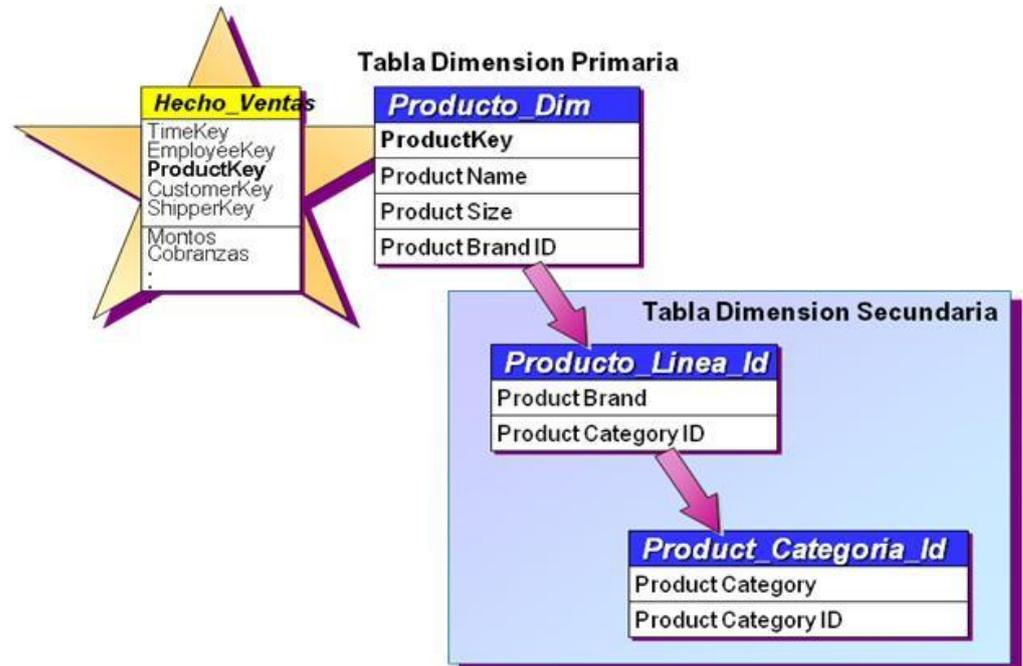


Figura 10: El modelo Copo de Nieve

Fuente: (Salcedo Parra, Milena Galeano, & Rodriguez B, 2010).

2.3.4.3. Ventajas y desventajas de los modelos:

Modelo en estrella:

- Este esquema es simple y veloz para ser usado en análisis multidimensionales. Permite acceder tanto a datos agregados como de detalle.
- El diseño de esquemas en estrella permite implementar la funcionalidad de una base de datos multidimensional utilizando una clásica base de datos relacional.
- Es simple desde el punto de vista del usuario final. Las consultas no son complicadas, ya que las condiciones y los joins sólo involucran a la tabla de hechos y a las de dimensiones.

- Son más simples de manejar que los modelos de copo de nieve.
- Es la opción con mejor rendimiento y velocidad pues permite indexar las dimensiones de forma individualizada sin que repercuta en el rendimiento de la base de datos en su conjunto.

Modelo copo de nieve:

- El único argumento a favor de los esquemas en copo de nieve es que al estar normalizadas las tablas de dimensiones, se evita la redundancia de datos y con ello se ahorra espacio.
- Se puede usar un esquema de copo de nieve en un Data Warehouse, aunque estos sean realmente grandes y complejos, pero nunca en sistemas donde el tiempo de respuesta sea un factor crítico para los usuarios.

2.3.5 DATA MART

Según los autores (Salcedo Parra, Milena Galeano, & Rodriguez B, 2010) en el artículo “Modelamiento Dimensional de Datos”, de la revista “Red de Revistas Científicas de América Latina, el Caribe, España y Portugal”, indican que un Data Mart es una parte de un Data Warehouse y que le permite construir en menos tiempo una solución de Soporte de Decisiones. Si el Data Warehouse integra los datos de toda la organización, el Data Mart se restringe a un determinado proceso de negocios o departamento.

Componentes Básicos de un Datamart

- **Sistemas Orígenes (Legacy System):** Buscan capturar, fundamentalmente, las transacciones del negocio de

forma consistente. Las bases de datos OLTP constituyen el insumo básico para que un Data Mart tenga “vida”.

- **Área de Depuración de Data (Data Staging Area):** constituye un área de almacenamiento en donde se realizan los procesos de limpieza, y consistencia de datos a ser usada en el Data Mart a partir de los sistemas operacionales. En algunos casos la data fuente puede estar constituido por archivos planos o XML, que igual podría ser depurada. Aquí es donde se produce el proceso de ETL (Extract, Transformation, Load). La idea del ETL es extraer los datos de los sistemas orígenes, realizar transformaciones y llevar la información a los Data Mart respectivos.
- **Servidor de Presentación:** es el equipo o servidor físico en donde los datos del Data Mart estarán organizados y almacenados y a partir del mismo se pueda analizar datos. El Data Mart puede ser implementado en un Gestor de Base de Datos Relacional el cual para ser aprovechado en toda su magnitud, se incorpora la tecnología OLAP, representada por los Cubos.
- **Herramientas OLAP:** constituyen la forma de explotar como está estructurado el Data Mart. Mediante esta tecnología podemos analizar datos y transformarlos en información, proporcionando una vista multidimensional de los datos. Los usuarios pueden navegar de arriba hacia abajo o viceversa (dril down and dril up). Así como construir vistas dinámicas de los datos. Su estructura de almacenamiento interno puede ser MOLAP o ROLAP.
- **Aplicaciones de usuario final:** Constituyen las herramientas que se enlazan a un Cubo OLAP y que permiten al tomador de decisiones realizar los análisis respectivos por medio de información tabular y grafica que

muestran estas herramientas. Hay una serie de fabricantes como COGNOS, PROCLARITY, PENTAHO y MICROSOFT, entre otros.

En la siguiente figura se muestran los componentes básicos de un Data Mart.

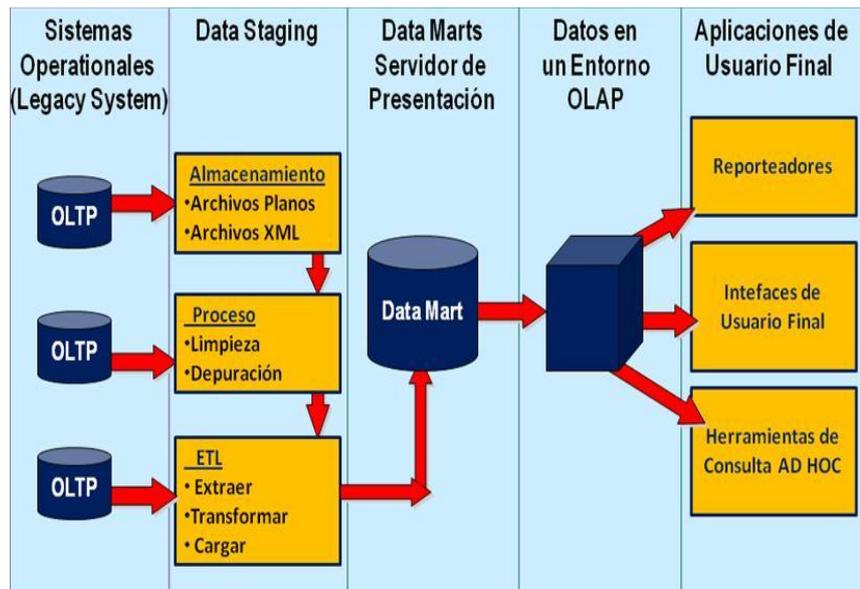


Figura 11: Componentes Básicos de un Data Mart.

Fuente: (Salcedo Parra, Milena Galeano, & Rodriguez B, 2010).

2.3.6 DATA MART Y DATAWAREHOUSE

Según los autores (Salcedo Parra, Milena Galeano, & Rodriguez B, 2010) en el artículo “Modelamiento Dimensional de Datos”, de la revista “Red de Revistas Científicas de América Latina, el Caribe, España y Portugal”, indican que Un Data Mart cumple los mismos principios que un Data Warehouse, construir un repositorio de datos único, consistente, fiable y de fácil acceso.

Entonces la diferencia entre ambos es su alcance. El data mart está pensado para cubrir las necesidades de un grupo de trabajo o de un determinado departamento dentro de la organización, en cambio, el ámbito de un data warehouse es la organización en su conjunto.

El DataMart supone una buena opción para pequeñas y medianas empresas que no puedan afrontar el coste de poner en marcha un Data Warehouse. La escalabilidad de los data marts hacia el data warehouse puede ser una solución si el número de data marts aumenta considerablemente. Los datos que se utilizan para poblar el data mart provienen de los sistemas operacionales y/o fuentes externas.

2.3.7 PROCESO ANALÍTICO EN LÍNEA (OLAP)

Según los autores (Salcedo Parra, Milena Galeano, & Rodriguez B, 2010) en el artículo “Modelamiento Dimensional de Datos”, de la revista “Red de Revistas Científicas de América Latina, el Caribe, España y Portugal”, indican que OLAP es una tecnología que permite sacar provecho a como está estructurado un Data Mart, presentando textos y números bajo el concepto dimensional. Presenta las siguientes características:

- **Esta optimizado para realizar consultas rápidas de los usuarios:** los cubos OLAP manejan una serie de niveles sumarios de datos altamente optimizados para consultas.
- **Poseen un motor robusto para realizar análisis numéricos:** generando simples reportes ejecutando cálculos complejos dentro de su motor.

- **Es un modelo de datos conceptual y altamente intuitivo:** los usuarios pueden comprender fácilmente el modelo.
- **Proporciona una vista de datos multidimensional:** permitiendo una vista flexible de datos, análisis y navegación, por lo tanto los usuarios pueden navegar alrededor de los datos, produciéndose de esta forma las sumalizaciones respectivas, partiendo del resumen hacia el detalle, a su vez se pueden crear vistas dinámicas incorporando dimensiones a nivel de fila, columna y filtros. Los mismos que pueden ser cambiados rápidamente por los mismos usuarios.

Según los autores (Kommineni Sivaganesh & Suresh Chandra, 2012) en su artículo “Optimization of ETL Work Flow in Data Warehouse” de la revista “*International Journal on Computer Science and Engineering.*” afirman que esta tecnología de manera interna maneja dos cosas: Datos y Agregaciones. Este último es una forma de organizar la información internamente con el objetivo de mejorar el tiempo de respuestas de las consultas que se realice.

La arquitectura OLAP presenta diferentes tipos de estructura de almacenamiento como son:

ROLAP (Relational OLAP) se organizan tanto Datos como Agregaciones dentro del Data Mart como tablas relacionales.

MOLAP (Multidimensional OLAP) se organizan Datos y Agregaciones de una manera multidimensional.

En la siguiente figura se muestra el esquema de un Cubo OLAP.

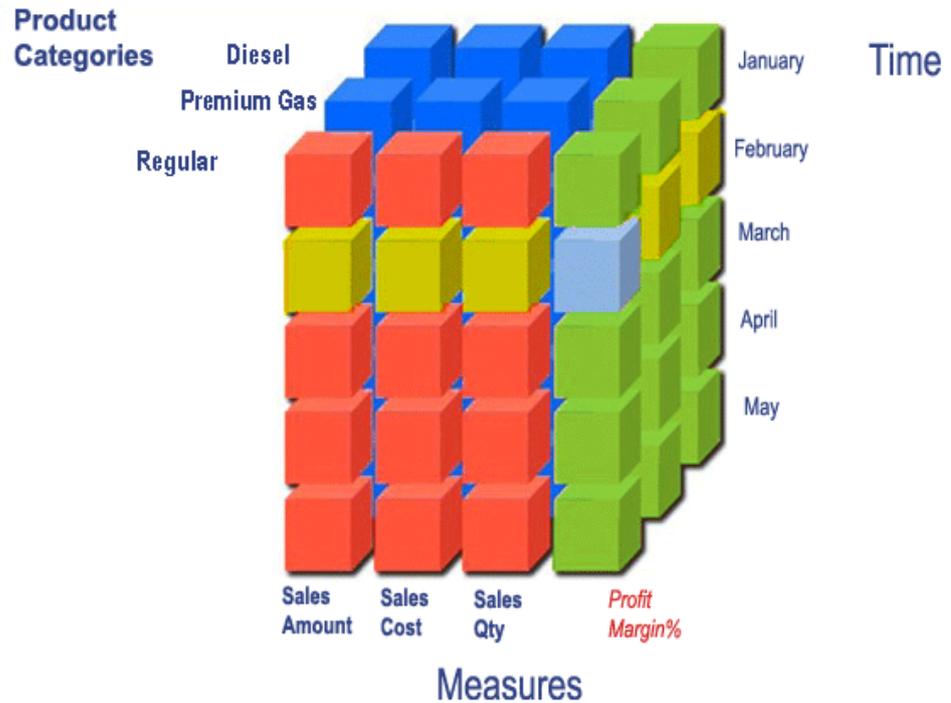


Figura 12: Cubo OLAP.

Fuente: (Kommineni Sivaganesh & Suresh Chandra, 2012).

2.4. BIG DATA ANALYTICS

2.4.1. Definición

Según los autores (Hurwitz, Nugent, Halper, & Kaufman, 2013) en su libro *“Big Data for Dummies”* mencionan que Big Data no es una sola tecnología, sino una combinación de viejas y nuevas tecnologías que ayuda a las empresas a obtener una visión práctica, brindando la capacidad de manejar un gran volumen de datos dispares, a la velocidad adecuada, y dentro del marco de tiempo adecuado para permitir el análisis en tiempo real.

Es necesario identificar la cantidad y los tipos de datos que pueden ser analizados para impactar los resultados del negocio, este tipo de gestión de datos requiere que las empresas aprovechan tanto sus datos estructurados y no estructurados. Big Data va más allá de una simple cuestión de tamaño, el Big Data es una oportunidad para encontrar ideas nuevas y emergentes en los tipos de datos y contenidos, para que un negocio sea más ágil, y para responder a preguntas que antes se consideraban fuera del alcance.

Por otro lado podemos decir que Big Data es una de las nuevas tendencias de Business Intelligence, entre ellas cabe destacar:

- ✓ Real Time BI: Cloud Big Data, las decisiones no pueden esperar.
- ✓ Mobile BI: Decisiones inteligentes desde cualquier lugar.
- ✓ Integración Social Media - Social Intelligence: Redes Sociales.
- ✓ Web Analytics: Proceso de recolección, análisis y toma de decisión entorno a las estadísticas que genera un web site en cuanto a sus visitantes y sus hábitos de uso de la página.
- ✓ Web Mining: Metodología de recuperación de la información que usa herramientas de la minería de datos y de analítica web para extraer información tanto del contenido de las páginas, de su estructura de relaciones y de los registro de navegación de los usuarios.

Uno de los puntos que han cambiado la forma en que se hace y hará Business Intelligence es la gran cantidad de datos que anteriormente no se analizaban. Ahora es posible combinar y analizar de forma conjunta, tanto datos estructurados (relacionales, legacy, dbcolumn, etc.) con no estructurados (Hadoop, MapReduce, NoSQL), permitiendo alcanzar cantidades enormes de datos. Uno de los sistemas de almacenamiento que están teniendo mayor atención es Hadoop. (DataPrix, 2013).

Con Big Data el concepto de base de datos que se utilizaba anteriormente cambió:

- ✓ Distintos tipos de bases de datos, para distintos tipos de necesidades.
- ✓ Bases de datos transaccionales: para almacenar información crítica del negocio, con origen en las aplicaciones de siempre (CRM, ERP, etc.)
- ✓ NoSQL: para capturar de manera segura y escalable, grandes volúmenes de información continua generados por eventos.
- ✓ Analíticas: para permitir a múltiples usuarios contestar rápidamente preguntas de negocio que requieran de grandes volúmenes de información.
- ✓ Hadoop (HDFS & Map Reduce & Hive): para almacenar y procesar grandes volúmenes de información estructurada, no estructurada o semi-estructurada.

Por lo tanto Big Data también representa nuevos tipos de Información. Nuevos tipos de datos multi estructurados con relaciones desconocidas, requiriendo procesamiento

especializado independiente de su volumen para descubrir ideas. (Acis, 2013).

En la siguiente figura se muestra el incremento de la variedad de datos proporcionados por múltiples fuentes y el grado de complejidad para analizarlas.

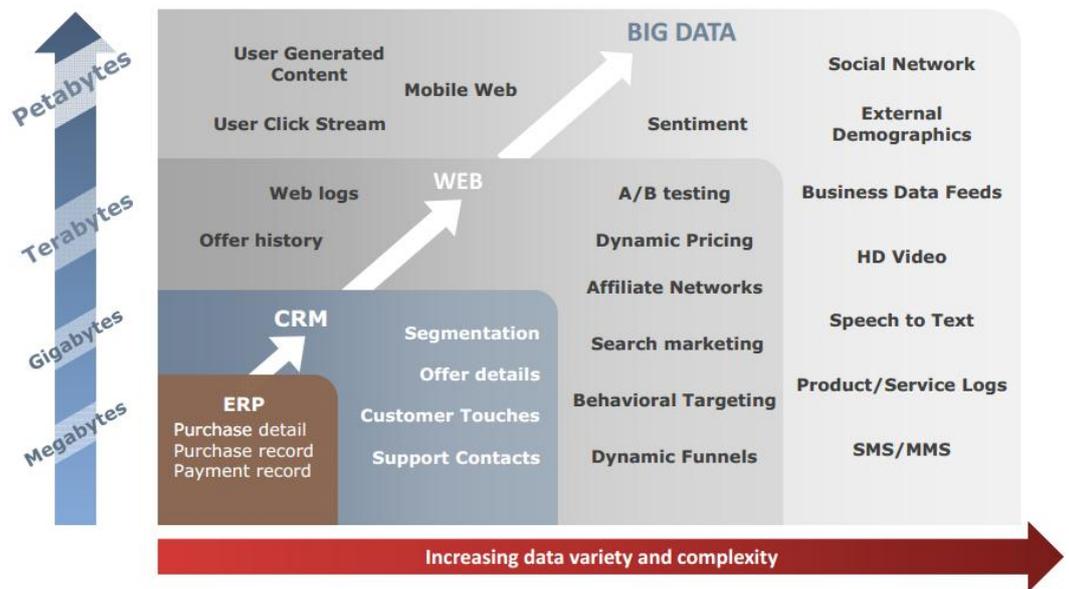


Figura 13: Explosión de la Información.

Fuente: (Acis, 2013).

Siempre que hagamos una búsqueda, enviemos un email, usemos un teléfono móvil, actualicemos una red social, usemos una tarjeta de crédito o activemos el GPS dejamos detrás de nosotros una montaña de datos y registros que ofrecen una información muy valiosa, la enorme cantidad de datos que generan empresas, usuarios y dispositivos, ha experimentado un crecimiento explosivo que requiere su análisis para obtener ventajas competitivas. Sin embargo no sería correcto olvidar otros retos, al margen de los tecnológicos, que nos encontraremos al llevar a cabo esta tarea. En primer lugar, los datos por sí mismos son incapaces

de producir un beneficio. Solo los humanos que sepan explotarlos podrán hacerlo. Es por esto que además de la tecnología necesaria para mover tal volumen de datos, es necesario afinar al máximo los modelos, metodologías y procesos que se utilizan para acceder y explotar esta información.

En conclusión el Big Data puede definirse como los procesos por los cuales se recopilan y analizan grandes cantidades de datos de forma estructurada y no estructurada, procedente de diversas fuentes, de modo que seamos capaces de conseguir información relevante de gran valor estratégico y táctico para nuestros objetivos, sean de negocio, sociales, o de índole particular y específico para nuestra empresa.

2.4.2. Características de Big Data

Según los autores (Mayer-Schönberger & Cukier, 2013) en su libro "Big Data: La Revolución de los datos masivos" mencionan que lo que hay que hacer es saber cómo estructurar los datos a largo plazo, en un entorno controlado, solamente estructurando bien los datos, obtendremos resultados. Para Big Data existe la regla de las cuatro V, o cuatro dimensiones del Big data, tal y como determinó IBM (Zikopoulos, Eaton, Deroos, Lapis, & Deutsch, 2012) en el libro "*Understanding Big Data*". Las cuales representan las cuatro características que definen a Big Data:

- ✓ Volumen: es la cantidad de datos que recogemos de cada fuente que aplicamos.
- ✓ Velocidad: es el tiempo que pasa desde que se recogen los datos hasta que se saca un análisis y una

conclusión, es decir es qué tan rápido se procesa esos datos.

- ✓ Variedad: se refiere a los distintos tipos de datos, el poder obtener los datos de diferentes formas y técnicas. Cuanta más información de diferentes lugares mejor.
- ✓ Veracidad: Se refiere a que los datos que recogemos, realmente son los que nos sirven para tomar decisiones.

En la siguiente figura se muestra las 4 V de Big Data:

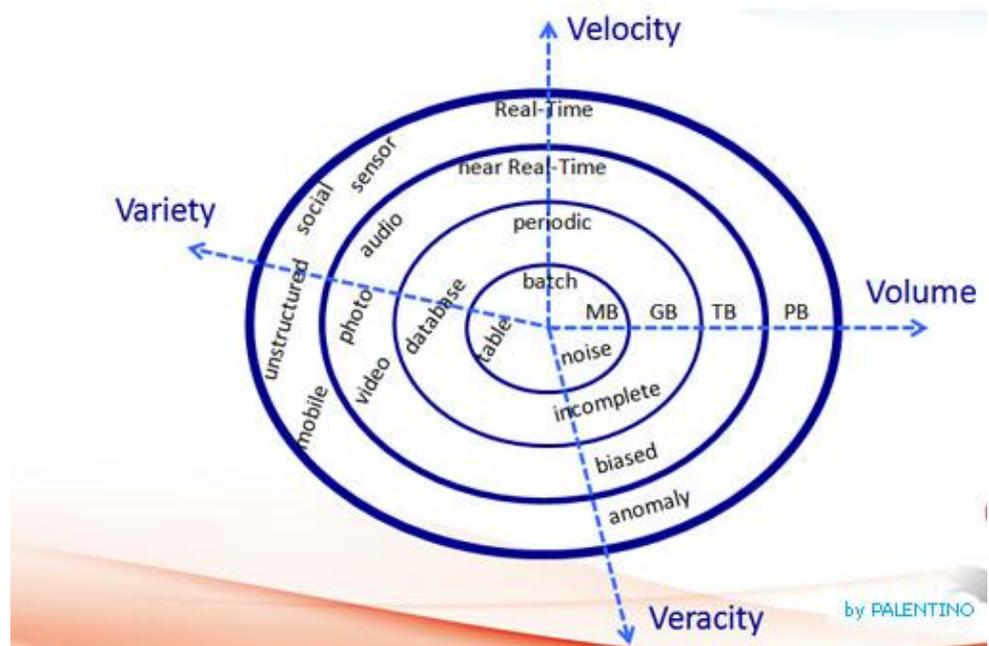


Figura 14: Las cuatro V de Big Data.

Fuente: (Palentino, 2013).

Los principales motivos por el cual Big Data está siendo tan importante son:

- ✓ Se trata, en definitiva de poder aprovechar una ventaja competitiva por parte de las compañías que mejor sepa aprovechar todo esta gran cantidad de información en su propio provecho.
- ✓ Surgen nuevos entornos para almacenamiento, trabajo y computación distribuida, como Hadoop. Hadoop se ha convertido en el estándar de Big Data y es la solución que están adoptando infinidad de empresas, con muchas adaptaciones proporcionadas por multitud de compañías.
- ✓ Se ha producido una evolución de las bases de datos hacia el mundo analítico. Precisamente, las necesidades de reporting, análisis en profundidad, web mining y social mining son uno de los motivos por los que el Big Data y las tecnologías que se están desarrollando avanzan tan aprisa.
- ✓ Ya no se trata sólo de almacenar una gran cantidad de datos y de información, sino de dejarla estructurada de forma que sea accesible de forma ágil y rápida para hacer todo tipo de consultas: rankings, medias, tendencias, etc.
- ✓ La gran importancia del Big Data va muy unida al desarrollo de las tecnologías y soluciones Open Source, puesto que para el manejo de esta gran cantidad de datos y por un número tan elevado de personas, requeriría de gastos en licencias de software de Base de Datos, analíticas, etc. que haría inviable su desarrollo o estaría limitado a muy pocas y grandes corporaciones. Por ello, es habitual ver que las empresas que están alrededor de estos movimientos también son habituales del uso y desarrollo de soluciones Open Source.

- ✓ Mientras que la tecnología (hardware, software, arquitectura) es cada vez más barata, al día de hoy no encontramos muchos especialistas o expertos que puedan desarrollar y desplegar soluciones Big Data, por lo que el factor humano es uno de los puntos más importantes para estar bien posicionados en estas nuevas tecnologías. (DataPrix, 2013).

2.4.3. Tipos de Implementación de Big Data

La implementación de Big Data, dependerá de los siguientes factores:

- ✓ La madurez del Business Intelligence de la empresa.
- ✓ La infraestructura de que la organización disponga para la implementación.
- ✓ Su nivel de conocimiento del nuevo entorno.

Existen dos modelos diferentes de implementación de Big Data el revolucionario y el Híbrido.

La implementación revolucionaria de Big Data

Quienes optan por esta forma de implementar Big Data han decidido romper con todo y empezar de cero. Es obvio que todas las empresas ya están funcionando, han generado datos y los han ido guardando y procesando a su manera. El pasado queda atrás y desde el primer momento todo comenzará de cero en esta nueva plataforma. Mediante su táctica de "divide y vencerás" permiten trabajar con volúmenes inmensos de información en poco tiempo y a bajo coste y que se complementan con la posibilidad de moverlo todo al cloud.

La alternativa híbrida a la implantación de Big Data

Este sistema alterna el uso de una y otra tecnología en función del objetivo perseguido. Para determinado tipo de información como por ejemplo de datos estructurados se seguiría usando la BI existente tanto como para la extracción y el análisis, mientras que para los datos no estructurados, se usaría Big Data para la extracción y conversión a data estructurada para su posterior análisis. Para lograrlo, bastaría con establecer un par de puntos de integración, consiguiendo que los datos no estructurados se conviertan a datos estructurados para que luego sean enviados al repositorio que tiene almacenada la data estructurada, el cual será visto por el motor de análisis como un DataMart, integrando de esta forma ambas fuentes de datos.

En esta opción, Big Data no sólo es un anexo al BI tradicional que permite ahora ver datos que antes no podía, sino que además es una plataforma que sirve para hacer análisis avanzado mezclando datos del tradicional que siguen allí, con los nuevos como por ejemplo Redes Sociales, es decir los datos no estructurados que los BI de antes no contemplaban a la hora de trabajar. (PowerData, 2013).

2.4.4. Ventajas y Desventajas de implementar Big Data Analytics

Ventajas:

Una de las ventajas de Big Data Analytics es que puede coexistir con el BI tradicional, como se explicó en los tipos de implementación. Su poder reside en la facultad que otorga de tener la posibilidad de trabajar con grandes volúmenes de información y mayor precisión, algo que redundaría en una toma

de decisiones impecable, por lo actualizado y profundo de su enfoque. (PowerData, 2013).

Los beneficios más importantes de Big Data Analytics para la empresa son:

- ✓ Tiempo de acceso:
 - Con Big Data Analytics se puede trabajar con millones de datos a la vez.
 - El tiempo de procesamiento se reduce exponencialmente.
 - Se gana en perspectiva.

- ✓ Almacenamiento:
 - Aumenta la capacidad de almacenar los datos pasando de terabytes a petabytes.
 - Se tiene acceso al detalle entero.
 - Se gana en granularidad.
 - Se logra una mejor comprensión de la información y, por tanto, del entorno y de la propia empresa.

Desventajas:

Según el autor (Jara, 2012) en su libro “Big Data & Web Intelligence” mencionan que como toda tecnología en surgimiento y desarrollo, Big Data no es todo color de rosas, presenta complicaciones y problemas relacionados a distintos factores, desde el hecho de hacer cambiar las infraestructuras y formas de pensar de los desarrolladores que hoy están acostumbrados a tecnologías como BI y Data Mining, utilizando estilos tradicionales de desarrollo, hasta saber qué tipo de datos son los adecuados para buscar información para estas implementaciones.

Entre los problemas más comunes podemos citar los siguientes:

Skills: Este problema trata básicamente la capacidad de las personas a cargo del manejo de la información recolectada. Al ser una tecnología en desarrollo, la cantidad de personas que tengan el know how o conocimiento para poder procesar de manera correcta el volumen de información es relativamente poco, lo que dificulta el desarrollo de proyectos.

La tecnología: Lo interesante es que Hadoop es ideal para el procesamiento por lotes a gran escala, que es como las operaciones de agregación o computo. El problema es que Hadoop no es una tecnología en tiempo real o muy dinámico en absoluto. La ejecución de consultas en un cluster Hadoop suele tener una gran latencia ya que hay que distribuir cada consulta individual, luego, hacer su etapa de reducción, que está trayendo todos los datos de nuevo juntos. Así que es una tecnología de alto rendimiento, pero de alta latencia.

Privacidad: Junto con la obtención de volúmenes de datos incalculables, viene una cantidad de datos que podríamos considerar intrusiva, podrá darse ejemplos como Facebook, Twitter, Google que manejan grandes volúmenes de datos de clientes, con esta capacidad de Big Data de intentar analizar absolutamente todo, podrá darse una examinación inapropiada de los datos de usuarios, conllevando rupturas en la privacidad de los datos de los usuarios. (Si bien esta problemática no es nueva, podrá agravarse con la capacidad avanzada de procesamiento que se obtiene con Big Data).

Volumen, Variedad, Velocidad, Veracidad: La capacidad de encontrar un equilibrio entre todas ellas depende de la capacidad de plantear un desarrollo sustentable y un plan

acorde a las posibilidades tecnológicas de la empresa que desarrolla con esta tecnología.

2.4.5. Big Data en el Marketing

Entendemos por Big Data como toda la información de gran volumen que no podemos analizar de forma tradicional, por ejemplo de todos los datos que puede recibir una tienda online de su mercado y sus clientes, que deberá de analizar para poder sacar conclusiones. Como bien sabemos, en cualquier estrategia de marketing, este pilar es primordial, ya que no nos sirve de nada llevar a cabo muchas técnicas si no analizamos qué funciona y que no.

Big Data permite medir el éxito del feedback con nuestros clientes y si las campañas que llevamos a cabo van por buen camino. Además, analizando estos datos, nos permite ofrecer por ejemplo a nuestros clientes de nuestra tienda virtual una oferta de los productos más personificada. (Lynko, 2013).

2.4.6. Diferencia entre datos estructurados y no estructurados

Los datos no estructurados son los que no tienen categoría, y por lo tanto difícilmente serán útiles para realizar un análisis. Caso contrario son los datos estructurados, estos tienen categoría y nos permiten tomar las decisiones oportunas. Por ejemplo, los datos no estructurados se pueden conseguir de Social Media: cuantos emails han abierto nuestros clientes, rebotes, clicks, cuantas visitas, horario que más se conectan los usuarios en las redes sociales. Los datos estructurados son almacenados de una manera perfectamente identificable, la más universal de las formas de dato estructurado se encuentra en una base de datos relacional que permite, a través de SQL (Structured Query Language), seleccionar

piezas específicas de información desde una tabla organizada en filas y columnas.

Por otro lado también existen los datos semi estructurados tradicionalmente incluyen imágenes, documentos de texto, hojas de cálculo y otros objetos que no son parte de una base de datos. La mayoría de los datos en una empresa moderna pueden ser considerados semi estructurados, para ejemplificar es suficiente decir que los mensajes de correo electrónico caen en esta categoría. A pesar de que los correos-e puedan estar almacenados en una base de datos, el cuerpo mismo del mensaje realmente contiene texto sin una estructura fija. Igual sucede con un documento escrito en Word: aunque deba cumplir con los formatos que impone cada empresa, el texto que contiene es de forma libre.

Entonces surge el problema de organizar estos datos semi estructurados o no estructurados de alguna manera que permita hacer consultas sobre ellos y, en consecuencia, se los pueda incorporar en los análisis que llevan a la toma de las mejores decisiones de negocios.

Por lo tanto, es muy importante lograr que esa información que ahora tenemos sin categoría se convierta en datos estructurados, para así poderlos usar en conjunto con otros datos ya estructurados. De esta manera haremos un análisis más completo. (Lynko, 2013).

2.4.7. Arquitectura Big Data: Hadoop y OLAP

Se puede aprovechar lo que se conoce como Big Data, utilizando herramientas tales como Hadoop (HDFS, Map Reduce) y Hive. Hadoop puede procesar eficazmente grandes

cantidades de datos a bajo precio por dividir los datos en nodos más pequeños. También proporciona la ventaja de ser capaz de almacenar los datos no estructurados. Sin embargo, a diferencia del tradicional DW, la analítica a través de Hadoop y Hive es todavía relativamente primitiva y en la actualidad carece de herramientas amigables para el usuario empresarial, que pueden no querer escribir sus propias consultas HiveQL, es por esto que se recomienda combinar los esfuerzos de Big Data para la explotación de los datos con herramientas más maduras en este rubro como por ejemplo es el caso de la plataforma Pentaho. (ZooskDev, 2012).

No existe una sola propuesta estándar con respecto a la arquitectura en la cual debe trabajar Big Data, cada empresa dedicada a desarrollar Big Data y cada proveedor como el caso de Pentaho proporcionan una propuesta de arquitectura. En la siguiente figura se muestra una primera propuesta de la arquitectura de Big Data basándose en las herramientas mencionadas, para luego explicar en qué consiste cada una de las etapas involucradas en esta arquitectura, es importante mencionar que esta arquitectura correspondería según lo comentado anteriormente al modelo de implementación de Big Data Híbrido.

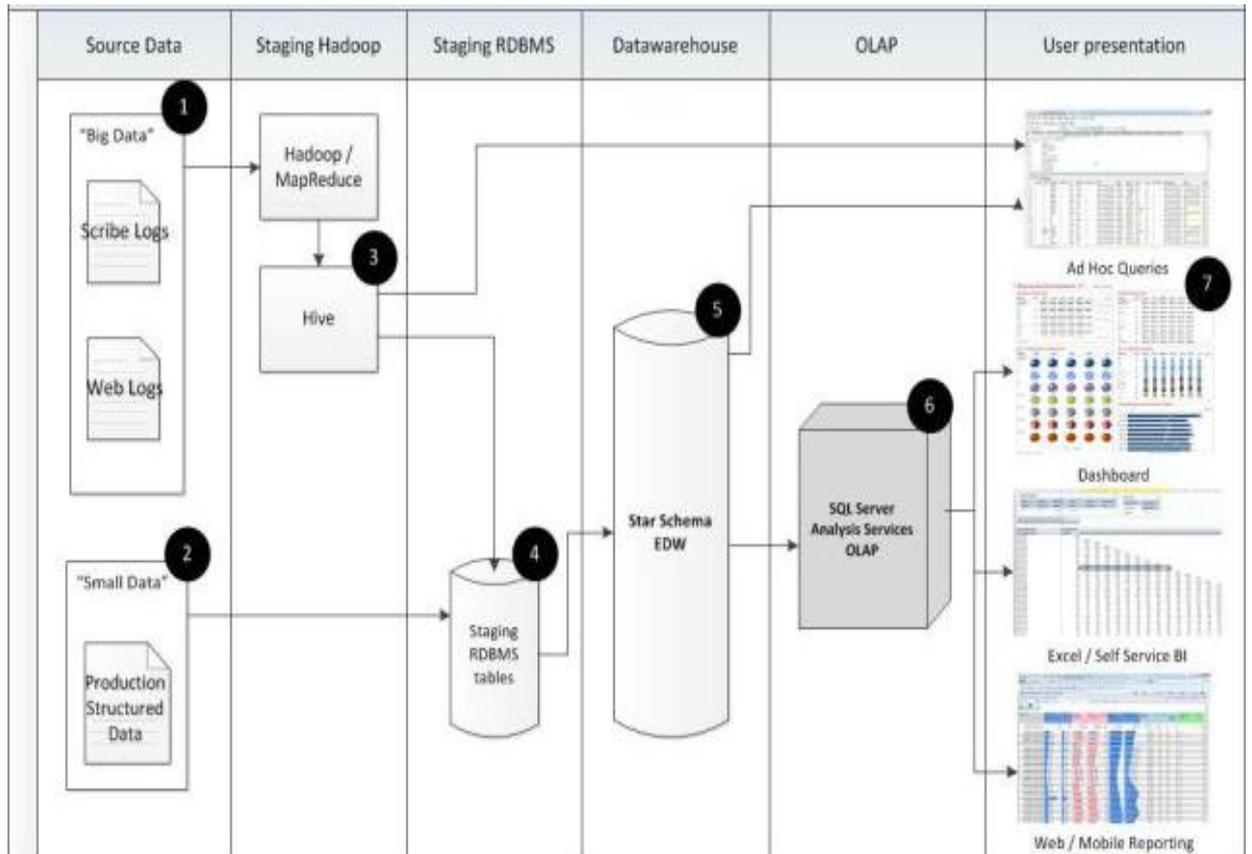


Figura 15: Arquitectura de Big Data.

Fuente: (ZooskDev, 2012).

1. En primer lugar, se comienza con la recolección de datos externos que necesitamos. Nuestro sitio web escribe los datos de comportamiento del usuario en un registro de trazado (Scribe log), mientras que el servidor web recopila estadísticas web vitales en sus registros web (web logs). Los weblogos son considerados parte de nuestro "Big Data" de nuestros datos de origen.
2. Al mismo tiempo, necesitamos los datos de producción críticos, estructurados como por ejemplo las bases de datos transaccionales. Estos grupos de base de datos incluyen los datos de perfil de usuario, la información financiera y otra información basada en transacciones.

Estos se consideran la parte "Data Small" de nuestros datos de origen.

3. Para los usuarios avanzados los registros de trazado / logs web se instanciaran en Hadoop, donde también tenemos a Hive en la parte superior de la misma para permitir directamente el acceso ad hoc de análisis de los datos subyacentes a través de la visualización por parte de los usuarios de los datos por medio de las interfaces para consultas ad hoc, aunque muy limitado ya que Hive si bien es considerado el Data warehouse de Hadoop aun en la actualidad no brinda todas las herramientas necesarias para realizar un análisis de manera sencilla. También existe otra opción que brinda esta arquitectura para usuarios menos experimentados en Big Data el cual es que puede actuar como una primera etapa del proceso de integración de datos que permita convertir nuestra "Big Data" en información más manejable, es decir estructurada para luego exportarlos a nuestro DW, en lugar de acceder directamente desde consultas ad hoc, de esta forma se logra un integración con la data estructurada proveniente de nuestras fuentes internas.
4. Secuencias de comandos de base de datos para extraer, transformar y cargar los datos (ETL), se utilizan para cargar ambos "Big Data" y " Small Data " que se establecerán en nuestras tablas de importación RDBMS (Sistema de gestión de bases de datos relacionales).
5. Mediante el uso de metodologías probadas de la industria en el modelado dimensional de los datos, se desarrollan las tablas de datos para almacenarlos en el Data Warehouse corporativo. (EDW Enterprise Datawarehouse). El EDW se puede utilizar para el análisis ad hoc de consulta también, pero su principal objetivo es poblar

nuestros Data Marts de datos multidimensionales para su posterior análisis por parte de los Cubos OLAP.

6. Los Data Marts que componen el EDW, tienen todas las estructuras dimensionales, KPI y métricas definidas para que los usuarios sólo tengan que navegar a través de los cubos OLAP para luego generar informes en la web.
7. Esta arquitectura permite a los usuarios obtener una visión de los datos a través de diferentes interfaces complementarias, como tablas de análisis, reportes, tableros de mando, etc. Los usuarios avanzados pueden realizar análisis ad hoc directamente de Hive o de SQL EDW dependiendo de la opción que hayan elegido al momento de implementar la arquitectura.

Por otra parte la herramienta que está demostrando hasta el momento ser la más eficaz al momento de integrar y aplicar Big Data, es la plataforma de software libre Pentaho, en la cual destaca su herramienta estrella de integración llamada Pentaho Data Integration que proporciona grandes ventajas al momento de trabajar con Big Data como por ejemplo la facilidad de poder conectarse a hadoop y a bases de datos como hive, Hbase entre otras. Pentaho también propone su propia arquitectura para el desarrollo de Big Data en las organizaciones, es importante mencionar que esta arquitectura correspondería según lo comentado anteriormente al modelo de implementación de Big Data Revolucionario, en la siguiente figura se muestra la arquitectura propuesta por Pentaho.

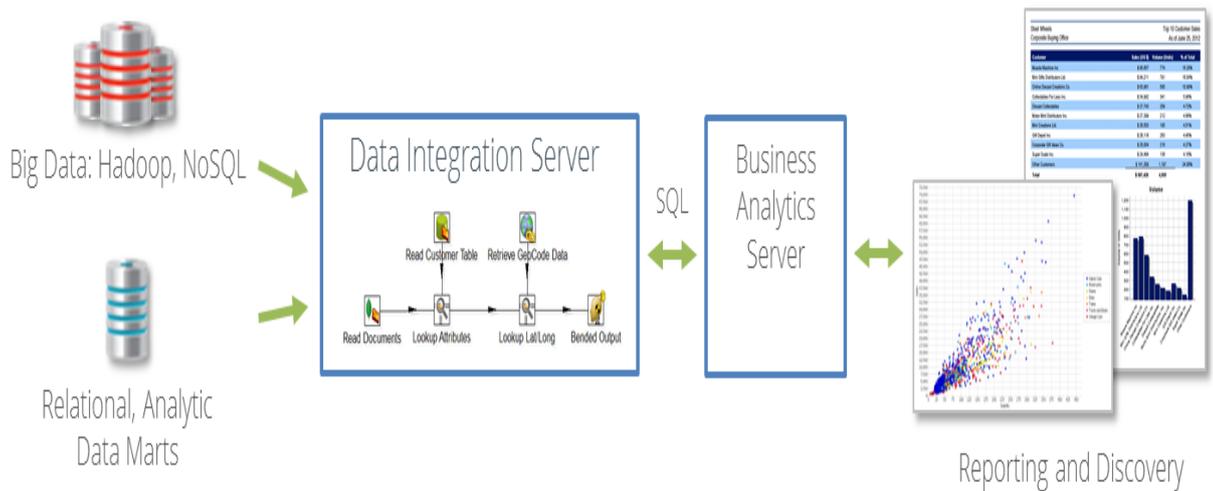


Figura 16: Arquitectura Big Data propuesta por Pentaho.

Fuente: (Pentaho, 2013)

2.5. METODOLOGIAS DE DESARROLLO DE INTELIGENCIA DE NEGOCIOS.

En la presente sección se estudian las metodologías más reconocidas de Datawarehouse y Big Data como:

2.5.1. METODOLOGÍA DE RALPH KIMBALL.

Según los autores (Kimball & Ross, 2010) en el libro *“The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence”* mencionan que esta metodología de desarrollo presenta las siguientes fases:

1. Planeación y administración del Proyecto

En esta primera fase se determina la preparación de la empresa para un proyecto de Data Warehouse, se desarrolla el enfoque preliminar, la justificación del negocio y las evaluaciones de factibilidad.

2. Definición de los Requerimientos del Negocio

Para esta fase es importante tener en cuenta que un factor determinante en el éxito de un proceso de Data

Warehouse es la interpretación correcta de los diferentes niveles de requerimientos expresados por los distintos grupos de usuarios.

3. Modelado Dimensional

La definición de los requerimientos del negocio determina los datos necesarios para cumplir los requerimientos analíticos de los usuarios. Diseñar los modelos de datos para soportar estos análisis requiere un enfoque diferente al usado en los sistemas operacionales.

4. Diseño Físico

El diseño físico de la base de datos se focaliza sobre la selección de las estructuras necesarias para soportar el diseño lógico. Un elemento principal de este proceso es la definición de estándares del entorno de la base de datos.

5. Diseño y Desarrollo de la Presentación de Datos

Esta etapa es típicamente la más subestimada de las tareas en un proyecto de Data Warehouse. Las principales actividades de esta fase del ciclo de vida son: la extracción, la transformación y la carga (ETL). Se definen como procesos de extracción aquellos requeridos para obtener los datos que permitirán efectuar la carga del Modelo Físico diseñado.

6. Diseño de la Arquitectura Técnica

Los entornos de Data Warehouse requieren la integración de numerosas tecnologías. Se deben tener en cuenta tres factores: los requerimientos del negocio, los actuales entornos técnicos y las directrices técnicas y estratégicas futuras planificadas por la compañía para poder establecer el diseño de la arquitectura técnica del entorno de Data Warehouse.

7. Selección de Productos e Instalación

Utilizando el diseño de arquitectura técnica como marco es necesario evaluar y seleccionar los componentes específicos de la arquitectura, como la plataforma de hardware, el motor de base de datos, la herramienta de ETL, las herramientas de acceso, etc.

8. Especificación de Aplicaciones para Usuarios Finales

No todos los usuarios del Data Warehouse necesitan el mismo nivel de análisis. Es por ello que en esta etapa se identifican los roles o perfiles de usuarios para los diferentes tipos de aplicaciones necesarias en base al alcance de los perfiles detectados (gerencial, analista del negocio, vendedor, etc.)

9. Desarrollo de Aplicaciones para Usuarios Finales

A continuación de la especificación de las aplicaciones para usuarios finales, el desarrollo de las aplicaciones de los usuarios finales involucra configuraciones de los metadatos y construcción de reportes específicos.

10. Implementación

La implementación representa la convergencia de la tecnología, los datos y las aplicaciones de usuarios finales accesibles para el usuario del negocio.

11. Mantenimiento y crecimiento

Como se remarca siempre, la creación de un Data Warehouse es un proceso (de etapas bien definidas, con comienzo y fin, pero de naturaleza espiral) que acompaña a la evolución de la organización durante toda su historia. Se necesita continuar con las actualizaciones de forma constante para poder seguir la evolución de las metas por conseguir.

12. Gestión del Proyecto

La gestión del proyecto asegura que las actividades del ciclo de vida se lleven a cabo de manera sincronizada.

En la siguiente figura se muestra el esquema de esta metodología (ciclo de vida) propuesta por Kimball para la construcción de una solución de DW/BI (Datawarehouse/Business Intelligence).

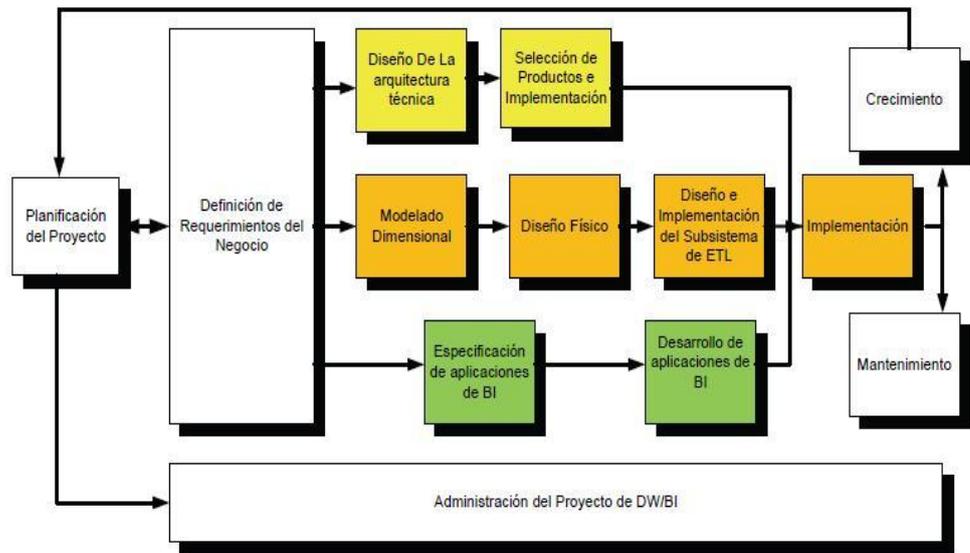


Figura 17: Etapas de la metodología de Kimball, denominada Business Dimensional Lifecycle.

Fuente: (Kimball & Ross, 2010).

El utilizar esta metodología brinda beneficios para el desarrollo de una Solución de Inteligencia de Negocios ya que parte por el desarrollo de los Data Marts, para satisfacer las necesidades específicas de un departamento o área dentro de la empresa, permitiendo así un mejor control de la información que se está abarcando, priorizando algunos procesos específicos del negocio, y que además la unión de estos, establecen el Data Warehouse, lo cual permite obtener implementaciones más rápidas, y menores riesgos, porque

hay menos dependencia entre áreas. A esta metodología también se le conoce como Bottom-Up.

2.5.2. METODOLOGÍA BILL INMON.

Según los autores (Inmon, Strauss, & Neushloss, 2010) en el libro *“DW 2.0: Architecture for The Next Generation of Data Warehousing”* mencionan la necesidad de transferir la información de los diferentes OLTP (Sistemas Transaccionales) de las organizaciones a un lugar centralizado donde los datos puedan ser utilizados para el análisis. Insiste además en que ha de tener las siguientes características:

- **Orientado a temas:** Los datos en la base de datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí.
- **Integrado:** La base de datos contiene los datos de todos los sistemas operacionales de la organización, y dichos datos deben ser consistentes.
- **No volátil:** La información no se modifica ni se elimina, una vez almacenado un dato, éste se convierte en información de sólo lectura, y se mantiene para futuras consultas.
- **Variante en el tiempo:** Los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones.

La información debe estar a los máximos niveles de detalle. Los Data WareHouses departamentales o Data Marts son tratadas como subconjuntos de este Data WareHouse

corporativo, son construidos para cubrir las necesidades individuales de análisis de cada departamento, y siempre a partir del Data Warehouse central.



Figura 18: Bill Inmon: Data Warehouse Corporativo.

Fuente: (Inmon, Strauss, & Neushloss, 2010).

La metodología Inmon también se referencia normalmente como Top-down. Los datos son extraídos de los sistemas operacionales por los procesos ETL y cargados en las stage área, donde son validados y consolidados en el Data Warehouse corporativo, donde además existen los llamados metadatos que documentan de una forma clara y precisa el contenido del Data Warehouse. Una vez realizado este proceso, los procesos de refresco de los Data Mart departamentales obtienen la información de este, y con las consiguientes transformaciones, organizan los datos en las estructuras particulares requeridas por cada uno de ellos, refrescando su contenido.

Al tener este enfoque global, es más difícil de desarrollar en un proyecto sencillo (pues se intentara abordar el “todo”, a partir del cual luego se ira al “detalle”).

Esta metodología conocida como el Business Intelligence Roadmap, incluye las fases Justificación, Análisis del Negocio, Diseño, Construcción, Despliegue y las siguientes actividades:

1. Evaluación del negocio.
2. Definición de requerimientos.
3. Análisis de datos.
4. Prototipo de aplicación.
5. Diseño de la base de datos.
6. Diseño del ETL.
7. Desarrollo del ETL.
8. Desarrollo de la aplicación.
9. Certificación.
10. Implementación.
11. Evaluación post producción.

En la siguiente figura se muestra las diferentes fases y actividades de esta metodología.

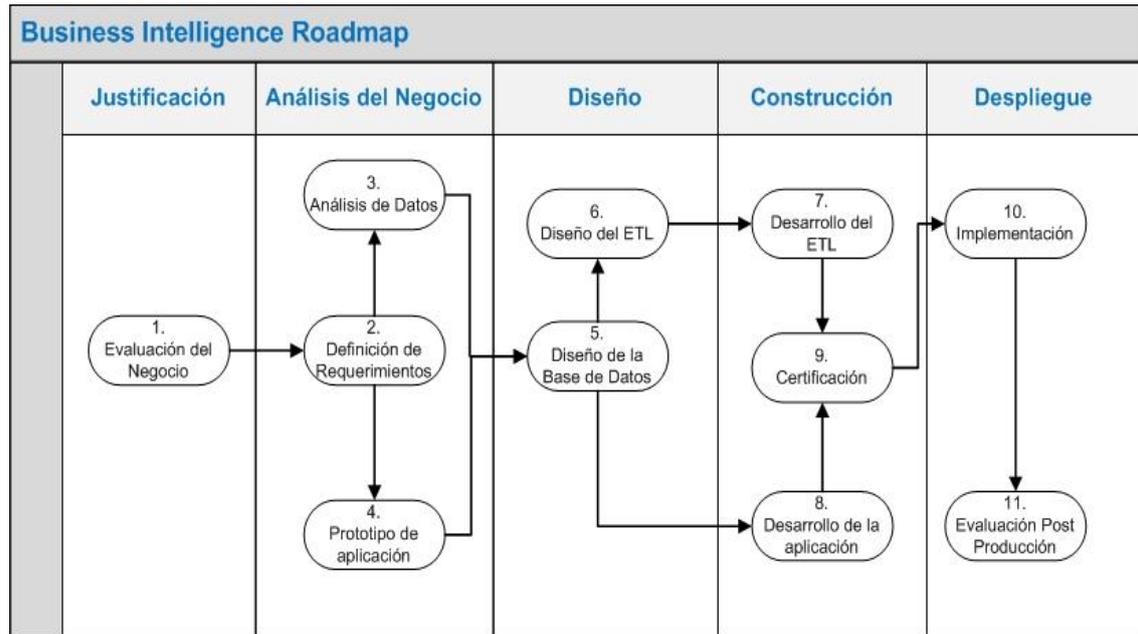


Figura 19: RoadMap BI.

Fuente: (Inmon, Strauss, & Neushloss, 2010).

Esta metodología muestra que tiene un mayor alcance porque aplica un enfoque global, ya que indica que debe partir primero por el desarrollo del Data Warehouse y de éste generar los Data Marts por lo tanto requiere de un esfuerzo general por parte de la organización y de más tiempo porque se necesita un mayor esfuerzo corporativo y de un conocimiento más profundo del negocio.

2.5.3. METODOLOGÍA DE JOSEP CURTO.

Según los autores (Curto Díaz & Conesa Caralt, 2010) en el libro “Introducción al Business Intelligence” mencionan que las fases de un proyecto de inteligencia de negocio son las siguientes:

- a) Análisis y requerimientos
- b) Modelización
- c) Desarrollo

- d) Producción
- e) Formación y Documentación

En la siguiente figura se muestra el esquema de esta metodología propuesta por Curto.

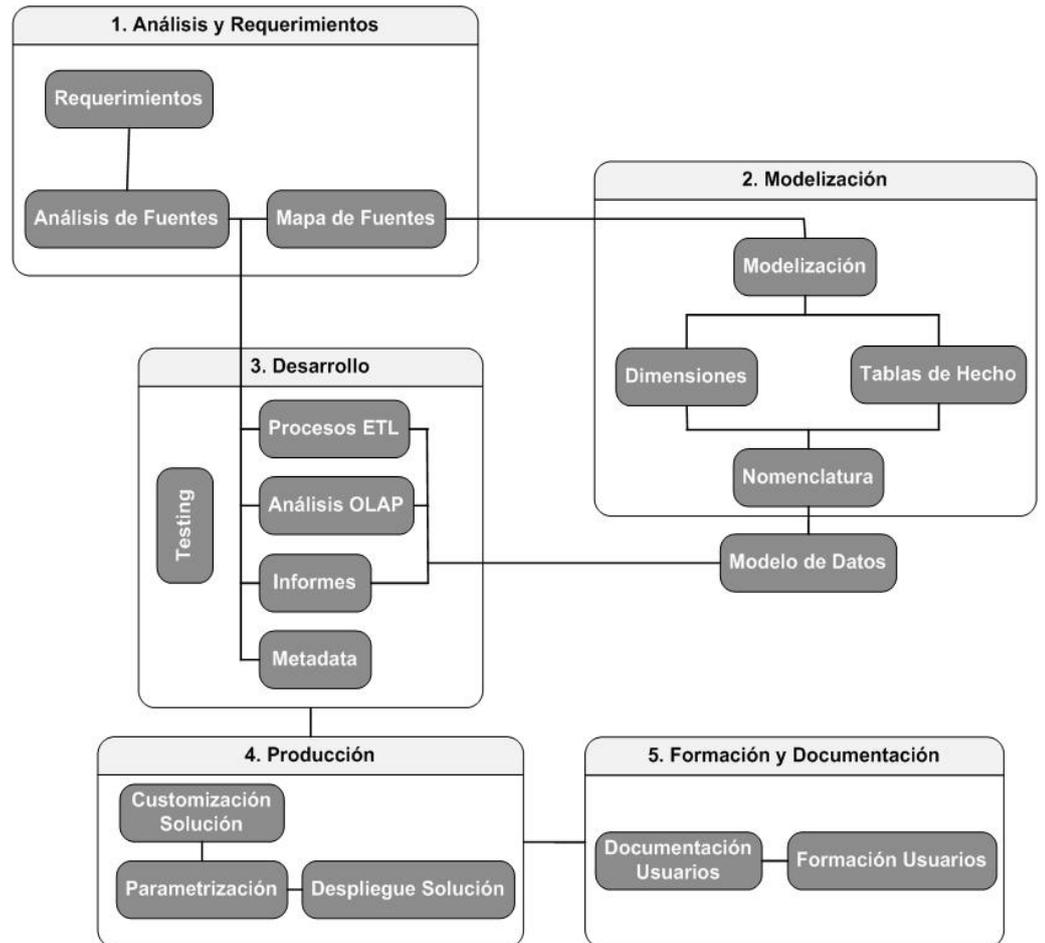


Figura 20: Fases de un Proyecto BI.

Fuente: (Curto Díaz & Conesa Caralt, 2010).

El enfoque de la metodología que propone Josep curto es el de la metodología ya consolidada en múltiples proyectos y sobre la que se sustentan todas las evoluciones actuales.

En esta metodología se muestra una principal ventaja que es la forma de desarrollar una aplicación de inteligencia de negocios la cual es bastante práctica, es decir está orientado a la aplicación inmediata y al software libre.

2.5.4. METODOLOGÍA DE BIG DATA.

Según The Computing Research Association (Association, 2012), en el paper “*Challenges and Opportunities with Big Data*”, menciona que existen cinco fases para el desarrollo de Big Data, en la siguiente figura se muestra las fases en el análisis de grandes volúmenes de datos las cuales se muestran en la parte superior del flujo y en la parte inferior se encuentran las necesidades de Big Data que hacen a estas tareas desafiantes.

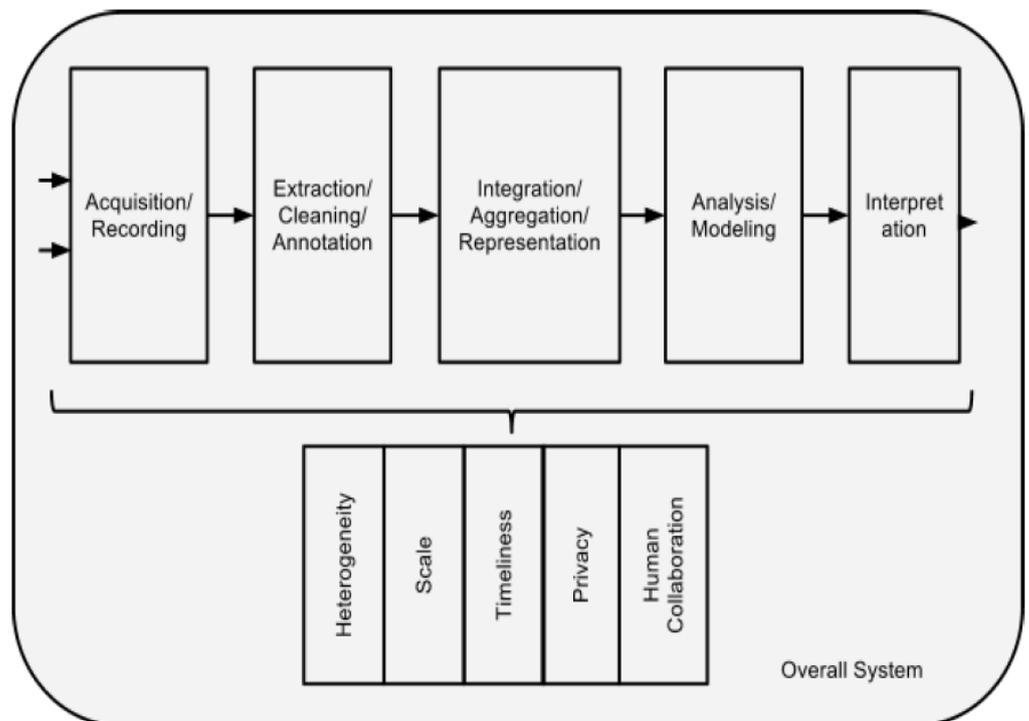


Figura 21: Fases de un Proyecto Big Data.

Fuente: (Association, 2012)

Fases

A continuación se explicará cada una de las fases de Big Data:

- 1. Adquisición de datos y grabación.** Es la primera fase de cualquier solución Big Data. Sin datos que consumir no tendríamos problema que tratar. Para ello es necesario definir qué datos vamos a necesitar desde nuestra fuente de información, como puede ser los de nuestras aplicaciones, proveniente de una red social o a través de simples ficheros de datos. Una vez definidos estos datos necesitaremos grabarlos en nuestros sistemas para tratarlos en fases posteriores.
- 2. Extracción y pre procesamiento de la información.** Una vez adquiridos los datos a tratar se necesita una segunda fase durante la cual debemos de estructurar estos datos para que sean posteriormente procesados y analizados. Para ello debemos 'limpiar' los datos recogidos durante la primera fase de manera que se obtenga un subconjunto estructurado de los mismos con la información relevante a tratar.
- 3. Representación, agregación e integración de datos.** En la fase anterior se ha hablado de estructurar los datos como algo necesario en cualquier solución Big Data, esto es así, pero el siguiente paso es encontrar las relaciones entre los distintos datos que conforman el conjunto de la información que tenemos disponible en nuestro problema. Previamente a realizar el análisis de los datos se necesitará desde almacenar nuestros datos estructurados en bases de datos o sistemas que se utilizarán en la siguiente fase.
- 4. Procesamiento de peticiones, modelado de datos y análisis.** Esta área se encarga de una multitud de tareas,

como manipular, procesar, modelar, analizar y extraer la información que se necesite dado un problema determinado.

5. Interpretación de los datos. Una vez extraída la información que se haya creído relevante de fases anteriores se necesita interpretar estos datos. La interpretación de estos datos dará el resultado final del problema.

En esta metodología se muestra una principal ventaja que es el manejo de grandes volúmenes de datos estructurados y no estructurados y realizar un análisis más profundo de la información para la toma de decisiones y de esta forma generar una ventaja competitiva en la organizaciones.

2.6. HERRAMIENTAS DE INTELIGENCIA DE NEGOCIOS.

Las herramientas de inteligencia de negocios están categorizadas por alguna función en particular por ejemplo:

- Optimizadores de consultas
- Extracción
- Reportes
- Tablero de mandos
- Servidores OLAP
- Optimizadores de servidores
- Herramientas de publicación
- Suits integradas

A continuación en la siguiente tabla se muestra las soluciones BI más representativas tanto para software libre como Propietario:

Open Source	Propietario
Pentaho	Information Builder

JasperSoft	Cognos IBM
LucidDB	SAP
SpagoBI	Microstrategy
OpenReports	Oracle
BeeProject	Panorama
OpenI	Apesoft
MonetCB	Actuate
Ingres	SAS –minería de datos
Infobright	Delta miner
PMML	Microsoft BI
Rapid miner	
Octopus	
Palo BI	
Joost	
BIRT	
Xineo	

Tabla 1: Soluciones BI de software libre y propietario.

2.6.1. Herramientas De Inteligencia De Negocio De Software

Libre (Open Source)

Según los autores (Vega Torres, Rojas Díaz, & Placeres Villar, 2009) en el libro “La inteligencia de negocio. Su implementación mediante la plataforma Pentaho” menciona que la comunidad “Open Source” hoy en día ha incursionado prácticamente en todas las áreas de la informática y existen algunas donde su supremacía es indudable. En el área de la Inteligencia de Negocio también se ha producido un despegue en el desarrollo de soluciones.

Una solución basada en Inteligencia de Negocio, de acuerdo a sus características debe de estar soportada por un conjunto de herramientas donde se establece una cooperación entre ellas para transitar por las diferentes etapas del proceso de análisis de los datos, desde la adquisición hasta la visualización de los resultados.

Entre las principales herramientas disponibles están:

- **Herramientas ETL:** Kettle, Clover, Enhydra Octopus, etc.
- **Desarrollo OLAP:** Mondrian y JPivot.
- **Minería de Datos:** WEKA, YALE y otras herramientas con versiones libres limitadas como: Tiberius, WizWhy, CART y See5 / C5.0.
- **Motores de Reportes y Gráficos:** JFreeReport, BIRT, JasperReport y JFreeChart.
- **Entorno de desarrollo para Cuadros de Mando (Dashboards):** JetSpeed y JBoss Portal.
- **Gestores de Bases de Datos:** MySQL y PostgreSQL.
- **Soluciones completas:** Pentaho y SpagoBI.

2.6.1.1 Pentaho BI

Según los autores (Bouman & Dongen, 2010) en el libro *“Pentaho Solutions: Business Intelligence and Data Warehousing with Pentaho and mysql”* menciona que Pentaho es una completa herramienta Business Intelligence de código abierto que permite construir almacenes de datos y aplicaciones ricas y poderosas de BI a una fracción del costo de una solución propietaria.

Pentaho ofrece una amplia gama de herramientas orientadas a la integración de información y al análisis inteligente de los datos de una organización. Cuenta con potentes capacidades para la gestión de procesos ETL (extracción, transformación y carga de datos), informes interactivos, análisis multidimensionales de información (OLAP). En la siguiente figura se muestra el esquema de la

arquitectura estructurada de los diferentes componentes que forman parte de Pentaho.

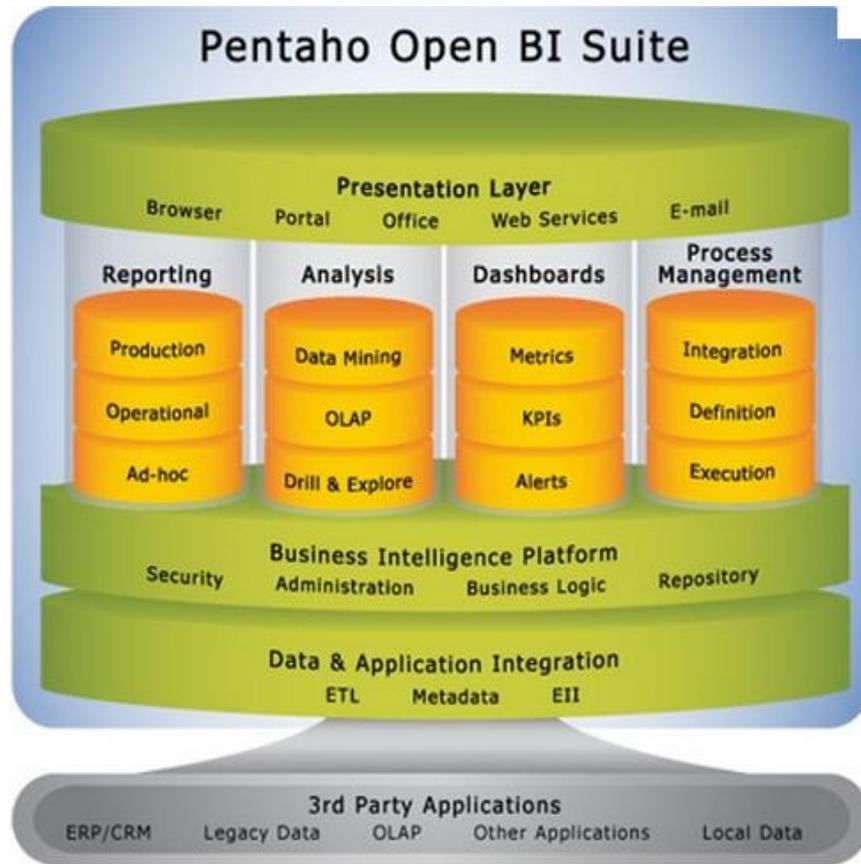


Figura 22: Arquitectura de componentes Pentaho.

Fuente: (Bouman & Dongen, 2010).

Todos estos servicios están integrados en una plataforma web, en la que el usuario puede consultar la información de una manera fácil e intuitiva. Los módulos incluidos por Pentaho BI, pueden utilizarse de manera conjunta o de forma separada según las necesidades de la organización. A continuación se detalla algunos de los módulos que conforman la suit de inteligencia de negocios Pentaho:

a) Pentaho Data Integration

Transforma e integra datos entre sistemas de información existentes y los Data Marts que compondrán el sistema BI. Estas son algunas de sus características:

- Entorno gráfico de desarrollo.
- Uso de tecnologías estándar: Java, XML, JavaScript.
- Fácil de instalar y configurar.
- Multiplataforma.
- Basado en dos tipos de objetos: Transformaciones (colección de pasos en un proceso ETL) y Trabajos (colección de transformaciones).
- Herramientas para Big Data.
- Incluye cuatro herramientas:
 - ✓ SPOON: para diseñar transformaciones ETL usando un entorno gráfico.
 - ✓ PAN: para ejecutar transformaciones diseñadas con Spoon.
 - ✓ CHEF: para crear trabajos.
 - ✓ KITCHEN: para ejecutar trabajos.

b) Pentaho Reporting

Obtiene y muestra informes de los indicadores de la organización. Pueden ser confeccionados por un usuario final, o estar predefinidos para una consulta directa. Estos informes pueden ser visualizados en formatos estándar como html, pdf o excel. Algunas de sus características son:

- Proporciona funcionalidad crítica para usuarios finales como:
 - ✓ Acceso vía web
 - ✓ Informes parametrizados.
 - ✓ Scheduling.
 - ✓ Suscripciones.
 - ✓ Distribución (bursting).

- Proporciona ventajas a expertos en informes:
 - ✓ Acceso a fuentes de datos heterogéneos.
 - ✓ Capacidad de integración en aplicaciones o portales.
 - ✓ Definición modular de informes (presentación y consulta).

- Diseño de informes flexible
 - ✓ Entorno de diseño gráfico.
 - ✓ Capacidad de uso de templates.
 - ✓ Acceso a datos relacionales, OLAP y XML.

2.6.1.2 Versiones de Plataforma Pentaho Open Source

BI

Existen dos versiones disponibles de Pentaho:

- **Pentaho Community Edition**
Completamente gratuita.
- **Pentaho Enterprise Edition**
Se ofrece bajo modalidad de suscripción.

Pentaho Community Edition

Según la información obtenida en (Pentaho Community, 2013), las características que brinda esta edición de Pentaho son las siguientes:

- ✓ Versión comunitaria, sin costos de licencia, sin servicios de soporte asociados.
- ✓ Es una suite completa con todas las funcionalidades necesarias para el correcto desarrollo de proyectos de BI.
- ✓ Pentaho soló la recomienda para:
 - Aficionados al BI.
 - Académicos.
 - Programadores de Software Abierto.
- ✓ Es totalmente software de código abierto, la licencia en su mayoría bajo la versión GNU.

Pentaho Enterprise Edition

Según la información obtenida en (Pentaho, 2013), las características que brinda esta edición de Pentaho son las siguientes:

- ✓ Versión empresarial, sin costos de licencia.
- ✓ Tiene asociados servicios de soporte y mantención que se pagan a través de una suscripción anual.
- ✓ Es una versión certificada que además posee algunas funcionalidades mejoradas para la consola de administración, y para la construcción de tableros de control.

Comunity o Enterprise

En el contexto Open Source, nos encontramos que existe una versión open source comunitaria con un conjunto de funcionalidades bien definidas, acotadas y plenamente operativas, y una versión profesional que presenta o bien más funcionalidades que la otra versión o bien una versión mejorada de las mismas funcionalidades. Pentaho soporta la creación de Dashboards, pero sólo la versión profesional presenta un Dashboard Designer Ad-hoc.

Pentaho recomienda la versión Enterprise para entornos empresariales en donde se gestione grandes volúmenes de datos a analizar, y la versión Community es utilizado mayormente para trabajos académicos, en donde solo es necesario trabajar con un prototipo.

En la siguiente figura se muestra las diferencias entre ambas versiones:

Software and Services	Community Edition	Enterprise Edition
Reporting	Open Source	Certified
Analysis	Open Source	Certified
Dashboards	Open Source	Certified
Data Integration / ETL	Open Source	Certified
Business Intelligence Platform	Open Source	Certified
Data Mining	Open Source	Certified (Add-On)
Community Forums Interaction	✓	✓
Community Web Documentation (wiki)	✓	✓
Professional Support		
• Telephone support (toll-free)		✓
• E-mail support		✓
• Service Level Agreement		✓
• Unlimited support cases		✓
Software Maintenance		
• Software maintenance	By in-house staff	✓ By Pentaho Engineers
• Patch releases		✓
• Fixes included in future releases		✓
Enhanced Functionality		
• Pentaho Enterprise Console		✓
• Single Sign-On		✓
• Streamlined security configuration		✓
• Application diagnostics		✓
• Repository utilities		✓
• Lifecycle management		✓
• Audit reports		✓
• Automated content expiration		✓
• Clustering		✓
• Performance monitoring		✓
• ETL management and monitoring		✓
Certified Software		
• Stabilized software		✓
• Managed release cycle		✓
• Optimized builds		✓
Product Expertise		
• Professional documentation		✓
• Knowledge base		✓
• Consultative support		✓
• Remote assistance packages		✓
• Installation/configuration packages		✓
• Design and integration packages		✓
• Troubleshooting and optimization packages		✓
• Enterprise Edition online forum		✓
• Web based training		✓
Software Assurance		
• Intellectual Property Indemnification		✓
• Warranty for services		✓

Figura 23: Comparativa de Pentaho Enterprise y Comunidad.

Fuente: (Pentaho, 2013).

Pentaho Enterprise proporciona software bajo una licencia de suscripción, donde el factor principal es el número de usuarios, dependiendo de ello los precios varían, generalmente Pentaho EE proporciona paquetes destinados a un número establecido de usuarios o también se puede realizar la suscripción a un solo usuario, por ejemplo para la configuración de pequeñas empresas se utiliza mayormente el nivel "Silver", que impone una restricción de no más de 25 usuarios.

Precio de soportes:

- Pentaho EE BI Suite "Silver", 25 Usuarios (Por año) \$10.000,00
- Pentaho EE BI Suite "Gold", 6 CPU para multicore (Por año) \$29.000,00

En la siguiente figura se muestra los precios para el nivel silver.

User type	Product Component Description	Units	Software Price	Support Price	First Year Cost	3 Year Total
Support	Pentaho Silver edition	1	\$ -	\$ 10,000	\$ 10,000	\$ 30,000
Admin / IT	<i>covered in line 1</i>	1	\$ -	\$ -	\$ -	\$ -
Professional	<i>covered in line 1</i>	1	\$ -	\$ -	\$ -	\$ -
Expert user	<i>covered in line 1</i>	3	\$ -	\$ -	\$ -	\$ -
Basic user	<i>covered in line 1</i>	20	\$ -	\$ -	\$ -	\$ -
Dev environment	<i>no charge</i>	1	\$ -	\$ -	\$ -	\$ -
Test environment		0	\$ -	\$ -	\$ -	\$ -
Totals					\$ 10,000	\$ 30,000

Figura 24: Detalle de Precios de Pentaho BI para el nivel Silver

Fuente: (Pentaho, 2013).

En la siguiente tabla se muestra una comparativa a nivel de costos de licencia entre BI comercial y Open Source.

Diferencias	ORACLE BISE 1	SQL SERVER 2008 R2	PENTAHO
País Creación	Estados Unidos	Estados Unidos	Orlando
Fabricado por	Oracle Corporation	Microsoft	Pentaho
Licencia	-	-	GNU General Public License versión 2 (GPLv2).
Costo	Min 5 Usuario Por usuario \$ 180.00 Por procesador (Usuario ilimitados) \$ 5800.00	Por Procesador Datacenter Server SQL * Licencia \$ 54,990.00 Garantía de Software \$ 13,748.00 SQL Server Enterprise Licencia \$ 27,495.00 Garantía de Software \$ 6,874.00 Estándar SQL Servidor Licencia \$ 7,171.00 Garantía de Software \$ 1,793.00 licencias de acceso de cliente adicionales (CAL) SQL Server Enterprise Licencia \$ 8,592.00 Garantía de Software \$ 2,148.00 Estándar SQL Servidor	Gratis: Versión Community Edition . Enterprise Edition proporciona una mayor funcionalidad del software, amplio soporte técnico profesional, experiencia en productos, software certificado y mantenimiento de software, y más. Sin Costo de licencia en las dos Versiones. Solo suscripción anual dependiendo del Número de

		Licencia \$ 898.00 Garantía de Software \$ 224.00	Usuario. \$ 10. 000 por el Soporte para el nivel Silver y \$ 174 por usuario.
Lenguaje de desarrollo	JAVA	-	JAVA J2EE
Año de lanzamiento	2007	2010	2004
Plataforma Windows, Linux	Multiplataforma	Windows	Multiplataforma (tanto a nivel de cliente como servidor): Mac, linux/unix y Windows

Tabla 2: Diferencias De Bi Comercial Frente A Open Source Pentaho.

Fuente: (Pentaho, 2013), (Oracle, 2013), (Microsoft, 2013).

En la siguiente tabla se muestra el costo promedio por cada usuario, dependiendo del tamaño de la organización.

Soporte Profesional	Pequeño	Mediano	Grande
Precio por un único usuario.	\$174	\$870	\$1200
Total usuarios	25	100	500
Total administradores	1	2	3
Desarrolladores profesionales	1	3	15
Usuarios expertos	2	10	50
Usuarios básicos	21	85	432

Tabla 3: Detalle de costos por tipo de usuario.

Fuente: (Pentaho, 2013).

En la siguiente figura se muestra el costo promedio para un único usuario para los tres diferentes tipos de organización.

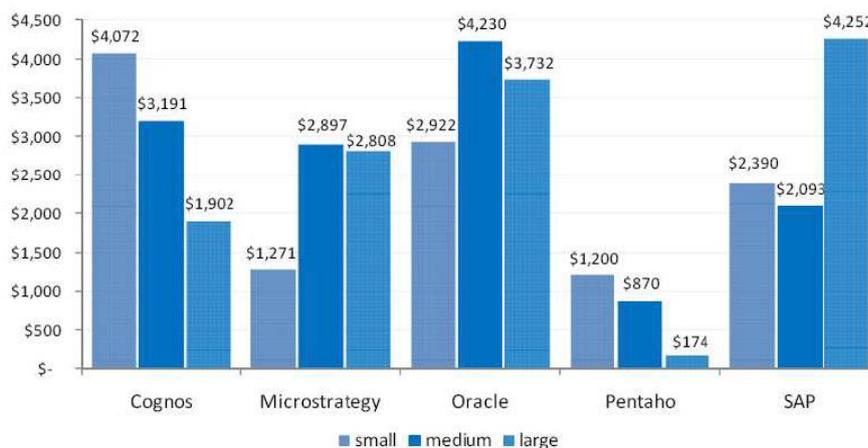


Figura 25: El costo promedio por usuario, por el proveedor y el escenario.

Fuente: (Pentaho, 2013).

Como podemos apreciar Oracle sigue siendo la herramienta más cara en la actualidad.

Servicios incluidos en la suscripción:

- ✓ Diseño, desarrollo o asistencia con cualquier componente del Pentaho Business Analytics platform.
- ✓ Asistencia en la configuración o ayuda resolviendo un cuello de botella con el desempeño.
- ✓ Asistencia para cambiar de una versión a otra.

- ✓ Soporte rápido en momentos críticos para aplicaciones críticas en los negocios.
- ✓ Alertas acerca problemas en el sistema o parches existentes en temas de seguridad. Tips para optimizar y más.

2.6.2. Herramientas Open Source para BIG DATA

Big Data no se debe a los datos en sí, sino al crecimiento exponencial que estos están teniendo, es por esto que se han desarrollado herramientas que permitan obtener el mejor conocimiento de esta gran cantidad de información (DatosIntensos, 2012).

Estas son algunas de las herramientas más utilizadas en el mundo Open Source para el manejo de BIG DATA:

Plataformas y herramientas para el análisis:

- ✓ MapReduce: Un modelo de programación que hace honor a la frase célebre de Julio César "divide y vencerás", y que permite utilizar el procesamiento paralelo de datos en computadoras distribuidas.
- ✓ Hadoop: Simplemente no se habla de Big Data sin mencionar a Hadoop.
- ✓ Storm: Conocido como el Hadoop en tiempo real.
- ✓ Pentaho: Una de las herramientas más utilizadas para la comunidad Open Source y que se integra eficientemente con Big Data.

Bases de datos y Datawarehouse:

- ✓ HBASE: Es el sistema de almacenamiento no relacional para Hadoop.
- ✓ CASSANDRA: Otro sistema de almacenamiento NoSQL desarrollado originalmente por Facebook.

- ✓ MongoDB: Una base de datos documental NoSQL.
- ✓ Neo4j: El sistema líder de bases de datos de grafos.
- ✓ Riak: Declara ser el mejor sistema en producción de bases de datos distribuida Open Source.
- ✓ HyPertable: Otro sistema de almacenamiento NoSQL.
- ✓ Hive: Es el Datawarehouse de Hadoop.

Sistemas de archivos:

- ✓ Hadoop Distributed File System: El primer sistema de almacenamiento para Hadoop.
- ✓ Gluster: Un sistema de archivo que permite ir más allá de las limitaciones del sistema de archivos de hadoop, tiene la capacidad de manejar hasta 72 Brontobytes.

Leguajes de programación.

- ✓ Pig/Pig Latin: Un proyecto del ecosistema de Big Data de la fundación apache que utiliza un lenguaje textual el Pig Latin con la plataforma de análisis Pig.

2.6.2.1. Apache Hadoop

La historia de Hadoop está necesariamente unida a la de Google. De hecho, podría decirse que Hadoop nace en el momento en que Google precisa urgentemente de una solución que le permita continuar procesando datos que repentinamente estaban creciendo de forma exponencial. Google se ve incapaz de poder indexar la web al nivel que exige el mercado y por ello decide buscar una solución, que se basa en un sistema de archivos distribuidos, haciendo suyo el lema “divide y vencerás”. (PowerData, 2013).

Esta solución, que posteriormente se denominará Hadoop, se basa en un gran número de pequeños ordenadores, cada uno de los cuales se encarga de procesar una porción de información. La grandiosidad del sistema es que, a pesar de que cada uno de ellos funciona de forma independiente y autónoma, todos actúan en conjunto, como si fueran un solo ordenador de dimensiones increíbles. Hadoop permite almacenar y procesar grandes volúmenes de cualquier tipo de información lo que lo hace ideal para grandes volúmenes de datos no estructurados (Emails, tweets, logs, etc.). Sus ventajas son muchas:

- ✓ Aísla a los desarrolladores de todas las dificultades presentes en la programación paralela.
- ✓ Es capaz de ejecutar procesos en paralelo en todo momento.
- ✓ Dispone de módulos de control para la monitorización de los datos.
- ✓ Implica un costo relativamente bajo.
- ✓ Proporciona un rápido retorno de la inversión.
- ✓ Brinda la posibilidad de afrontar nuevos retos y dar solución a problemáticas que antes no se podían asumir, o que quedaban sin respuesta.

Hoy en día Hadoop es el proyecto estrella de Apache soportado por una gran comunidad de desarrollo de código abierto. Hadoop es un sistema

que se puede implementar sobre hardware a un costo relativamente bajo, siendo a su vez totalmente gratuito para software. Los componentes básicos de Hadoop son el HDFS y el MapReduce. Por lo tanto Hadoop es igual a decir:

Almacenamiento + procesamiento (consultas)
HDFS + Map/Reduce (Java)

2.6.2.2. Hadoop y su Arquitectura

El Big Data nace para solucionar estos problemas:

- ✓ Como almacenar y trabajar con grandes volúmenes de datos.
- ✓ Y la más importante, como poder interpretar y analizar estos datos, de naturaleza muy dispar.

Hadoop aparece en el mercado como una solución para estos problemas, dando una forma de almacenar y procesar estos datos.

Apache Hadoop es un framework que permite a través de su modelo simple de programación y su diseño, pasar de pocos nodos a miles de nodos de forma ágil. Hadoop es un sistema distribuido que usa una arquitectura Master-Slave, para almacenar su Hadoop Distributed File System (HDFS) y algoritmos de MapReduce para hacer cálculos. (DataPrix, 2013).

En la siguiente figura se muestra la arquitectura de Hadoop:

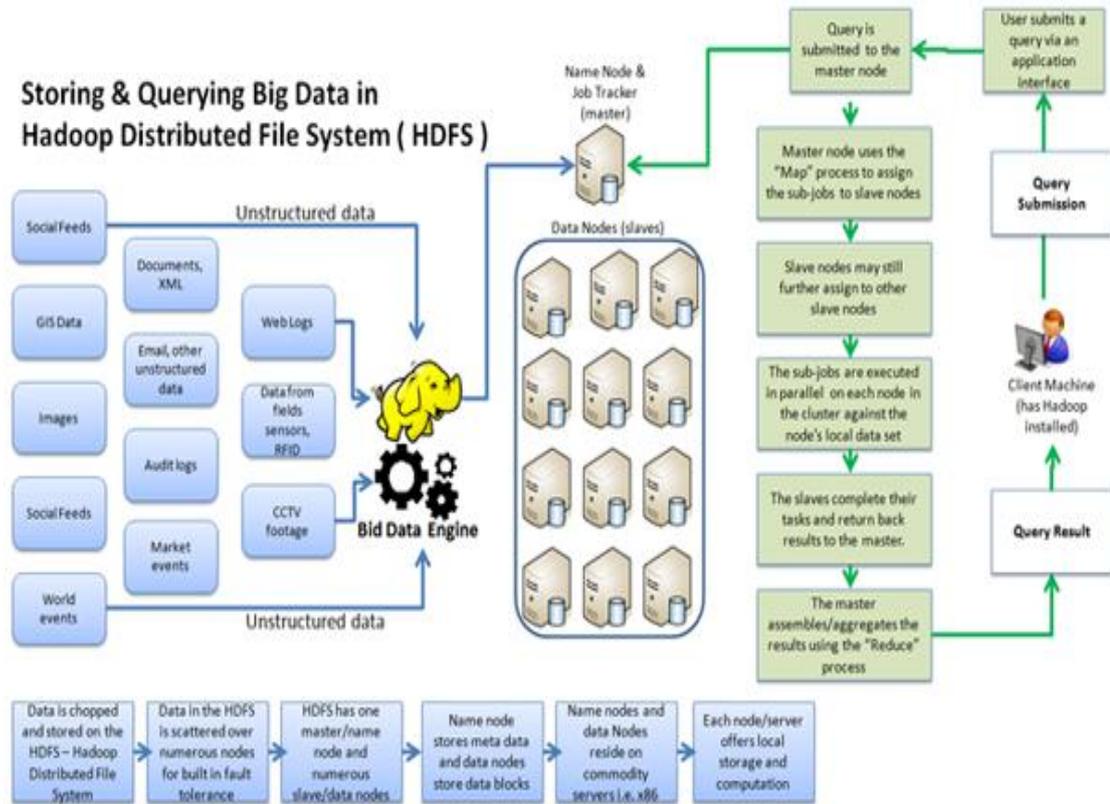


Figura 26: Arquitectura de Hadoop.

Fuente: (DataPrix, 2013).

2.6.2.3. HDFS

HDFS es el sistema de almacenamiento, es un sistema de ficheros distribuido. Fue creado a partir del Google File System (GFS). HDFS se encuentra optimizado para trabajar con grandes flujos y con ficheros grandes en sus lecturas y escrituras. La escalabilidad y disponibilidad son otras de sus claves, gracias a la replicación de los datos y tolerancia a los fallos (Jara, 2012).

HDFS (Hadoop Distributed File System), gestiona la información a procesar y el tráfico entre nodos. Este componente, escrito también en java, es el

encargado de repartir las unidades de trabajo a los nodos MapReduce. Este componente también trabaja a dos niveles, tiene un nodo principal que gestiona al resto, el namenode que envía los fragmentos de información a los nodos secundarios y guarda la información de qué datos tiene cada uno. Los nodos secundarios se llaman datanodes, y cada nodo de TaskTrackers tendrá un datanode que le de soporte. (White, 2012).

Además, existe un secondary namenode, que se encarga de sacar snapshots periódicos del namenode de modo que si hay una caída, sea posible recuperar el estado inmediatamente anterior. En la siguiente figura se muestra la arquitectura de HDFS:

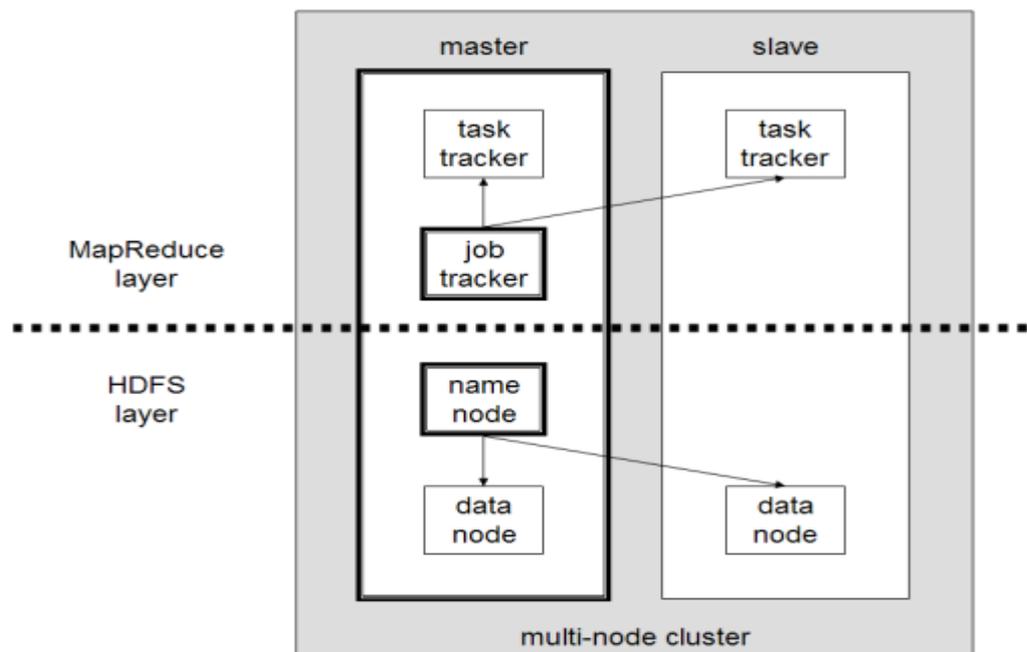


Figura 27: Arquitectura de HDFS.

Fuente: (losutron, 2013).

2.6.2.4. MapReduce

Se trata de un framework que hace posible aislar al programador de todas las tareas propias de la programación en paralelo. Es decir, permite que un programa que ha sido escrito en los lenguajes de programación más comunes, se pueda ejecutar en un cluster de Hadoop. (White, 2012).

MapReduce permite de una forma simple, paralelizar el trabajo sobre los grandes volúmenes de datos. El modelo de MapReduce simplifica el procesamiento en paralelo, abstrayéndonos de la complejidad que hay en los sistemas distribuidos. Básicamente las funciones Map transforman un conjunto de datos a un número de pares key/value. Cada uno de estos elementos se encontrará ordenado por su clave, y la función reduce es usada para combinar los valores (con la misma clave) en un mismo resultado. Un programa en MapReduce, se suele conocer como Job, la ejecución de un Job empieza cuando el cliente manda la configuración de Job al JobTracker, esta configuración especifica las funciones Map, Combine (shuttle) y Reduce, además de la entrada y salida de los datos. MapReduce Provee:

- ✓ Paralelismo automático
- ✓ Tolerancia a Fallos
- ✓ Monitoreo

2.6.2.5. HIVE

Apache Hive es un sistema de Data Warehouse para Hadoop, que facilita el uso de la agregación de los datos, ad-hoc queries, y el análisis de grandes datasets almacenados en Hadoop. Hive proporciona métodos de consulta de los datos usando un lenguaje parecido al SQL, llamado HiveQL.

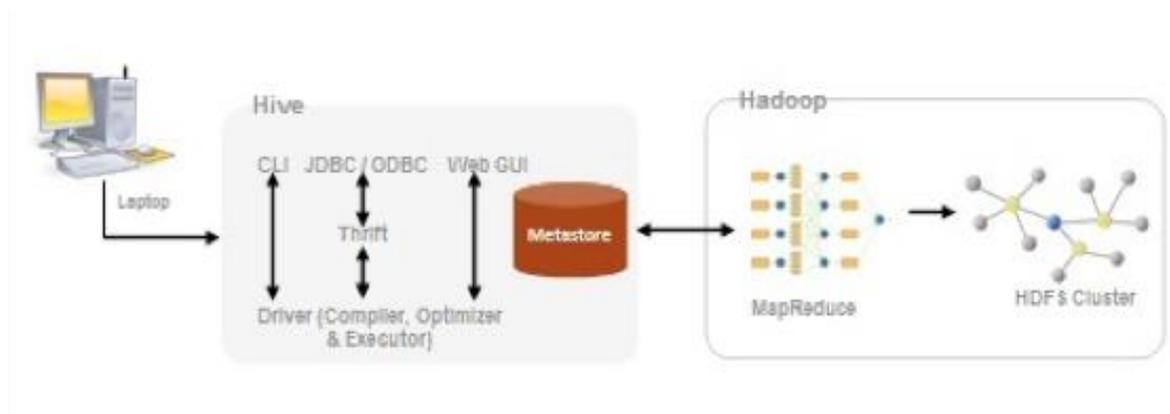


Figura 28: Arquitectura de HIVE.

Fuente: (DataPrix, 2013)

2.6.3. Pentaho y su integración con Apache Hadoop

Apache hadoop, es la tecnología de preferencia para empresas que necesitan reunir, almacenar y procesar efectivamente grandes cantidades de información estructurada y compleja. Aunque hadoop es muy poderoso, en su forma cruda, carece de interfaces fáciles de usar para análisis efectivos en tiempo y costo. (DataPrix, 2013).

Por su parte Pentaho reduce drásticamente el tiempo para diseñar, desarrollar e implementar soluciones de Big Data llegando a ser hasta 15 veces más rápido:

- ✓ Desarrollo fácil 100% visual, 0% programación Java.

- ✓ Procesos de análisis simplificado de punta a punta.
- ✓ Alta performance, incluyendo capacidad de ejecución dentro de Hadoop.
- ✓ Toda la solución (ETL, cubos, reportes) se conecta con las distribuciones más importantes de Hadoop y las principales bases NoSQL y Analíticas.
- ✓ Cargar/extraer archivos de Hadoop en un cluster local, en la nube, etc.
- ✓ Ejecutar desde PDI jobs de Map/Reduce.
- ✓ PDI está desarrollado en Java: puede ejecutarse dentro del cluster de Hadoop,

En la siguiente figura se muestra la integración de Hadoop con Pentaho:

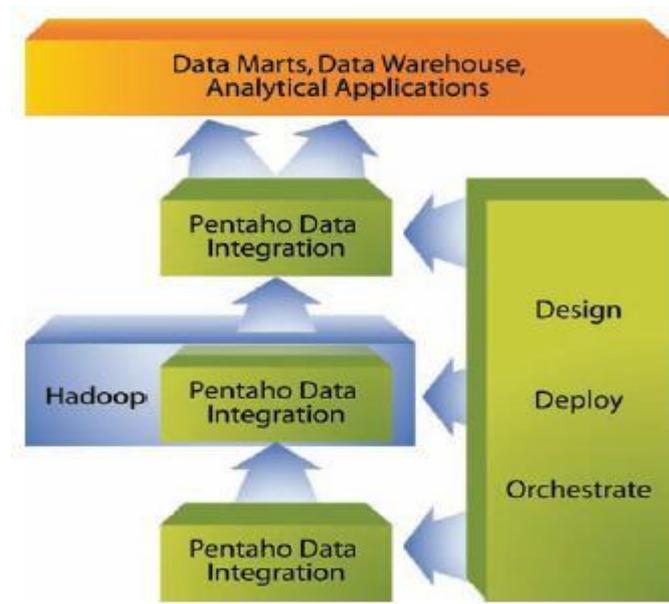


Figura 29: Hadoop y Pentaho.

Fuente: (DataPriX, 2013).

CAPITULO III: MATERIAL Y METODOS

En este capítulo se plantean las técnicas, métodos e instrumentos utilizados para desarrollar el modelo de inteligencia de negocios, definiendo una metodología a seguir.

3.1. MATERIAL Y MÉTODOS

3.1.1. Material

3.1.1.1. Población

En el presente trabajo de investigación la población de estudio son las áreas de la empresa las cuales son 8 (Compras, Contabilidad, Sistemas, Recursos Humanos, Almacén, Tesorería, ventas, producción).

3.1.1.2. Muestra

La muestra es solo una, el área de ventas.

3.1.1.3. Unidad de Análisis

La unidad de análisis es la toma de decisiones.

3.1.2. Método

3.1.2.1. Tipo de Estudio

No experimental – descriptivo; investigación Aplicada.

3.1.2.2. Diseño de Investigación

El método más adecuado para este tipo de investigación es el inductivo, se considera inductivo porque los resultados obtenidos en la muestra se pueden proyectar en la población sin mayor diferencia significativa.

a) Entrevistas.

- b) Observación.
- c) Modelo de Inteligencia de Negocios.
 - ✓ Ralph Kimball.
 - ✓ Josep Curto
 - ✓ Bill Inmon.
 - ✓ Big Data.
- d) Herramientas BI
 - ✓ Pentaho BI.
 - ✓ SpagoBi
 - ✓ Java – BIRT
 - ✓ Business Objects
 - ✓ Java – Clover
 - ✓ Java - Octopus
- e) Método de Validación del modelo BI.
 - ✓ Encuesta (Juicio de Expertos).

3.1.2.3. Variables y operativización de variables

a) Variable independiente.

- El modelo de inteligencia de negocios.

Indicadores:

- Adaptación del modelo.
- Tiempo de desarrollo.
- Costo de desarrollo.
- Nivel de calidad.

b) Variable dependiente.

- La toma de decisiones en el proceso de venta.

Indicadores:

- Tiempo de respuesta en formulación de propuestas por reunión.
- Índice de Tiempo de respuesta en el proceso de análisis por propuesta formulada.
- Numero de propuestas formuladas por reunión.

c) Operativización de variables

VARIABLE	DEFINICION CONCEPTUAL	DEFINICION OPERACIONAL	INDICADORES	INSTRUMENTO	FORMULA	UNIDADES DE MEDIDA
Modelo de inteligencia de negocios.	Es la combinación de tecnología, herramientas y procesos que permiten transformar datos almacenados en información, esta información en conocimiento y este conocimiento puede generar escenarios, pronósticos y reportes que apoyen a la toma de decisiones con la información correcta, en el momento y lugar correcto.	El modelo de inteligencia de negocios presenta los siguientes elementos operativos: a) Adaptación del modelo. b) Tiempo de desarrollo. c) Costo de desarrollo. d) Nivel de calidad del modelo de inteligencia de negocio.	Adaptación del modelo.	Encuesta.	-----	Escala de Liker.
			Tiempo de desarrollo.	Encuesta.	-----	Escala de Liker.
			Costo de desarrollo.	Encuesta.	-----	Escala de Liker.
			Nivel de calidad del modelo de inteligencia de negocio.	Encuesta.	-----	Escala de Liker.

Tabla 4: Operativización de la variable independiente.

VARIABLE	DEFINICION CONCEPTUAL	DEFINICION OPERACIONAL	INDICADORES	INSTRUMENTO	FORMULA	UNIDADES DE MEDIDA
Toma de decisiones en el proceso de venta.	Es el proceso que consiste en realizar una elección entre dos o más alternativas, cualquier toma de decisiones debería incluir un amplio conocimiento del problema ya que al analizarlo y comprenderlo, será posible darle una solución. La calidad de las decisiones tomadas marca la diferencia entre el éxito o el fracaso.	La toma de decisiones en el proceso de venta será medido tomando en cuenta los valores de los siguientes parámetros: a) Tiempo de respuesta en formulación de propuestas por reunión. b) Índice de tiempo de respuesta en el proceso de análisis por propuesta formulada. c) Numero de propuestas formuladas por reunión.	Tiempo de respuesta en formulación de propuestas por reunión.	Cronometro.	$RFP = \frac{(HIL + HID)}{NR}$	Hora.
			Índice de tiempo de respuesta en el proceso de análisis por propuesta formulada.	Actas de reunión.	$PAP = \frac{((HIP * DT) / JL) * 100}{}$	%
			Numero de propuestas formuladas por reunión.	Actas de reunión.	-----	Escalar.

Tabla 5: Operativización de la variable dependiente.

Leyenda:

- ✓ **RFP:** Tiempo de respuesta en formulación de propuestas por reunión.
- ✓ **HIL:** Horas invertidas en lanzamiento de ideas.
- ✓ **HID:** Horas invertidas en discusiones de las ideas.
- ✓ **NR:** Número de reuniones.
- ✓ **PAP:** Índice de tiempo de respuesta en el proceso de análisis por propuesta formulada.
- ✓ **HIP:** Horas invertidas en el proceso de análisis.
- ✓ **JL:** Jornada laboral.
- ✓ **DT:** Días trabajados

3.1.2.4. Instrumentos de recolección de datos

Las técnicas de recolección de información para la presente investigación son entrevistas y observación, análisis documental las cuales se utilizarán con el fin de recopilar los datos sobre una situación existente, cada una ayudará a asegurar una completa investigación.

a) Entrevistas

Las entrevistas se utilizan para recopilar información en forma verbal, a través de preguntas que propone el entrevistador. Quienes responden son las personas involucradas en la toma de decisiones los cuales serán los usuarios del modelo de inteligencia de negocios, los cuales respondieron a la encuesta mostrada en el ANEXO 07.

b) Observación

Es el registro visual de lo que ocurre en las reuniones de toma de decisiones, clasificando consignando los acontecimientos pertinentes de acuerdo con algún esquema previsto y según el problema que se estudia, la ficha de observación se muestra en el ANEXO 08.

c) Análisis documental

Es el proceso de lectura, síntesis y representación del contenido y la forma del texto de un documento para facilitar su consulta y recuperación, el análisis documental tiene como objetivo describir un documento en sus partes esenciales para su posterior identificación y recuperación. Con esta herramienta se obtendrá datos de la realidad de la organización por ejemplo planes estratégicos, informes, etc. Para ello se utilizara fichas de análisis documental el cual se muestra en el ANEXO 09.

3.1.2.5. Metodología y análisis estadístico de datos

Metodología

Para el desarrollo de la presente investigación se va a utilizar la metodología conformada de las siguientes etapas.

1. DESCRIPCION DEL PROCESO DE TOMA DE DECISIONES EN EL AREA DE VENTAS DE LA EMPRESA SAN ROQUE S.A.: Inicialmente se elaborara una entrevista de preguntas abiertas a los responsables de la toma de decisiones del área

de ventas, luego se aplicara un cuestionario en función de los datos obtenidos en la entrevista.

2. ANÁLISIS COMPARATIVO DE LAS METODOLOGIAS DE DESARROLLO DE INTELIGENCIA DE NEGOCIOS: Luego con ayuda del asesor se validaran las metodologías de Inteligencia de Negocio para posteriormente realizar la propuesta del modelo.
3. ELABORACIÓN DE LA PROPUESTA DEL MODELO DE INTELIGENCIA DE NEGOCIOS: Se realiza la propuesta en base a un marco metodológico que representa el nuevo modelo de inteligencia de negocios.
4. APLICACIÓN DEL MODELO PROPUESTO: Posteriormente se desarrollaran los artefactos de Inteligencia de Negocio aplicando la Suit Pentaho y los componentes de Big Data, se ejecuta el modelo propuesto en el área de ventas de la empresa San Roque S.A.
5. OBTENCIÓN DE LOS RESULTADOS: en base a la aplicación del modelo propuesto en el área de ventas de la empresa San Roque S.A.
6. DISCUSIÓN DE RESULTADOS: en base a estadística descriptiva sobre los indicadores del modelo de inteligencia de negocios propuesto. Se validara el modelo aplicándolo a un caso de estudio y a través de juicio de expertos.
7. CONCLUSIONES Y RECOMENDACIONES: de acuerdo a los resultados obtenidos en la etapa previa se redactan los puntos concluyentes y se plantean las recomendaciones necesarias.

Análisis estadístico de datos

Se tabularon, ponderaron e interpretaron todos los datos obtenidos, para lo cual se utilizó la herramienta informática Microsoft Excel 2010 y SPSS versión 20, los cuales permitieron evaluar el comportamiento de las variables y permitir indicar un análisis estadístico que fueron presentados en cuadros y gráficos descriptivos.

CAPITULO IV: PROPUESTA DEL MODELO DE INTELIGENCIA DE NEGOCIO

En el presente capítulo, el punto de partida para establecer la propuesta del modelo de inteligencia de negocios a desarrollar, es determinar cuál de las metodologías revisadas debemos seleccionar para que sirva de base para la propuesta metodológica a ser desplegada en el modelo, en tal sentido se realiza una explicación de cómo desarrollar cada una de las etapas que corresponden a la metodología propuesta.

4.1. ANÁLISIS COMPARATIVO DE METODOLOGÍAS.

Las metodologías de desarrollo de proyectos de inteligencia de negocios candidatas para realizar este proyecto son:

- ✓ Ralph Kimball
- ✓ Bill Innom
- ✓ Josep Curto
- ✓ Big Data

A continuación se realizara un análisis comparativo entre dichas metodologías.

4.1.1. Metodologías

Entre las mejores prácticas más conocidas para el desarrollo de una solución BI se tiene a “La Metodología de Ralph Kimball”, el “Bussines Intelligence Roadmap” por Bill Innom, y a la metodología de Josep Curto. Todas ellas orientan en los pasos a seguir en la construcción de un Data Warehouse o un Datamart.

Enfoque “Top-Down”

El BI Roadmap de Innom propone que un DWH organizacional puede hacer crecer los Data Marts departamentales, es decir establece que el desarrollo de una Solución de Inteligencia de Negocios debe partir por el desarrollo del Data Warehouse y de éste derivar los Data Marts. Esta perspectiva es conocida como “Top – Down”.

Enfoque “Bottom-Up”

La Metodología de Ralph Kimball, describe como realizar un desarrollo teniendo como premisa que Data Marts departamentales pueden constituir un DWH de toda la organización, es decir establece que el desarrollo de una solución de inteligencia de negocios debe partir por el desarrollo de los Data Marts y que el conjunto de estos constituyen el Data Warehouse. Esta perspectiva es conocida como “Bottom – up”.

Ventajas

a) Top-Down

- Requiere de un esfuerzo corporativo.
- Vista empresarial del negocio.
- Se debe definir una arquitectura empresarial.
- Se trabaja con un único almacén central, se tiene un control centralizado.
- Se puede obtener un resultado rápido si se trabaja con iteraciones.

b) Bottom-Up

- Se pueden obtener implementaciones más rápidas porque la unidad de trabajo son los Data Marts.

- Existen menores riesgos, porque hay menos dependencia entre áreas.
- Se obtienen resultados rápidos.

Desventajas

a) Top-Down

- Requiere más tiempo porque se necesita mayor esfuerzo corporativo.
- Requiere un conocimiento más profundo del negocio.

b) Bottom-Up

- Cada Data Mart es una vista de una parte de los datos.
- Pueden existir datos redundantes o inconsistentes entre los Data Marts.

Kimball vs. Inmon

En la siguiente tabla se muestra las principales diferencias de ambas metodologías las cuales se identifican en los siguientes puntos:

	Kimball	Inmon
Filosofía de desarrollo data warehouse	<p>Se da en base a la priorización de algunos procesos específicos del negocio.</p> <p>Desarrollo directo de data marts en los procesos seleccionados del negocio.</p> <p>Uso exclusivo de modelos dimensionales desnormalizados</p>	<p>Se da en base al modelo de datos de toda la empresa.</p> <p>Desarrollo de un data warehouse empresarial basado en un esquema de base de datos normalizado.</p> <p>El desarrollo de data marts, se basa en datos obtenidos del data warehouse.</p>

	(esquema estrella)	
Definición de data mart	<p>Un data mart mantiene los datos al menor nivel de detalle, los cuales se refieren a un proceso de negocio.</p> <p>Un data mart se construye mediante la extracción de datos directamente desde los sistemas operacionales.</p> <p>Los data marts están vinculados entre sí.</p> <p>Un data mart mantiene todos los datos históricos</p>	<p>Un data mart mantiene los datos agregados que se relacionan a la unidad de negocio.</p> <p>Un data mart se construye mediante la extracción de datos del data warehouse de la empresa (también llamados data marts dependientes).</p> <p>Los data marts no están vinculados entre sí.</p> <p>Un data mart mantiene una historia limitada, ya que ésta se mantiene en el data warehouse de la empresa.</p>
Enfoque de desarrollo por etapas	<p>Las etapas de desarrollo de un data mart se basan en procesos específicos del negocio y están vinculadas a las dimensiones, que forman la arquitectura de bus data warehouse.</p>	<p>El diseño de un data warehouse para toda la empresa se basa en su modelo de datos. Es una aplicación progresiva de las áreas temáticas, de acuerdo con las prioridades establecidas.</p>

Tabla 6: Cuadro comparativo Kimball vs Inmon.

Kimball vs. Josep Curto

En la siguiente tabla se muestra las principales diferencias de ambas metodologías las cuales se identifican en los siguientes puntos:

	Kimball	Josep Curto
Concepción de un Data Warehouse	El data warehouse es un proyecto a largo alcance formal.	El data warehouse es una tecnología modular, preliminar del BI.
Definición de la metodología	Define la metodología de una manera formal con bastante rigidez, su metodología es más conceptual, genérica.	Define la metodología de una forma práctica, está orientado a la aplicación inmediata, más tecnológica, por el hecho que es más contemporáneo que kimball.
Plataforma de desarrollo	Desde el punto de vista tecnológico kimball guarda una independencia de algún tipo de plataforma	Está orientado al software libre.

Tabla 7: Cuadro comparativo Kimball vs Josep Curto.

Business Intelligence Tradicional vs Big Data

Las tres metodologías mencionadas anteriormente representan al BI tradicional. Por otro lado Big Data representa una de las tendencias del BI tradicional que nace como respuesta al constante crecimiento de la data de diferentes tipos por parte de fuentes externas, en este contexto el BI tradicional se encuentra limitado al momento de realizar un análisis de una mayor cantidad de datos dispersos y no estructurados. En la siguiente tabla se muestra las principales diferencias de ambas metodologías las cuales se identifican en los siguientes puntos:

	BI Tradicional	Big Data
Definición de la metodología	Define la metodología como un sistema que analiza datos, estructurados en un DataWarehouse, generados por una empresa, y que muestra cómo está funcionando el negocio en sus distintas áreas para poder tomar las mejores decisiones.	Define la metodología como un sistema que recoge datos de diferentes fuentes (redes sociales, mails, vídeos, archivos PDF, etc), de volumen ilimitado, estructurados, semiestructurados o sin estructura, y proporciona un análisis que permitirá adelantarse a las tendencias del mercado.
Visión	Aporta al usuario una experiencia más estructurada.	Ofrece un análisis profundo y una visión global de los datos.
Datos y Eventos	Queries SQL contra información estructurada: RDBMS, DWH, Data Mining.	Queries contra eventos de múltiples fuentes de información estructurada o no estructurada: Web, RSS, RDBMS, Fichero, JMS, Web Service.
Almacenamiento de gran cantidad de datos	A un alto costo, al tratar de manejar gran cantidad de datos y que puede requerir el uso de diversas tecnologías.	A un bajo costo al usar herramientas como Hadoop, MapReduce y bases de datos NoSQL para el control sobre los volúmenes crecientes de grandes datos. Aumenta la capacidad de almacenar los datos en un orden de magnitud, pasando de

		terabytes a petabytes.
Objetivos	BI se centra en ayudar a la organización a la toma de decisiones de una manera rápida y ágil.	Big Data se centra en el proceso y visualización de gran cantidad de datos.
Integración	BI basado en metodologías maduras.	Puede coexistir con el BI tradicional.
Granularidad y Perspectivas	Nivel de detalle limitado a conocer solo fuentes estructuradas, y de limitadas perspectivas.	Permite trabajar a un nivel de detalle muy superior (granularidad) y con millones de datos a la vez. Se gana en perspectiva por enfoque global de la data. Se logra una mejor comprensión de la información y, por tanto, del entorno y de la propia empresa.
Tiempo de acceso	Velocidad relevante al momento de acceder a los datos.	Mayor velocidad de acceso a los datos comparados con el BI tradicional considerando la gran cantidad de información que Big Data accede.
Análisis	Se basa en el uso de cubos OLAP (On-Line Analytical Processing). Fácil implementación y uso.	Se basa en nuevos métodos de trabajar con grandes volúmenes de datos, nuevas herramientas tales como Hadoop (HDFS y MapReduce) y Hive, y métodos especializados de

		<p>programación tipo no-SQL.</p> <p>Implementación y uso complicado.</p>
--	--	--------------------------------------------------------------------------

Tabla 8: Cuadro comparativo BI Tradicional vs Big Data.

4.1.2. Definición de la metodología de BI

Como parte de la definición de la metodología de BI, se van a analizar dos puntos:

- Elección del Enfoque de Inteligencia de Negocios.
- Elección de la Metodología para el desarrollo de la solución BI.

Enfoque de Inteligencia de Negocios

Se han definido los dos enfoques de Inteligencia de Negocios:

- ✓ Enfoque Top-Down, el cual es presentado como el Modelo de Inmon.
- ✓ Enfoque Bottom-Up, el cual es presentado como el Modelo de Kimball.

Ambos modelos presentan ventajas y desventajas frente a su implementación en un proyecto de Inteligencia de Negocios. Es por ello que se presentó una comparación en donde se plasman las principales ventajas y desventajas de cada uno de los modelos y sus principales diferencias.

Para el desarrollo de este proyecto se ha decidido utilizar el modelo de Kimball, el enfoque Bottom – Up, la metodología de Josep Curto y la metodología de Big Data como base para

la propuesta de la metodología de desarrollo del modelo de inteligencia de negocio del presente trabajo de investigación. Ya que sus características son aplicadas a este proyecto de la siguiente manera:

- El presente proyecto considera la implementación de un Data Mart, con proyección a la implementación de un Data Warehouse Corporativo. Esta es una gran ventaja que presenta el enfoque Bottom – Up, que plantea que las implementaciones se realicen por áreas temáticas, por Data Marts. Los cuales constituyen el Data Warehouse.
- Los usuarios del área de ventas, requieren obtener resultados rápidos y precisamente el enfoque Bottom – Up, permite obtener una implementación rápida que cubre este requerimiento, porque se centra específicamente en la implementación del Data Mart que atiende los requerimientos de un área de la organización.

Con respecto a la metodología de Inmon, tiene mayor alcance ya que aplica un enfoque global, sin embargo el proyecto por ahora está delimitado solo al área de ventas.

La metodología de Ralph Kimball propone, además como arquitectura de una solución de soporte de decisiones, armar primero Data Marts para satisfacer las necesidades específicas de un departamento o área dentro de la empresa, permitiendo así un mejor control de la información que se está abarcando.

En la mayoría de los casos los proyectos de inteligencia de negocios comienzan como un esfuerzo dentro de un área específica, la cual se inicia con la construcción de Data Marts, así se contempla que en un futuro se pueda llegar a construir un Data Warehouse donde se almacene la totalidad de datos de la empresa, esto se acerca más a la idea de Ralph Kimball, por otro lado en esta metodología destaca la fase de mantenimiento y crecimiento la cual no está presente en las metodologías de Inmon y Josep Curto el cual indica que una vez que se ha construido e implantado el Data Warehouse no hay tiempo para el descanso, rápidamente debemos estar preparados para administrar el mantenimiento y crecimiento del mismo. Kimball brinda entonces una serie de puntos a tener en cuenta para mantener exitosamente el Warehouse. Entre ellos se destacan: el continuo soporte y la constante capacitación a usuarios de negocios, el manejo de la infraestructura (monitoreo de base de datos, tráfico, etc.), tuning de rendimiento sobre las consultas, mantenimiento del metadata y procesos ETLs.

Por su parte la metodología de Josep Curto es más práctica y tecnológica y está orientado al software libre en el cual destaca del mercado de soluciones Open Source Business Intelligence (OSBI) a la solución Pentaho, esta es una característica muy importante de esta metodología porque en la actualidad el open source ya no es una tendencia emergente sino es un enfoque que tiene un impacto profundo y que tiene una presencia importante en todos los sectores, en los últimos años el mercado de BI se ha enriquecido con soluciones open source que cubren todo el espectro de necesidades de una organización para la explotación de la información. Algunas de estas herramientas tienen ya a sus

espaldas varios años de recorrido y actualmente se hallan respaldadas por organizaciones. Pentaho es una de las suites más completas y maduras del mercado OSBI que existe desde el año 2006, está compuesta por diferentes módulos que cubren todo el proceso de análisis de la data, incluso incluye muchas herramientas para trabajar con la nueva tendencia de BI llamada Big Data. Este es un punto favor ya que se pretende trabajar con la plataforma Pentaho para el desarrollo de la solución de inteligencia de negocios para el presente proyecto por las razones ya indicadas.

Por su parte la metodología de Big Data nos muestra que las soluciones tradicionales de Business Intelligence no son suficientes para obtener información de los actuales datos que proporcionan fuentes externas como redes sociales, web log, etc, lo que ha provocado el concepto Big Data, que para muchos es una evolución del BI. No obstante, esta analítica avanzada y el BI son perfectamente complementarios, ya que mientras la primera ofrece un análisis profundo y una visión global de los datos, el BI aporta al usuario una experiencia más estructurada. Para que los datos sean realmente útiles, hay que ser capaces de conjugar datos de clientes con financieros, de ventas, de productos, de marketing, de redes sociales, de la competencia, etc y extraer de ellos un análisis valioso y efectivo.

Pero a pesar de la rápida aparición de las tecnologías de grandes datos como alternativas a las plataformas de los Sistemas de Gestión de Base de Datos Relacionales (RDBMS), el EDW no está en vías de extinción en un corto plazo, sino más bien transformándose, ya que las empresas buscan ampliar sus actuales sistemas de almacenamiento de

datos y procesos para ayudar a gestionar los nuevos tipos de datos.

Por lo tanto se prevé una versión modificada del EDW, en la que los conjuntos de datos estructurados y no estructurados se almacenan y gestionan en donde tienen más sentido como parte de una arquitectura extendida pero bien coordinada. Asimismo, la EDW se puede ver como un sistema híbrido que combina prácticamente múltiples tecnologías de procesamiento de datos.

BI trabaja con información que necesita ser estructurada y los mecanismos para hacer posible el tratamiento de esa información es lo que ofrece Big Data, una evolución a los mecanismos tradicionales que utiliza BI. Lo ideal es utilizar ambas tecnologías persiguiendo un objetivo común: ayudar a la organización a ser más eficiente, más competitiva y en definitiva mejor ante sus competidores.

El Big Data no significa la sustitución del BI tradicional pues son técnicas y procesos que los consultores BI pueden usar para ganar rendimiento y velocidad. Big Data es un nuevo conjunto de recursos y procesos para optimizar nuestra solución BI con datos no estructurados nuevos o datos estructurados de toda la vida. Por ello, no se debe ver a Big Data como algo sustituyente sino algo complementario al BI, pues sin BI no hay análisis, y sin Big Data no hay como absorber esa cantidad nueva e ingente de datos. Por lo tanto Big Data es una buena opción a elegir como base para la metodología del modelo de BI a proponer en el presente proyecto, por las razones ya indicadas y además porque se pretende analizar la información de una fuente externa como es el WebLog del servidor de la web Site de la organización a quien se aplicara el modelo propuesto.

Según las características de las metodologías comentadas anteriormente, se justifica y se decide tomar en cuenta las metodologías de Kimball, Josep Curto y Big Data como referencia para la propuesta metodológica del modelo de inteligencia de negocio a desarrollar, el cual será un híbrido entre las metodologías seleccionadas.

4.1.3. Metodología Propuesta para el Desarrollo del Modelo

Luego del resultado de la investigación de las metodologías explicadas en el capítulo II y con la comparación realizadas entre ellas en el presente capítulo, a continuación propongo la siguiente metodología para el desarrollo del modelo de inteligencia de negocios del presente proyecto, el cual tiene como base la metodología para el desarrollo de proyectos de inteligencia de negocios de Ralph Kimball, Josep Curto Diaz y de Big Data.

Fases de la Metodología Propuesta

- Fase de Planificación
 - Plan del Proyecto
- Fase de Modelo del Negocio
 - Comprensión del Dominio del Negocio
- Fase de Análisis
 - Definición de Requerimientos y Análisis de la Data.
 - Estructuración y Modelado de Datos de Fuentes Externas.
 - Análisis Dimensional.
- Fase de Diseño
 - Diseño de la Arquitectura de la Solución.
 - Diseño Dimensional Físico.
 - Limpieza de Datos y Diseño de Extracción.
 - Diseño de Explotación.

- Fase de Construcción
 - Evaluación y Selección de la plataforma BI de Software Libre
 - Instalación y Configuración de Software
 - Construcción de DataMart
 - Ejecución de ETL
 - Creación de Cubo
 - Construcción de Interfaces
 - Pruebas
- Fase de Implantación
 - Implantación de la Solución

Estas fases se presentan como resumen tomando como referencia las mejores prácticas recomendadas por las metodologías base seleccionadas para la implementación del modelo de inteligencia de negocios. En la siguiente figura se presenta el esquema de la metodología propuesta para el desarrollo del modelo de inteligencia de negocios.

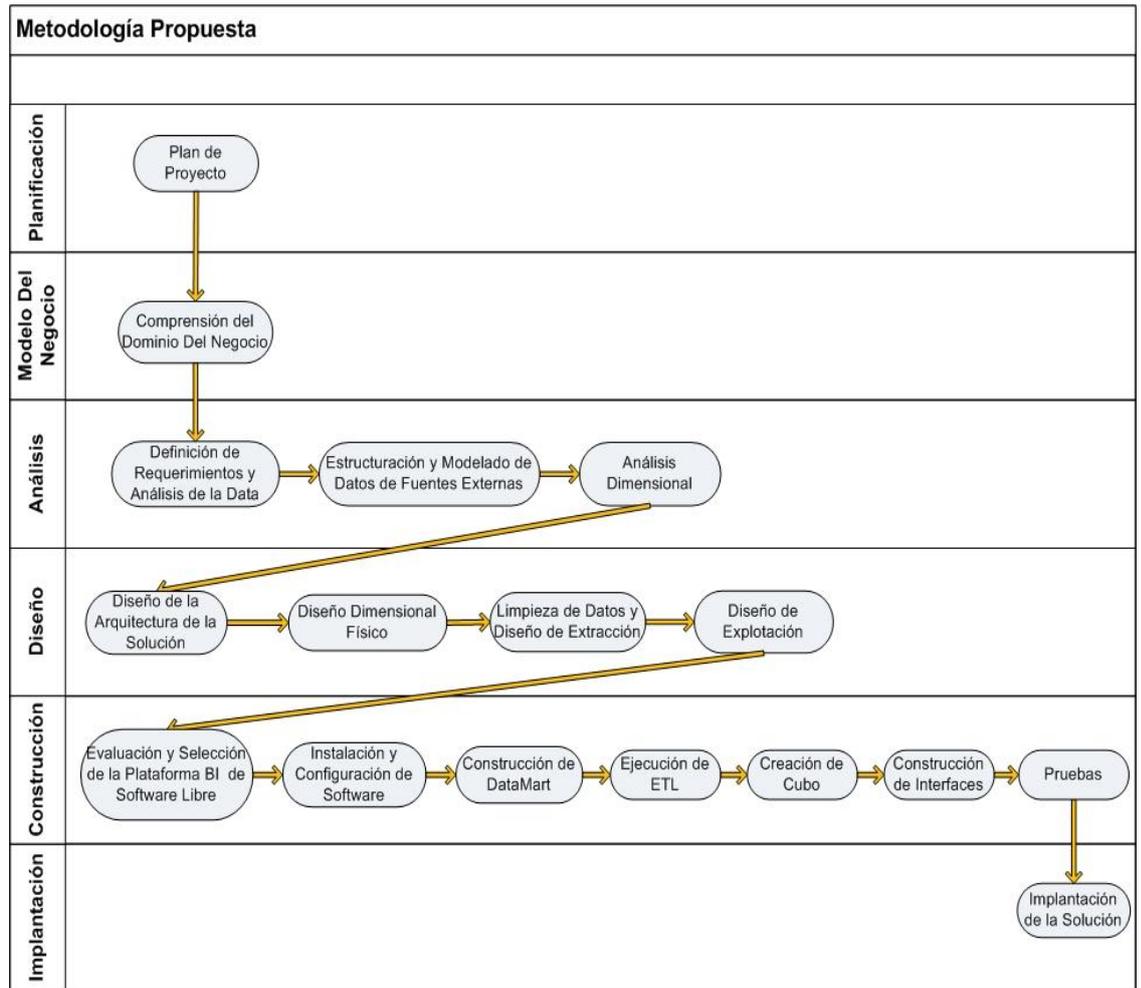


Figura 30: Metodología Propuesta de la Solución BI.

Esta metodología plantea realizar el desarrollo de la solución del modelo de inteligencia de negocios en 6 fases, a continuación se explica cada una de ellas:

4.1.3.1 Planificación

4.1.3.1.1 Plan De Proyecto

Visión del Proyecto: Documento que muestra de manera general los objetivos del proyecto, su alcance, que es lo que se necesita para ejecutarlo y los posibles riesgos que podría tener el

proyecto. En el ANEXO 10 se tiene el formato del documento.

En este documento también se muestra a los integrantes del equipo responsable de la solución, tanto del lado de los usuarios de gestión como del personal de Tecnología de Información. Basándose en la propuesta de KIMBALL en lo referente a la conformación de equipos, los diferentes roles que desempeñarían el equipo del proyecto para la presente propuesta serían los siguientes:

- ✓ **Representante del Negocio:** Es la persona que recibe el beneficio directo del producto. El que toma las decisiones directamente. Generalmente puede ser el gerente o jefe de área.
- ✓ **Experto del Negocio:** Es el conocedor de la forma como se obtienen los datos o lo que realmente se desea analizar.
- ✓ **Asistente de Proyecto:** Es la persona encargada del desarrollo de la documentación del proyecto.
- ✓ **Especialista de Datos:** Es el Administrador de la Base de Datos (DBA). Conocedor de la actual base de datos transaccional que servirá como insumo principal para la generación de datos.

- ✓ **Líder Tecnológico:** Responsable del lado de Tecnologías de Información de gestionar el equipo de tecnología y elaborar el producto deseado.

- ✓ **Líder Técnico:** Es el responsable de brindar el acceso a las fuentes de datos internas y externas necesarias para el producto deseado.

- ✓ **Analista Dimensional:** Es la persona que identificará los requerimientos estratégicos. Debe conocer y dominar la técnica del “pensamiento dimensional”.

- ✓ **Desarrollador DW/BI:** Conoce los temas de ETL, construcción de aplicaciones BI, manejo de controles para lectura de información con cubos y Lenguaje MDX.

4.1.3.2 Modelo del Negocio

4.1.3.2.1 Comprensión del Dominio del Negocio

El objetivo en esta etapa es descubrir la información relevante que otorgue a los usuarios y principalmente a los desarrolladores del proyecto, tener la suficiente y adecuada comprensión del dominio de negocio de la organización a la cual se le realiza el proyecto, con la finalidad de consolidar los necesarios lazos de confianza entre

los clientes y desarrolladores del proyecto, obtener de ambos una visión común de la organización y relacionarse con los problemas de la organización que se requiere resolver, también en esta etapa se determina el o las áreas temáticas en donde se desarrollara el proyecto.

Esta es la etapa de todo proyecto de Inteligencia de Negocios, en la cual se realiza el levantamiento de información de la empresa. A continuación se propone el siguiente esquema para desarrollar esta etapa:

- La organización
 - ✓ Nombre de la Organización
 - ✓ Ubicación
 - ✓ Actividad Económica
 - ✓ Producto
 - ✓ Misión
 - ✓ Visión
 - ✓ Estructura Organizacional
 - ✓ FODA
 - ✓ Objetivos Estratégicos
 - ✓ Mapa Estratégico
 - ✓ Mercado
 - ✓ Entorno Competitivo

- Modelo de negocios Decisional Actual
 - ✓ Actores Organizacionales del Proceso Decisional
 - ✓ Información para la toma de decisiones

- ✓ Tecnología utilizada para tomar decisiones
- ✓ Descripción del Proceso de toma de decisiones actual
- ✓ Indicadores de gestión del proceso de toma de decisiones
- ✓ Resultados actuales de la toma de decisiones

4.1.3.3 Análisis

4.1.3.3.1 Definición de Requerimientos y Análisis de la Data.

Los requerimientos están orientados a identificar:

- Medidas
- Dimensiones
- Indicadores

En este punto se presentan las definiciones más importantes que presenta el modelo. Esto servirá para tener una línea base de definiciones acerca del modelo dimensional.

Se presenta inicialmente el modelo de datos de las fuentes de información internas como por ejemplo del sistema transaccional de la organización, etc. y luego se determina cuáles serán las fuentes de información externa a considerar para la siguiente etapa.

Para luego identificar las dimensiones y la tabla de hechos de acuerdo a los requerimientos solicitados.

Para la identificación de las medidas y dimensiones se pueden considerar las siguientes fuentes:

Fuentes de información interna:

- Plan Estratégico.
- Entrevistas
- Reportes de Gestión
- Base de Datos

Fuentes de información externa:

- Redes Sociales.
- Youtube.
- Mailing.
- Web Site de la institución.
- Blogs.

Veamos ahora en la siguiente tabla lo que nos devuelve cada una de las fuentes internas:

Tipo Función	Fuente de Información	Encontrar
Gestión (Definir el deseo del contenido del sistema)	Plan Estratégico	Medidas en función a los indicadores
	Entrevistas	Medidas y Dimensiones
	Reportes Gestión.	Medidas y Dimensiones
TI (verificar la	Base de Datos	Verificar si las

realidad)	Transaccional	medidas y dimensiones están contenidas en la BD
-----------	---------------	-------------------------------------------------

Tabla 9: Fuentes de información interna.

Se deben tener las siguientes consideraciones al momento de revisar las siguientes fuentes de información:

a) El Plan Estratégico

Se debe revisar el Plan Estratégico identificando indicadores de gestión que permitan determinar el seguimiento de los objetivos y/o metas establecidas.

Como producto final de los requerimientos debemos obtener los indicadores, bajo el siguiente formato:

Indicador	Medidas
Indicador de Recaudación	Montos Vendidos Montos Cobrados

Tabla 10: Ejemplo de Indicador.

b) Entrevistas

Se puede aplicar entrevistas diferenciadas:

- A los responsables de gestión.
- Al personal de Tecnologías de Información, especialmente al experto de la base de Datos.

✓ **Personal de Gestión:** las preguntas se clasificarán en:

Conocer la gestión del negocio: Aquí es vital 3 puntos

- Identificar los objetivos
- Conocer las estrategias
- Definir los indicadores y Medidas.

Identificar los requerimientos analíticos:

Para ello enfocado en el proceso de negocios de estudio hay que aplicar la técnica dimensional.

Técnica Dimensional	Descripción	Ejemplo (Ventas)
¿Qué?	Referido a la razón de ser del proceso. Que es lo que realiza	Producto
¿Quién?	Define quienes son los que demandan o realizan el que	Cliente
¿Dónde?	Responde al lugar, físico o virtual, donde se produce el contacto o transacción entre: Quien y el Que.	Organización
¿Cuándo?	Es el momento en que se realiza la transacción.	Tiempo
¿Cómo?	Cuál es el medio por el cual, el quien demanda el Que	Forma Pago

Tabla 11: Técnica Multidimensional.

Esto nos lleva a encontrar las dimensiones que son las preguntas pre-concebidas a

realizar las cuales se muestran en el ANEXO 11.

✓ **Personal de Tecnologías de Información:** las preguntas se orientan a conocer:

- **Análisis de Data.**

Determinar el nivel de dependencia del personal de gestión respecto a las IT. Sería altamente conveniente ubicar los reportes que IT prepara, claro si los realizan, para el personal de Gestión.

- **Disponibilidad de la Data.**

Conocer las características del motor actual donde se almacena la BD transaccional.

- **Calidad de la Data.**

Determinar si se han aplicado conceptos de integridad de datos lo cual es fundamental para aplicar posteriormente el proceso de ETL.

Esto nos lleva a encontrar las dimensiones que son las preguntas pre-concebidas a realizar las cuales se muestran en el ANEXO 12.

c) Reportes de Gestión

La revisión de un reporte de este tipo implica encontrar:

- Medidas
- Dimensiones
- Jerarquías

A continuación se plantea una serie de actividades que pueden ayudar a encontrar: medidas, dimensiones y jerarquías en dichos reportes.

Identificar dimensiones

- Visualice la clasificación del cuadro e identifique las cabeceras de las columnas con menor nivel.
- Visualice la clasificación del cuadro e identifique las cabeceras de las filas con menor nivel.
- Visualice el título del reporte e identifique el nivel más bajo

Identificar medidas

- Cuál es la intersección entre las columnas y filas de menor nivel.

Identificar Jerarquías

- Identifique los acumulados a nivel de Columna.
- Identifique los acumulados a nivel de Fila
- Identifique alguna clasificación en el título

d) Datos Transaccionales

Aquí es donde debemos incluir el Modelo de Datos Actual, así mismo se debe indicar las tablas involucradas en el proceso a analizar en el Modelo de Datos Actual, el formato para realizar este análisis se encuentra en el ANEXO 13.

Fuentes Externas para el Análisis de Datos

Veamos ahora a manera de ejemplo en la siguiente tabla lo que nos puede devolver algunas de las fuentes externas que podemos utilizar:

Fuente de Información	Encontrar	Ejemplo
Facebook	Medidas en función a los indicadores.	Se obtienen estadísticas de los fans que han visto una publicidad: N° de “Likes”, N° de veces que ha sido compartida una publicidad, N° de comentarios, sentimiento de los comentarios, perfil de los usuarios que han interactuado por ejemplo con una campaña.
Twiter	Medidas en función a los indicadores.	Se obtienen estadísticas de los followers y otros usuarios que han visto este link: N° de Re“tweets”, N° de respuestas, perfil de los usuarios, sentimiento de las respuestas, alcance de una campaña.
Mailing	Medidas en función a los indicadores.	Se obtienen estadísticas de la interacción de los

		contactos con el e-mail de la organización. Se obtendrán datos como por ejemplo: N° de contactos que han abierto el e-mail, N° de rebotes, N° de clicks en la URL.
WebLog Del Servidor.	Medidas y Dimensiones	Se obtienen estadísticas de la interacción de los usuarios con la web como por ejemplo: N° de visitas, N° de páginas vistas de la web, N° de visitas por ubicación (países), Motores de búsqueda (con qué palabras clave le buscan en Internet), Sitios web de referencia (Qué páginas enlazan a su web), N° de visitas por sección de la web.

Tabla 12: Fuentes de información externa.

En este punto también se debe mencionar las ventajas que se obtendrán por de cada una de las fuentes externas que se hayan seleccionado para el análisis de la data. Luego por último se deben definir los Requerimientos Organizacionales y Requerimientos del Modelo de Inteligencia de Negocios.

4.1.3.3.2 Estructuración y Modelado de Datos de Fuentes Externas

Luego de haber seleccionado la o las fuentes externas a utilizar, se realiza la estructuración y modelado de datos de fuentes externas, para cumplir con este objetivo se utiliza Big Data, ya que nos permitirá convertir nuestra data no estructurada en estructurada para luego proceder con el modelado final. Antes de explicar a detalle lo que se debe realizar es importante tener claro cómo funcionan los principales componentes de Big Data para poder aplicarlo sin ningún inconveniente en este punto:

Apache Hadoop

Hadoop es un framework que implementa un paradigma computacional llamado Map Reduce, donde se divide la aplicación en muchos fragmentos pequeños de trabajo, cada uno de los cuales pueden ser ejecutados o re ejecutados en cualquier nodo del clúster. Además, proporciona un sistema de archivos distribuido (HDFS) que almacena los datos en los nodos de cómputo, proporcionando ancho de banda agregado muy alto en todo el clúster. Tanto Map Reduce, como el sistema de archivos distribuidos están diseñados para que los fallos de nodo se gestionen automáticamente por el framework.

MapReduce

MapReduce se compone de dos partes, el mapeo y la reducción. El mapeo consiste en “traducir” una serie de información de entrada en una salida

basada en pares clave/valor. La reducción consiste en tomar esos pares clave/valor, y aplicarles una función de agregación.

Por ejemplo, tenemos 3Tb de datos no estructurados en un archivo log (Un log es un registro de actividad de un servidor web, que generalmente se guarda en un fichero de texto, al que se le van añadiendo líneas a medida que se realizan acciones sobre este servidor) de un sitio web con mucho tráfico, y queremos saber qué tipo de respuestas ha dado el servidor a las peticiones de los clientes, y cuántas respuestas ha habido de cada tipo, el resultado final tiene que ser una tabla donde para cada tipo de respuesta se indique el número acumulado, es decir, cuantas respuestas 200 que significa OK, cuantas 302 REDIRECT (re direccionadas), cuantas 404 NOT FOUND (no encontradas), etc.

LOGS -> (3Tb)	(mapeo)->	404,1 ->	(reducción)->	404,2
		404,1		200,3
		200,1		500,1
		500,1		302,1
		302,1		
		200,1		
		200,1		

Figura 31: Ejemplo de Mapeo y Reducción.

La función de mapeo tomará como entrada el log, y producirá como salida pares que contienen el código de respuesta y el número del mismo tipo encontrado. La función de agregación tomará

todas las salidas que se generen desde los mapeos, y sumará los totales por cada tipo de respuesta. Otra forma común de realizar este proceso es a través de la identificación de los caracteres de delimitación en la data no estructurada que permitan separar los datos. Por ejemplo se podría realizar la división de los datos del archivo log a través de un separador que podría ser (- o |) que se encuentre separando cada uno de los datos del archivo log, esto nos permitirá asignar los valores para los campos clave/valor mencionados anteriormente y así realizar la reducción de los datos.

La magia de Hadoop es que permite fragmentar los datos de entrada, y lanzar en paralelo múltiples mapeos y reducciones. Proporciona capacidad de computación utilizando además hardware de bajo coste, y sobre todo un control sobre el tiempo de ejecución de las tareas:

- Procesar estos 3Tb de datos del log con un único nodo se procesaría en 240 horas, al poner dos nodos, se procesarían en 120horas; con cuatro nodos serían 60 horas, con ocho 30 horas y con 240 nodos tardaría una hora.
- Por el contrario, con un script secuencial no sería posible reducir el tiempo de ejecución del mismo a menos que se mejore el hardware.

JobTrackers y TaskTrackers

Para ejecutar los procesos MapReduce, Hadoop hace uso de dos tipos de nodos:

- JobTrackers, son el punto de acceso a la funcionalidad MapReduce: los clientes envían los trabajos a los JobTrackers, y éstos los despachan a un segundo nivel (los TaskTrackers) que son los que realmente realizan el trabajo. Los JobTrackers conocen a todos los TaskTrackers, su estado y su topología, de modo que pueden optimizar el envío de las tareas
- TaskTrackers, son los que realmente realizan el trabajo (ya sean mapeos o reducciones). Proporcionan todos los mecanismos posibles para minimizar el riesgo de caída del proceso, pero si este ocurre, los JobTrackers reenviarán la misma tarea a otro TaskTracker de modo que no se pierda.

HIVE

Apache HIVE es una infraestructura de almacenamiento de datos construida sobre Hadoop para proveer resumen de datos, consultas y análisis de grandes volúmenes de datos almacenados en Hadoop o en sistemas de archivos distribuidos compatibles.

Hive define un lenguaje tipo SQL llamado HiveQL que da soporte a los usuarios familiarizados con SQL para realizar consultas de los datos. Al mismo tiempo, también permite a los programadores que trabajan con el framework de MapReduce el poder conectarse con sus mappers y reducers de tal manera que puedan realizar un análisis más sofisticado de la información.

HiveQL tiene un alcance limitado a consultas tipo insert, select y create table, pero carece de soporte para transacciones y vistas, se interrelaciona con Hadoop a través de un compilador que traduce la sentencia en un DAG de los jobs de MapReduce.

Cabe destacar que existe una herramienta de Open Source de la Suite Pentaho llamada Pentaho Data Integration que nos permite trabajar con las herramientas anteriormente descritas, integrándose con ellas de una manera eficiente con el objetivo de desarrollar procesos ETLs los cuales nos van a ayudar a extraer, normalizar y dejar todos los datos en un almacén de datos o base de datos optimizada para el análisis de estos grandes volúmenes de datos, incluso simplifica el trabajo con MapReduce ya que Pentaho Data Integration permite realizar configuraciones para el procesamiento de la data, como por ejemplo para el caso del log de un servidor web solo sería necesario identificar el tipo de separador utilizado en el archivo log, y con esto se logra estructurar la data.

Para el desarrollo de esta etapa se propone desarrollar las siguientes actividades que han sido definidas en base a la metodología de Big Data propuesta en el capítulo dos del presente proyecto:

a) Recolección, Extracción y Procesamiento de los Datos. Dependiendo de la fuente de información externa seleccionada en la etapa anterior es decir en función del tipo y el origen de los datos, podemos encontrar varias librerías que podrían utilizarse en esta primera actividad. Si hablamos de datos no estructurados, como pueden ser ficheros de logs, el framework más utilizado probablemente sea Hadoop (HDFS y Map Reduce). Sin embargo si estamos tratando con datos provenientes de una base de datos relacional Sqoop es la herramienta más utilizada. Map Reduce en primer lugar se encarga de la limpieza de los datos (definición de campos que recibirán la información de la fuente externa, en este caso del log), y de extraer y estructurar los datos realizando un filtrado previo de los mismos dando además la estructura conveniente para que sirvan de entrada a HDFS. Esto por motivo que necesitamos que los datos pre procesados sean almacenados en un sistema distribuido de ficheros. Este es el rol que cumple HDFS como componente core dentro de Hadoop.

Luego se envía todos los datos a Hive desde HDFS para el análisis y modelado de datos (creación de tablas).

b) Análisis y Modelado de datos.

Hive es un framework que permite crear tablas, insertar datos y realizar consultas con un lenguaje similar a como se podría realizar con queries SQL llamado HiveQL. La diferencia fundamental con SQL es que utiliza MapReduce para ejecutar todas esas queries. Por lo tanto en esta actividad se modela los datos (creación de tablas con los campos obtenidos en HDFS) a través de lenguaje de consultas similar a SQL, cuyo resultado final es una base de datos con diversas tablas, luego se envía la data a un gestor de base de datos relacional para crear la base de datos relacional final.

c) Integración e interpretación de datos. Luego de obtener el modelo relacional final estructurado este queda listo para aplicar la siguiente etapa análisis dimensional en ambas fuentes y finalizando con la interpretación de los datos a través de una herramienta BI en la cual quizá la herramienta más reconocida y la más apropiada a recomendar es la Suit Pentaho BI. Al finalizar esta última actividad debe presentarse el Modelo de datos relacional final como resultado de esta etapa.

Utilizando la herramienta Pentaho Data Integration, este proceso se realizaría en cinco

pasos, para representar que resultados se estaría obteniendo internamente en cada paso se tomara como ejemplo el trabajo con un log de servidor. En primer lugar tendríamos el fichero log con la siguiente data no estructurada:

```
121.41.111.60 | 800 | GET | /index.html | Mozilla/4.0 |  
131.43.113.50 | 800 | GET | /newsite/productos.html |  
Mozilla/4.0 | 121.41.111.60 | 900 | POST | /index.html  
| Mozilla/4.0
```

Figura 32: Datos no estructurados.

1. Creación de la transformación Pentaho

Mapper: Permite la limpieza, extracción y estructuración de los datos a través del mapeo (traducción) de la información de entrada en una salida basada en pares clave/valor. En la data no estructurada se puede apreciar que existe un delimitador “|” por lo tanto en esta transformación se especifica que los datos se dividan en base al delimitador, luego se generan los campos que recibirán la información proveniente del archivo log para mapearlos con los datos obtenidos de la ejecución de esta transformación. De esta manera el resultado de la ejecución de esta primera transformación solo para el campo definido como IP sería el siguiente:

121.41.111.60, 1
121.41.111.60, 1
131.43.113.50, 1

Figura 33: Resultado de Pentaho Mapper.

2. Creación de la transformación Pentaho

Reducer: Permite la reducción o filtrado de los datos a partir de los valores asignados para los pares campo/valor para luego aplicarles una función de agregación, la reducción se realiza para que pueda procesar rápidamente los datos. El resultado de la ejecución de esta segunda transformación solo para el campo definido como IP sería la siguiente:

121.41.111.60, 2
131.43.113.50, 1

Figura 34: Resultado de Pentaho Reducer.

3. Creación del Job para ejecutar el

proceso Map Reduce: Esta tarea ejecuta las dos primeras transformaciones. El resultado de la ejecución de este Job es la data pre procesada almacenada en HDFS. En la siguiente figura se muestra el resultado final de como los datos quedarían organizados en los campos que nosotros hayamos definido en la primera transformación:

IP 121.41.111.60 131.43.113.50
PAGINA /newsite/productos.html /index.html
ANCHO_DE_BANDA 0800 0900
METODO GET POST
NAVEGADOR Mozilla/4.0

Figura 35: Resultado de Job MapReduce.

- 4. Creación del Job para cargar los datos desde Hadoop a Hive:** Permite enviar los datos pre procesados desde HDFS a Hive para el modelado de datos (creación de base de datos y creación de tablas).

Como resultado en Hive (a través de las consultas HiveQL) se obtiene los campos mencionados anteriormente estructurados en tablas, en la siguiente figura se muestra la estructuración de los campos IP, página y ancho de banda correspondiente a la tabla Host y páginas:

HOST IP
PAGINAS Pagina Ancho_De_Banda

Figura 36: Resultado del Job Hadoop - Hive.

- 5. Creación del Job para extraer los datos desde Hive al RDBMS:** Permite enviar los datos desde Hive a un gestor de base de datos relacional para la creación de la base de datos relacional final, para el posterior análisis dimensional de ambas fuentes de información (interna y externa) el cual se realizara en la siguiente etapa de la presente fase.

Para realizar todos estos pasos en Pentaho solo es necesario realizar configuraciones en cada nodo a utilizar dentro de las transformaciones, esto facilita el trabajo con Big Data, sin embargo la cantidad de configuraciones a realizar por nodo aun es alta, en próximas versiones de Pentaho se espera que mejoren este inconveniente.

Al tener ambas fuentes interna y externa como información estructurada, es posible aplicar las siguientes etapas sin ningún tipo de inconveniente a ambas por igual, a continuación a manera de ejemplo en las siguientes fases y etapas se explica cómo se trabajaría con las fuentes de datos internas como por ejemplo la BD

transaccional, pero es importante dejar claro que también es perfectamente aplicable a la información proveniente del modelo de datos relacional estructuradas de las fuentes externas.

4.1.3.3.3 Análisis Dimensional

El propósito del análisis dimensional es consolidar y formalizar las necesidades de información analítica encontradas desde las distintas fuentes consultadas. Los documentos de trabajo a utilizar en este punto tienen 2 objetivos:

- Buscar un compromiso por parte del personal de gestión, que son los beneficiarios directos, acerca de la definición del producto.
- TI asume el reto de apoyar al equipo de desarrollo que implementará lo definido por la gente de gestión a fin de elaborar el producto final.

Para ayudar a la definición del producto y determinar el compromiso del personal de gestión e IT definiremos 4 formatos.

Formato	Descripción
Hoja de Gestión	Permite definir los indicadores y las medidas respectivas
Hoja de Análisis	Define Medidas y Dimensiones finalmente encontradas
Cuadro Dimensiones vs	Especifica las jerarquías

Jerarquías	necesarias a fin de permitir un análisis de lo más genérico a lo más detallado
Cuadro de Medidas vs Dimensiones	Permite definir que dimensiones serán analizadas por tales medidas.

Tabla 13: Formatos Análisis Dimensional.

a) Hoja De Gestión

El formato que se presenta en el ANEXO 14, incluye los items:

- **Proceso:** definir el nombre del proceso de negocios materia de estudio
- **Objetivos:** resumir los objetivos del proceso de negocios encontrados propuestos como para gestionar adecuadamente el proceso.
- **Estrategias:** especificar las acciones que se toman en cuenta con el fin de satisfacer los objetivos planteados
- **Indicadores:** Permiten realizar un seguimiento de la consecución de los objetivos planteados. Ejemplo:

Indicador	Medidas	Estado	Inicial	Final	
Indicador de Nivel de Ventas	MontosVendidos	Bueno	>85%		Green
		Regular	70% a	85%	Yellow
		Malo		<70%	Red

Tabla 14: Ejemplo de Hoja de Gestión.

Indicador: corresponde a identificar el indicador. Por ejemplo: Indicador de Nivel de Ventas.

Medidas: determina las variables a partir de las cuales se tendrá el valor del indicador. Por ejemplo: Montos Vendidos.

Estado: determina el resultado actual del indicador. Los valores del estado son administrados por los expertos en gestión y pretenden mostrar el avance del indicador respecto a una meta establecida. Se sugiere manejar 3 estados:

- ✓ **Bueno:** indica la parte ideal a ser alcanzada. Es la meta propiamente dicha.
- ✓ **Regular:** es un estado intermedio el cual nos muestra que estamos acercándonos a la meta y saliendo del estado negativo.
- ✓ **Malo:** muestra un estado negativo del indicador y un peligro latente de no alcanzar la meta en los plazos establecidos.

Para los estados mencionados deben de indicar el valor o valores mínimo y máximo. A esta técnica de definir un estado al indicador se le conoce como la técnica del semáforo.

b) Hoja De Análisis

El formato que se presenta en el ANEXO 15 incluye los items:

- **Proceso:** especificar el nombre del proceso de negocios materia de estudio.
- **Medidas:** definir las medidas encontradas a partir de los indicadores de gestión definidos en el plan estratégico o las entrevistas tenidas. Así mismo si existiera alguna medida más encontrada en los requerimientos definidos en la etapa anterior también debe ser incluida.
Ejemplo: Montos Vendidos
- **Dimensiones y Niveles:** incluir las dimensiones encontradas, así como expresar los puntos de análisis o niveles encontrados por cada dimensión. No hay ningún orden definido en los niveles, lo importante es incluir todos los puntos de análisis.

Dimensiones	Niveles
Cliente	RazonSocial, Ciudad, Region, TipoCliente, Pais
Producto	Descripcion, Marca, Linea, Proveedor

Tabla 15: Ejemplo Dimensiones y Niveles.

c) Cuadro Dimensiones y Jerarquías

El formato que se presenta en el ANEXO 16 incluye los items:

- **Proceso:** especificar el nombre del proceso de negocios materia de estudio
- **Jerarquías:** organizar los niveles de las dimensiones, para permitir el análisis de lo más genérico a lo detallado o viceversa. Estos niveles son los mismos encontrado en el punto anterior, pero de una manera organizada. El nivel 1 debe incluir el dato más detallado hasta alcanzar el más genérico.

Dimensiones	JERARQUIAS				
	Nivel 1	Nivel 2	Nivel 3	Nivel 4	Nivel 5
<i>Cliente</i>	<i>RazonSocial</i>	<i>TipoCliente</i>			
	<i>RazonSocial</i>	Ciudad	Region	Pais	
<i>Producto</i>	<i>Descripcion</i>	Marca	Linea		
	<i>Descripcion</i>	Marca	Proveed or		

Tabla 16: Ejemplo Dimensiones vs Jerarquías.

La forma de representarlo en el análisis dimensional es la siguiente:



Figura 37: Ejemplo de Dimensiones y Jerarquías.

d) Cuadro Dimensiones y Medidas

El formato que se presenta en el ANEXO 17 incluye los items:

- **Proceso:** especificar el nombre del proceso de negocios
- **Cuadro Dimensiones y Medidas:** considerar la información de las medidas y dimensiones encontradas.

Medidas	DIMENSIONES				
	Cliente	Producto	Tiempo	Organización	Forma Pago
Monto Vendidos	X	X	X	X	X
Cantidad Vendida	X	X	X	X	

Tabla 17: Ejemplo de Cuadro de Dimensiones vs Medidas.

Para efectuar este análisis elegir una medida y definir en qué transacción se genera esta medida. Luego de ello preguntar si en la transacción hay un dato que represente a la dimensión o si es factible que la medida sea analizada por tal dimensión.

e) Análisis Dimensional Final

En las siguientes figuras se muestra cómo desarrollar nuestro análisis dimensional aplicando la técnica dimensional tanto para dimensiones y medidas.

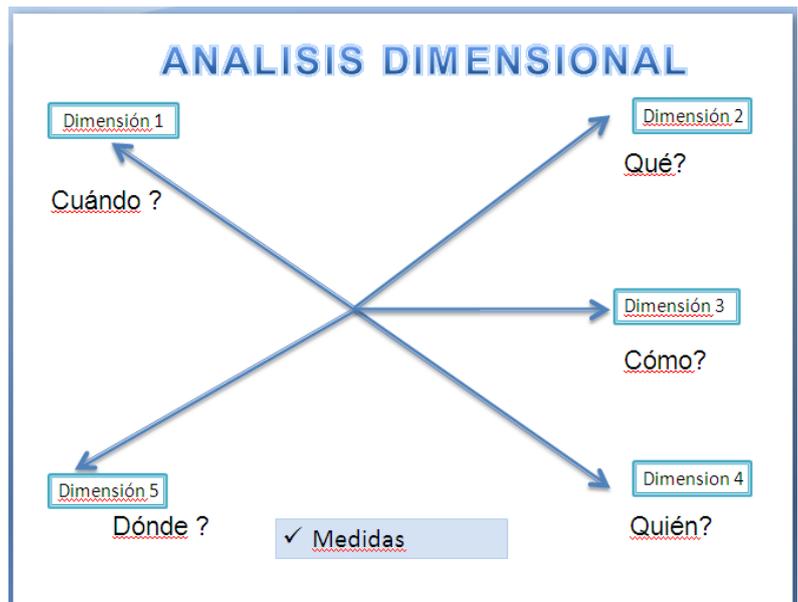


Figura 38: Ejemplo de Análisis Dimensional Aplicando la Técnica Dimensional.

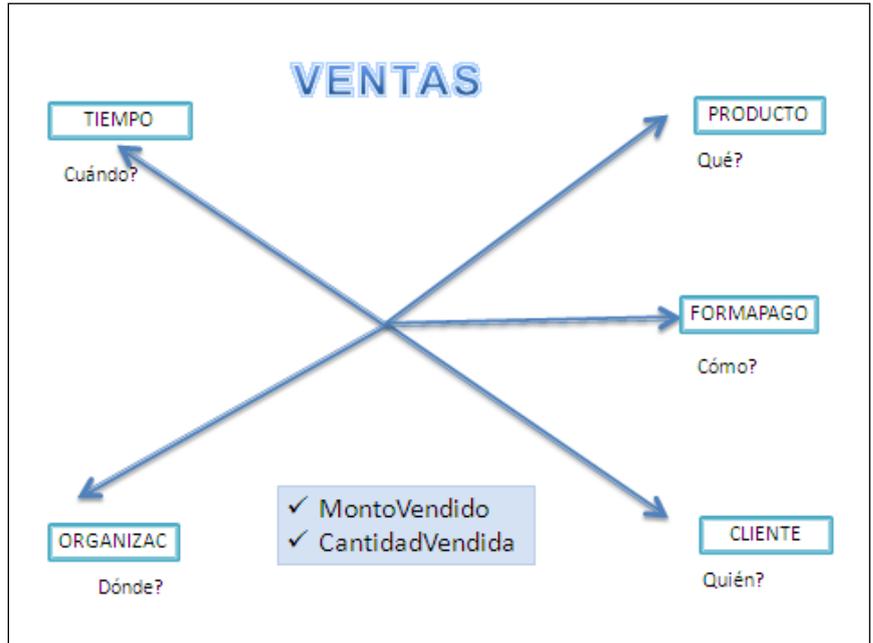


Figura 39: Ejemplo Análisis Dimensional Ventas. Medidas y Dimensiones.

Tomando como muestra las tablas de dimensiones y medidas del punto anterior tendríamos el siguiente análisis dimensional final.

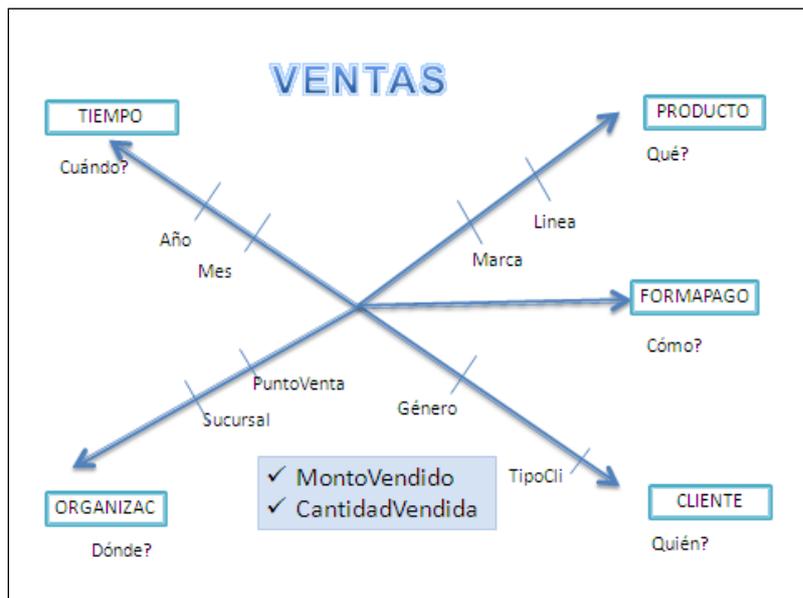


Figura 40: Ejemplo de Análisis Dimensional Final.

4.1.3.4 Diseño

4.1.3.4.1 Diseño de la Arquitectura de la Solución

En este punto se presenta la arquitectura de la solución. Esta constituye el diseño de una aplicación a alto nivel; es decir, se realiza el estudio de la estructura de la aplicación desde el punto de vista de componentes que interactúan entre sí, se refiere a la plataforma donde descansaran los datos y los procesos. Es la planta física del modelo, los canales y plataforma que mantienen y transportan los datos y se ejecutan las aplicaciones. Por lo tanto se definen los componentes que formaran parte de la nueva solución propuesta. En esta etapa es importante indicar que los componentes de Big Data utilizados en la etapa de “Estructuración y Modelado de Fuentes Externas” de la fase de Análisis, también deben ser considerados en el diseño de la arquitectura de la solución en la presente etapa.

Los sistemas de soporte de decisiones usan sistemas OLAP en su arquitectura. En general, estos sistemas soportan requerimientos complejos de análisis, permiten analizar datos desde diferentes perspectivas y soportar análisis complejos contra un volumen muy grande de datos entre otras funcionalidades.

La funcionalidad de los sistemas OLAP se caracteriza por facilitar un análisis multidimensional de datos corporativos, brindando posibilidades de navegación sobre la información de interés del usuario. A continuación en la siguiente tabla se realiza una comparación entre dos tipos de

arquitectura OLAP, arquitectura MOLAP y arquitectura ROLAP:

	MOLAP	ROLAP
VENTAJAS	<p>Excelente rendimiento: los cubos MOLAP son contruidos para tener una rápida recuperación de datos y esta optimizado para operaciones de <i>slicing & dicing</i> (rebanar y dividir).</p> <p>Puede realizar cálculos complejos: ya que todos los cálculos han sido pre generados cuando el cubo se crea. Por lo tanto los cálculos complejos se almacenan y regresan su resultado rápidamente.</p>	<p>Puede almacenar grandes cantidades de datos. La limitante de tamaño en la tecnología ROLAP es la limitante de la base de datos relacional, es decir ROLAP en sí misma no está limitada.</p> <p>Puede cubrir funcionalidad inherente a las bases de datos relacionales, las cuales ya vienen con un set de funciones que son heredables ya que esta tecnología se monta sobre esta base de datos.</p>
DESVENTAJAS	<p>Limitado en la cantidad de datos a ser manejados. Porque todos los cálculos son contruidos cuando se genera el cubo, no es posible incluir grandes cantidades de datos en el</p>	<p>Bajo rendimiento. Ya que cada informe ROLAP es esencialmente una o múltiples consultas en la base de datos relacional, el tiempo de respuesta se alarga si el tamaño de la base de datos es mayor.</p> <p>Está limitada a la</p>

	<p>cubo en sí mismo. Esto no quiere decir que los datos del cubo no deriven de una gran cantidad de datos. Sí es posible, pero en este caso, sólo la información de alto nivel puede ser incluida en este.</p>	<p>funcionalidad SQL ya que la tecnología ROLAP utiliza básicamente sentencias SQL o queries (consultas) de la base de datos relacional, y SQL no aporta todas las necesidades de consultas multidimensionales.</p>
--	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Tabla 18: Arquitectura MOLAP vs arquitectura ROLAP.

Dependiendo del proyecto se deberá elegir una de las dos arquitecturas OLAP.

4.1.3.4.2 Diseño Dimensional Físico

Para llevar a cabo este punto, se decide cuál es el RDBMS (Relational Data Base Management System, Sistema de gestión relacional de bases de datos) que se va a utilizar, para que el esquema físico se adapte a él. Una vez cumplida la etapa de análisis en donde formalmente se definió por el lado del personal de gestión sus requerimientos estratégicos, ahora el personal de tecnología inicia sus tareas para ir construyendo el producto basado en el análisis dimensional. Por lo tanto esta etapa constituye el implementar las dimensiones en tablas dimensionales y las medidas en tablas hecho a partir de los análisis dimensionales finales encontrados, en la etapa de análisis.

Se tienen dos modelos para la arquitectura de almacenamiento de datos:

- Modelo Estrella (Desnormalizado)
- Modelo Copo de Nieve (Normalizado)

Para la presente propuesta se ha elegido el Modelo Dimensional tipo Estrella, como a continuación describiremos:

Definir el Grano

Consiste en identificar el nivel mínimo de detalle de análisis de la información de acuerdo al análisis efectuado anteriormente. La idea es definir qué nivel de la dimensión está presente cuando se genera la medida, ejemplo: de la dimensión tiempo el grano sería Fecha.

El grano permite:

- Escoger el nivel más bajo de detalle.
- Proporciona capacidad de análisis del detalle de los datos.

Diseñar Dimensiones

Cada dimensión encontrada se convierte en tabla dimensional, la misma que debe contener:

Una Clave primaria de tipo surrogate key (es un dato sin interpretación). Normalmente es un dato auto numérico.

Por cada nivel encontrado en la dimensión crear un atributo. Este atributo debe ser descriptivo. Vamos a tomar como ejemplo la dimensión cliente:

En la siguiente figura se muestra cómo quedaría el diseño físico de la dimensión cliente:



Figura 41: Ejemplo de Diseño Dimensional Físico Cliente.

Finalmente se debe realizar la descripción de la dimensión de acuerdo al formato que se muestra en la siguiente tabla:

Nivel	Descripción	Ejemplo
RazonSocial	Nombre de la empresa.	ALFILCOMP E.I.R.L.
Ciudad	Lugar donde reside	Lima

Tabla 19: Ejemplo de descripción de Dimensiones.

La Tabla Hecho

Las medidas comunes encontradas en el análisis dimensional se almacenarán en una tabla hecho.

Una tabla hecho debe contener:

- Una Clave Primaria Compuesta, que es el resultado de la relación de cada clave primaria de las dimensiones.

- Las medidas encontradas en el análisis dimensional a implementar.

En la siguiente figura se muestra un ejemplo del diseño físico de una tabla hechos.



Figura 42: Ejemplo de diseño físico de la Tabla Hecho.

Diseño Dimensional Físico Completo

En este punto se presenta el diseño dimensional físico completo. Las dimensiones deben unirse con su respectiva tabla hecho, de acuerdo al análisis dimensional final presentado en la fase anterior.

4.1.3.4.3 Limpieza de Datos y Diseño de Extracción

El diseño de extracción de datos comprende la carga del Data Mart desde las tablas fuentes. Se siguen los siguientes pasos en el diseño de extracción:

- Descripción de las tablas fuentes:** Son las fuentes desde donde se extraen los datos para alimentar la dimensión correspondiente. Se utiliza el siguiente formato:

Tipo	de	Nombre	de	Descripción
------	----	--------	----	-------------

Fuente	Fuente	

Tabla 20: Formato de Tablas Fuentes.

b) Fuente de Datos: Se muestra las características de las tablas fuentes. Se debe mencionar las consideraciones importantes que se tomen en cuenta para la transformación de los datos. Se utiliza el siguiente formato:

Tabla	Nombre	Llave	Tipo	Formato	Consideración Importante

Tabla 21: Formato de Fuente de Datos.

c) Estandarización y Limpieza de Datos: Especifica la nomenclatura y transformación que deben tener los datos de las dimensiones y la tabla hechos, antes de almacenarlos en el Data Mart. El ETL verifica los datos que ingresan al Data Mart, realizando una limpieza de este si se requiere. Se utiliza el siguiente formato:

Nombre	Llave	Tipo	Formato	Limpieza	Valor Por Defecto

Tabla 22: Formato de Limpieza de Datos.

d) Mapeo de Datos: El mapeo de los datos se realiza partiendo de los datos una vez “limpios”. Se transforma los datos de acuerdo con las reglas de negocio y los estándares establecidos. La transformación incluye: cambios de formato, sustitución de códigos, valores derivados y agregados. Los agregados, como por ejemplo la suma de las venta, normalmente se pre-calculan y se almacenan para conseguir mayores rendimientos cuando lanzamos las consultas que requieren el cálculo de totales al Data Mart. Se utiliza el siguiente formato:

Tabla:		
Campo	Tipo	Mapeo

Tabla 23: Formato de Mapeo de Datos.

e) Proceso: Se describe el proceso que se realiza para la extracción, transformación y carga de las dimensiones correspondientes y la tabla hecho.

f) Esquema de Extracción Final: Luego de realizar el diseño de extracción para cada una de las dimensiones y para las tablas hechos, se realiza el esquema de extracción final, el cual

permite conocer el orden en que se realiza el proceso de extracción. Se utiliza el siguiente formato:

N°	Proceso	Dependencia

Tabla 24: Formato de Esquema de Extracción.

4.1.3.4.4 Diseño de Explotación

En este punto se diseñan los reportes analíticos según la información que necesita analizar el usuario final. Estos reportes se caracterizan por facilitar la obtención de información en forma rápida y oportuna llegando incluso, si se quisiera, a un mayor detalle (drill down o uso de filtros). Por lo tanto son aquellos reportes que servirán de ayuda para el proceso de toma de decisiones.

Las actividades relacionadas al Diseño de Explotación son:

- **Diseño de Reportes:** Diseñar reportes según las necesidades de los usuarios.
- **Diseño de Tableros de Mando:** Realizar prototipos de la interfaz gráfica de la solución tales como tableros de mando.

4.1.3.5 Construcción

4.1.3.5.1 Evaluación y selección de la plataforma BI de Software Libre

En este punto se evalúa las diferentes plataformas de inteligencias de negocios de software libre existentes, las más representativas, se realiza un análisis comparativo entre ellas, clasificados en herramientas de extracción y explotación de datos, con el objetivo de seleccionar la mejor herramienta para la solución de inteligencia de negocios, luego se debe especificar las aplicaciones a utilizar para cada uno de los componentes mencionados en la etapa “Diseño de la arquitectura de la solución” de la fase de Diseño. Para desarrollar esta etapa se propone el siguiente esquema:

- ✓ Análisis Comparativo de Herramientas BI.
 - Herramientas de extracción
 - Herramientas de explotación.
- ✓ Software a utilizar.

4.1.3.5.2 Instalación y Configuración de Software

En esta etapa se muestran las instrucciones a tomar en cuenta para la instalación y configuración de las herramientas de inteligencia de negocios necesarias para el diseño y desarrollo de la base de datos (Data Mart) y la plataforma de Inteligencia de Negocios a utilizar para generar la herramienta BI.

4.1.3.5.3 Construcción de DataMart

En este punto se realiza la creación de la base de datos y las estructuras diseñadas para el soporte del proceso ETL, a través de una herramienta de

diseño y creación de base de datos. Esta base de datos sirve como una estructura intermedia en donde se almacenan los datos que son extraídos de las fuentes de datos, para luego ser procesadas por la herramienta OLAP.

4.1.3.5.4 Ejecución de ETL

Es la sección donde se agrupan una serie de procesos que llevan a cabo tareas relacionadas con la extracción, manipulación, control, integración, limpieza de datos, carga y actualización de los Data Marts, es decir todas las tareas que se hagan desde que se toman los datos de la fuente, hasta que se carguen en el Data Mart, para su posterior utilización por parte del cubo OLAP y reportes.

La etapa de construcción de ETL consta de tres sub etapas principales: extracción, transformación y carga de datos.

1. Extracción: Durante esta sub etapa se siguen los procesos necesarios para obtener los datos que permiten efectuar la carga del modelo físico.

2. Transformación: Es esta sub etapa se siguen los procesos para convertir los datos fuente a fin de calcular las métricas y mantener un formato estándar de los datos.

3. Carga: Durante la carga de datos, se siguen los procesos necesarios para poblar el Data Mart.

Carga de dimensiones:

- ✓ Construir la carga de una tabla dimensional estática. La principal meta de este paso es resolver los problemas de infraestructura que pudieran surgir entre ellos: conectividad, transferencia, seguridad.
- ✓ Construir los procesos de actualización de una dimensión.
- ✓ Construir las cargas de las restantes dimensiones.

Carga de la Tabla de hechos:

Construir la carga histórica de la fact table (carga masiva de datos).

Tablas Temporales

Las tablas temporales se usan para lograr una mayor calidad de datos en el manejo de un gran volumen de datos.

Estrategia de Poblamiento

Como resultado final se obtiene un esquema de poblamiento que es la secuencia ordenada que muestra cómo se realizara todo el proceso de poblamiento que generalmente se realiza a través de una tarea o job. Luego de finalizar este proceso, el Data Mart queda listo para su utilización por parte del cubo OLAP.

4.1.3.5.5 Creación De Cubo

En este punto se realiza la creación del cubo OLAP necesario para el análisis de los datos obtenidos por el proceso ETL en el Data Mart. El

cubo OLAP permitirá la explotación de la información a través de un amplio conjunto de herramientas de consulta y análisis de la información. Las actividades que se desarrollan en esta etapa son las siguientes:

Identificar los orígenes de datos

Consiste en conectarse al Data Mart.

Crear Cubo

- ✓ Seleccionar tablas dimensionales
- ✓ Seleccionar Tabla Hecho y medidas.

Personalizar Cubo

Va por el lado de la creación de jerarquías. De acuerdo a lo que se ha definido en el modelo dimensional.

Procesar Cubo

Este proceso permite el procesamiento de importantes volúmenes de información.

Publicar Cubo

El proceso de creación del cubo OLAP finaliza con la publicación del cubo a través del servidor OLAP.

4.1.3.5.6 Construcción de Interfaces

Se construyen a partir de los diseños realizados en la etapa “Diseño de Explotación“, por lo tanto las interfaces se realizan para visualizar los reportes y tableros de mando.

Las interfaces deben tener las siguientes características:

- **Fáciles de usar:** permitiendo la construcción del reporte analítico usando la técnica “drag and drop”.
- **Tiempo de respuesta adecuado:** para los cambios en la elaboración del mismo el sistema responderá rápidamente.
- **Permitir la autoconstrucción del reporte analítico:** combinando dimensiones y medidas en busca de nuevas relaciones.
- **Permitir la navegación a diferente nivel de detalle:** es importante la posibilidad de analizar la información a diferente nivel de detalle.
- **Permitir funcionalidades adicionales:** como ordenamientos, filtros, etc.

Las secciones recomendadas mínimas que deben formar parte de la interfaz son los siguientes:

a) Componentes del Cubo

Es el lugar en donde se encuentran las dimensiones, jerarquías y medidas proporcionadas por el cubo OLAP, las cuales deben ser arrastradas dentro del área de construcción del reporte. Es la sección a partir del cual se pueden ir creando el reporte analítico.

b) Área de Construcción

En este lugar es donde deben ir apareciendo las dimensiones (filas, columnas y filtros) y medidas.

Esta área cuenta con dos sub áreas:

- **Área de Dimensiones**

Es donde se pueden “arrastrar” o incluir las dimensiones respectivas a partir de la sección de los componentes del cubo la cual se va reflejando en el área de Construcción y Visualización del Reporte Dinámico, se pueden ubicar en 3 secciones:

- ✓ Filas
- ✓ Columnas
- ✓ Filtros

- **Área de Medidas**

Es donde se van incluyendo las medidas desde los componentes del cubo.

c) Visualización de Reporte Dinámico

Es aquí donde se visualizan los datos y debe permitirse la navegabilidad por medio de las jerarquías definidas en cada dimensión o análisis de niveles simples de cada dimensión, medidas y filtros. Generalmente en esta sección se muestra la información a través de tablas de análisis y diversos tipos de gráficos.

d) Barra de Herramientas

Incluye funcionalidad complementaria al área de Construcción y Visualización de Reporte Dinámico como:

- Ordenamientos
- Opción de guardar, editar.
- Exportaciones a hojas de cálculo.
- Impresión de reportes, entre otras.

e) Lista de Reportes o Tableros Predefinidos

Esta sección permitirá al usuario del sistema registrar lo que ha elaborado en el área de construcción a fin de poder volver a analizarlo posteriormente, por lo tanto se visualizaran en esta lista estos archivos y luego al seleccionarlo se podrán visualizar pero ya con los datos actualizados.

4.1.3.5.7 Pruebas

En este punto se analizan las pruebas que permitan detectar errores para su corrección oportuna. El tipo de prueba utilizada es de caja negra, el cual consiste en el ingreso de datos que luego deben responder al resultado esperado sin importar el proceso que se realice con los datos. Estas pruebas se realizan sobre el producto final mediante: pruebas de ejecución de procesos de carga y pruebas de interfaces.

4.1.3.6 Implantación

4.1.3.6.1 Implantación de la Solución

En este punto se deben realizar las actividades destinadas a la implantación de la Solución de Inteligencia de Negocios en un ambiente de producción real, para que los usuarios accedan a trabajar con la información.

La implantación representa el contar con el producto final dentro de las instalaciones del negocio.

Existen varios factores extras que aseguran el correcto funcionamiento del producto, entre ellos se encuentran la capacitación al usuario, el soporte técnico, la comunicación y las estrategias de feedback. Todas estas tareas deben ser consideradas antes que cualquier usuario pueda tener acceso a los Data Marts. Se considera importante la fase de capacitación del usuario, ya que es necesario que los usuarios conozcan cómo utilizar de manera adecuada la solución de inteligencia de negocios implementada.

CAPITULO V: APLICACIÓN DEL MODELO, CASO: EMPRESA SAN ROQUE S.A.

En el presente capítulo, se presenta el desarrollo de la metodología propuesta, cuya finalidad es desarrollar un modelo de inteligencia de negocios como soporte a la toma de decisiones para ello se ha tomado como caso de estudio el área de ventas de la empresa San Roque S.A. Este modelo permitirá conocer las tendencias de los productos en las ventas y suministrar información relevante acerca de determinados indicadores que permiten tomar decisiones pertinentes.

La presente metodología está basada en características de la metodología de desarrollo de Soluciones de BI tradicional de Ralph Kimball, Joseph Curto Diaz y de la metodología Big Data. El ciclo de vida de este proyecto consiste, según lo ya descrito en el capítulo anterior, en seis fases: planificación, modelo del negocio, Análisis, Diseño, Construcción, Implantación. El desarrollo de la solución de inteligencia de negocios será íntegramente desarrollado utilizando la Suit Pentaho en su versión Enterprise, el cual es una herramienta integrada de inteligencia de negocios bajo plataforma Open Source.

5.1 APLICACIÓN DE LA METODOLOGÍA.

5.1.1. FASE 1: PLANIFICACIÓN

5.1.1.1. PLAN DE PROYECTO

Documento Visión del Proyecto:

Título del Proyecto

Modelo de inteligencia de negocio para la toma de decisiones en la Empresa San Roque S.A.

Introducción del Proyecto

El presente documento pretende resumir que objetivos persigue el proyecto, hacia quien va dirigido, quienes son los involucrados, así como los recursos necesarios que se utilizaran en el desarrollo del proyecto.

Objetivos del Proyecto

Los objetivos que se persiguen son los siguientes:

- Generar información analítica para los tomadores de decisiones.
- Presentar información dinámica en forma tabular y gráfica, soportando el análisis dimensional.
- Reducir los tiempos de generación de información consolidada para toma de decisiones.
- Contar con información sólida que brinde un sustento a las propuestas de las promociones.
- Reducir el número de personas involucradas en la toma de decisiones.

Alcance del Proyecto

El alcance considerado para la solución de inteligencia de negocios, involucra el análisis de los datos de las ventas de los productos San Roque en sus diferentes puntos de ventas. Las áreas que recibirán un beneficio directo de la propuesta son el área de ventas de la empresa comercializadora de dulces tradicionales San Roque S.A., específicamente el periodo de tiempo analizado son entre los años 2007 y 2012, también se analiza la información

proveniente de la web site de la organización considerada como fuente externa.

Descripción de los Stakeholders

Los involucrados son los siguientes:

Stakeholder	Cargo	Función
Jorge Piscoya Madueño	Gerente General	Proporciona Información de la problemática de la empresa San Roque S.A. con respecto a la toma de decisiones en los procesos de ventas.
Ana Mendoza Saavedra	Jefe de Costos	Información de la problemática de la empresa San Roque S.A. con respecto al análisis de los costos de las propuestas emitidas en las reuniones de Gerencia.
Víctor Seclén Esquivés	Jefe de Sistemas	Proporciona información con respecto al uso de los sistemas de información, al estado actual de la base de datos y al archivo weblog del servidor de la web site.
Fabiola Díaz	Jefe de Distribución y Ventas	Proporciona información acerca de las fuentes de datos como por ejemplo el nivel de ventas, las promociones lanzadas en el año, etc. Datos que serán utilizados en el

		proyecto.
Mariella Hague	Asesora Comercial	Información de los objetivos estratégicos del negocio, la estructura organizacional y datos e información importante para la realización del proyecto.

Tabla 25: Stakeholders del Proyecto.

Requisitos del Proyecto

a) Equipo del Proyecto

✓ Gestión

Persona	Cargo	Rol	Función
Mariella Hague Asesora Comercial	Jefe de Proyecto	Representante del Negocio.	Coordinar que los recursos organizacionales requeridos por el ejecutante del proyecto, estén disponibles de acuerdo al plan de trabajo propuesto.
Fabiola Díaz Jefe de Distribución y Ventas	Jefe Operativo del Proyecto	Experto del Negocio	Tiene a su cargo la conducción del proceso de captura y validación de los requisitos de negocios, asesorando la decisión sobre qué datos deben ser utilizados en el proyecto

Omar Antonio Sánchez Guevara	Ejecutante de Proyecto	Desarrollador DW/BI	Tiene a su cargo el desarrollo de la solución de inteligencia de negocios.
------------------------------	------------------------	---------------------	----------------------------------------------------------------------------

Tabla 26: Equipo Gestión para el Proyecto.

✓ Tecnologías Información

Persona	Rol	Función
Omar Antonio Sánchez Guevara	Líder Tecnológico	Gestionar el equipo de tecnología y elaborar el producto deseado.
	Analista Dimensional	Identificar requerimientos estratégicos y aplicar la técnica dimensional.
	Especialista de Datos	Conocer la Base de Datos actual.
	Asistente del Proyecto	Generar Documentación.
Víctor Seclén Esquivés Jefe de Sistemas	Líder Técnico	Proporciona acceso a las fuentes de datos internas y externas: BD Transaccional, web institucional, etc.

Tabla 27: Equipo de Tecnología de Información Para el Proyecto.

b) Oportunidad de Negocio

Mantener el liderazgo de la empresa San Roque, incrementando el nivel de ventas mediante estrategias de ventas personalizadas.

c) Restricciones

Se tomará como datos de referencia para el análisis de las ventas del presente proyecto los

que corresponden al año 2007 hasta el año 2012, es decir los datos históricos de ventas de seis años y a los datos provenientes del weblog de la web site del año 2012.

d) Riesgos

- La no disponibilidad de los datos, al no contar con la información necesaria para el desarrollo del presente proyecto por parte de una determinada área de la empresa.
- No contar con la disponibilidad que se requiere del personal involucrado para el proyecto.
- No culminar el proyecto en el tiempo estimado.

e) Objetivos no considerados

- Crear un Data Warehouse o una base de datos centralizada con la información integrada de toda la empresa.
- Realizar minería de datos.

5.1.2. FASE 2: MODELO DEL NEGOCIO

5.1.2.1. COMPRESION DEL DOMINO DEL NEGOCIO

5.1.2.1.1 La Organización

Nombre de la Organización

San Roque S.A

Ubicación

Carretera Panamericana Norte Km. 780 en la ciudad de Lambayeque.

Actividad Económica

Producción y comercialización de dulces tradicionales.

Producto

Cuentan con una gran variedad de productos entre los cuales destaca el King Kong, alfajor de gran tamaño y de diversos sabores.

Misión

Producir y comercializar bienes del sector alimentos, con altos estándares de calidad para el mercado Mundial, buscando satisfacer la necesidad del consumidor y contribuir al proceso de desarrollo del Perú. Ser una empresa flexible al cambio en función a los requerimientos del mercado Nacional e Internacional.

Visión

Mantener el liderazgo de nuestros productos entre los dulces tradicionales del Norte del Perú, con reconocimiento a nivel internacional; constituyéndonos como una empresa moderna, con tecnología de punta, innovación permanente, personal calificado y con una alta cultura de responsabilidad, comprometidos con la gestión eficiente de sus recursos para la sostenibilidad de la empresa.

Estructura Organizacional

La Estructura Organizacional, está conformada por: el Gerente General, la Gerente Administrativa y Financiera, que tiene a su cargo las áreas de Compras, Contabilidad, Recursos Humanos, Sistemas, Tesorería y Almacén; el Gerente de Producción, Ing. Oscar Linares Rosas, a cargo de Producción, Investigación y Desarrollo, y Control de Calidad; y la Gerente Comercial, que tiene a su cargo las áreas de Distribución, Marketing, Almacén de Productos Terminados (APT), Costos, Supervisión de Tiendas y Exportación. En la siguiente figura se ilustra el organigrama de la empresa.

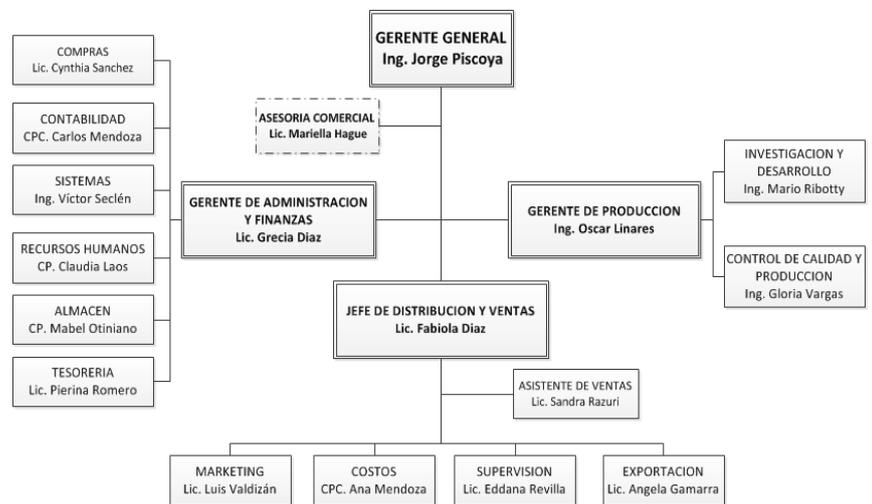


Figura 43: Organigrama de la empresa San Roque S.A.

FODA

En las siguientes tablas se muestra el análisis interno y externo de la organización.

FORTALEZAS	DEBILIDADES
Se diseña la imagen publicitaria y empaques de los productos.	Desorganización
Introducción de productos nuevos al mercado.	Atención deficiente al cliente
Estudios de mercado antes de introducir un producto.	Personal no identificado con la empresa
Los precios relacionados directamente a la calidad ofrecida de los productos.	Poco conocimiento de la competencia
Comercialización directa de los productos.	Personal con perfil académico no adecuado a su puesto de trabajo.
Sólida presencia de los productos en el mercado.	No hay investigación y desarrollo
Marca de reconocido prestigio.	Marketing y Publicidad inadecuados
Control de calidad.	Áreas gerenciales desempeñados y ocupados por dueños no permitiendo la intervención de terceros.
Abastecimiento del insumo principal (fabrica propia).	Obstaculización en la toma de decisiones.
Pioneros en la producción de King Kong	Escucha al personal pero estas ideas no son puestas práctica.
	Escaso número de lanzamientos de promociones de productos y publicidad.
	Falta de implementación de equipos.
	Aun no logra el posicionamiento de sus productos en el sur del país.
	El equipo requiere ser modernizado de acuerdo a las exigencias del mercado.

Tabla 28: Análisis Interno De San Roque S.A.

OPORTUNIDADES	AMENAZAS
Expandir el negocio, abriendo sucursales o anexos.	Resistencia al cambio tecnológico
Otorgamiento de incentivos en los precios, a aquellos clientes que adquieran volúmenes importantes del producto.	Exigencia de nuevos estándares tecnológicos
Otorgamiento de recursos para optimizar mayor producción del producto.	Posibilidad de competencia, con similar producto a bajo precio.
Otorgamiento de incentivos al personal con la finalidad de obtener un mejor producto.	Utilización de estrategias publicitarias indebidas de los competidores.

El Gobierno Provincial de Lambayeque apoya el Desarrollo de las empresas de producción y comercialización de dulces.	BOICOT de los competidores en los productos exitosos, para restar su calidad.
Incremento del turismo en el departamento de Lambayeque	Posibilidad de pérdida, si es que, se efectúan ventas a crédito sin respaldo.

Tabla 29: Análisis Externo De San Roque S.A.

Objetivos Estratégicos

- Posicionarse como líderes en el mercado nacional e internacional, en el segmento de dulces tradicionales, con nuestro producto King Kong.
- Participar en eventos gastronómicos para que el King Kong sea considerado como un producto tradicional en virtud a las características, ingredientes y procesos de producción que aún se mantiene hasta nuestros días.
- Desarrollar nuevas estrategias en el proceso de ventas que generen expectativas de beneficios en los clientes de la empresa San Roque.
- Realizar capacitaciones para el personal que desempeñan un cargo dentro de la empresa.
- Implementar programas laborales para ser una fuente generadora de trabajo para la gran familia lambayecana.

Mapa Estratégico

En la siguiente tabla se muestra el mapa estratégico de la empresa San Roque S.A, el cual se realiza a través de la clasificación de estrategias por perspectiva:

PERSPECTIVA	ESTRATEGIA	CODIGO
CLIENTES	Programa de capacitación para atención al cliente	C1
	Programa de publicidad	C2
	Acreditación de calidad	C3
FINANCIERA	Programa de inversiones en investigación y desarrollo.	F1
	Atracción de inversiones en infraestructura.	F2
	Programa de Incremento del volumen de ventas	F3
	Programa de disminución de costos.	F4
	Programa de Incremento de la cantidad de ventas.	F5
PROCESOS	Ordenamiento de funciones de los trabajadores	P1
	Manual de Normas y Procedimientos	P2
	Programa para la innovación	P3
	Establecer estrategias de ventas	P4
	Programa para brindar servicio post venta	P5
APRENDIZAJE	La satisfacción del empleado	A1
	Productividad del empleado	A2
	La retención del empleado.	A3
	Programa de rotación de puestos (personal operacional)	A4

Tabla 30: Mapa Estratégico San Roque S.A.

En la siguiente tabla se muestra los Indicadores y Metas de la organización.

CODIGO	INDICADOR	META
C1	Nº de cursos	4 cursos anuales
C2	Nº de campañas publicitarias	4 campañas anuales (TV, impresos)
C3	Acreditación de calidad	Etapas 1
F1	Inversiones en Investigación y desarrollo	10% anual
F2	Atracción de inversiones para infraestructura	Emitir bonos o acciones
F3	Incrementar las ventas	80% de lo normal
F4	Disminución de costos	25% de lo normal
F5	Incrementar cantidad	60% de lo normal
P1	MOF actualizado	Unidad de mando, empowerment
P2	MNP	Empleados sustituibles fácilmente
P3	Encuestas sobre nuevos productos y servicios	3 anuales

P4	N ^a de estrategias de ventas	Aumentar 60% de lo normal
P5	N ^a de reclamos de los clientes	Disminuir 10% anual
A1	N ^o de deserciones	Disminuir 10%
A2	N ^a de cumplimiento de metas	Aumentar 10%
A3	N ^a de deserciones	10% de lo normal
A4	N ^o de rotaciones	04 anuales por persona

Tabla 31: Indicadores Y Metas De San Roque S.A.

El objetivo estratégico definido por la empresa: “Desarrollar nuevas estrategias en el proceso de ventas que generen expectativas de beneficios en los clientes de la empresa San Roque”, se realiza a través de:

- Identificar el volumen de ventas.
- Identificar la cantidad de ventas.
- Establecer estrategias de ventas (propuestas de promoción).
- Analizar los costos de las estrategias, este último con el objetivo de identificar la rentabilidad de la propuesta.

Las estrategias de ventas para la empresa San Roque S.A, son las promociones (descuentos de producto en una determinada fecha, asociación de productos, etc), realizadas por el área de marketing en conjunto con la información brindada por el área de ventas, la cual dicha información es imprescindible para el proceso de toma de decisiones acerca de que producto o productos deben ser

lanzados a través de una promoción, esto se realiza evaluando las siguientes medidas:

- Total de ventas
- Cantidad de productos vendidos
- Total de costos de producción de las ventas.

Los indicadores obtenidos para el presente proyecto a partir del plan estratégico, a través de sus estrategias y perspectivas son los siguientes:

Meta	Medidas	Estado	Valores		
Incrementar Volumen de ventas en un 80% con respecto al año anterior.	Total de Ventas	Bueno		>88%	
		Regular	80%	88%	
		Malo	<80%		
Incrementar Cantidad de ventas en un 60% con respecto al año anterior.	Cantidad de productos vendidos	Bueno		>65%	
		Regular	55%	65%	
		Malo	<55%		
Disminuir los Gastos generales al 25% del volumen de ventas con respecto al año anterior.	Total de Costos de producción de las ventas	Bueno		>30%	
		Regular	20%	30%	
		Malo	<20%		

Tabla 32: Perspectiva Financiera.

Meta	Medidas	Estado			Valores
		Bueno	Regular	Malo	
Incrementar el número de estrategias de ventas a un 65% con respecto al año anterior.	Cantidad de estrategias de ventas ejecutadas.	Bueno		>70%	
		Regular	60%	70%	
		Malo	<60%		

Tabla 33: Perspectiva Proceso.

Según la información obtenida a través de las entrevistas realizadas mencionadas en el primer capítulo, el número promedio de estrategias ejecutadas (número de promociones) en los años 2011 al 2012 es de 5.

Mercado

La empresa San Roque, no tiene un segmento establecido de clientes, sí se puede mencionar, pero según la distribución de sus puntos de ventas dentro del país las preferencias están en el paladar de la costa norte y centro del Perú.

Entorno Competitivo

Los principales competidores de la empresa SAN ROQUE S.A son:

- Dulceria y Pasteleria Castañeda
- King kong Lambayeque
- King kong La piuranita
- King kong Virgen de La Puerta
- King kong Tradicion Lambayecana
- King kong La Española

- King kong Llampayec E.I.R.L
- King kong Bruning
- King kong Sipan

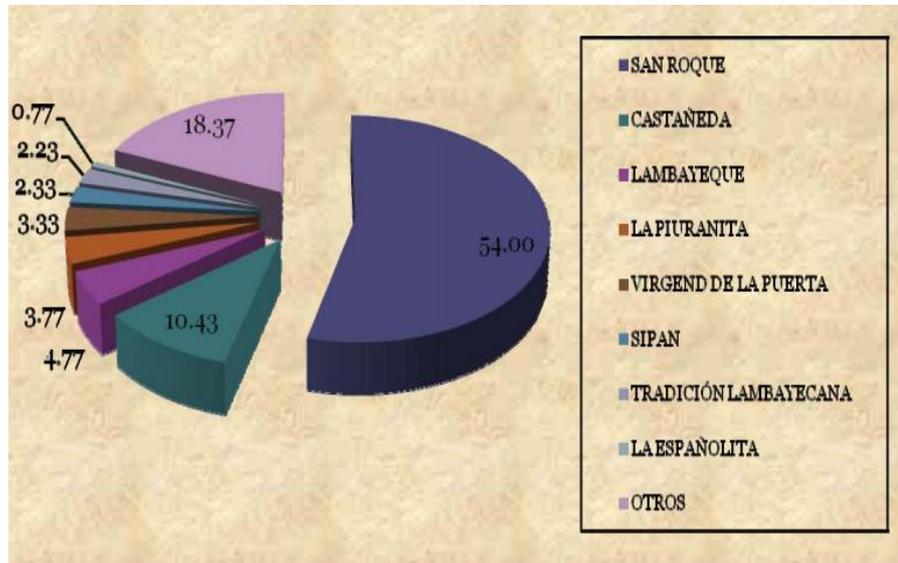


Figura 44: Participación en el Mercado.

Fuente: Jefe de Distribución y Ventas. Fabiola Díaz Orbegoso.

El área temática, es el área de ventas de la entidad; la realización de este proyecto tiene por objetivo automatizar el procedimiento que realizaban los usuarios, para obtener los reportes que contienen información importante, para la toma de decisiones en el área de ventas de la empresa. La realización de este proyecto ofrece a los usuarios una solución a sus necesidades de información, brinda información confiable, rápida y de fácil acceso.

5.1.2.1.2 Modelo De Negocios Decisional Actual

Actores Organizacionales del Proceso Decisional

En este punto se identifican los actores que intervienen directamente, denominados actores primarios, en el proceso decisional. Luego se identifican a los actores secundarios, que son los actores que cumplen roles de mucho valor para proporcionar la información o el conocimiento que requieren los actores primarios. En la siguiente tabla se detalla a los actores involucrados en el proceso.

Rol	Personal	Tipo
Gerente General	Jorge Piscoya	Primario
Asesora Comercial	Mariella Hague	Secundario
Jefe de Distribución y Ventas	Fabiola Díaz	Secundario
Contadora de Costos	Ana Mendoza	Secundario
Jefe de Sistemas	Víctor Seclén	Secundario

Tabla 34: Listado de actores involucrados en el proceso decisional.

Información para la Toma de Decisiones

Aquí se descubre la información que requiere el actor primario para poder tomar las decisiones:

- ✓ Análisis de la influencia en el nivel de ventas presentado por la Jefe de

Distribución y Ventas a la Asesora Comercial.

- ✓ Análisis de los costos de las propuestas, presentado por la Contadora de Costos al gerente general.

Tecnología utilizada para tomar Decisiones

La tecnología utilizada para tomar decisiones son únicamente los reportes del sistema de información emitidas a través de hojas de cálculo (Excel) y, lo cual actualmente esta tecnología genera demora al analizar los costos de la posibles promociones siendo el tiempo de análisis de hasta 4 días promedio.

Descripción Del Proceso de toma de decisiones actual

Se realizan reuniones todos los días martes que tienen por duración de 3 horas aproximadamente. Se establece las posibles promociones de ventas a lanzar, a través de la técnica “Brainstorming”, con opiniones de los involucrados pero sin basarse en datos históricos de ventas, el número promedio de propuestas establecidas por reunión es de tres. Desde que se inicia la idea hasta que se lanza el producto al mercado demora entre dos a tres meses.

Luego estas propuestas se envían a la asesora comercial, la cual solicita la información necesaria a utilizar para el análisis de las propuestas al jefe de distribución y ventas quien se encarga del obtener la información a través de una solicitud al área de sistemas, luego la asesora comercial, envía esta información para el respectivo análisis de costos. La contadora de costos es la encargada de analizar la propuesta de promoción, donde el tiempo utilizado es de 4 días en promedio por propuesta, esta propuesta es analizada en primer lugar según la rotación y ventas del producto o los productos de la promoción por tiendas en los dos últimos años y los meses anteriores.

Toda esa información la consigue utilizando los reportes del sistema el cual se obtiene a través de hojas de cálculo, analizando los datos obtenidos de la rotación y venta del producto para establecer de esta manera el posible descuento de la promoción calculando si la rentabilidad satisface al indicador ya determinado.

Una vez definido el análisis de los costos, las propuestas son evaluadas y seleccionadas para realizar el informe general para ser presentada en la siguiente reunión de Gerencia para tomar la decisión si la promoción se lanza al mercado o si se

cambia a otra propuesta de promoción, si es la primera opción se enviará el informe a la asesora externa de marketing para la publicidad respectiva, a logística para el pedido de envases e insumos y a producción para que se encargue de la elaboración de la promoción, caso contrario se repite el proceso de análisis de la contadora de costos, hasta que la promoción sea la aceptada por Gerencia. El gerente general es quien se encarga de informar los resultados de la propuesta analizada. Las promociones realizadas no son muy frecuentes ya que se han realizado un promedio de 5 promociones por año.

Las actas de reunión revisadas indican que las horas invertidas en el proceso de análisis de las propuestas es de 3 horas al día en un periodo de 4 días a la semana como ya se mencionó anteriormente.

El personal requerido específicamente para el proceso de toma de decisiones es de cinco: Gerente General, Asesora Comercial, Jefe de Distribución y Ventas, Jefe de Sistemas y Contadora de Costos.

Todo el proceso de elaboración de la promoción del producto es de tres meses pero el proceso de toma de decisiones una

vez lanzada las propuestas es de 4 días en promedio por propuesta.

En la siguiente figura se describe el proceso de manera gráfica luego en la siguiente tabla se muestra los indicadores de gestión del proceso:

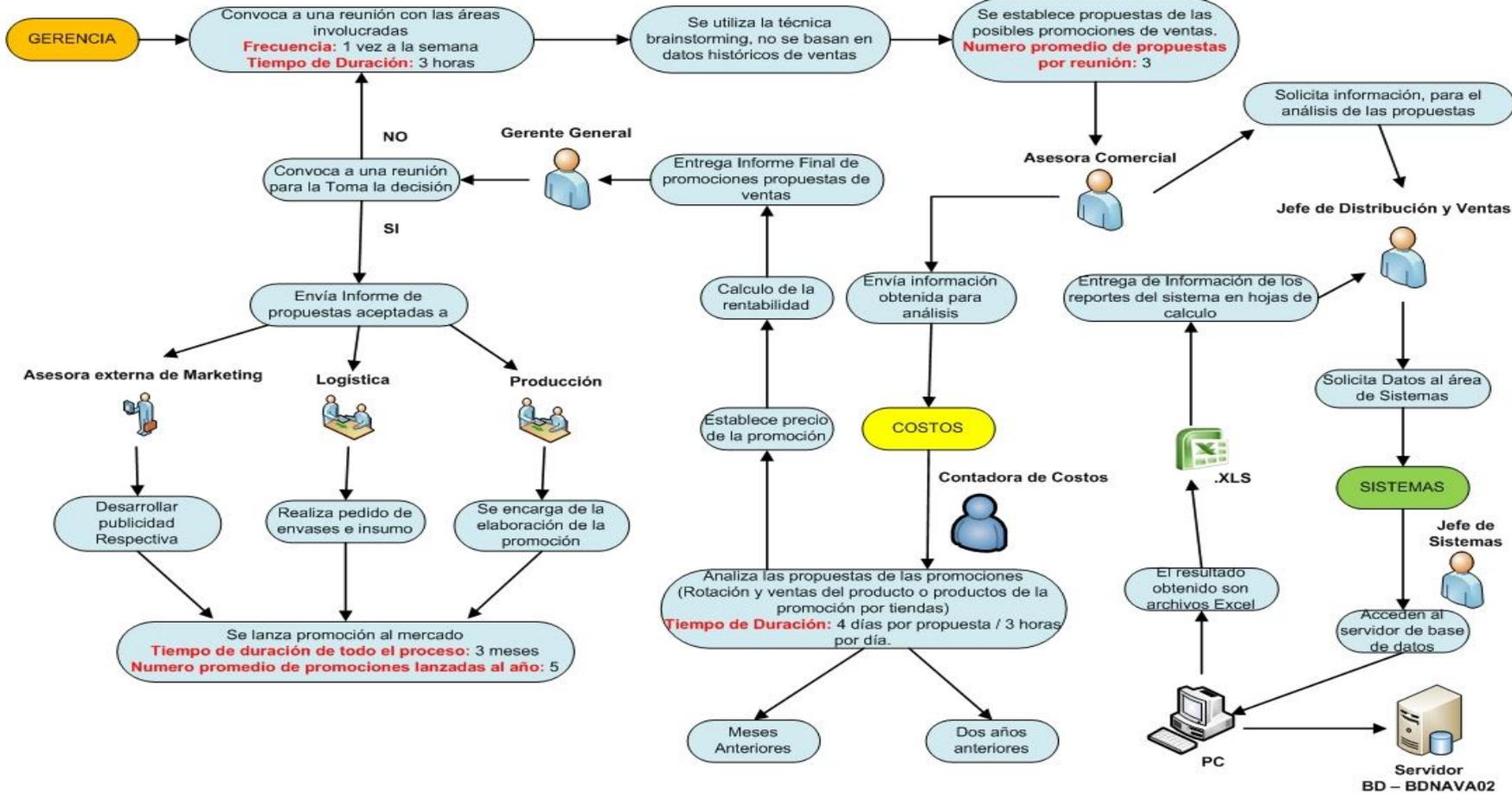


Figura 45: Proceso de toma de decisiones actual de las propuestas de promociones de productos de ventas y lanzamiento al mercado.

INDICADORES	INSTRUMENTO	FORMULA	UNIDADES DE MEDIDA	APLICACIÓN DE FORMULA	PERIODO	RESULTADOS
Tiempo de respuesta en formulación de propuestas por reunión.	Cronometro.	$RFP = \frac{(HIL + HID)}{NR}$	Hora.	$RFP = \frac{(2 + 1)}{1}$	Semana	3 Horas
Índice de tiempo de respuesta en el proceso de análisis por propuesta formulada.	Actas de reunión.	$PAP = \frac{((HIP * DT) / JL) * 100}{100}$	%	$PAP = \frac{((3 * 4) / 48) * 100}{100}$	Semana	25
Numero de propuestas formuladas por reunión.	Actas de reunión.	-----	Escalar.	-----	Semana	3

Tabla 35: Indicadores de gestión del proceso de toma de decisiones.

Leyenda:

- ✓ **RFP:** Tiempo de respuesta en formulación de propuestas por reunión.
- ✓ **HIL:** Horas invertidas en lanzamiento de ideas.
- ✓ **HID:** Horas invertidas en discusiones de las ideas.
- ✓ **NR:** Número de reuniones.
- ✓ **PAP:** Índice de tiempo de respuesta en el proceso de análisis por propuesta formulada.
- ✓ **HIP:** Horas invertidas en el proceso de análisis.
- ✓ **JL:** Jornada laboral.
- ✓ **DT:** Días trabajados

Resultados actuales de la toma de decisiones

La eficacia de la toma de decisiones en la empresa no es adecuada ya que utilizan la técnica “Brainstorming” para definir las promociones de ventas, con opiniones de los involucrados pero sin basarse en datos históricos de ventas.

La mayoría de las decisiones son erróneas, ya que el tiempo que demoran en tomar decisiones es demasiado largo, desde que se inicia la idea hasta que se lanza el producto al mercado demora entre dos a tres meses, por lo tanto el tiempo invertido antes del lanzamiento de las promociones, que mayormente son en meses por fechas

festivas, es amplio, esto a su vez origina escasas promociones de ventas lanzadas anualmente. Lo cual genera insatisfacción por parte de los involucrados en el desarrollo del proceso de toma de decisiones. Es por esto que se propone como solución un modelo de inteligencia de negocios que brinde a los usuarios acceso a la información con muy poca demora y que pueda ser visualizada en un entorno amigable y de fácil uso, para que cuenten con información consolidada y disponible que atienda sus requerimientos de información y que de esta manera puedan tomar decisiones con una base de información sólida y confiable.

5.1.3. FASE 3: ANÁLISIS

5.1.3.1 DEFINICIÓN DE REQUERIMIENTOS Y ANÁLISIS DE LA DATA

5.1.3.1.1. FUENTES DE INFORMACIÓN INTERNAS

a) Plan Estratégico

Los indicadores obtenidos a partir del plan estratégico son los siguientes:

Indicador	Medidas
Indicador Cantidad de ventas	Cantidad
Indicador Total de	Precio Unitario

ventas	Valor de Venta Soles Valor de Venta Dólares Total Soles Total Dólares
Indicador Total de Costos	Costo Soles Costo Dólares

Tabla 36: Indicadores de Gestión del Proyecto.

b) Entrevistas

Dentro de las entrevistas realizadas al personal de gestión de la empresa tenemos los siguientes resultados:

✓ Jefe de Distribución y Ventas / Asesora Comercial:

Objetivos

- Mejorar el Análisis del estudio de estrategias de ventas basándose en datos históricos.
- Identificar asociaciones en los productos según la información de ventas.

Indicadores

- Indicador de Cantidad Vendida
- Indicador de Monto Vendido
- Indicador Total de costos.

Medidas

- Precio Unitario
- Cantidad

- Costo Soles
- Costo Dólares
- Valor de Venta Soles
- Valor de Venta Dólares
- Total Soles
- Total Dólares

Dimensiones

- Cliente
- Producto
- Tiempo
- Punto de Venta

✓ Jefe del Centro de Informática y Sistemas (Especialista de Datos)

Análisis de la Data

- Existe dependencia del área de TI, por parte del personal de gestión para obtener los reportes necesarios para el análisis de ventas.

Disponibilidad de la Data

- Se encuentra en línea el 100% de la data.
- Usa como RDBMS al SQL Server 2008

Calidad de la Data

- Integridad de Dominio.
- Integridad Referencial
- Integridad Entidad

- Manejo de NULL: es consistente se puede observar que los campos obligados han sido correctamente definidos con la posibilidad de obligar siempre a grabar un dato dentro del mismo.

c) Reportes de Gestión

- Reporte de Total de Ventas Por Año

San Roque S.A		Fecha: 16 - 05 -12			
RESUMEN DE TOTAL DE VENTAS POR AÑO					
SUCURSAL: LAMBAYEQUE					
TIENDA: FABRICA					
Categoría	Productos	2009	2010	2011	2012
KingKong					
	Piña, Maní y Manjarblanco	6500	2000	1500	10000
	Manjarblanco	9750	3000	2250	15000
	Piña y Manjarblanco	3250	1000	750	5000
	TOTAL	19500	6000	4500	30000
Barras					
	Manjarblanco	9750	3000	2250	15000
	Piña y Manjarblanco	7800	2400	1800	12000
	Lúcuma	5200	2600	1200	8000
	Chirimoya	6500	3200	1600	11000
	TOTAL	29250	11200	6850	46000
Latas					
	Lata Turística	13000	4000	3000	15000
	Manjarblanco	9750	3000	2250	12000
	Natilla	12500	2300	4250	8000
	TOTAL	35250	9300	9500	35000
Chocotejas					

Figura 46: Muestra de Cuadro de Total de Ventas Por Año.

Fuente: Jefe de Distribución y Ventas. Fabiola Díaz Orbegoso.

- Reporte de total de cantidad vendida por año.

San Roque S.A		Fecha: 23 - 06 -12			
RESUMEN DE TOTAL DE CANTIDAD VENDIDA POR AÑO					
SUCURSAL: CHICLAYO					
TIENDA: REAL PLAZA					
		2009	2010	2011	2012
Categoría	Productos				
KingKong					
	Piña, Maní y Manjarblanco	3500	930	2300	7500
	Manjarblanco	2400	2200	740	8200
	Piña y Manjarblanco	1300	1500	690	9500
	TOTAL	7200	4630	3730	25200
Barras					
	Manjarblanco	400	6300	1500	9000
	Piña y Manjarblanco	560	4500	4500	8000
	Lúcuma	750	2300	3400	6000
	Chirimoya	2400	2400	3500	9000
	TOTAL	4110	15500	12900	32000
Latas					
	Lata Turística	5000	5600	4500	13000
	Manjarblanco	7500	4500	3500	10000
	Natilla	3450	3400	5250	9500
	TOTAL	15950	13500	13250	32500
Chocotejas					

Figura 47: Muestra de Cuadro de Total de cantidad vendida por Año.

Fuente: Jefe de Distribución y Ventas.
Fabiola Díaz Orbegoso.

d) Base De Datos Transaccional

El acceso a esta base de datos, fue autorizado por el Gerente General de San Roque S.A. En la siguiente tabla, se muestran las tablas involucradas en el proceso de ventas de la base de datos BDNVA02.

Nombre de Tabla	Descripción Tabla
tbl01dis	Distrito
tbl01pro	Provincia
tbl01dep	Departamento
tbl01zon	Zona
mst01fac	Ventas
dtt01fac	Detalle de Venta
mst01cli	Cliente
tbl01fam	Familia
tbl01sbf	Subfamilia
tbl01grp	Grupo
tbl01alm	Almacén
tbl01pto	Puntos Venta
tbl_tienda	Tienda
tbl_sucursal	Sucursal
tbl01tcl	Tipo Cliente
tbl01cac	Categoría Cliente
tbl01act	Actividad Cliente
prd0101	Padrón de Productos

Tabla 37: Tablas involucradas en el proceso de ventas de la base de datos.

Para este proceso fue necesario disponer de los datos que se recolectan en el proceso de ventas. Todos estos datos, están localizados en la base de datos en SQL: BDNAVA02 que contiene datos de las ventas, desde Enero del 2007 a Diciembre del 2012 tal y como se muestra en la siguiente figura:

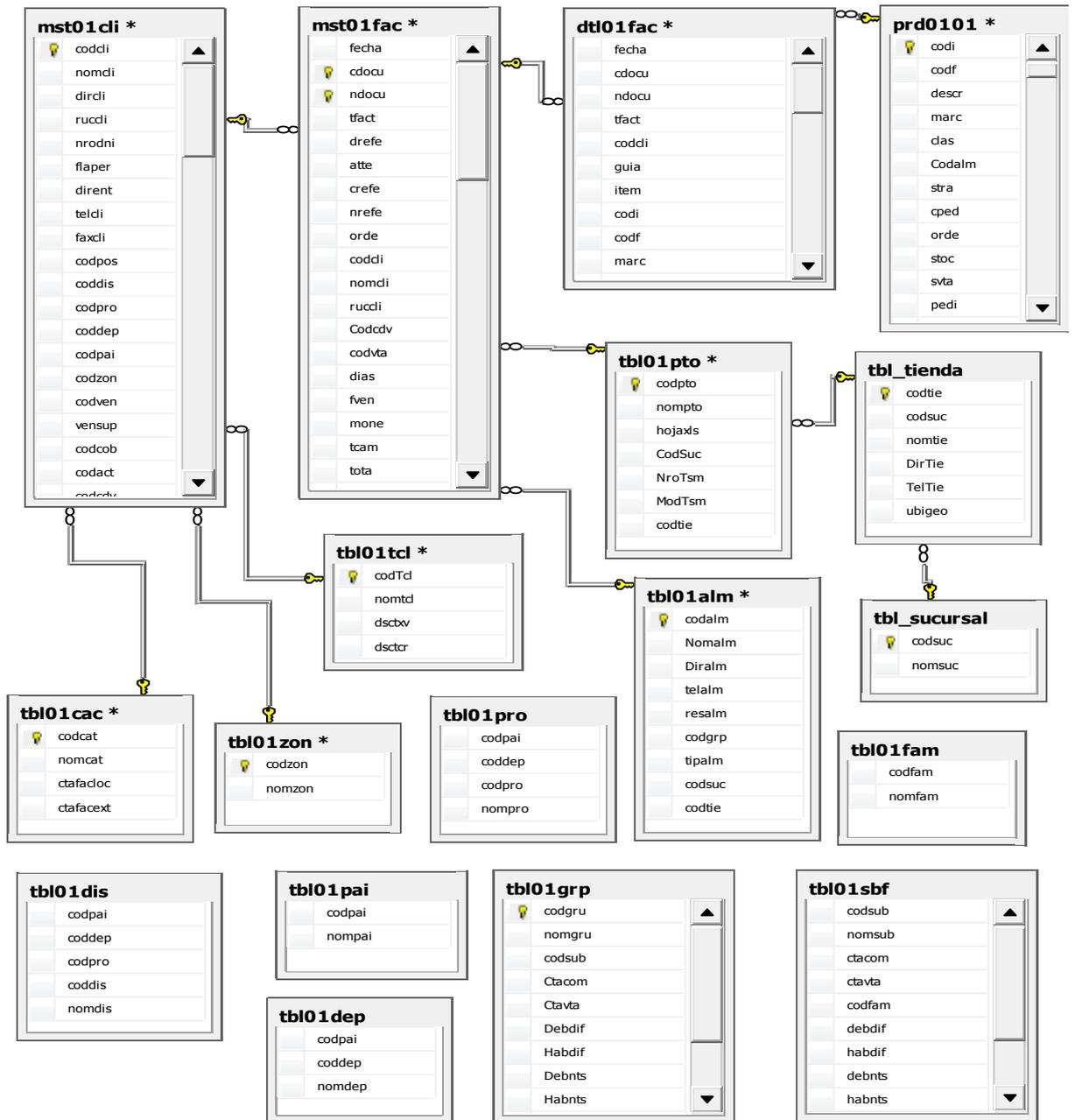


Figura 48: Base de Datos BDNVA02 (Proceso Ventas).

5.1.3.1.2. FUENTES DE INFORMACIÓN EXTERNAS

La fuente de información externa a considerar en el presente proyecto es el archivo weblog del servidor web de la WebSite de la empresa San Roque S.A.

Ventajas de analizar el weblog

Permite responder preguntas como estas:

- ✓ ¿He mejorado mis visitas con respecto al mes anterior?
- ✓ ¿Cuántas visitas he recibido a través de buscadores?
- ✓ ¿A través de qué palabras claves me encuentra la gente? Es decir que palabra clave o Keyword ha escrito un usuario en un motor de búsqueda para llegar a nuestra web.
- ✓ ¿Cuáles son las mejores palabras claves?
- ✓ ¿Desde qué sitios llegan visitas a mi página web? Sitios web de referencia (Qué páginas enlazan a su web).
- ✓ ¿Qué secciones de mi página web son las más visitadas: productos nuevos, promociones, etc.? Es decir cuáles son las páginas más populares y más visitadas, esto nos permitirá saber cuál de nuestras páginas es más rentable.
- ✓ ¿Cuántas personas visitan la Web?
- ✓ ¿Cómo orientar la empresa hacia un perfil de cliente determinado?
- ✓ ¿Cómo conseguir un mejor posicionamiento?
- ✓ ¿Desde qué países se realiza la visita?
- ✓ ¿Cuántas visitas he tenido por días, semanas, meses, etc?

Todas las respuestas que se encuentren ayudan a optimizar las estrategias de Ventas y Marketing y a la web, con el objetivo de seguir incrementando ingresos y rentabilidad.

Medidas o Indicadores

Algunas de las posibles medidas a obtener a partir del análisis del web log son:

- ✓ Número de visitas.
- ✓ Número de páginas vistas de la web.
- ✓ Numero de Url previas.

Requerimientos Organizacionales

El objetivo estratégico es el siguiente: **“Desarrollar nuevas estrategias en los procesos de ventas que generen expectativas de beneficios en los clientes de la empresa San Roque”**, exactamente lo que se requiere es obtener información que permita tomar decisiones eficaces destinados a, **“Mejorar la elaboración de las estrategias de ventas de la empresa San Roque”**.

Elementos que se relacionan con el problema definido, son las relacionadas con las propuestas de promociones que están involucradas en los procesos de ventas. Lo que se quiere lograr es optimizar el proceso de ventas en la empresa San Roque. Para lograr ello es necesario definir requerimientos específicos del modelo de inteligencia de negocio propuesto.

Requerimientos del Modelo de Inteligencia de Negocios

- Información acerca del volumen de ventas de productos en un determinado periodo de tiempo y lugar.
- Información acerca de los productos más vendidos por periodo de tiempo.
- Información de los productos que más les interesan a los usuarios de la web de la organización.
- Información acerca de los clientes que han generado más ventas.
- Reducir el tiempo para la toma de decisiones.
- Contar con información sólida que brinde un sustento a las propuestas de las promociones.
- Reducir el número de personas involucradas en la toma de decisiones.
- Presentar información dinámica en forma tabular y gráfica, soportando el análisis dimensional.
- Reducir los tiempos de generación de información consolidada para toma de decisiones.

5.1.3.2 ESTRUCTURACIÓN Y MODELADO DE DATOS DE FUENTES EXTERNAS

Para el desarrollo de esta etapa se trabajó con la suite de BI Open Source Pentaho.

5.1.3.2.1. CREACIÓN DE LA TRANSFORMACIÓN PENTAHO MAPPER

A continuación se describen a manera referencial cada uno de los pasos o “steps” que se ejecutan en esta transformación:

- a) **MapReduce Input:** En este paso se determina el tipo de dato para los campos clave/valor.
- b) **Split Fields:** Este paso sirve para crear y determinar los tipos de dato para los campos que recibirán los datos del archivo log, es decir se prepara la estructura que tendrá la data.
- c) **Filter Rows:** Permite filtrar las filas de cada campo basándose en condiciones y comparaciones. Permite asignar un campo por defecto a cada campo.
- d) **Split URL:** Permite delimitar cada campo por caracteres de delimitación.
- e) **Hadoop File Input - Section:** Se realiza la división de los datos desde la fuente en este caso el archivo log a través de un separador que se haya identificado en el archivo log, en este caso el guion (-), para encontrar los campos que se pueden obtener del archivo log. Luego se selecciona los campos a utilizar para realizar el mapeo en el siguiente paso, con los datos definidos en el paso Split Fields, a cada campo se le asigna un nombre.

#	Field1	Field2	Field3	Field4	Field5	Field6
1	1.514.812.370	05/11/2013:00:00:41	0800	GET	/order/?ref=002 HTTP/1.1	200 554
2	1.514.812.370	151.48.123.70	0800	GET	/order/?ref=002 HTTP/1.1	200 554
3	1.514.812.370	151.48.123.70	0800	GET	/order/?ref=002 HTTP/1.1	200 554
4	1.514.812.370	05/11/2013:00:00:42	0800	GET	/img/tm_home.gif HTTP/1.1	200 554
5	1.514.812.370	05/11/2013:00:00:42	0800	GET	/img/tm_sep.gif HTTP/1.1	200 411
6	1.514.812.370	05/11/2013:00:00:42	0800	GET	/img/tm_ssp.gif HTTP/1.1	200 733
7	1.514.812.370	05/11/2013:00:00:42	0800	GET	/img/tm_dl.gif HTTP/1.1	200 690
8	1.514.812.370	05/11/2013:00:00:43	0800	GET	/img/tm_sup.gif HTTP/1.1	200 597
9	1.514.812.370	05/11/2013:00:00:43	0800	GET	/img/tm_on.gif HTTP/1.1	200 656
10	1.514.812.370	05/11/2013:00:00:43	0800	GET	/img/tm_pp.gif HTTP/1.1	200 787
11	1.514.812.370	05/11/2013:00:00:43	0800	GET	/img/tm_cu.gif HTTP/1.1	200 654
12	1.514.812.370	05/11/2013:00:00:43	0800	GET	/img/shull.gif HTTP/1.1	200 411

Figura 49: Vista Previa de los datos pre procesados.

- f) **Stream Lookup:** En este paso se mapean todos los datos tanto el campo clave como los campos normales.
- g) **User Defined Java Expression:** Es necesario para concatenar campos y convertirlo en clave y de esta forma generar nuevos valores.
- h) **MapReduce Output:** Sirve para seleccionar la nueva llave y el nuevo valor.

En la siguiente figura se muestra el esquema final de Pentaho Mapper.

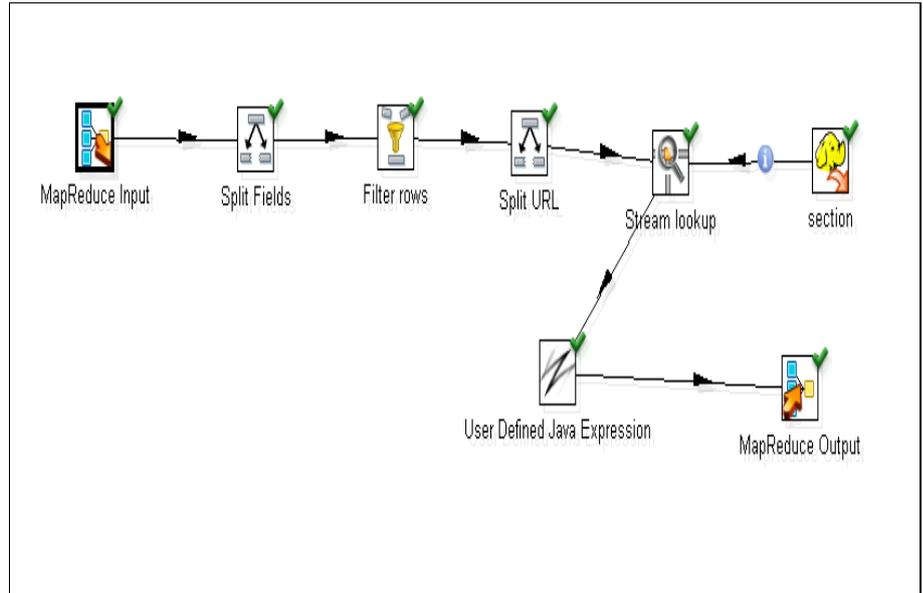


Figura 50: Esquema Final de Pentaho Mapper.

5.1.3.2.2. CREACIÓN DE LA TRANSFORMACIÓN PENTAHO REDUCER

A continuación se describen a manera referencial cada uno de los pasos o “steps” que se ejecutan en esta transformación:

- a) **MapReduce Input:** Reduce los registros ordenados en un orden de clave. Nuevamente se define los tipos de datos para los campos clave/valor.
- b) **Group By:** Permite sumar los valores de cada llave.
- c) **User Defined Java Expression:** Nuevamente se vuelve a concatenar como en la transformación anterior.
- d) **MapReduce Output:** Nuevamente se define los valores para llave y valor.

En la siguiente figura se muestra el esquema final de Pentaho Reducer.

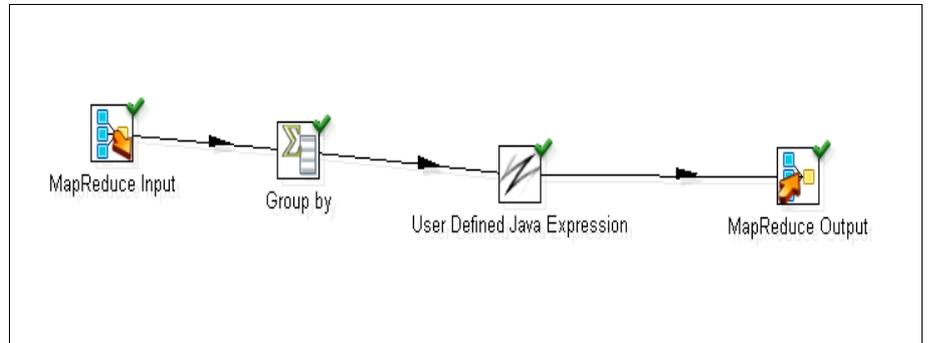


Figura 51: Esquema Final de Pentaho Reducer.

5.1.3.2.3. CREACIÓN DEL JOB PARA EJECUTAR EL PROCESO MAP REDUCE

Esta tarea o trabajo ejecuta las transformaciones de mapeo y reducción. En la siguiente figura se muestra el esquema final del Job Map Reduce.

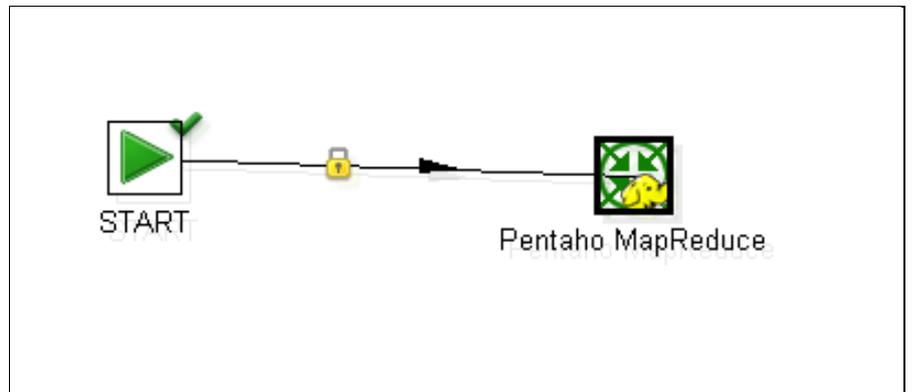


Figura 52: Esquema Final del Job Map Reduce.

5.1.3.2.4. CREACIÓN DEL JOB PARA CARGAR LOS DATOS DESDE HADOOP A HIVE:

Ya que Hadoop (HDFS) no tiene las herramientas necesarias para modelar los datos (creación de base de datos y tablas)

de manera adecuada, se exporta los datos a Hive, el DataWarehouse de Hadoop, para ello primero se tiene que crear tablas en Hive para que almacenen todos los datos encontrados en hadoop y que se encuentran almacenados como campos en HDFS, luego se procede a ejecutar este Job, una vez que los datos se cargan en las tablas, se podrán ejecutar instrucciones HiveQL para consultar estos datos. En la siguiente figura se muestra el esquema final del Job Hadoop - Hive.

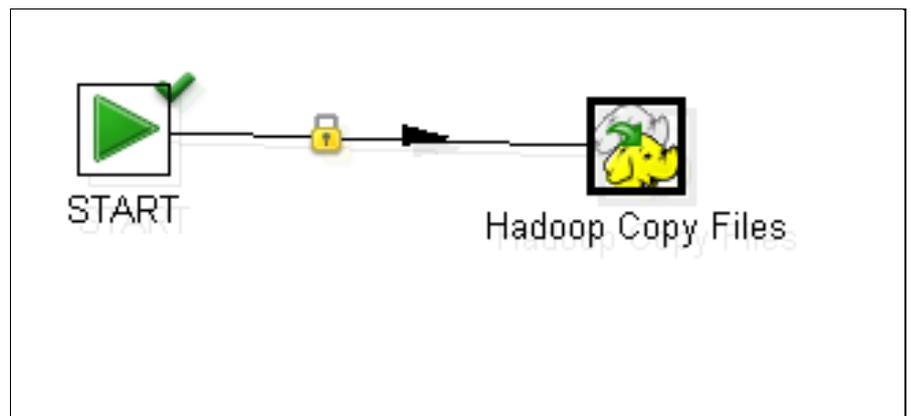


Figura 53: Esquema Final del Job Hadoop - Hive.

5.1.3.2.5. CREACIÓN DEL JOB PARA EXTRAER LOS DATOS DESDE HIVE AL RDBMS:

Una vez modelado los datos en Hive se envían estos datos al gestor de base de datos relacional (RDBMS) en el cual se procede a estructurar la data obteniendo como resultado la base de datos relacional final. Para realizar esto primero se crea una base de datos y las tablas necesarias en el RDBMS, luego una conexión a Hive y al

RDBMS que en este caso sería el repositorio del gestor de base de datos MySQL Workbench, y a través de los steps table input y output se realiza la exportación de los datos para cada tabla, luego ya en MySQL se estructura la data de manera relacional. Luego queda todo listo para aplicar el análisis dimensional. En las siguientes figuras se muestra las configuraciones y el esquema final de este último Job llamado Hive - RDBMS.

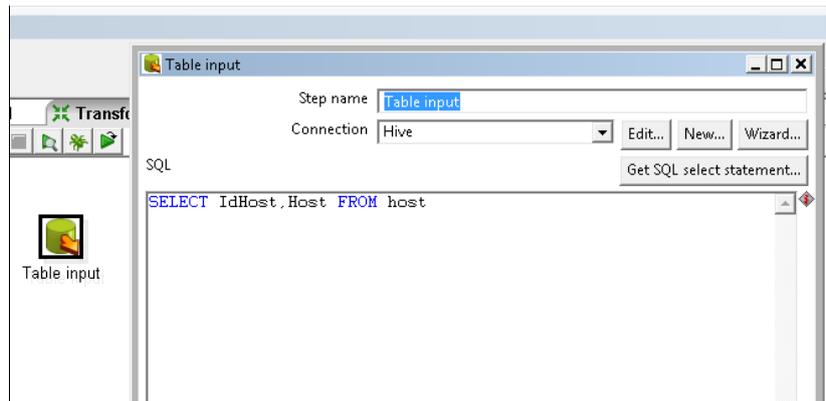


Figura 54: Configuración de los datos de entrada desde HIVE.

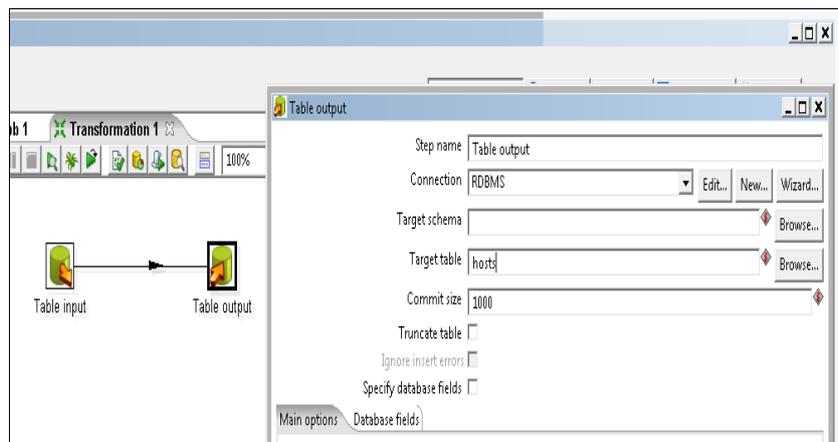


Figura 55: Configuración de los datos de salida hacia cada tabla del RDBMS.



Figura 56: Esquema Final del Job Hive - RDBMS.

5.1.3.2.6. MODELO DE DATOS RELACIONAL FINAL

En la siguiente figura se muestra el modelo de datos relacional final como resultado de esta etapa, conteniendo los datos ya estructurados en la BD llamada weblog:

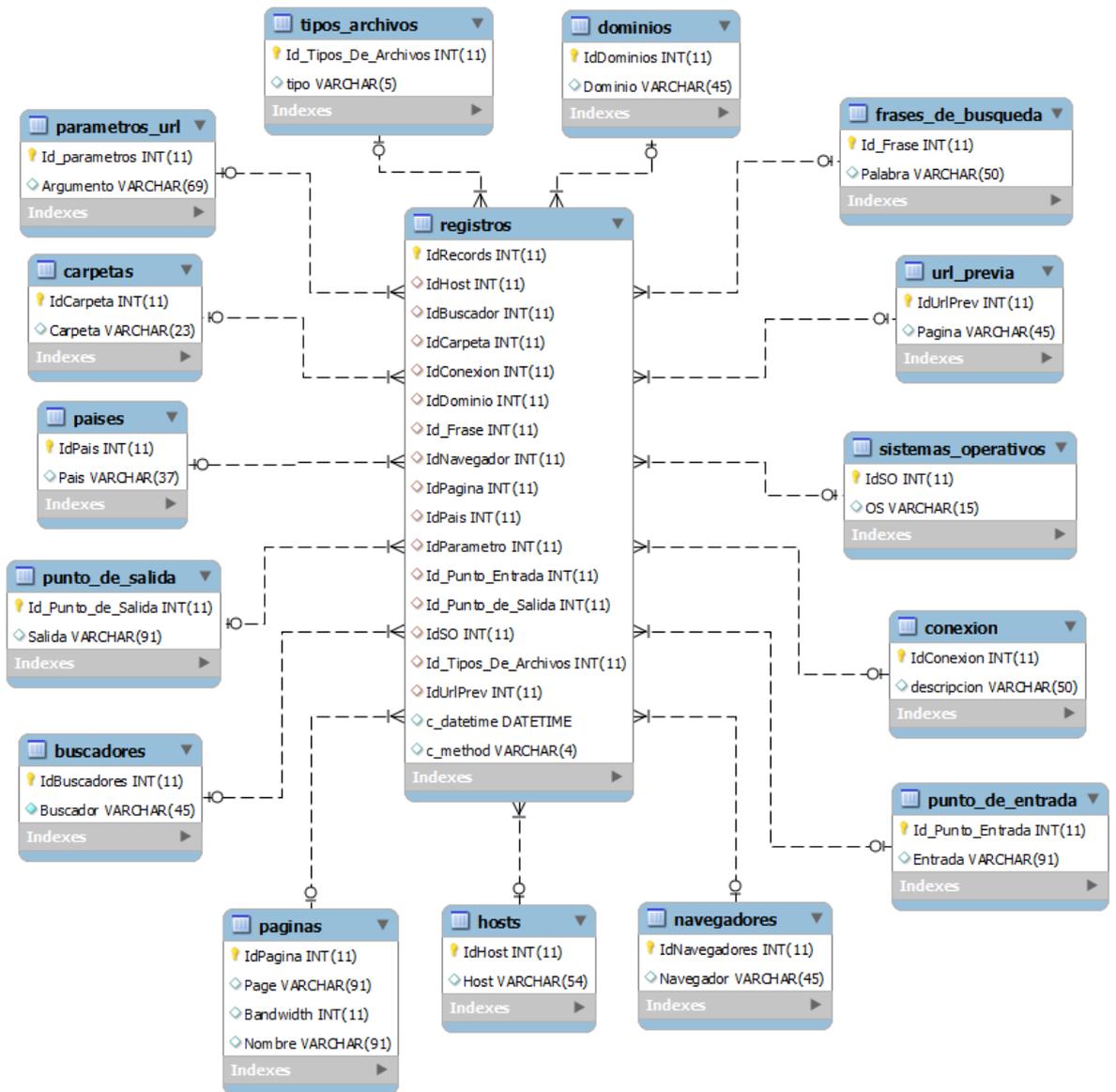


Figura 57: Base de Datos Relacional WebLog.

Las tablas que forman parte de la base de datos WEBLOG son:

- buscadores
- carpetas
- conexión
- dominios
- frases_de_busqueda

- hosts
- navegadores
- paginas
- países
- parámetros_url
- punto_de_entrada
- punto_de_salida
- registros
- sistemas_operativos
- tipos_archivos
- url_previa

5.1.3.3 ANALISIS DIMENSIONAL

A continuación a modo de ejemplo se realiza el análisis al detalle de la fuente interna y luego se muestra el resultado del análisis dimensional final tanto para la fuente interna como externa.

a) Resumen Hoja de Gestión

✓ **Proceso**

Proceso de Ventas

✓ **Objetivo**

- Mantener el liderazgo de la empresa San Roque, incrementando el nivel de ventas mediante estrategias de ventas personalizadas.

✓ **Estrategia**

- Identificar y analizar los indicadores que influyen en el nivel de ventas de los productos de la empresa San Roque S.A.

✓ **Indicadores**

- Cantidad Vendida
- Total de Ventas
- Total de Costos

Indicador	Medidas	Estado Valores			
		Bueno	Regular	Malo	
Indicador Cantidad Vendida	Cantidad	Bueno		>65%	Green
		Regular	55	65%	Yellow
		Malo	<55%		Red
Indicador Total de Ventas	Precio Unitario	Bueno		>88%	Green
	Valor de Venta Soles	Regular	80	88%	Yellow
	Valor de Venta Dólares	Malo	<80%		Red
Indicador Total de Costos	Total Soles	Bueno		>30%	Green
	Total Dólares	Regular	20%	30%	Yellow
	Costo Soles	Malo	<20%		Red

Tabla 38: Hoja de Gestión del Proyecto.

b) Hoja de Análisis

Proceso: Ventas

• Medidas

- ✓ Precio Unitario
- ✓ Cantidad
- ✓ Costo Soles
- ✓ Costo Dólares
- ✓ Valor de Venta Soles
- ✓ Valor de Venta Dólares
- ✓ Total Soles

✓ Total Dólares

• Dimensiones y Niveles

Dimensiones	Niveles
Cliente	Cliente, Distrito, Provincia, Departamento, País, Zona, Actividad, Tipo de Cliente, Categoría, Condición, Crédito, Descuento
Producto	Producto, Grupo, Tipo, Sub-Familia, Familia
Tiempo	Fecha, Mes, Año, Semestre, Trimestre, Bimestre, Día Semana, Fin de Semana
Punto de Venta	Punto de Venta, Tienda, Almacén, Sucursal

Tabla 39: Hoja de Análisis Ventas.

c) Cuadro de Dimensiones y Jerarquías

Proceso: Ventas

Dimensiones	JERARQUIAS							
	Nivel 1	Nivel 2	Nivel 3	Nivel 4	Nivel 5	Nivel 6	Nivel 7	Nivel 8
Cliente	Cliente	Tipo de Cliente	Actividad	Categoría				
	Distrito	Provincia	Departamento	Zona	País			
	Condición	Crédito	Descuento					
Producto	Producto	Tipo	Grupo	Sub-Familia	Familia			
Tiempo	Fecha	Día Semana	Fin de Semana	Mes	Bimestre	Trimestre	Semestre	Año
Punto de Venta	Punto de Venta	Tienda	Almacén	Sucursal				

Tabla 40: Cuadro Dimensiones y Jerarquías del Proyecto.

d) Cuadro de Dimensiones y Medidas**Proceso: Ventas**

Medidas	DIMENSIONES			
	Cliente	Producto	Tiempo	Punto De Venta
Precio Unitario	X	X	X	X
Cantidad	X	X	X	X
Costo Soles	X	X	X	X
Costo Dólares	X	X	X	X
Valor de Venta Soles	X	X	X	X
Valor de Venta Dólares	X	X	X	X
Total Soles	X	X	X	X
Total Dólares	X	X	X	X

Tabla 41: Cuadro de Medidas y Dimensiones del Proyecto.

e) Análisis Dimensional Final

De acuerdo al análisis del cuadro de dimensiones y medidas se puede deducir el siguiente análisis dimensional final:

Gestión Ventas

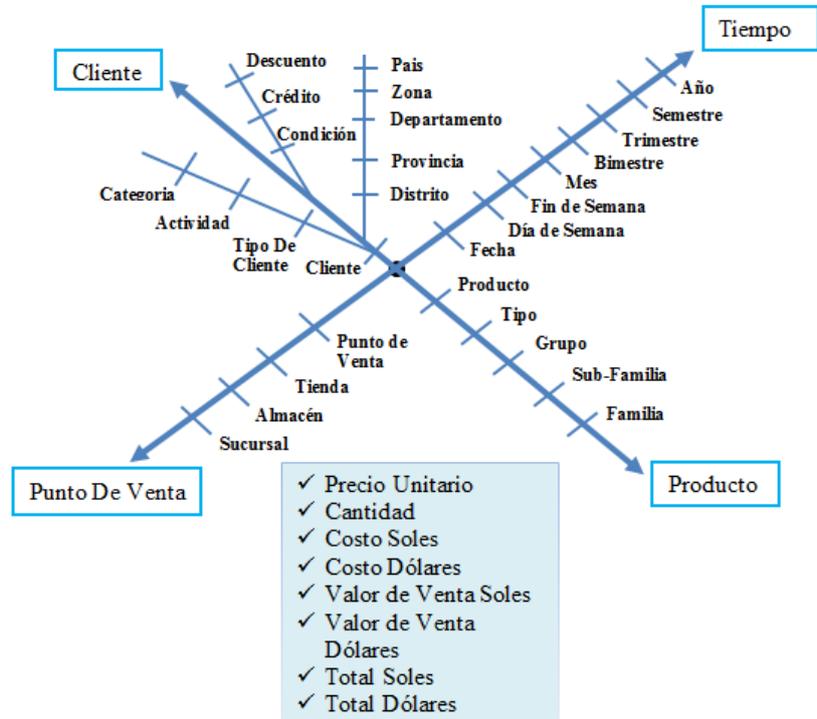


Figura 58: Análisis Dimensional Ventas.

En la siguiente figura se muestra el resultado final del mismo proceso de análisis dimensional realizado anteriormente pero esta vez aplicado a los datos obtenidos de la web site de la empresa:

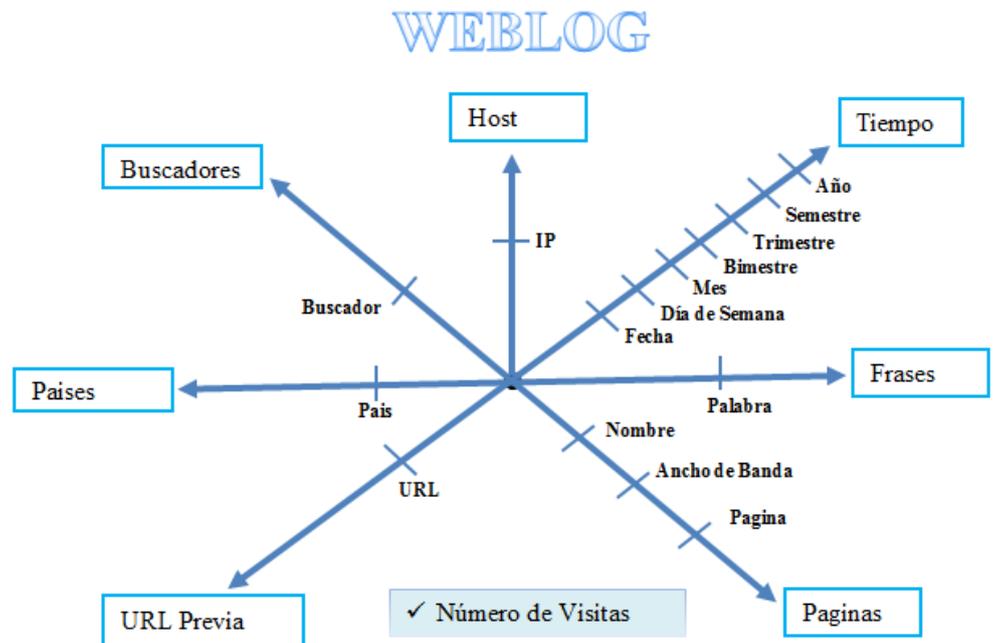


Figura 59: Análisis Dimensional WebLog.

5.1.4. FASE 4: DISEÑO

5.1.4.1. DISEÑO DE LA ARQUITECTURA DE LA SOLUCIÓN

La arquitectura a seguir en la solución del presente proyecto es la arquitectura ROLAP, la cual soporta agregación, cálculos y categorización de datos de bases de datos relacionales estándares, como Oracle, MySQL, Microsoft SQL Server, IBM DB2, y otros.

Se descarta utilizar la arquitectura MOLAP debido a que limita la cantidad de datos. Dado que el proyecto está orientado a una comercializadora de dulces tradicionales a nivel nacional dicha arquitectura no soporta el volumen de información requerido.

La arquitectura de la solución para el presente proyecto está compuesta por la fuente de datos interna para el Data Mart que en este caso será la

base de datos transaccional de la organización. Para los procesos ETL se realizarán todas las operaciones y transformaciones que necesitan los datos previamente a ser almacenados al Data Mart, también permitirá integrar en un mismo repositorio (Data Mart) a los datos estructurados y a los datos no estructurados proveniente de la fuente externa WebLog, para cumplir con este objetivo de estructurar la data en una base de datos relacional, se utiliza herramientas de Big Data.

Los usuarios tendrán acceso a los cubos OLAP para el análisis de la información interna y externa mediante herramientas de explotación.

En la siguiente figura se muestran los componentes que intervienen en el diseño de la arquitectura de la solución. Se muestra cada componente y la interacción con el resto. Es la arquitectura donde descansará el modelo propuesto:

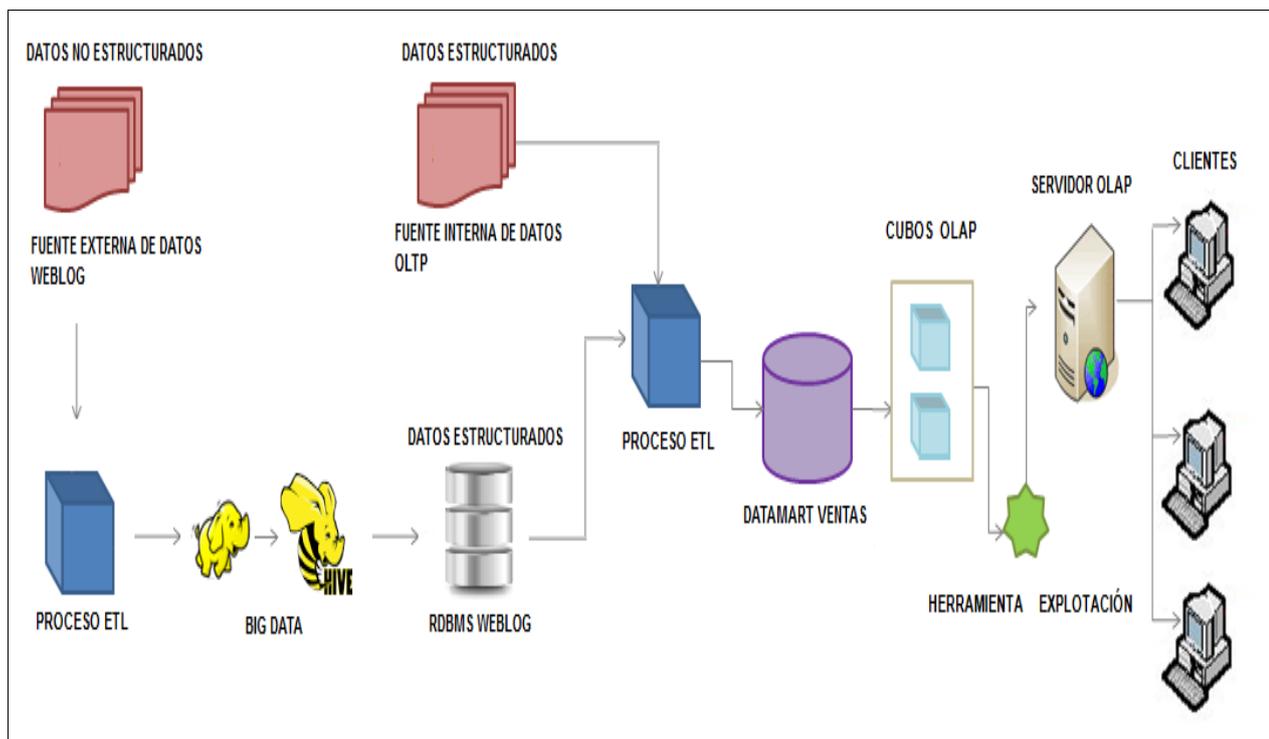


Figura 60: Arquitectura de la solución de Inteligencia de Negocios.

Componentes

A continuación se detalla cada uno de los componentes a utilizar en el presente proyecto:

a) Fuente de datos externa WEBLOG

WebLog proviene de WEB (Nos indica el uso de una página web) y LOG: hace referencia a un diario o bitácora. Este archivo WebLog de la web site de San Roque S.A es un archivo de registro Web creado y mantenido automáticamente por el servidor web. Cada "Hit" para el sitio Web, incluyendo cada vista de un documento HTML, imagen u otro objeto, se registra. Contiene información acerca de quién estaba visitando el sitio, de dónde venían, y que es exactamente lo que estaban haciendo en el sitio web. Se almacenan datos como la dirección IP, navegador, fecha, etc., de cada acceso al mismo.

b) Big Data

Nos permitirá a través del componente Hadoop procesar la gran cantidad de datos proporcionada por el weblog. A su vez el componente Hive nos permitirá modelar los datos.

c) RDBMS WEBLOG

Permite almacenar los datos ya modelados en una base de datos relacional estructurada para su posterior análisis dimensional y poblamiento del DataMart.

d) Fuente de datos interna OLTP

La fuente de datos interna será la base de datos BDNVA02 del sistema transaccional de la empresa, el cual trabaja con el gestor de base de datos SQL SERVER 2008.

e) Procesos ETL

El proceso ETL (Extracción, Transformación y Carga) del Data Mart permite subir información hacia las Base de Datos Dimensional, por otra parte también permite realizar la extracción y análisis de los datos de la fuente externa, en este caso el weblog de la web site a través de Big Data.

f) Data Mart Ventas

El Data Mart de ventas desarrollado es una base de datos que contiene las tablas que conforman las estructuras del modelo dimensional diseñado tanto para la información interna como externa. Esta base de datos es cargada mediante la ejecución de los procesos ETL. Optimizada para acceso OLAP.

g) Cubos OLAP

Este componente permitirá el procesamiento de importantes volúmenes de información, llevar a cabo análisis, para su posterior explotación de los datos, que permitan mejorar las operaciones productivas y tomar decisiones inteligentes. Se desarrollara dos cubos OLAP para el análisis de la fuente interna y externa.

h) Servidor OLAP

El servidor OLAP (OLAP Server) procesa consultas MDX, retorna resultados multidimensionales. Puede emitir consultas SQL al RDBMS (Sistema de Gestión de Base de Datos Relacional). Manejo de Cachés optimizado y construcción de agregados para mejorar rendimiento.

i) Herramienta Explotación

El proceso de explotación permite generar reportes dinámicamente, tablas de análisis, tableros de mando y así explotar la información que se tiene procesada.

j) Usuarios

Los usuarios pueden acceder de manera Online a la Aplicación Web que se encuentra en el Servidor. Desde esta aplicación los usuarios pueden acceder a los reportes que contienen el diseño base solicitado, así mismo los usuarios pueden explotar la información modificando la estructurada de estos reportes según lo requieran y las tablas de análisis de análisis proporcionados por la herramienta de explotación de datos creada.

5.1.4.2. DISEÑO DIMENSIONAL FÍSICO

Se usara el modelo estrella dado que su uso permite optimizar el tiempo de respuesta de los reportes en comparación al modelo copo de nieve debido a que resulta ser un modelo más complejo. Para el diseño dimensional físico de la base de datos, se ha decidido utilizar el Mysql Workbench 5.2. A continuación se

muestra al detalle el diseño dimensional físico de la fuente de datos interna, luego al final se muestra el diseño dimensional físico final de la fuente de datos interna y externa.

5.1.4.2.1. Definir el Grano

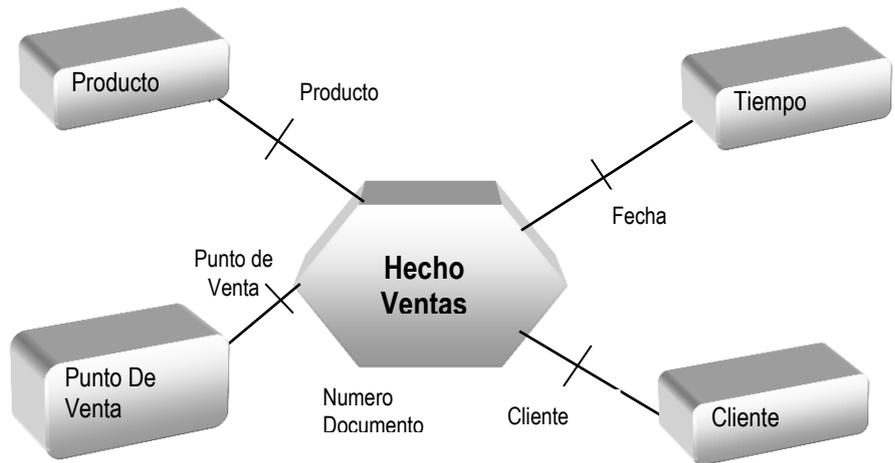


Figura 61: Determinando el grano.

5.1.4.2.2. Diseñar Dimensiones

Dimensión Producto

Contiene información de los productos tales como familia, subfamilia, grupo, etc



Figura 62: Dimensión Producto.

Nivel	Descripción	Ejemplo
Producto	Es la descripción del producto.	King Kong Mini, Galleta Paciencia
Tipo	Indica si es una Promoción.	Promoción, No Promoción.
Grupo	Representa la categoría del producto.	Chocotejas, Alfajores, Suspiros
SubFamilia	Representa la presentación del producto.	King Kong, Potes, Dulcería, Bebidas
Familia	Grupo de productos que se comercializan.	Mercadería

Tabla 42: Descripción Dimensión Producto.

Dimensión Cliente

Contiene información de los clientes tales como actividad, tipo de cliente, categoría, etc



Figura 63: Dimensión Cliente.

Nivel	Descripción	Ejemplo
Zona	Indica la zona del país a la que pertenece el cliente.	Norte, Centro, Oriente
Actividad	Es el rubro en el que se desempeña el cliente.	Supermercados, Persona Natural
TipoCliente	Clasificación del cliente según el volumen de compras.	Nuevo, Regular, Principal

Categoría	Clasificación para los clientes que compran al por mayor y al por menor.	Publico General, Mayorista
Condición	Indica si se le ha brindado crédito al cliente	Crédito 7 días, Contado
Crédito	Indica el monto del crédito otorgado al cliente	300, 110
Descuento	Indica el descuento brindado al cliente	5 , 10

Tabla 43: Descripción Dimensión Cliente.

Dimensión Punto de Venta

Contiene información de los puntos de venta tales como tienda, sucursal, etc.

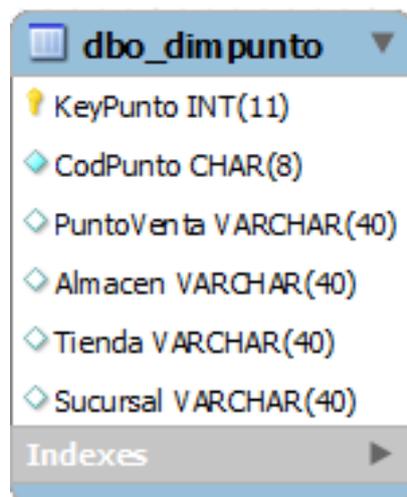


Figura 64: Dimensión Punto de Venta.

Nivel	Descripción	Ejemplo
PuntoVenta	Local donde se vende los productos a los clientes.	Oficina, Pto_SanJose, Pto_Lima

Almacén	Lugar que abastece al punto de venta	San José, Bolognesi
Tienda	Lugar donde se encuentra el punto de venta.	Real Plaza, Open Plaza
Sucursal	Ciudad donde se encuentra la tienda.	Lambayeque, Chiclayo, Trujillo

Tabla 44: Descripción Dimensión Punto de Venta.

Dimensión Tiempo

Contiene información del análisis del tiempo (año, mes, semana, etc).



Figura 65: Dimensión Tiempo.

Nivel	Descripción	Ejemplo
DíaSemana	Día de la semana en que se realiza la operación	7-Domingo, 1-Lunes

Mes	Mes en que se realiza la operación	01 -Enero, 04 - Abril
Año	Año en que se realiza la operación	2007, 2008
Semestre	Semestre en que se realiza la operación	1 -Semestre, 2 - Semestre
Trimestre	Trimestre en que se realiza la operación	3 -Trimestre, 4 - Trimestre
Bimestre	Bimestre en que se realiza la operación	5 -Bimestre, 6 - Bimestre
FinSemana	Es o no fin de semana en que se realiza la operación	SI, NO

Tabla 45: Descripción Dimensión Tiempo.

5.1.4.2.3. La Tabla Hecho

Hecho Ventas

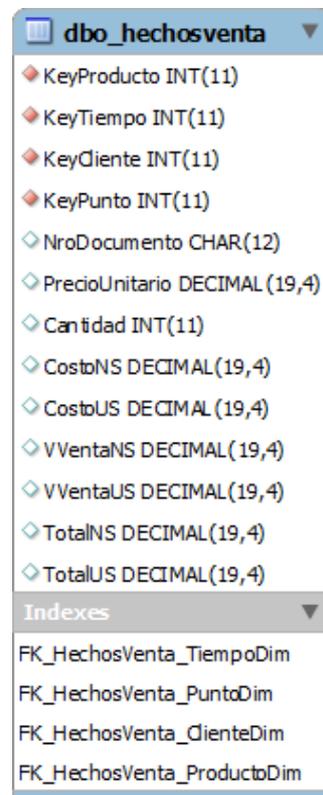


Figura 66: Tabla Hecho Ventas.

Medida	Descripción
PrecioUnitario	Precio de venta del Producto en moneda local
Cantidad	Número de productos vendidos
CostoNS	Monto total de producción en moneda local de los productos vendidos Campo Calculado: (costo de producción del producto x cantidad)
CostoUS	Monto total de producción en dólares de los productos vendidos
VVentaNS	Monto Total bruto de la venta en moneda local de los productos vendidos Campo Pre calculado: (precio unitario x cantidad)
VVentaUS	Monto Total bruto de la venta en dólares de los productos vendidos
TotalNS	Monto Total Neto de la venta en moneda local de los productos vendidos aplicando el IGV (19%). Campo Pre calculado: (Monto Total bruto * 19%)
TotalUS	Monto Total Neto de la venta en dólares de los productos vendidos aplicando el IGV (19%)

Tabla 46: Descripción Tabla Hecho Ventas.

5.1.4.2.4. Diseño Dimensional Físico Completo

El diseño dimensional físico final obtenido de la fuente interna (Base de datos transaccional de ventas) está compuesto por 4 dimensiones, 1 tabla hechos y sus

relaciones. En la siguiente figura se muestra el diseño dimensional final.

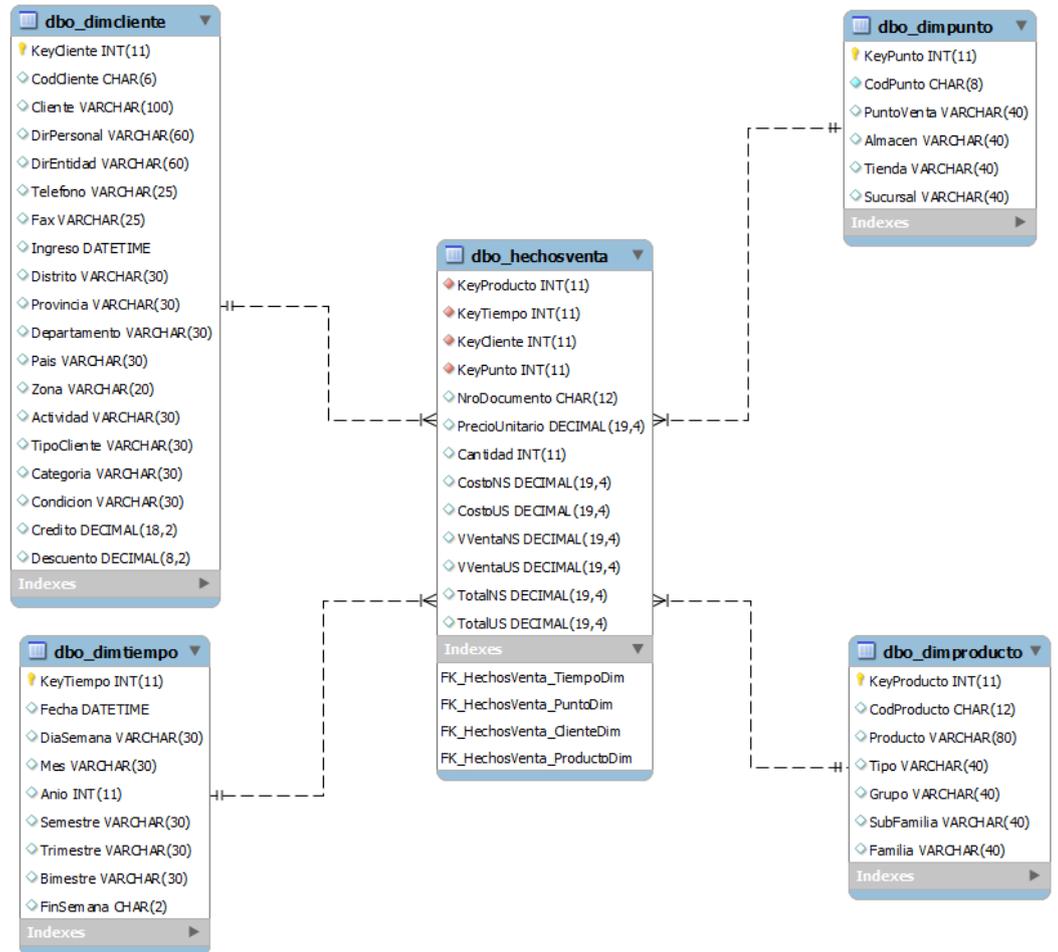


Figura 67: Diseño Dimensión Físico Final de la fuente interna.

Así mismo, el diseño dimensional físico final obtenido de la fuente externa (weblog) está compuesto por 7 dimensiones, 1 tabla hechos y sus relaciones. En la siguiente figura se muestra el diseño dimensional final.

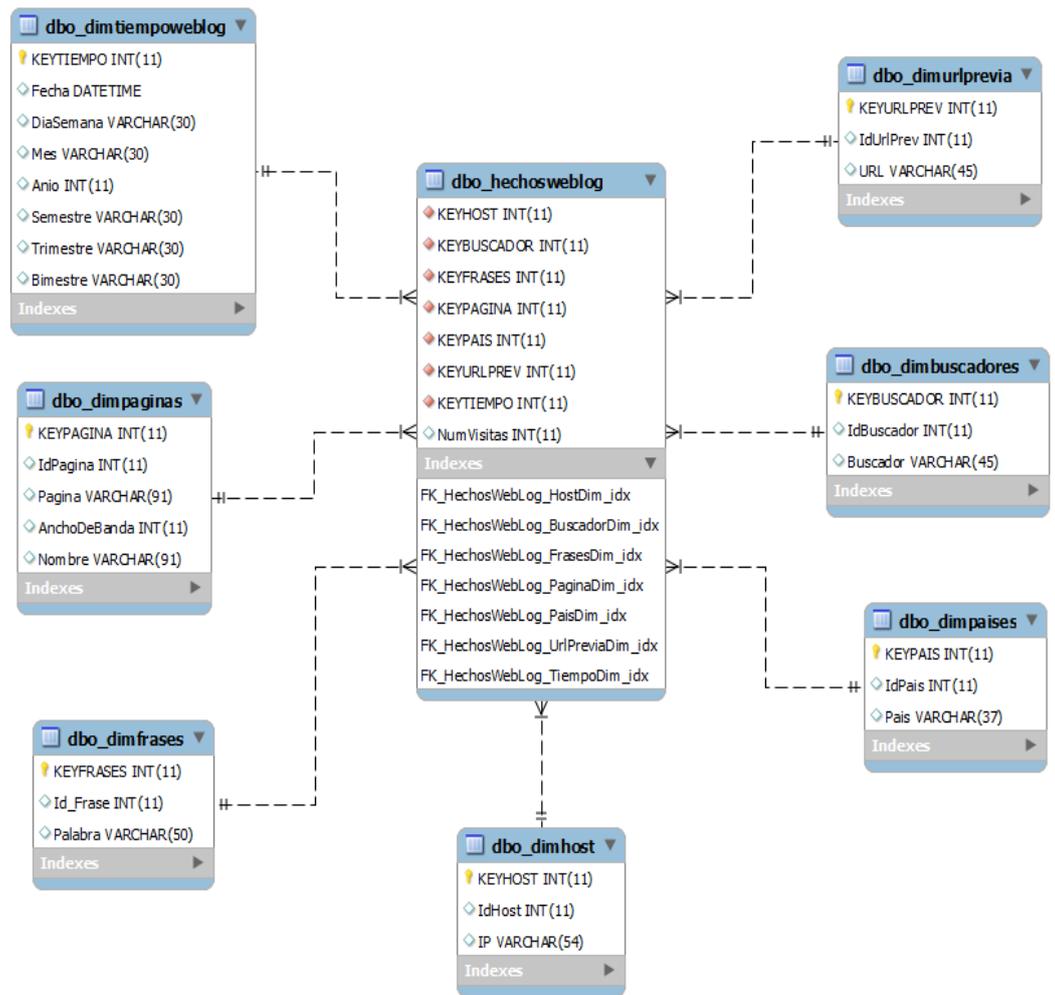


Figura 68: Diseño Dimensión Físico Final de la fuente externa.

5.1.4.3. LIMPIEZA DE DATOS Y DISEÑO DE EXTRACCIÓN

La fuente de datos para el proceso de extracción es la base de datos transaccional BDNVA02 y la base de datos relacional WebLog. A continuación a modo de ejemplo solo se muestra el diseño de extracción y carga de la dimensión producto, tiempo y de la tabla hechos venta correspondiente a la fuente de datos interna:

DIMENSION PRODUCTO

5.1.4.3.1. Descripción de las tablas fuentes

Tipo de Fuente	Nombre de Fuente	Descripción
Base de datos transaccional	prd0101	Padrón de Productos: Esta tabla contiene el listado de productos.
Base de datos transaccional	tbl01grp	Grupo: Esta tabla contiene el listado de los grupos de productos
Base de datos transaccional	tbl01sbf	Subfamilia: Esta tabla contiene el listado de las subfamilias de productos
Base de datos transaccional	tbl01fam	Familia: Esta tabla contiene el listado de las familias de productos

Tabla 47: Tablas Fuentes para la dimensión producto.

5.1.4.3.2. Fuente de Datos

Tabla	Nombre	Descripción	Llave	Tipo	Formato	Consideración Importante
prd0101	codi	Código de	PK	Char(12)	TEXTO	No Nulo

		Producto				
prd0101	descr	Descripción del Producto		Varchar(80)	TEXTO	
prd0101	tipro	Tipo de Producto		Varchar(40)	TEXTO	Puede ser solo "Promoción" o "No Promoción"
tbl01grp	nomgru	Nombre de Grupo		Varchar(40)	TEXTO	
tbl01sbf	nomsub	Nombre de subfamilia		Varchar(40)	TEXTO	
tbl01fam	nomfam	Nombre de Familia		Varchar(40)	TEXTO	

Tabla 48: Fuente de Datos de la dimensión producto.

5.1.4.3.3. Estandarización y Limpieza de Datos

Nombre	Llave	Tipo	Formato	Limpieza	Valor Por Defecto
KeyProducto	PK	Int(11)	Numero entero	No Nulo	0
CodProducto		Char(12)	Texto	No Nulo	
Producto		Varchar(80)	Texto		
Tipo		Varchar(40)	Texto	Solo puede contener "Promoción" o "No Promoción"	
Grupo		Varchar(40)	Texto		
SubFamilia		Varchar(40)	Texto		
Familia		Varchar(40)	Texto		

Tabla 49: Limpieza de Datos de la dimensión producto.

5.1.4.3.4. Mapeo de Datos

Tabla:	dbo_dimproducto	
Campo	Tipo	Mapeo
KeyProducto	Int(11)	Identificador Numérico Autogenerado
CodProducto	Char(12)	prd0101.codi
Producto	Varchar(80)	prd0101.descr
Tipo	Varchar(40)	prd0101.tipro
Grupo	Varchar(40)	tbl01grp.nomgru
SubFamilia	Varchar(40)	tbl01sbf.nomsb
Familia	Varchar(40)	tbl01fam.nomfam

Tabla 50: Mapeo de Datos de la dimensión producto.

5.1.4.3.5. Proceso

El proceso que se realiza para la extracción, transformación y carga de la dimensión producto es la siguiente:

1. **Extracción de los datos de las tablas prd0101, tbl01grp, tbl01sbf, tbl01fam:** Se toma los datos de las tablas OLTP mencionadas y se realiza la transformación de acuerdo al mapeo indicado.
2. **Carga de la Dimensión Producto:** Se transforma e ingresan los datos en la dimensión Producto. En el caso que sean nuevos productos se insertan, en

el caso que sean antiguos productos sólo se actualizan los datos.

El diseño de extracción de la dimensión tiempo y de la tabla hechos venta se encuentra en el ANEXO 18: Diseño de extracción.

5.1.4.3.6. Esquema de Extracción Final

N°	Proceso	Dependencia
1	Limpieza de Datos	
2	Carga de la dimensiones	1
3	Carga de la tabla temporal	2
4	Carga de las tablas Hechos	3

Tabla 51: Esquema de Extracción Final.

5.1.4.4. DISEÑO DE EXPLOTACIÓN

5.1.4.4.1. Diseño de Reportes

Para el análisis se pueden realizar los siguientes reportes:

- ✓ Ventas Realizadas
- ✓ Total de ventas por productos, por subfamilia, por tipos de cliente de una

determinada sucursal, y periodo de tiempo.

- ✓ Cantidad de ventas por producto, por periodo, por tienda
- ✓ Producto más vendido por sucursal y por periodo de tiempo
- ✓ Total de ventas por productos vendidos, por tiendas y por periodo de tiempo
- ✓ Total de ventas por cliente, por periodo, por producto
- ✓ Cantidad de ventas por cliente, por periodo y por producto
- ✓ Reporte de los montos de venta total por moneda local por período de tiempo
- ✓ Reporte de los montos de venta total por moneda en dólares por período de tiempo
- ✓ Reporte de la cantidad vendida por cliente y tipo de cliente en el tiempo.
- ✓ Reporte de Análisis de Ventas por clientes y por el producto A
- ✓ Reporte de Análisis de Ventas por clientes y por el producto B
- ✓ Número de visitas por página web

A manera de ejemplo se realizara el diseño del “reporte de ventas realizadas”, “total de ventas por productos y tipos de clientes” y “Numero de visitas por página web”:

a) Reporte “Ventas Realizadas”

Reporte de ventas realizadas en unidades y monto en soles. Este reporte permite usar filtros por periodo (año) y Sub Familia de producto.

Diseño

Este reporte nos muestra el número de ventas e ingresos por tienda, con esta información se puede identificar que productos tienen demanda y que productos tienen baja rotación por localidad. A partir de ello se puede impulsar las ventas de productos a través de campañas promocionales. A continuación el diseño del reporte de ventas realizadas:

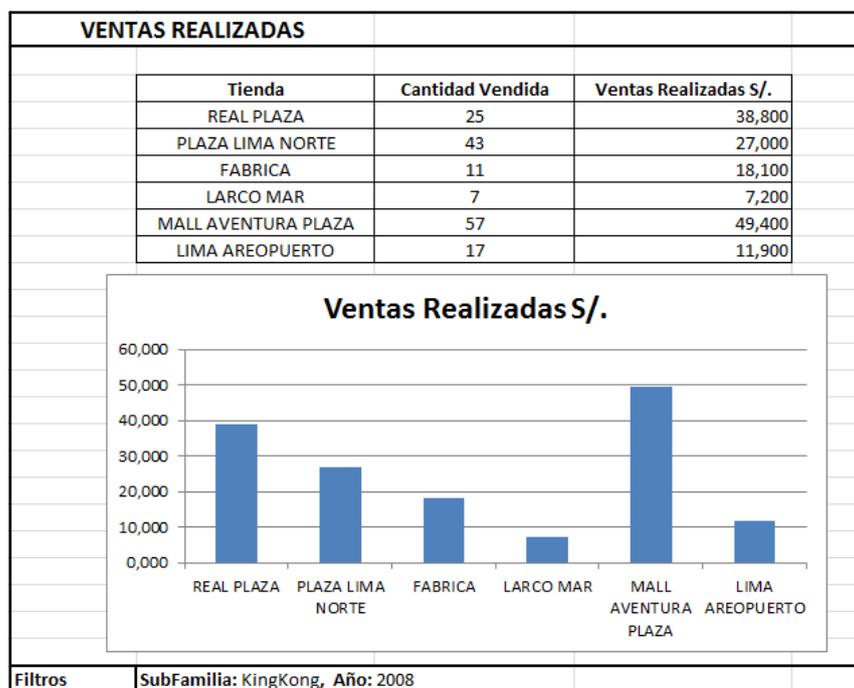


Figura 69: Diseño de reporte de ventas realizadas.

Tipo

El tipo del reporte es gráfico y tabla.

Filas

La tabla no contiene cabeceras por fila.

Columnas

Las columnas que tiene el reporte de ventas realizadas:

No.	Dimensión	Nivel
1	DimPunto	Tienda

Tabla 52: Columnas del Reporte 1.

Medidas

Los cálculos que se realizan para el reporte:

No.	Medida	Formato
1	Cantidad	Numérico
2	TotalNS	Moneda (S/.)

Tabla 53: Medidas del Reporte 1.

Filtros

Los filtros usados en el reporte:

No.	Operación
1	SubFamilia de producto = <SubFamilia de

	producto seleccionado>
2	Año=<Año ingresado>

Tabla 54: Filtros del Reporte 1.

b) Reporte “Total de ventas por productos, subfamilia, tipos de cliente de una determinada sucursal, y periodo de tiempo.”

Reportes de los tipos de clientes que han generado más ventas en soles por productos. Este reporte permite usar filtros por periodo (año) y Sucursal.

Diseño

Este reporte nos muestra los ingresos por tipo de cliente, con esta información se puede identificar que productos tienen demanda por determinado tipos de cliente. A partir de ello se puede impulsar las ventas de productos a través de campañas promocionales a un determinado tipo de cliente. A continuación el diseño del reporte:

TOTAL DE VENTAS POR PRODUCTOS Y TIPOS DE CLIENTE				
		TIPOS DE CLIENTE		
SUBFAMILIA	PRODUCTO	BUENOS	REGULAR	TOTAL
King Kong				
	King Kong Mediano 900 GR	5,600	4,500	10,100
	Mini Barra de Manjar 35 GR	5,403	7,200	12,603
	1/2 Barra de Manjar Y Piña	4,320	4,300	8,620
	Total King Kong	15,323	16,000	31,323
Potes				
	Lata Manjarblanco 250 GR	9,750	10,500	20,250
	Lata Natilla 250 GR	7,400	8,450	15,850
	Total Potes	17,150	18,950	36,100
Dulceria				
	Alfajores Pack De 14 UNI	13,200	9,500	22,700
	Suspiros Bolsa	11,150	11,400	22,550
	Coquitos	9,650	5,450	15,100
	Especial San Roque 705 GR	10,350	6,430	16,780
	Total Dulceria	44,350	32,780	77,130
Filtros	Sucursal: Trujillo, Año: 2010			

Figura 70: Diseño de reporte total de ventas por productos, subfamilia, tipos de cliente de una determinada sucursal, y periodo de tiempo.

Tipo

El tipo del reporte es Crosstab.

Filas

Las filas que tiene el reporte de ventas por productos y tipos de cliente:

No.	Dimensión	Nivel
1	DimProducto	SubFamilia
2	DimProducto	Producto

Tabla 55: Filas del Reporte 2.

Columnas

Las columnas que tiene el reporte de ventas por productos y tipos de cliente:

No.	Dimensión	Nivel
1	DimCliente	TipoCliente

Tabla 56: Columnas del Reporte 2.

Medidas

Los cálculos que se realizan para el reporte:

No.	Medida	Formato
1	TotalINS	Moneda (S/.)

Tabla 57: Medidas del Reporte 2.

Filtros

Los filtros usados en el reporte:

No.	Operación
1	Sucursal = <Sucursal seleccionada>
2	Año=<Año ingresado>

Tabla 58: Filtros del Reporte 2.

c) Reporte “Numero de visitas por página web.”

Reporte del número de visitas que se han generado por cada una de las páginas web que conforman el web site

de la empresa San Roque S.A. Este reporte permite usar como filtro el año.

Diseño

Este reporte nos muestra el número de visitas por página, con esta información se puede identificar que productos producen mayor interés por parte de los usuarios de la web y comparar con la información interna proveniente de la base de datos transacción de la organización del total de ventas por producto. A partir de ello se puede impulsar las ventas de dichos productos a través de campañas promocionales. A continuación el diseño del reporte:

NUMERO DE VISITAS POR PAGINA WEB	
PAGINAS	NRO VISITAS
/new_site/catalogo_nuevo/Barra_de_Piña_y_Manjarblanco_nuevo_empaque.html	25
/new_site/catalogo_nuevo/Manjarblanco_de_Lúcuma_nuevo_empaque_de_200g.html	24
/new_site/catalogo_chocotejas/Surtido_60GR.html	24
/new_site/catalogo_barras/Chirimoya.html	23
/new_site/catalogo_kingkong/Manjarblanco_900GR.html	20
/new_site/catalogo_kingkong/Piña_y_Manjarblanco_450GR.html	19
/new_site/promociones/promocion_klm_king_kong_lata_sr.html	17
/new_site/puntoventa.html	16
Total Nro Vistas	
168	
Filtros	Año: 2012

Figura 71: Diseño del reporte número de visitas por página web.

Tipo

El tipo del reporte es Crosstab.

Filas

Las filas que tiene el reporte de número de visitas por página:

No.	Dimensión	Nivel
1	DimPaginas	pagina

Tabla 59: Filas del Reporte 3.

Columnas

No presenta columnas.

Medidas

Los cálculos que se realizan para el reporte:

No.	Medida	Formato
1	Nro Visitas	Numero entero

Tabla 60: Medidas del Reporte 3.

Filtros

Los filtros usados en el reporte:

No.	Operación
1	Año=<Año ingresado>

Tabla 61: Filtros del Reporte 3.

5.1.4.4.2. Diseños de Tableros de Mando

A manera de ejemplo se presenta el diseño de tres tableros de mando para el presente proyecto:

a) Tablero de Mando Análisis de ventas

El diseño para este tablero de control organiza la información a mostrar en tres secciones:

- ✓ **Análisis de Ventas:** Muestra la información del total de ventas de

los dos últimos años por sucursal y subfamilia.

- ✓ **Indicador Cantidad de ventas:**
Muestra el nivel de cantidad de ventas a través de un rango específico representado a través de un gráfico tipo semáforo, cuyo filtro es el año (parámetro).
- ✓ **Indicador Total de Ventas**
Muestra el nivel de total de ventas a través de un rango específico representado a través de un gráfico tipo semáforo, cuyo filtro es el año (parámetro).

En la siguiente figura se muestra el prototipo para este tablero de mando.

CUADRO DE MANDO ANALISIS DE VENTAS				
Analisis de Ventas			Indicador Cantidad de ventas	
Sucursal	Subfamilia	Año		
		2011	2012	
	Bebidas	11540,00	13400,000	<div style="text-align: center;">13 500</div>
	Dulceria	9540,000	9750,000	
Lambayeque	Helados	8400,00	4320,00	
	King Kong	9500,00	7850,000	
	Potes	12640,00	9560,000	
	Miel de Abeja Ecologica	11300,00	11200,00	
	Total Soles	62920,000	56080,000	
			Indicador Total de Ventas	
Distribucion	Dulceria	9780,00	11340,00	<div style="text-align: center;">125 650</div>
	King Kong	10500,00	9540,00	
	Potes	12740,00	13256,00	
	Total Soles	33020,00	34136,00	
	Bebidas	12640,00	12900,00	
Lima	Dulceria	11230,00	11340,00	
	Helados	14320,00	9560,00	
	King Kong	12450,00	10420,00	
	Total Soles	50640,00	44220,00	
Filtro	Año : 2010			

Figura 72: Prototipo de panel de control de Análisis de Ventas.

b) Tablero de Mando Análisis de Ventas por Productos

El diseño para este tablero de control organiza la información a mostrar en tres secciones:

- ✓ **Análisis de Ventas por producto:** Muestra la información del total de ventas por producto de los dos últimos años.
- ✓ **Ventas por sucursal del Producto:** De la lista mostrada de productos en la sección “Análisis de ventas por producto” se selecciona un producto para mostrar en la presente sección las ventas por sucursal de dicho producto seleccionado.
- ✓ **Ventas Por Tienda de la Sucursal**
Muestra el nivel de total de ventas por tienda de la sucursal seleccionada en la sección anterior “Ventas Por Sucursal Del Producto”.

En la siguiente figura se muestra el prototipo para este panel de control.

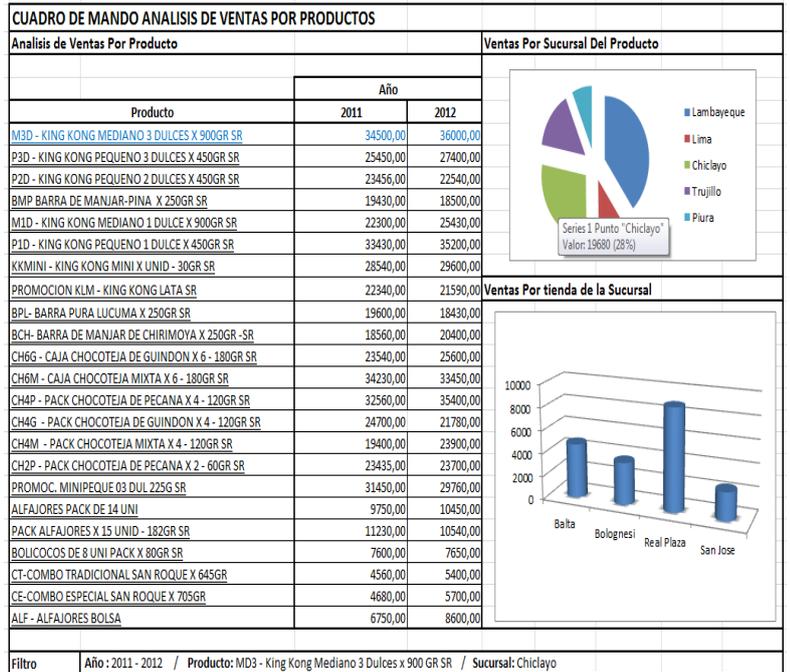


Figura 73: Prototipo de tablero de mando Análisis de Ventas por productos.

c) Tablero de Mando número de visitas a la página de un determinado producto

El diseño para este tablero de control organiza la información a mostrar en tres secciones:

- ✓ **Análisis de Ventas Por Productos:** Muestra la información del total de ventas por producto del año 2012.
- ✓ **Indicador número de visitas a la página del producto:** De la lista mostrada de productos en la sección “Análisis de ventas por producto” se selecciona un producto para mostrar en la presente sección el nivel del

número de visitas a la página de dicho producto seleccionado a través de un rango específico (valor mínimo, intermedio y máximo del número de visitas a las páginas de la web) representado a través de un gráfico tipo semáforo, cuyo filtro es el año 2012.

✓ **URL previas a la página del producto**

Muestra el top 5 del número de visitas a las URL previas a la página del producto seleccionado y filtrado por el año 2012.

En la siguiente figura se muestra el prototipo para este tablero de mando.

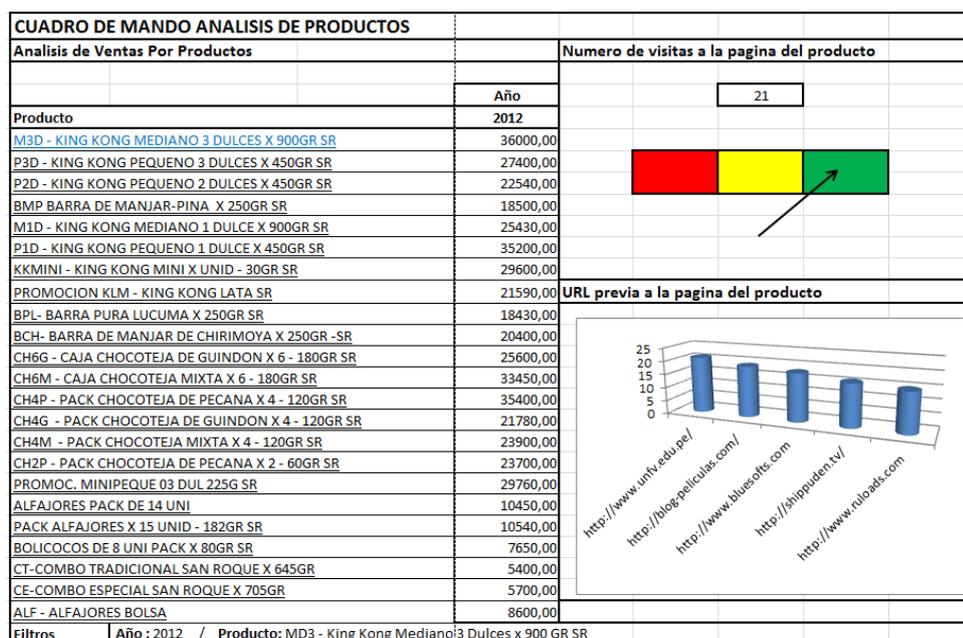


Figura 74: Prototipo de panel de control de Número de visitas a la página de un determinado producto.

5.1.5. FASE 5: CONSTRUCCIÓN

5.1.5.1 EVALUACIÓN Y SELECCIÓN DE LA PLATAFORMA BI DE SOFTWARE LIBRE

5.1.5.1.1 ANÁLISIS COMPARATIVO DE HERRAMIENTAS BI.

Para realizar el análisis se clasifica a las herramientas en dos grupos: Extracción y Explotación.

a) Herramientas de extracción (ETL) en el mercado

Una herramienta de extracción puede calificarse como una interfaz que facilita las actividades ETL relacionadas a cualquier solución de inteligencia de negocios. A continuación se muestran los requisitos necesarios para la elección de la herramienta y luego una comparación entre algunas de las principales herramientas.

Requerimientos a cumplir:

- Limpieza de datos.
- Interfaz gráfica para el desarrollo de componentes ETL.
- Formar parte de una plataforma integrada de BI.
- Es multiplataforma.
- Filtrado de datos

- Monitoreo y diseño del proceso ETL.
- Programación del proceso ETL para que se ejecute automáticamente.
- Trabajo con más de una fuente de datos.
- Trabajo con las principales bases de datos del mercado como: Microsoft SQL Server 2005/2008, Oracle, PostgreSQL, MySQL entre otras.
- Permite cargas desde ficheros excel, xml y planos.
- La plataforma posee una herramienta de explotación, herramientas de reporting, herramientas de consultas y análisis.
- Requiere una fácil instalación de la herramienta
- Brinda bastante información sobre el uso y soporte de la herramienta.
- Facilidad de Integración con Big Data.

Entre las más conocidas en el mercado destacan:

- Data Integrator de Business Objects.
- Kettle de Pentaho.
- Java – Clover
- Java – Octopus

Análisis Comparativo

Luego de revisar las principales características de las herramientas de

extracción candidatas se realizó un cuadro comparativo con los requerimientos que debe cumplir la herramienta seleccionada.

Para ello se le ha asignado una calificación a cada requerimiento que va de 0 (no cumple) a 5 (cumple totalmente).

En la siguiente tabla se muestra una comparativa de las características de cada uno de las herramientas de ETL presentadas.

Características	Pentaho Kettle	Business Objects	Java Clove	Java Octopus
Interfaz gráfica para el desarrollo de componentes ETL	5	4	3	4
Limpieza de Datos	5	4	3	3
Filtrado de Datos	4	4	3	3
Permitir el monitoreo del proceso ETL, así como el diseño del mismo.	4	5	4	4
Programación del proceso ETL para que se ejecute automáticamente	5	5	3	3
Trabajo con las principales bases de datos del mercado como: Microsoft SQL Server 2005/2008 y Oracle Database9g.	5	5	4	4
Brinda bastante	4	5	3	3

información sobre el uso y soporte de la herramienta				
Trabajo con más de una fuente de datos.	5	4	4	4
Permite cargas desde ficheros 242xcel, xml y planos	5	4	4	4
La plataforma posee una herramienta de explotación, herramientas de <i>reporting</i> , herramientas de consultas y análisis	5	5	3	3
Forma parte de una plataforma integrada de inteligencia de negocios	5	4	3	3
Requiere una fácil instalación de la herramienta	4	4	3	3
Es multiplataforma	5	5	5	5
Facilidad de Integración con Big Data	5	3	2	2
Puntaje Total	66	61	47	48

Tabla 62: Cuadro Comparativo – Herramientas Extracción.

Selección de la Herramienta ETL para el Proyecto

Luego de realizar un análisis y comparación entre las diversas herramientas de extracción se muestra que Pentaho Kettle obtuvo mayor puntaje, a pesar que actualmente no existe un vasto soporte, la

documentación que se encuentra disponible de forma oficial cubre la mayor parte de las principales necesidades del presente proyecto.

La herramienta Kettle, es intuitiva y de fácil aprendizaje ya que presenta un entorno gráfico para el diseño del ETL, en este aspecto destaca la característica facilidad de integración con Big Data, el cual es un aspecto importante para el presente proyecto. Por lo tanto se concluye que al cumplir Pentaho con la mayoría de características, es la herramienta más apropiada para el desarrollo de este proyecto.

b) Herramientas de explotación de la información en el mercado

Actualmente existen diversas herramientas en el mercado para poder montar el modelo OLAP sobre las estructuras de datos diseñadas. A continuación se muestran los requisitos necesarios para la elección de la herramienta y luego una comparación entre algunas de las principales herramientas.

Requerimientos a cumplir:

- Análisis multidimensional.
- Interfaz gráfica/ Drag & drop para la generación de reportes.

- Gestión de dashboards.
- Drill and Down: Detalle de los datos accediendo directamente a las fuentes multidimensionales.
- Filtros y búsquedas personalizables.
- Slice and Dice: Operaciones de corte y rotación de datos.
- Brinda información sobre el uso y soporte de la herramienta.
- Permite realizar gráficos en función a datos obtenidos en los reportes.
- Permite realizar la generación de reportes web, exportables a otros tipos de documentos.
- Ser multiplataforma.
- Facilidad de uso de la herramienta.
- Ofrece facilidad en la instalación de la herramienta.

A continuación se menciona algunas de las herramientas que tienen presencia en el mercado. La cantidad de herramientas existentes es mucho mayor, solo se presentan las más representativas.

- Business Objects
- Pentaho Bi Suite
- SpagoBi
- Java - BIRT

Cada una de ellas presenta funcionalidades similares: construcción de cubos OLAP, capacidad de reporting, alarmas, dashboards y plataforma web.

Análisis Comparativo

Luego de revisar las principales características de las herramientas de explotación candidatas se realizó un cuadro comparativo con los requerimientos que debe cumplir la herramienta seleccionada.

Al igual que para la selección de herramientas ETL, se ha asignado una calificación por cada requerimiento que debe cumplir la herramienta de 0 (no cumple) a 5 (cumple totalmente).

En la siguiente tabla se muestra una comparativa de las características de cada uno de las herramientas de explotación anteriormente mencionadas:

Características	Pentaho BI Suite	Business Objects	SpagoBI	Java-BIRT
Análisis multidimensional	5	5	4	4
Gestión de dashboards.	5	4	4	4
Drill and Down	5	4	4	4
Filtros y búsquedas personalizables.	5	4	4	4
Slice and Dice: Operaciones de corte y rotación de datos	5	4	4	4
Permite realizar la	5	5	4	5

generación de reportes web, exportables a otros tipos de documentos.				
Permite realizar gráficos en función a datos obtenidos en los reportes.	5	5	5	5
Brinda información sobre el uso y soporte de la herramienta.	4	5	3	3
Ofrece facilidad en la instalación de la herramienta	4	5	3	3
Facilidad de uso de la herramienta.	5	4	4	4
Es multiplataforma	5	5	5	5
Puntaje Total	53	50	44	45

Tabla 63: Cuadro Comparativo – Herramientas Explotación.

Selección de la Herramienta de explotación de Datos

Luego de realizar un análisis y comparación entre las diversas herramientas de extracción se concluye que al cumplir Pentaho con la mayoría de características, tales como: integración de reportes, análisis por web, interfaz gráfica para fácil manejo por parte de los usuarios, entre otros, y siendo la herramienta que obtuvo más puntaje, es por lo tanto la herramienta más apropiada para el desarrollo de este proyecto.

Pentaho es una plataforma de Inteligencia de Negocios que brinda

soluciones para ambos procesos: extracción y explotación.

Algunas ventajas y beneficios del uso de open source son:

- Al construir, integrar y desarrollar componentes open source dentro de una sola plataforma el coste de implementar soluciones Business Intelligence se reduce, beneficiando de esta manera principalmente a las pymes que no pueden acceder a la tecnología comercial actual.
- Si se reduce la inversión en licencias, se puede destinar ese dinero a mejorar el análisis, el diseño, la formación de usuarios, la toma de decisiones.
- Permite realizar prototipos de reportes de una forma más rápida y permanente en el tiempo.
- El sistema es fácil y óptimamente escalable, pudiendo extender la misma entre diferentes servidores sin limitaciones del número de usuarios concurrentes, es decir, no existe un número limitado de licencias para su uso.
- Se proporciona interfaces de usuarios muy intuitivos que son

fácilmente customizables por los propios usuarios.

Se descarta usar la versión comunidad de Pentaho ya que el presente proyecto se aplicara a una empresa real que maneja un gran volumen de datos por lo tanto un prototipo no soportaría el modelo propuesto. Para el presente proyecto se utilizara la versión Enterprise de pentaho 4.8 ya que es una versión que presenta facilidades de aprendizaje y de uso tanto para los desarrolladores como para los usuarios finales, tiene una arquitectura flexible y de calidad, más aun su beneficio más importante es el de bajo costo, ello ha sido un elemento importante para que la empresa muestre un gran interés por esta versión.

5.1.5.1.2 SOFTWARE A UTILIZAR

En este punto se indica nuevamente los componentes de la arquitectura de la solución de inteligencia de negocios pero esta vez indicando cuales son las herramientas que se utilizaran en cada una de ellas. En la siguiente figura se muestran las herramientas que intervienen en los diferentes componentes de la arquitectura de la solución. Se muestra cada componente y su respectiva herramienta.

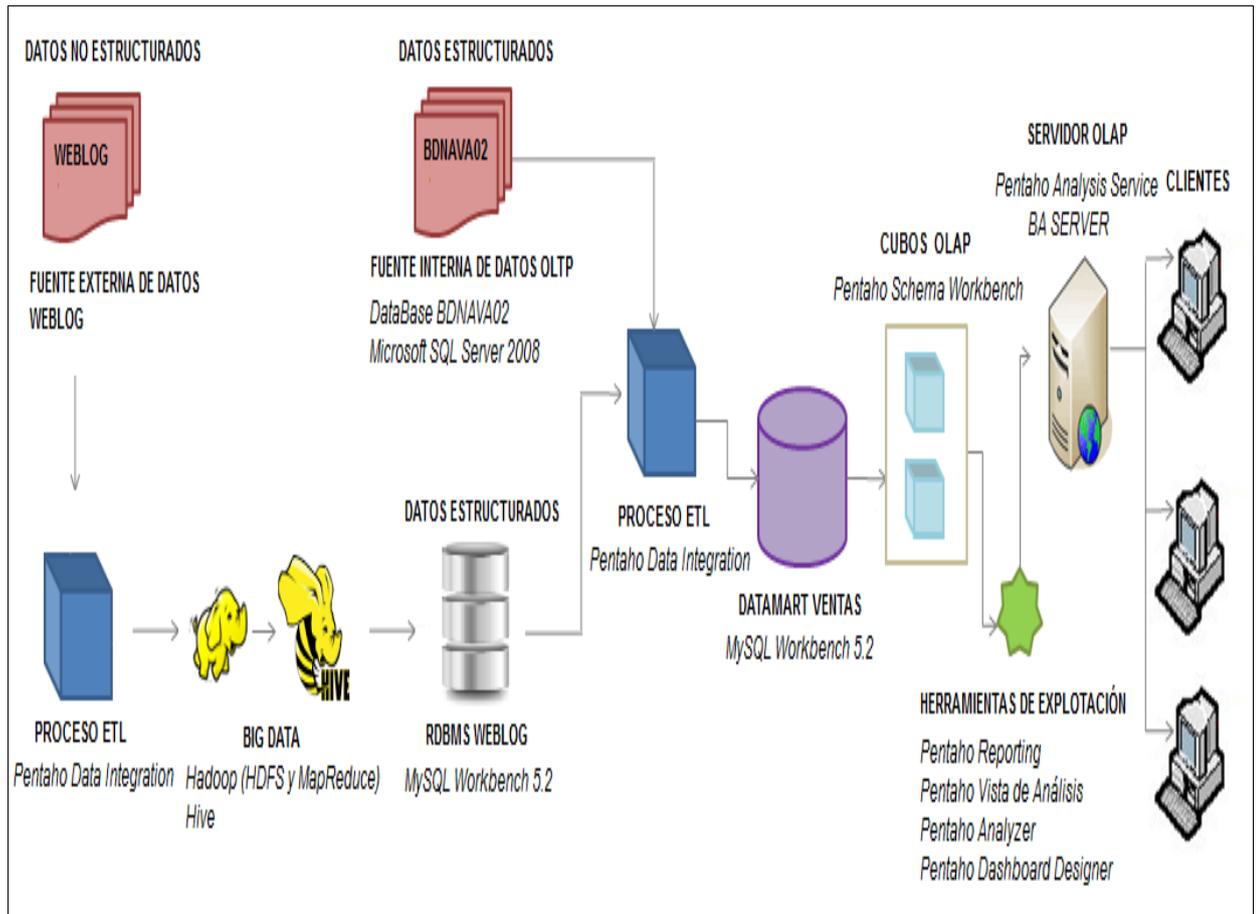


Figura 75: Herramientas BI para la arquitectura de la solución de Inteligencia de Negocios.

A continuación se detalla cada uno de las herramientas a utilizar en la arquitectura de la solución BI del presente proyecto:

a) Procesos ETL

Los procesos ETL serán desarrollados con la herramienta Pentaho Data Integration, mediante esta herramienta se va a crear un esquema ETL para los modelos dimensionales desarrollados, también permitirá la extracción y

análisis de los datos no estructurados provenientes del archivo weblog.

Pentaho Data Integration

Usa una interface gráfica “SPOON”, que permite diseñar “jobs” de transformación de datos para ser procesados bajo las herramientas PAN y KITCHEN.

PAN es un motor de transformación de datos permitiendo la lectura y escritura sobre diversas fuentes de datos.

KITCHEN es una aplicación que permite ejecutar “Jobs” planificados en modo “batch” (serie, lote) para correr automáticamente en intervalos de tiempo definidos.

b) Big Data

Permitira el manejo el procesamiento de la gran cantidad de información no estructurada proveniente de la fuente externa de datos a través de las herramientas Hadoop y Hive.

Apache Hadoop

Permitirá el mapeo y reducción de los datos no estructurados a través de sus componentes HDFS y MapReduce.

Apache Hive

Permitirá el modelado de los datos pre procesados (creacion de base de datos y creación de tablas) en Hadoop, a través de su lenguaje de consultas llamada HiveQL, luego esta base de datos será exportada al RDBMS.

c) RDBMS

Contiene la base de datos relacional final obtenida de la estructuración de los datos de la fuente externa weblog proveniente de Hive. El Modelador y Gestor de Base de Datos es el MySQL Workbench 5.2.

d) DataMart Ventas

El diseño físico y construcción del DataMart de ventas que contiene la información interna y externa ha sido desarrollado utilizando el Modelador y Gestor de Base de Datos: MySQL Workbench 5.2.

MySQL Workbench 5.2

Para el presente proyecto, el uso de un modelador de datos sirve de base principalmente para el proceso de análisis, por lo tanto para facilitar el diseño de la base de datos usada como repositorio se ha utilizado MySQL Workbench 5.2, así mismo se decide utilizar esta misma herramienta como

gestor de base de datos para el desarrollo del Data Mart, y así complementar junto a Pentaho BI, la solución OpenSource Business Intelligence (OSBI). Las principales características por las cuales se decidió utilizar este gestor de bases de datos son las siguientes:

- Fácil instalación y configuración.
- Es el gestor más usado en el mundo del software libre, debido a su gran rapidez y facilidad de uso.
- Aprovecha la potencia de sistemas multiprocesador.
- Soporta gran cantidad de tipos de datos para las columnas.
- Dispone de API's en gran cantidad de lenguajes (C, C++, Java, PHP, entre otros).
- Gestión de usuarios y passwords, manteniendo un muy buen nivel de seguridad en los datos.

e) Cubo OLAP

La herramienta OLAP que se usará para el procesamiento del Data Mart ventas es Schema Workbench 5.2, mediante esta herramienta se construirán los cubos OLAP (tanto para la información interna como externa) para su posterior explotación de los datos que contiene.

Pentaho Schema Workbench 5.2

Workbench es parte de la suite Pentaho, su finalidad es la de construir los cubos OLAP y administrar el lenguaje de consulta ROLAP llamado MDX (Multi-Dimensional Expression, Expresión Multidimensional). La aplicación genera y permite modificar archivos XML (los cuales contienen el esquema de los cubos) pudiendo agregar, eliminar o editar dimensiones y métricas.

f) Servidor OLAP

Para el presente proyecto se utilizara el servidor Mondrian Pentaho, el cual es una de las aplicaciones más importantes de la plataforma Pentaho BI.

Pentaho Analysis Services

Mondrian ahora bautizado como "Pentaho Analysis Service" forma parte del motor OLAP integrado en el Suite BI de Pentaho, es decir se instala en conjunto con la plataforma BI Suite Pentaho. Es un servidor OLAP de software libre que comunica la base de datos fuente (Data Mart) y la aplicación OLAP (escrita en Java).

g) Herramienta Explotación

El proceso de explotación es realizado haciendo uso de las herramientas Pentaho Reporting y Pentaho Dashboard.

Pentaho Reporting

Herramienta de reportes flexibles y con clase empresarial, basados en Web. La herramienta de reportes Pentaho permite comenzar desde sencillos reportes iniciales hasta formar complejos reportes ajustados a las necesidades de negocio. Permite la distribución de los resultados del análisis en múltiples formatos, todos los informes incluyen la opción de imprimir o exportar a formato PDF, XLS, HTML y texto.

Pentaho Reporting se divide en tres componentes o productos con diferentes enfoques y dirigidos a diferentes tipos de usuarios:

- **Pentaho Report Designer (New Report):** Permite la construcción de reportes desde cero. Es un Editor basado en eclipse con prestaciones profesionales y de calidad y con capacidad de personalización de informes a las

necesidades de negocio, destinado a desarrolladores. Está estructurado de forma que los desarrolladores pueden acceder a sus prestaciones de forma rápida: Incluye un editor de consultas para facilitar la confección de los datos que serán utilizados en un informe.

- **Pentaho Report Design Wizard:**

Esta herramienta utiliza un asistente de diseño de informes, que facilita el trabajo y permite a los usuarios obtener resultados de forma inmediata. Está destinada a usuarios con menos conocimientos técnicos. A través de pasos sencillos permite:

- ✓ Conectarse a todo tipo de bases relacionales.
- ✓ Integrar el resultado dentro del portal pentaho.

- **Web ad-hoc Reporting:**

Es similar a la herramienta anterior pero via web. Extiende la capacidad de los usuarios finales para la creación de informes a partir de plantillas pre configuradas y siguiendo un asistente de creación. Esta opción se encuentra

dentro de la interface de la consola de usuario de la suit Pentaho.

Según lo expuesto se decide que para la creación de los reportes en el presente proyecto se utilizara:

- Pentaho Report Designer
- Pentaho Report Design wizard

Pentaho dashboard Designer

Esta opción se encuentra dentro de la interface de la consola de usuario de la suit Pentaho. Todos los componentes del módulo Pentaho Reporting (Pentaho Report Designer, Pentaho Report Design, web ad-hoc reporting) y Pentaho Análisis (vista de análisis y analyzer report) pueden formar parte de un Dashboard. Pentaho proporciona la posibilidad de realizar cuadros de mandos a través de la interfaz web mediante el Dashboard Designer. Esta posibilidad existe sólo en la versión Enterprise. No existiendo una herramienta para realizar cuadros de mando Add-hoc en la versión Community. Pentaho dashboard designer permite diseñar páginas compuestas en modo web, con varias posibilidades de formato predefinidas donde las diferentes zonas pueden contener informes existentes o nuevas consultas ad hoc. En cada zona, el usuario puede

optar por la visualización de una tabla, un gráfico, un cubo, un informe, los reportes creados con la herramienta Pentaho Reporting, etc. En conclusión esta herramienta permite:

- Realizar cuadros de mando Add-hoc.
- Realizar cuadros de mandos basados en Metadata o en consultas SQL.
- Identificación de Indicadores (KPI's, Key Performance Indicators).

Herramientas Adicionales

Se utilizaran las siguientes herramientas adicionales que son componentes del módulo Pentaho Análisis para la explotación de los datos:

- **Vista de Análisis**

Ayuda a operar con máxima efectividad, ganar perspicacia y entender lo necesario para tomar optimas decisiones.

Las características generales son:

- ✓ Vista dimensional de datos.
- ✓ Navegar y explorar (Análisis Ad Hoc, Drill-down, etc.).
- ✓ Interactuar con alto rendimiento mediante tecnologías optimizadas para la rápida respuesta interactiva.

- **Pentaho Analyzer**

Un requerimiento frecuente de los usuarios de negocios es contar con una herramienta que permita visualizar datos en distintas dimensiones y niveles, que sea de fácil uso, que tenga funcionalidades avanzadas de visualización y que además sea capaz de exportar a otros formatos como Excel, pdf, CSV. Pentaho Analyzer es capaz de realizar todo lo anterior y más. La herramienta es 100% web, de fácil uso y de un rápido tiempo de implementación. Solo disponible en la versión Enterprise de Pentaho. Ofrece importantes mejoras de funcionalidad y facilidad de uso para los usuarios finales, entre los que podemos mencionar:

- ✓ Drag and Drop de Campos (se arrastran).
- ✓ Panel de filtros Ordenamiento y Filtrado (inclusión/exclusión) con un click.
- ✓ Medidas definidas por el usuario.
- ✓ Permite cambiar tamaño y formatos de columnas dinámicamente.
- ✓ Agregar/quitar subtotales.
- ✓ Gráficos Mejorados.

Todas las herramientas Pentaho BI descritas anteriormente se encuentran integradas en una única plataforma de Pentaho llamada: Pentaho Business Analytics Enterprise 4.8. La cual se utilizara para el desarrollo del presente proyecto. Así mismo se instalaran las aplicaciones de software libre apache Hadoop y Hive para el desarrollo del modelo propuesto.

5.1.5.2 INSTALACIÓN Y CONFIGURACIÓN DE SOFTWARE

Para el desarrollo de este proyecto se usa software libre. Como herramienta de análisis de los datos no estructurados se utilizó los componentes de Big Data apache Hadoop 2.2 y apache Hive 0.12, como diseñador y repositorio del Data Mart se usa MySQL workbench 5.2 y como herramienta de análisis se tiene la plataforma de Inteligencia de Negocios Pentaho Business Analytics Enterprise 4.8. Se realizó la instalación y configuración de cada una de las herramientas utilizadas en el presente proyecto.

5.1.5.3 CONSTRUCCIÓN DE DATAMART

Como se indicó anteriormente la herramienta seleccionada para la construcción del Data Mart es el Mysql Workbench 5.2, la cual es la misma que se utilizó para el diseño físico dimensional realizado en la de etapa de diseño dimensional es decir para el modelado de la base de datos. En base a este diseño

se procedió a crear el siguiente Data Mart que tiene como nombre “ventasmart”, en la siguiente figura se muestra el Data Mart ya implementado con sus respectivas dimensiones y tablas hechos, en espera del proceso ETL para ser poblada:

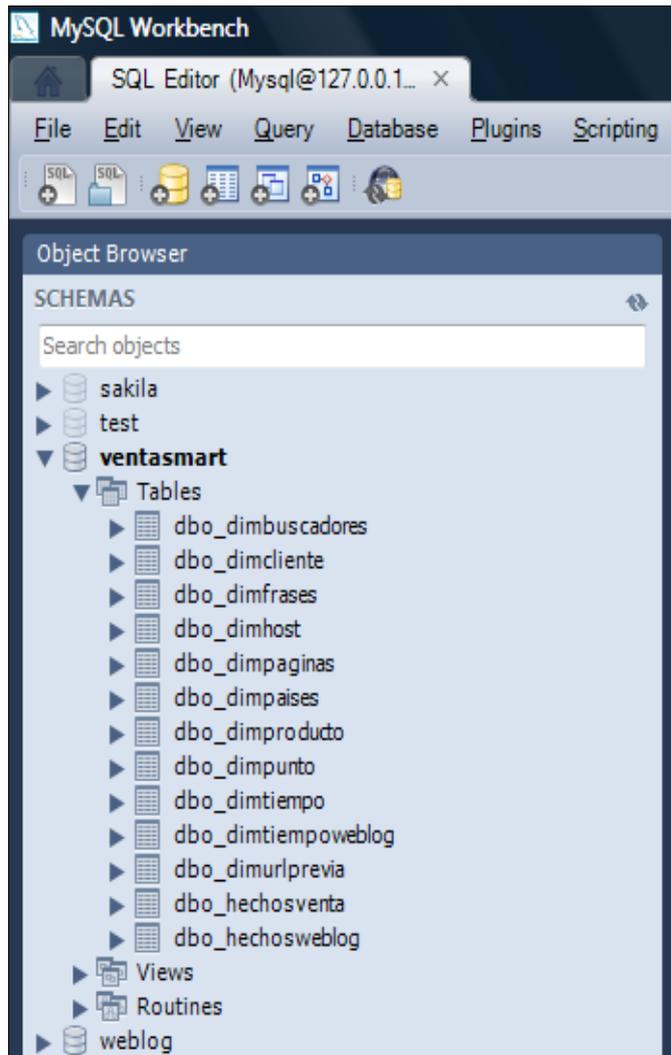


Figura 76: Data Mart VentasMart.

5.1.5.4 EJECUCIÓN DE ETL

Para la ejecución del ETL se utilizó la herramienta Data Integration incluida en la versión de Pentaho que se está utilizando:

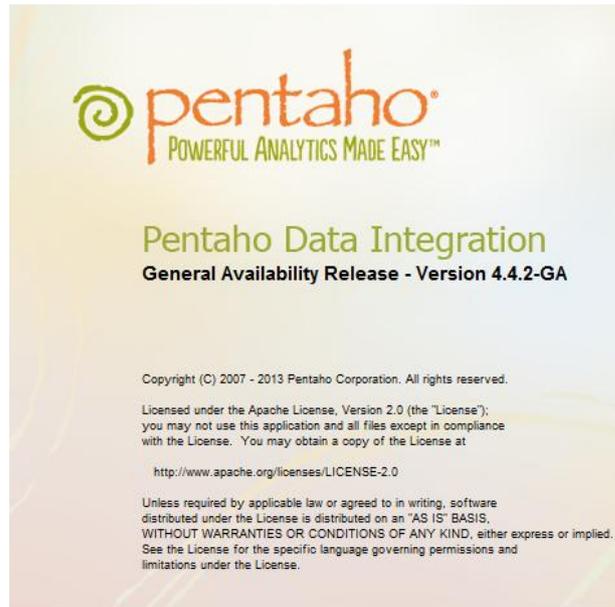


Figura 77: Pantalla de bienvenida de Pentaho Data Integration.

Estrategia De Poblamiento

La estrategia de poblamiento, consiste en la implementación de una tarea o “Job” (componente que crea una secuencia de actividades que brinda un orden de ejecución, son usados para un control de flujo), en la cual generamos las transformaciones necesarias para el poblamiento del Data Mart, dentro de las transformaciones se generan los pasos o “steps” (cada uno está diseñado para una función específica), las cuales se encuentran relacionadas a través de “hops” (son usados para transportar data de un step a otro), todos estos componentes realizan la carga ordenada de los datos requeridos en forma secuencial. La ejecución comienza en el nodo Start para luego continuar con el siguiente paso, la limpieza de las tablas dimensiones y las dos tablas Hechos correspondientes a través de un scripting el cual permite ejecutar una sentencia SQL, luego a través de las transformaciones se realiza la población de las

dimensiones, por motivo que para poblar la tabla hecho ventas es necesario una sentencia SQL que utilice tanto los datos de las cuatro dimensiones como de la base de datos transaccional y como no se puede realizar dicha consulta por estar en diferentes gestores de base de datos (SQL SERVER 2008 y MySQL Workbench), se tuvo que crear una tabla temporal llamada “apoyo ventas” que contenga la información obtenida de la BDNVA02 relacionada a las tablas ventas y detalle de ventas esto se realiza a través de la creación de un comando tal como se realizó para el caso de la limpieza de las tablas, luego se realiza el poblamiento de las dos tablas hechos respectivamente, se elimina la tabla temporal para luego finalizar con un mensaje que indica que el Job fue procesado con éxito, a continuación se muestra el esquema final de la estrategia de poblamiento bajo la herramienta Data Integration de Pentaho:

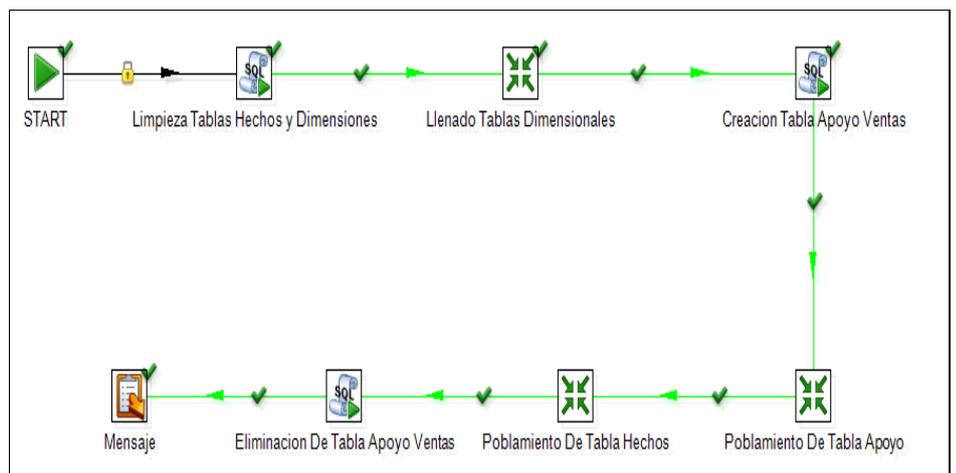


Figura 78: Esquema de poblamiento (Job).

5.1.5.5 CREACIÓN DE CUBO

El desarrollo de los cubos OLAP se hace a través de la herramienta Pentaho Schema Workbench. Al finalizar se obtiene el esquema final de ambos cubos, el primero cubo tiene como nombre “cuboventas” el cual analizara la información interna (BD Transaccional) de la empresa, el resultado final se muestra en la siguiente figura:

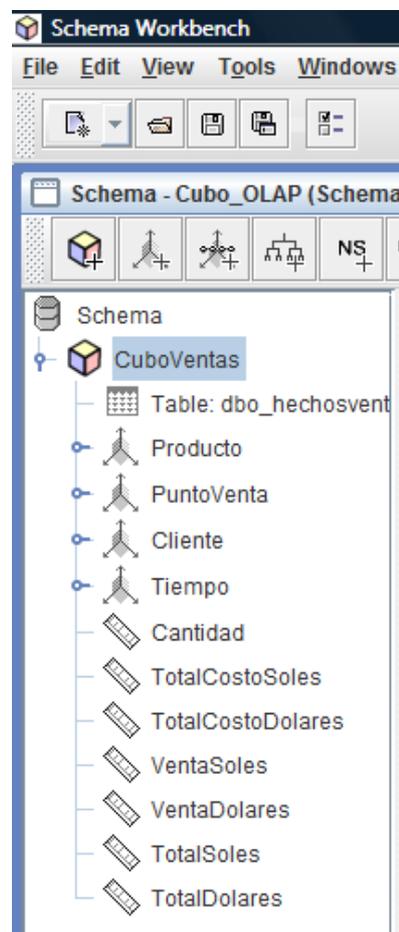


Figura 79: Cubo OLAP ventas.

El segundo cubo tiene como nombre “cuboweblog” el cual analizara la información externa (weblog de la web site) de la empresa, el resultado final se muestra en la siguiente figura:

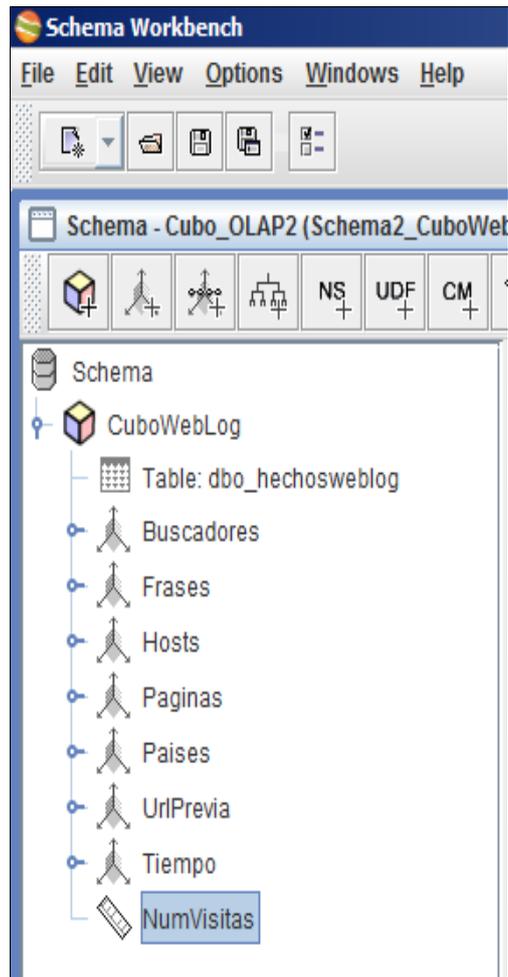


Figura 80: Cubo OLAP weblog.

Luego es necesario también publicar los cubos para que puedan ser utilizados por la herramienta Pentaho Análisis, para ello se requiere acceder al servidor a través de la clave configurada previamente que en este caso es “123456” y además los datos de credenciales de usuario en este caso: “joe” y “password” respectivamente. Si todo esta correcto debería mostrarse un mensaje indicando que la publicación fue exitosa y con esto terminaría el proceso de publicación del cubo realizado en schema

workbench y estaría lista para ser utilizada por la herramienta Pentaho Análisis, esta misma operación se realiza para publicar el cubo weblog, en la siguiente figura se muestra la publicación del cubo ventas:

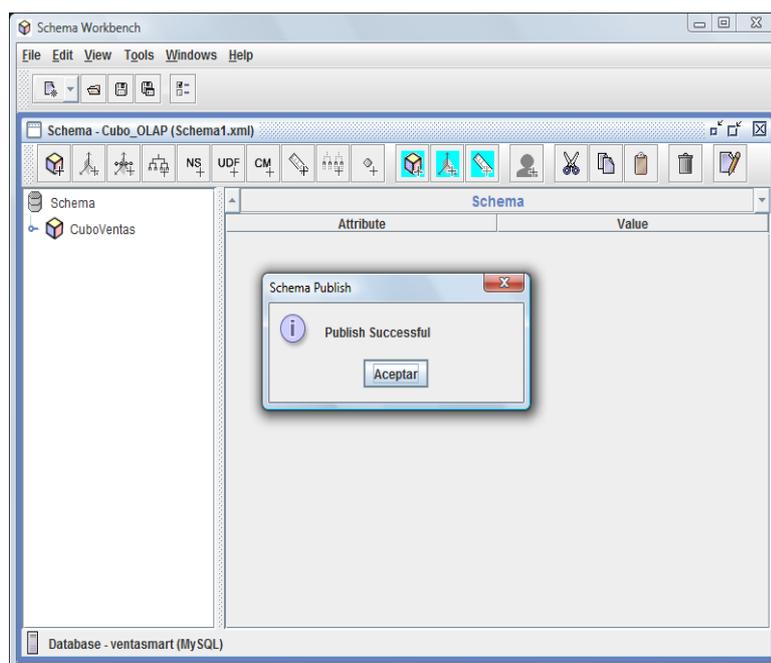


Figura 81: Publicación exitosa del Cubo Ventas.

5.1.5.6 CONSTRUCCIÓN DE INTERFACES

Las interfaces que proporciona la herramienta Pentaho para el presente proyecto se clasifican en Reportes, Tableros de Mando, Vista de Análisis y Analyzer Report.

5.1.5.6.1. Reportes

La herramienta usada para la creación de reportes es Pentaho Report Designer, para la creación de cada reporte se siguieron los siguientes pasos:

- Se asigna la conexión a la base de datos donde se encuentra cargado el Data Mart.
- Luego de establecer el origen de datos se realiza una static query de la cual se obtiene la consulta MDX utilizada para obtener los datos a presentar.
- Se seleccionan las columnas a mostrar.
- Luego se pasa a la etapa de personalización del reporte.
- El último paso es la publicación del reporte dentro del servidor para que pueda ser visualizado.

Todos los reportes se generaron en base a la misma estructura. A continuación a modo de ejemplo se muestra 3 reportes desarrollados en esta herramienta:

a) Reporte total y cantidad de ventas por año, tienda y subfamilia.

El reporte está compuesto por tres partes:

- La cabecera que contiene el nombre de la empresa.
- El agrupamiento de la cabecera que tiene información del código, producto, cantidad, total soles.
- El cuerpo genera información correspondiente a las cabeceras.
- Se puede filtrar el reporte por año, tienda, subfamilia y se puede

seleccionar el formato de salida que desee tenga el reporte.

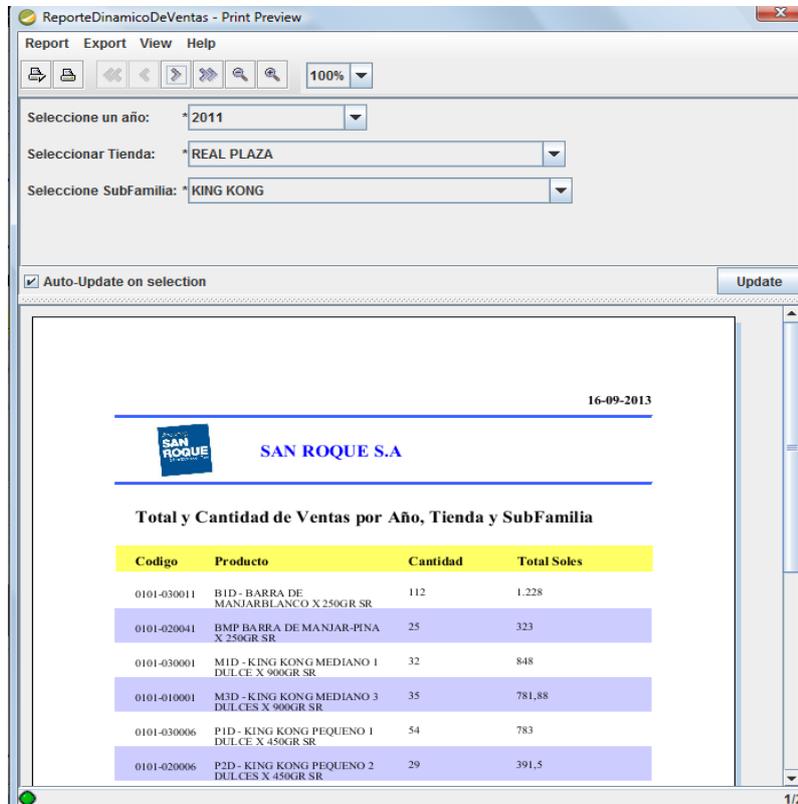


Figura 82: Reporte de total y cantidad de ventas.

Al final se realiza la publicación del reporte. En la siguiente figura se aprecia el reporte publicado en su vista final con la opción de poder filtrar por año, subfamilia y tienda.

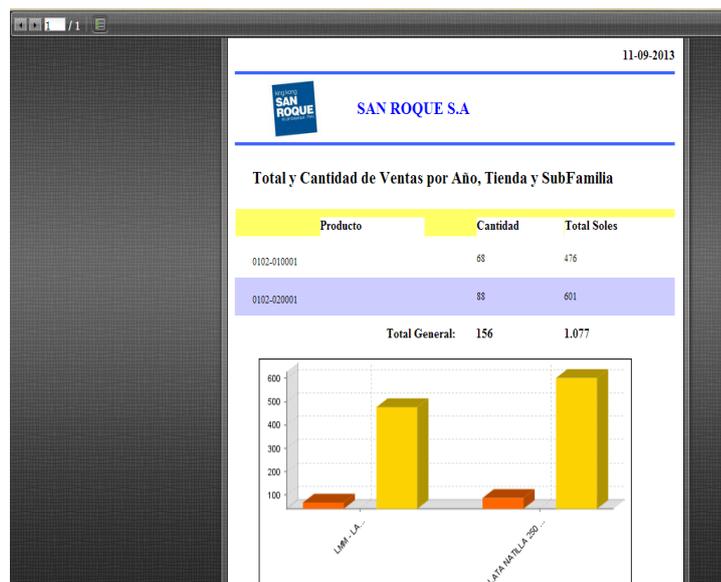


Figura 83: Vista Final del Reporte publicado.

b) Reporte total de ventas por cliente.

El reporte está compuesto por tres partes:

- La cabecera que contiene el nombre de la empresa.
- El agrupamiento de la cabecera que tiene información del cliente, tipo de cliente, departamento, total soles.
- El cuerpo genera información correspondiente a las cabeceras.
- Se puede filtrar el reporte por año y se puede seleccionar el formato de salida que desee tenga el reporte.

En la siguiente figura se muestra el reporte en su vista final.

ReporteDinamicoDeVentas2

septiembre 11, 2013 @ 10:57

San Roque S.A.

Total de Ventas Por Cliente

Tiendas

OPEN PLAZA

Cliente	TipoCliente	Departamento	Total Soles
CARMEN AMALIA PECSEN PEREZ	REGULAR	LAMBAYEQUE	314,6700

CENTRO CIVICO REAL P

Cliente	TipoCliente	Departamento	Total Soles
EMPRESA EDITORA LA INDUSTRIA DE CHICLAYO	REGULAR	LAMBAYEQUE	38,0000

FABRICA

Cliente	TipoCliente	Departamento	Total Soles
ESTUDIO REYNOSO E I R L	PRINCIPAL	LAMBAYEQUE	108,216,1700

LIMA AEROPUERTO

Cliente	TipoCliente	Departamento	Total Soles
INDUSTRIAS ABRAHAM F I R I	REGULAR	LAMBAYEQUE	190,4000

Figura 84: Vista final del reporte Total de Ventas por Cliente.

c) Reporte número de visitas por página web.

El reporte está compuesto por tres partes:

- La cabecera que contiene el nombre de la empresa.
- El agrupamiento de la cabecera que tiene información de las páginas y el número de visitas.
- El cuerpo genera información correspondiente a las cabeceras.

En la siguiente figura se muestra el reporte en su vista final.

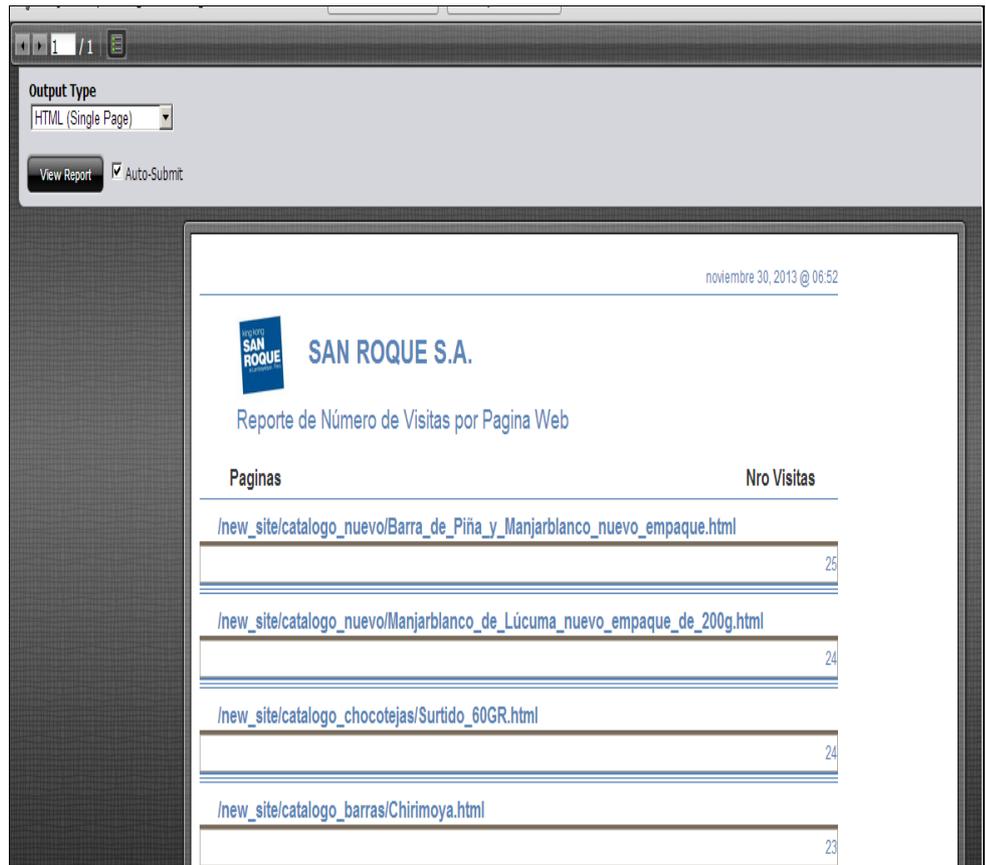


Figura 85: Vista final del reporte número de visitas por página.

5.1.5.6.2. Tableros de Mando

Para el desarrollo de los tableros de mando se tuvieron que realizar algunos reportes o vistas de análisis previamente, para luego ubicarlos en las respectivas secciones de los tableros de mando, también se tuvieron que crear conexiones que permitan crear orígenes de datos para que a través de consultas SQL se puedan crear gráficos de barras, o del tipo Dial (semáforo), en las cuales para este último caso fue necesario generar consultas bajo rangos específicos correspondientes por ejemplo al total de

ventas. Todos los tableros de mando desarrollados constan de 3 secciones.

A modo de ejemplo a continuación se muestra algunos de los tableros de mandos desarrollados:

a) Tablero de Mando Análisis de Ventas

Este tablero de mando se desarrolló en base al diseño generado en la fase de diseño. Consta de tres secciones, en la primera sección se encuentra un analyzer report, y en las otras dos secciones se encuentran dos gráficos del tipo dial que muestran la información correspondiente a los indicadores total de ventas y cantidad, ambas secciones interactúan según el parámetro seleccionado que en este caso es el año. A continuación se muestra en la siguiente tabla, el rango utilizado para medir los indicadores de cantidad y total de ventas en soles, en los respectivos gráficos del tipo dial, ubicadas en sus respectivas secciones.

Indicador Cantidad	
De 0 a 14000	
De 14000 a 18000	
De 18000 a 30000	

Indicador Total Soles	
De 0 a 130 000	
De 130 000 a 230 000	
De 230 000 a 300 000	

Tabla 64: Rango de valores para indicadores.

Fuente: Jefe de Distribución y Ventas.
Fabiola Díaz Orbegoso.

En las siguiente figura se muestra la vista final del tablero de mando mostrando los diferentes estados (rojo, amarillo, verde) dependiendo del valor del indicador cantidad y total de ventas del año seleccionado en este caso para el año 2011.

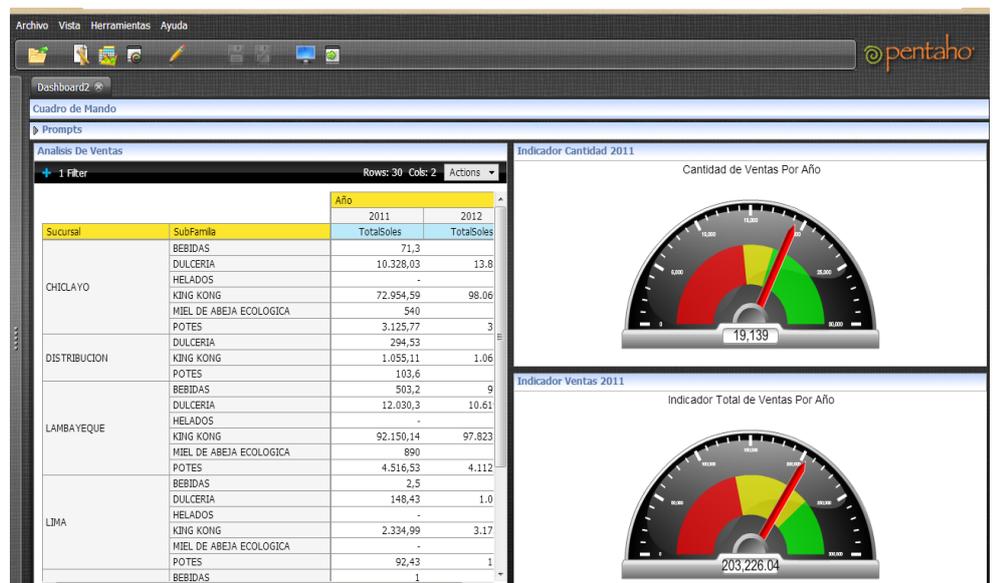


Figura 86: Vista final del tablero de mando
Análisis de Ventas para el año 2011.

b) Tablero de Mando Análisis de Ventas por Productos

Este tablero de mando se desarrolló en base al diseño generado en la fase de diseño. A continuación en la siguiente figura se muestra la vista final del tablero de mando filtrado por el producto seleccionado “M1D King Kong Mediano 1 Dulce x 900 GR”, mostrando el total de ventas por tienda para la sucursal Chiclayo correspondiente a la información de dicho producto seleccionado.

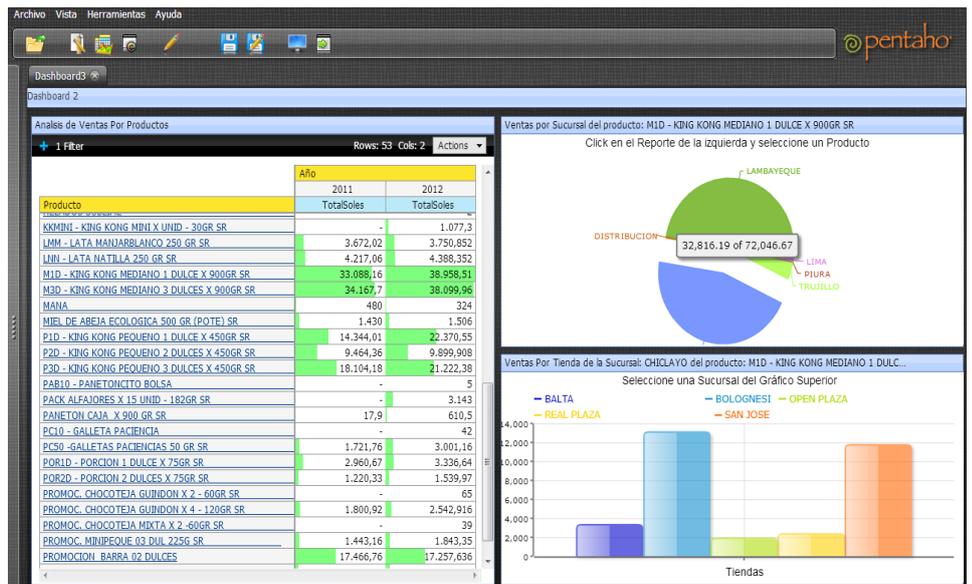


Figura 87: Tablero de Mando Total de Ventas por Producto para la sucursal Chiclayo.

c) Tablero de Mando Total de Ventas por Cliente

Este tablero de mando contiene en su primera sección un reporte dinámico generado previamente con la

herramienta pentaho reporting, el cual muestra el total de ventas por cliente del año seleccionado como parámetro. En la segunda sección el tablero muestra un gráfico de barras correspondiente a las ventas por departamento del mismo año seleccionado como parámetro en la sección anterior. En la tercera sección el tablero de mando muestra un gráfico del tipo Pie correspondiente a la información de las ventas por tipo de cliente del año seleccionado y del departamento seleccionado del gráfico de barras en la sección anterior.

A continuación en la siguiente figura se muestra la vista final del tablero de mando filtrado por el año 2007 y por el departamento Lambayeque.

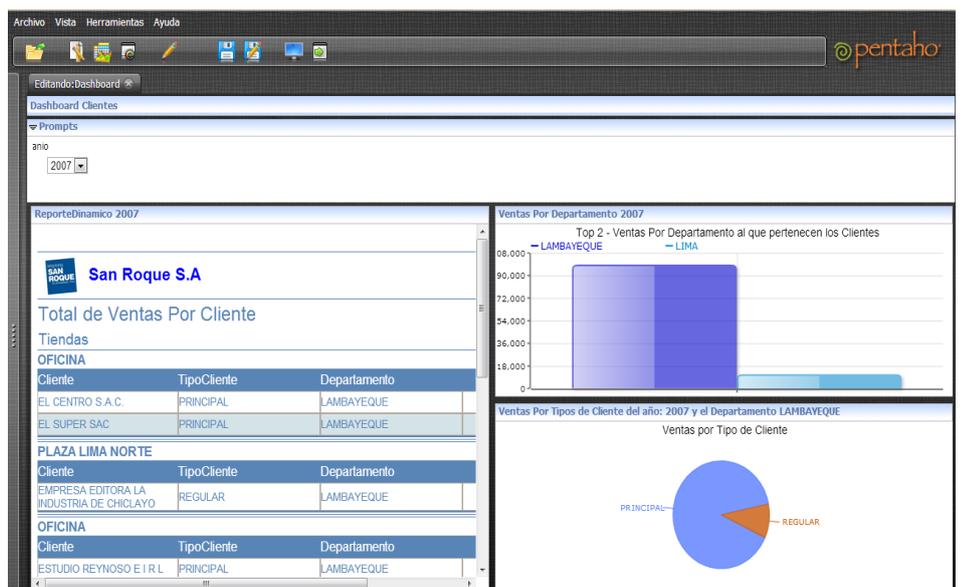


Figura 88: Vista Final del Tablero de Mando total de ventas por Cliente.

d) Tablero de Mando Ranking de las Frases más utilizadas

Este tablero de mando contiene en su primera sección un reporte generado previamente con la herramienta pentaho reporting, el cual muestra el número de visitas por páginas web. En la segunda sección el tablero muestra un gráfico de barras correspondiente al Top 10 de las frases más utilizadas por los usuarios en los buscadores para ubicar la página de la empresa. En la tercera sección el tablero de mando muestra una tabla de análisis correspondiente a la información interna de los productos más vendidos.

A continuación en la siguiente figura se muestra la vista final del tablero de mando.

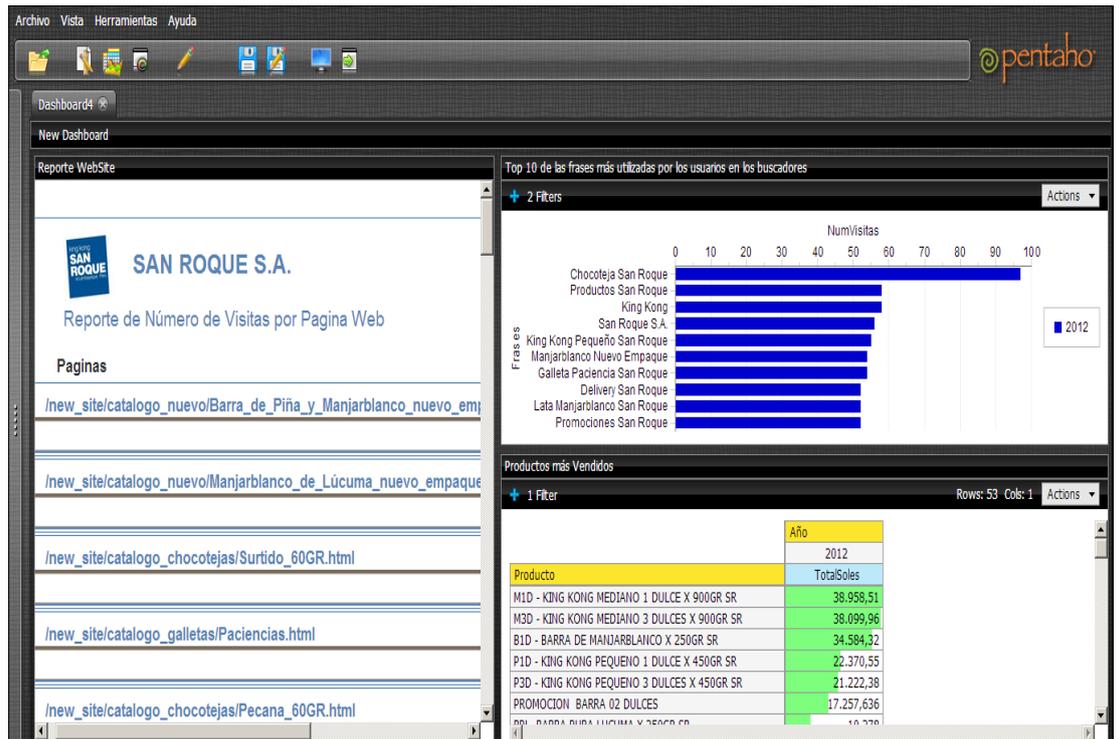


Figura 89: Vista Final del Tablero de Mando Ranking de las frases más utilizadas.

e) Tablero de Mando número de visitas a la página de un determinado producto

A continuación en las siguiente figura se muestra la vista final del tablero de mando filtrado por el producto seleccionado “M1D King Kong Mediano 1 Dulce x 900 GR”, mostrando en la primera sección el número de visitas a la página donde se encuentra el producto seleccionado y en la siguiente sección se muestra el número de visitas de las Top 5 URL previas a la página del producto seleccionado en la tabla de análisis.

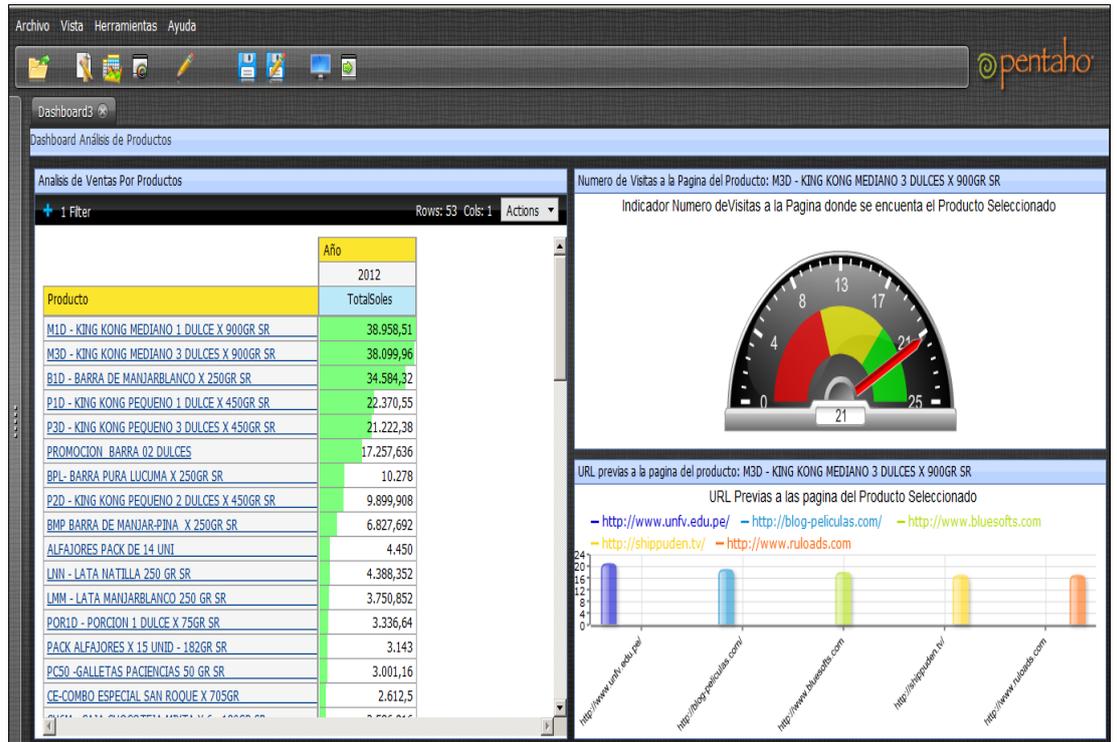


Figura 90: Vista Final del Tablero de Mando Número de visitas a la página de un determinado producto.

5.1.5.6.3. Vista de Análisis

Permite construir una vista personalizada en la cual el reporte muestre diferentes enfoques de análisis según las necesidades específicas, consta de una barra de herramientas para personalizar el reporte, en la siguiente figura a manera de ejemplo se muestra el análisis de la cantidad de ventas realizadas por año.

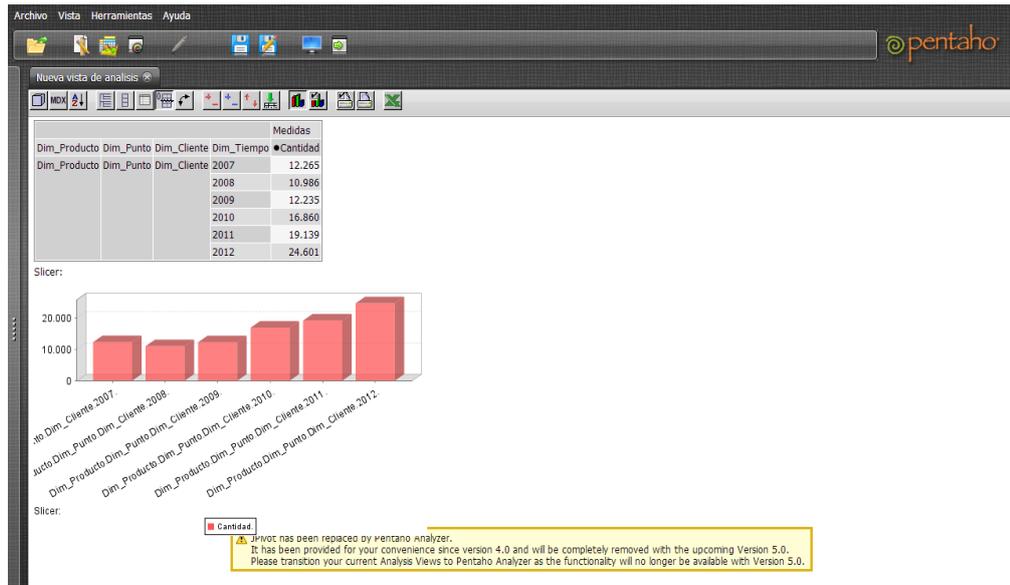


Figura 91: Cantidad de Ventas por año.

En la siguiente figura se muestra el análisis de la cantidad de ventas por sucursal y por año.

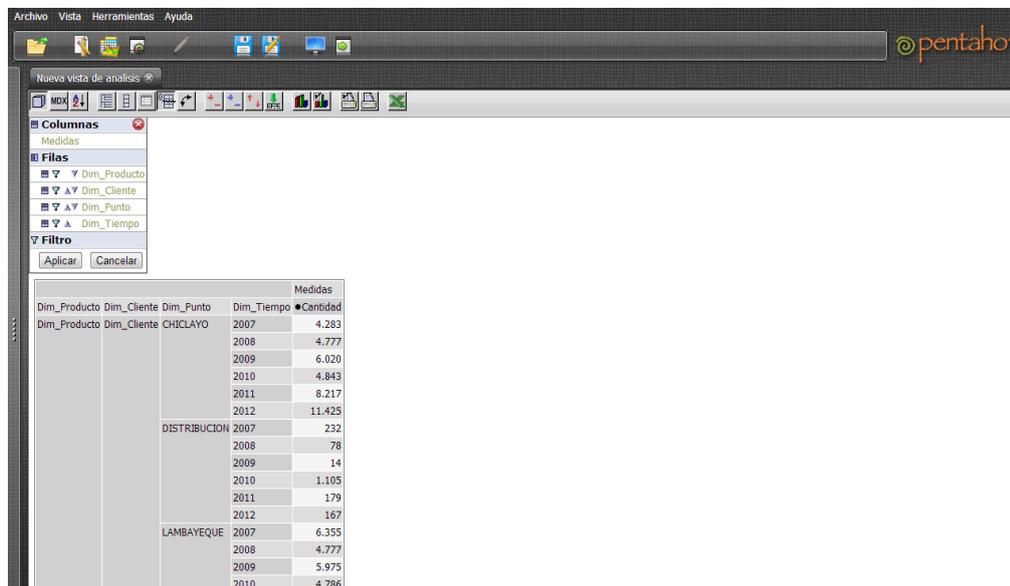


Figura 92: Cantidad de Ventas por sucursal y por año.

5.1.5.6.4. Analyzer Report

El desarrollo de Analyzer Report permite construir una vista multidimensional con el ratón, arrastrando y soltando las dimensiones e indicadores, y a continuación filtrando y excluyendo valores. Permite a los usuarios finales crear campos calculados, introduciendo la fórmula y, por consiguiente, sin un límite de complejidad, es decir les proporciona una herramienta de análisis que combina facilidad de acceso y potencia. A continuación se muestra algunos de los Analyzer Report creados para el presente proyecto:

a) Total de Ventas por Sub Familia, Producto, Tienda de los años 2011 y 2012.

The screenshot shows the Analyzer Report interface with the following configuration:

- Layout:** SubFamilia, Producto, Tienda, Año.
- Columns:** Año.
- Measures:** TotalDolares, TotalSoles.

The resulting data table is as follows:

SubFamilia	Producto	Tienda	Año	
			2011	2012
KING KONG	M3D - KING KONG MEDIANO 3 DULCE	FABRICA	15.359,88	14.643,5
		FERIAS	45	-
		LARCO MAR	22,5	877
		LIMA AEROPUERTO	20,5	82
		MALL AVENTURA PLAZA	136,33	-
		OFICINA	269,33	108,78
		OPEN PLAZA	940	1.670,95
		PLAZA LIMA NORTE	279,25	140
		REAL PLAZA	781,88	1.943,56
		SAN JOSE	5.220	7.344,67
		TRUJILLO ADP	175,5	965
		BALTA	260,28	1.087,07
		BOLOGHESI	3.173,04	3.860,5
		CENTRO CIVICO REAL P	67,8	54
P1D - KING KONG PEQUEÑO 1 DULCE	DISTRIBUCION	DOS DE MAYO	797,5	997
		FABRICA	6.479,5	9.211
		LARCO MAR	36,48	168,5
		LIMA AEROPUERTO	-	27
		OFICINA	13,5	132,98
		OPEN PLAZA	391,5	537,5
		PLAZA LIMA NORTE	72,5	-

Figura 93: Analyzer Report - Total de Ventas.

b) Cantidad de Ventas por Sub Familia, Producto, Sucursal de los años 2011 y 2012.

The screenshot shows a Pentaho Analyzer report with the following data:

SubFamilia	Producto	Sucursal	Año	
			2011	2012
KING KONG	M3D - KING KONG MEDIANO 3 DULCE	CHICLAYO	628	831
		DISTRIBUCION	-	15
		LAMBAYEQUE	860	765
	P1D - KING KONG PEQUEÑO 1 DULCE	LIMA	20	41
		TRUJILLO	15	39
		CHICLAYO	458	712
	P2D - KING KONG PEQUEÑO 2 DULCE	DISTRIBUCION	5	11
		LAMBAYEQUE	503	693
		LIMA	13	16
	P3D - KING KONG PEQUEÑO 3 DULCE	TRUJILLO	12	61
		CHICLAYO	315	308
		DISTRIBUCION	11	-
	P2D - KING KONG PEQUEÑO 2 DULCE	LAMBAYEQUE	354	364
		LIMA	24	9
		PIURA	-	6
TRUJILLO		2	26	
CHICLAYO		600	704	
DISTRIBUCION		16	16	
P3D - KING KONG PEQUEÑO 3 DULCE	LAMBAYEQUE	821	855	
	LIMA	2	24	
	TRUJILLO	15	42	
		CHICLAYO	590	604

Figura 94: Analyzer Report - Cantidad de Ventas

c) Total de costo y venta en soles por producto y por tienda de los años 2011 y 2012.

The screenshot shows a Pentaho Analyzer report with the following data:

Producto	Tienda	Año			
		2011		2012	
		TotalCostoSoles	TotalSoles	TotalCostoSoles	TotalSoles
ALFAJORES PACK DE 14 UNI	SAN JOSE	521,504	1,032	357,882	726
	REAL PLAZA	262,589	504	245,961	498
	OPEN PLAZA	134,445	258	102,205	204
	DOS DE MAYO	121,28	240	44,57	90
	BALTA	75,024	144	56,774	114
	TRUJILLO ADP	12,128	24	36,617	96
	PLAZA LIMA NORTE	2,568	4,5	29,39	60
	DISTRIBUCION	18,198	36	-	-
	CENTRO CIVICO REAL P	9,096	18	-	-
	LARCO MAR	3,032	6	2,938	10
B1D - BARRA DE MANJARBLANCO X 250GR SR	FABRICA	9,737,454	13,588,7	8,709,383	13,125,5
	BOLOGNESI	4,256,28	5,930,43	3,759,488	5,560,42
	SAN JOSE	3,507,494	4,892,36	4,039,689	6,042
	DOS DE MAYO	2,075,15	2,902	1,896,237	2,843,76
	REAL PLAZA	994,244	1,228	1,337,636	1,958
	BALTA	763,844	1,034	1,072,207	1,616
	OPEN PLAZA	436,608	594	782,145	1,148,8
	TRUJILLO ADP	7,89	11	940,38	1,655
	MALL AVENTURA PLAZA	252,604	354,5	54,788	82
	PLAZA LIMA NORTE	263,89	354	34,614	62,5
DISTRIBUCION		126,808	170,16	84,798	123,84
	LARCO MAR	141,876	189,75	45,24	99
	OFICINA	126,112	170,5	60,57	90

Figura 95: Analyzer Report – Total de Costo y Ventas.

d) Total de ventas por subfamilia, producto, tipo de cliente, del año 2008 y 2010.

SubFamilia	Producto	Año					
		2008			2010		
		NUEVO TotalSoles	PRINCIPAL TotalSoles	REGULAR TotalSoles	NUEVO TotalSoles	PRINCIPAL TotalSoles	REGULAR TotalSoles
HELADOS	Total	-	3	116	-	1,5	75
KING KONG	M1D - KING KONG MEDIANO 1 DULCE	111,55	18.967,12	1.598,1	3.266,92	14.719,54	5.910,
	M3D - KING KONG MEDIANO 3 DULCE	158,73	15.459,89	1.521,13	4.530,7	13.882,27	5.117,
	B1D - BARRA DE MANJARBLANCO X 2	184,38	15.726,88	1.454,79	2.127,87	14.595,12	3.575,
	P1D - KING KONG PEQUEÑO 1 DULCE	93	9.401,29	779,88	1.422,15	8.580,22	2.990,
	P3D - KING KONG PEQUEÑO 3 DULCE	-	9.114,35	438,35	1.443,75	7.599,57	2.601,
	PROMOCION BARRA 02 DULCES	-	6.325	776,72	1.037,59	7.375,23	3.749,
	PZD - KING KONG PEQUEÑO 2 DULCE	-	4.010,34	307,67	474,21	5.369,43	2.119,
	BMP BARRA DE MANJAR-PINA X 250G	-	3.612,5	272	85,5	4.481,1	1.638,
	BCH- BARRA DE MANJAR DE CHIRIMO	-	4.890	-	-	-	-
	BPL- BARRA PURA LUCUMA X 250GR	60	4.125	-	-	-	-
	POR1D - PORCION 1 DULCE X 75GR S	15	1.210,5	22	-	426,4	-
	KKMINI - KING KONG MINI X UNID - 3	-	1.189,9	-	-	-	-
	PROMOC. MINIPEQUE 03 DUL 225G S	6	615	18	-	228,8	-
	POR2D - PORCION 2 DULCES X 75GR	9	441,5	10	-	196,5	-
PROMOCION KLM - KING KONG LATA	-	-	-	-	598,78	-	
KING KONG	Total	637,66	95.089,27	7.198,64	14.388,69	78.052,96	27.777,4
MIEL DE ABEJA ECOLOGICA	MIEL DE ABEJA ECOLOGICA 500 GR	-	418	-	-	-	-
MIEL DE ABEJA ECOLOGICA	Total	-	418	-	-	-	-
POTES	LNN - LATA NATILLA 250 GR SR	-	928,5	108	-	880,86	-
	IMM - LATA MANJARBLANCO 250 GR	-	766	126	38,48	786,61	-

Figura 96: Analyzer Report – Total de Ventas Por Tipo de Cliente

e) Total de ventas por tienda, por año y por un producto específico del año 2010 y 2011

Esta herramienta permite al usuario final personalizar de manera sencilla la forma de mostrar los valores de las medidas como se aprecia en la siguiente figura en la cual se muestra el valor de la medida “Total en Soles” en forma de barra.

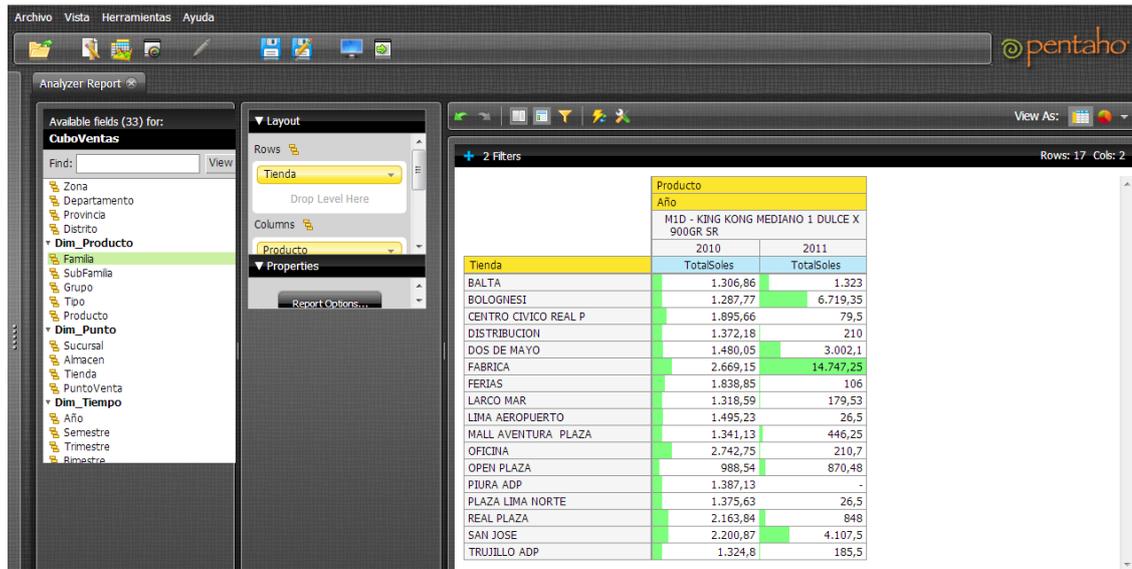


Figura 97: Analyzer Report – Total de Ventas de un producto específico.

f) Total de ventas en Soles por año y por un producto específico.

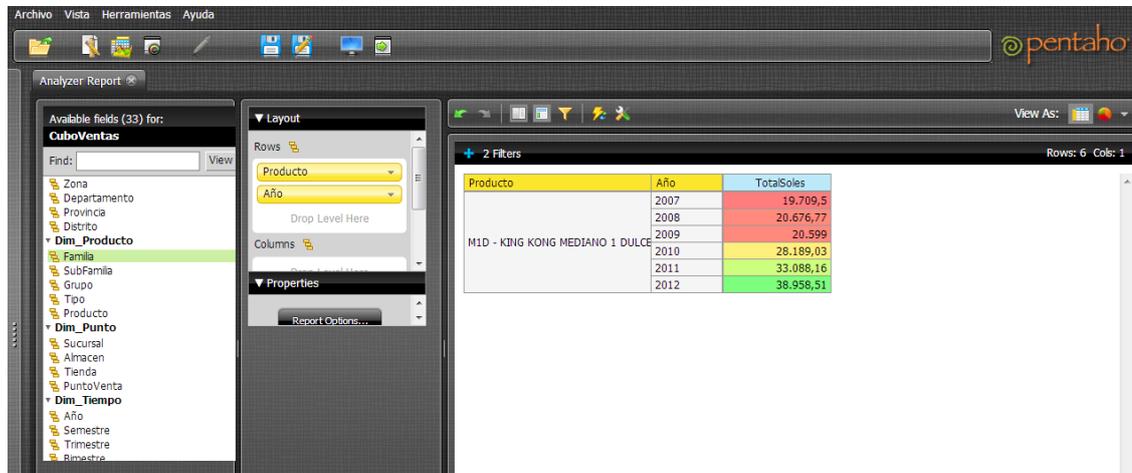


Figura 98: Analyzer Report – Total de Ventas por año.

g) Número de visitas por URL Previa.

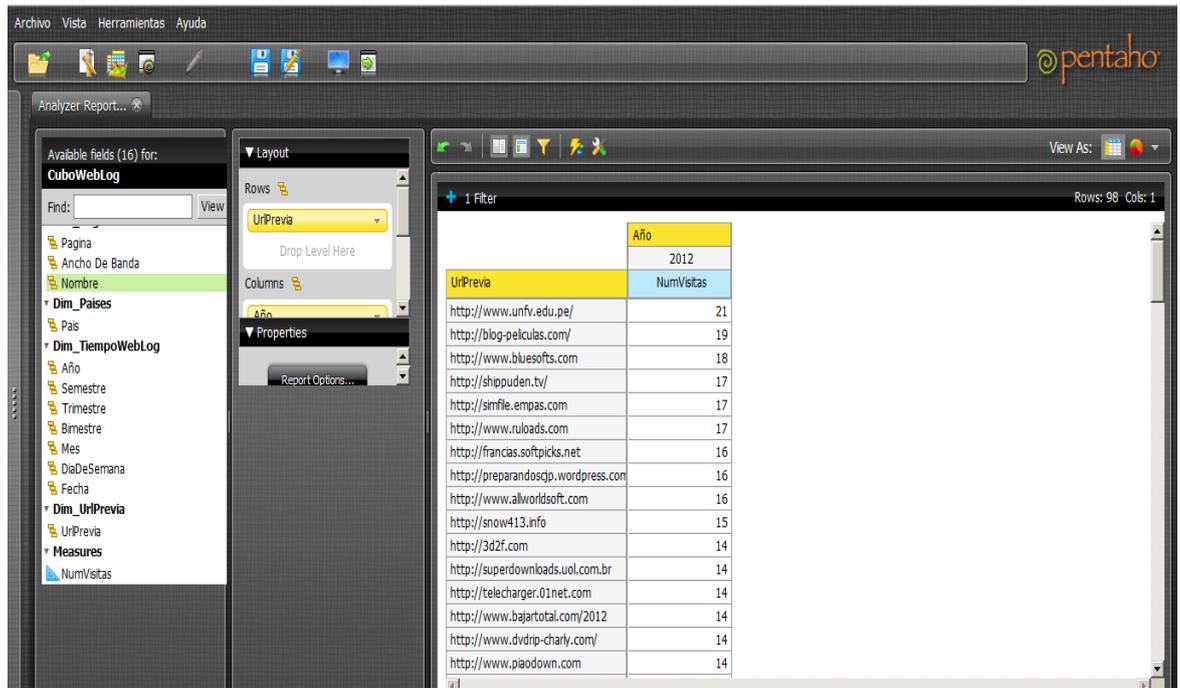


Figura 99: Analyzer Report – Nro de Visitas por URL Previa.

h) Número de visitas por frases utilizadas.

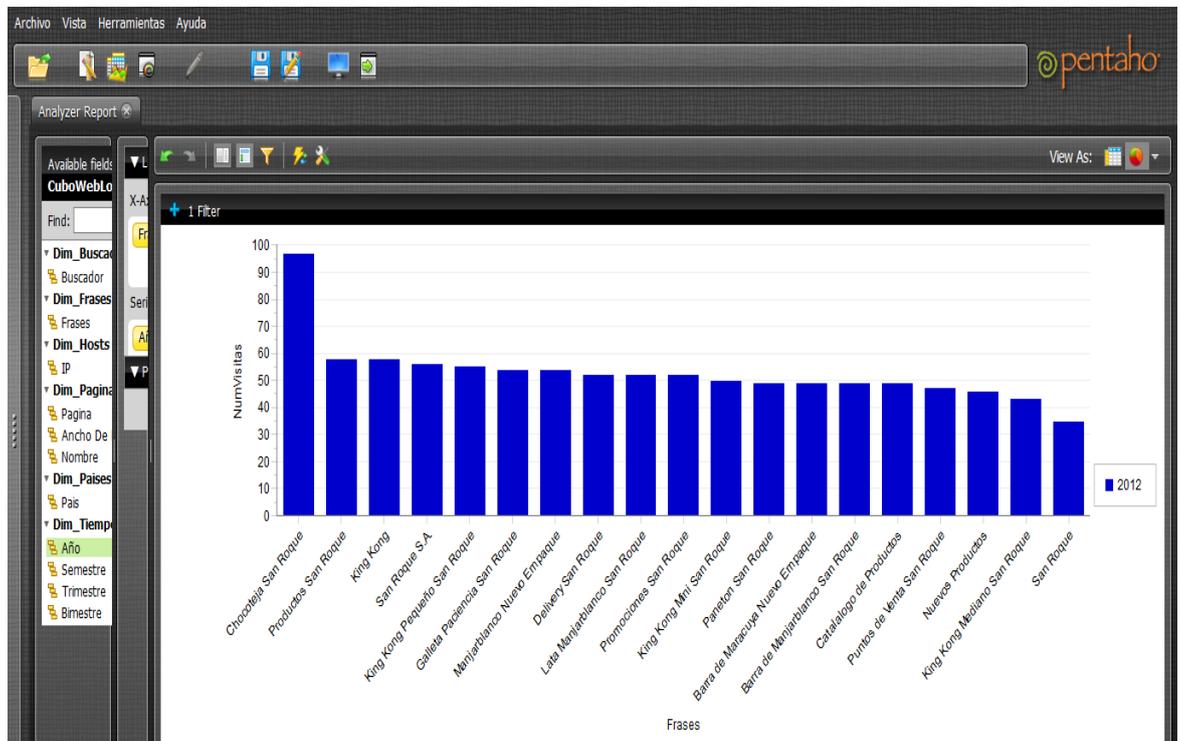


Figura 100: Analyzer Report – Nro de visitas por frases utilizadas.

i) Número de visitas por páginas.

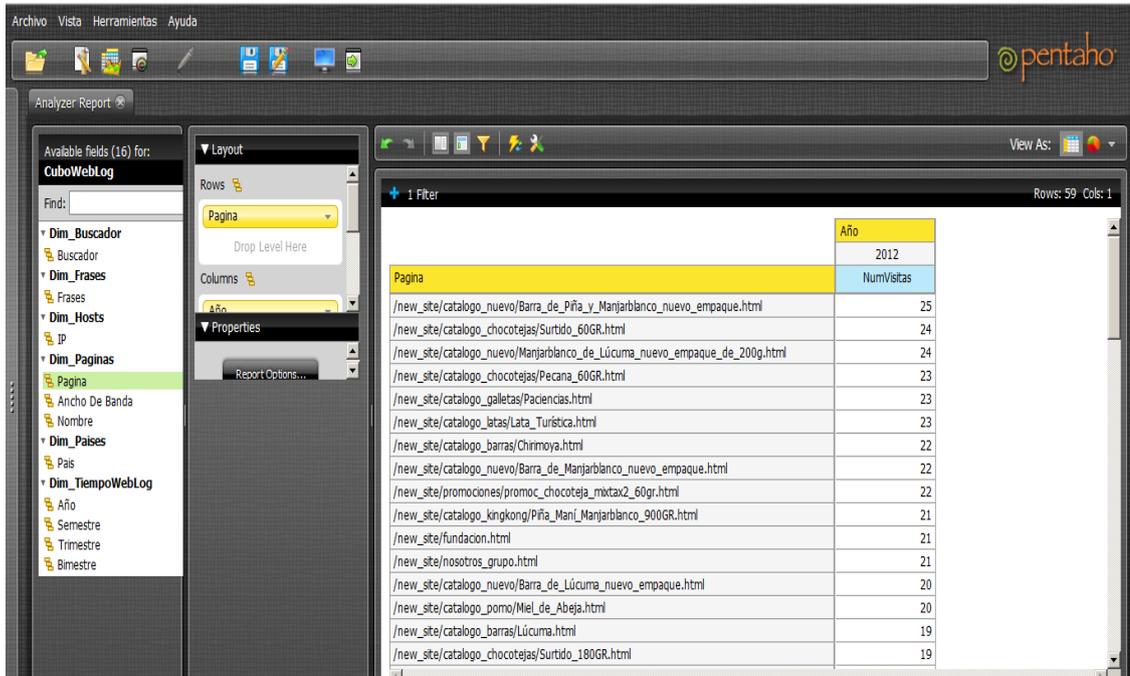


Figura 101: Analyzer Report – Nro de visitas por páginas.

j) Top 10 del Número de visitas por URL y nombre de la página web.

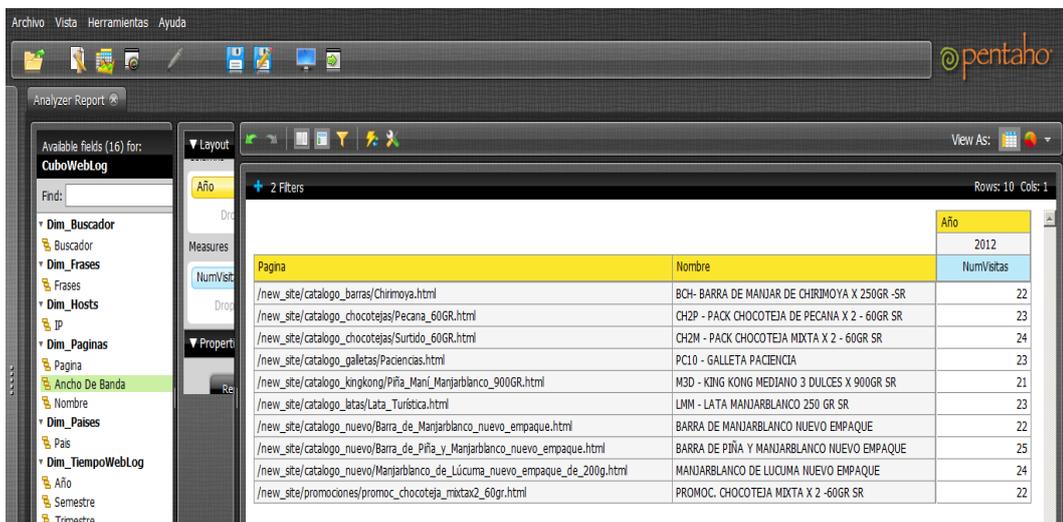


Figura 102: Analyzer Report – Top 10 del Nro de visitas por URL y nombre de la página web.

k) Número de visitas por meses de un determinado año.

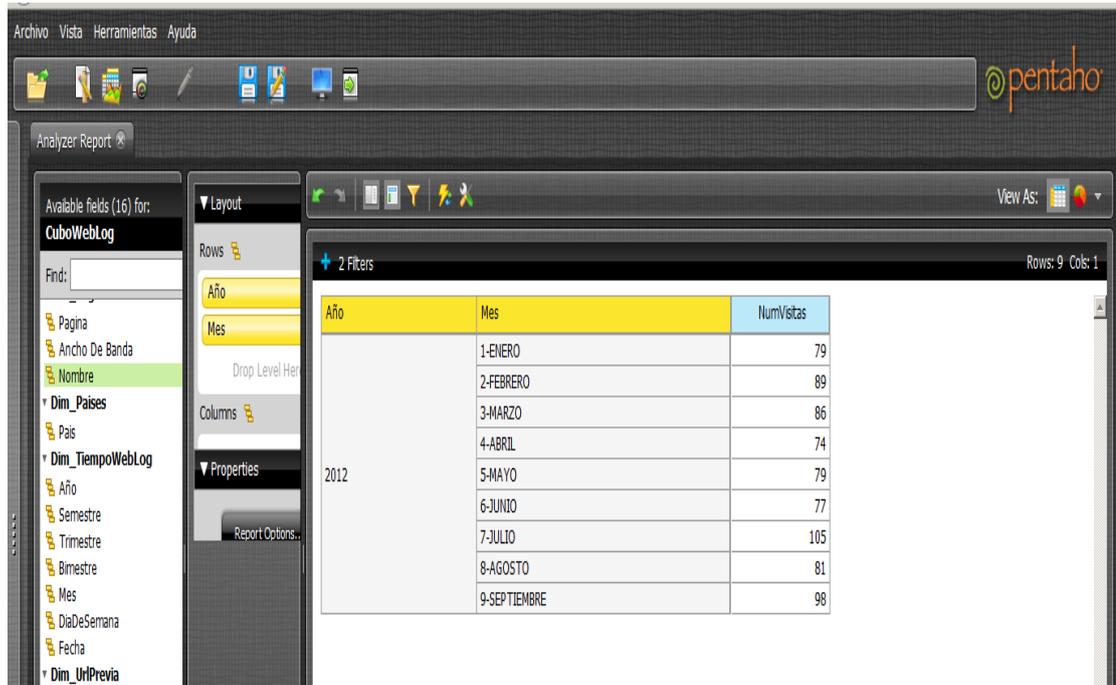


Figura 103: Analyzer Report – Nro de visitas por meses de un determinado año.

l) Top 5 de los meses con mayor número de visitas de un determinado año.

Archivo Vista Herramientas Ayuda

Analyzer Report

Available fields (16) for: CuboWebLog

Find:

Pagina

Ancho De Banda

Nombre

Dim_Paises

Pais

Dim_TiempoWebLog

Año

Semestre

Trimestre

Bimestre

Mes

DiaDeSemana

Fecha

Dim_UrlPrevia

UrlPrevia

Measures

NumVisitas

Layout

Rows

Año

Mes

Drop Level Here

Columns

Properties

Report Options...

2 Filters

Año includes 2012

Top 5 Mes by NumVisitas

Rows: 5 Cols: 1

Año	Mes	NumVisitas
2012	7-JULIO	105
	9-SEPTIEMBRE	98
	10-OCTUBRE	94
	2-FEBRERO	89
	3-MARZO	86

Figura 104: Analyzer Report – Top 5 de los meses con mayor número de visitas.

5.1.5.7 PRUEBAS

a) Pruebas de Ejecución de Procesos de Carga.

Luego de la construcción de los procesos de carga se procede a su ejecución mediante la opción de la herramienta Spoon.

A modo de prueba se ejecuta el proceso ETL y al terminar de correr el proceso la herramienta indica que este finaliza de manera satisfactoria como se muestra en la siguiente figura:

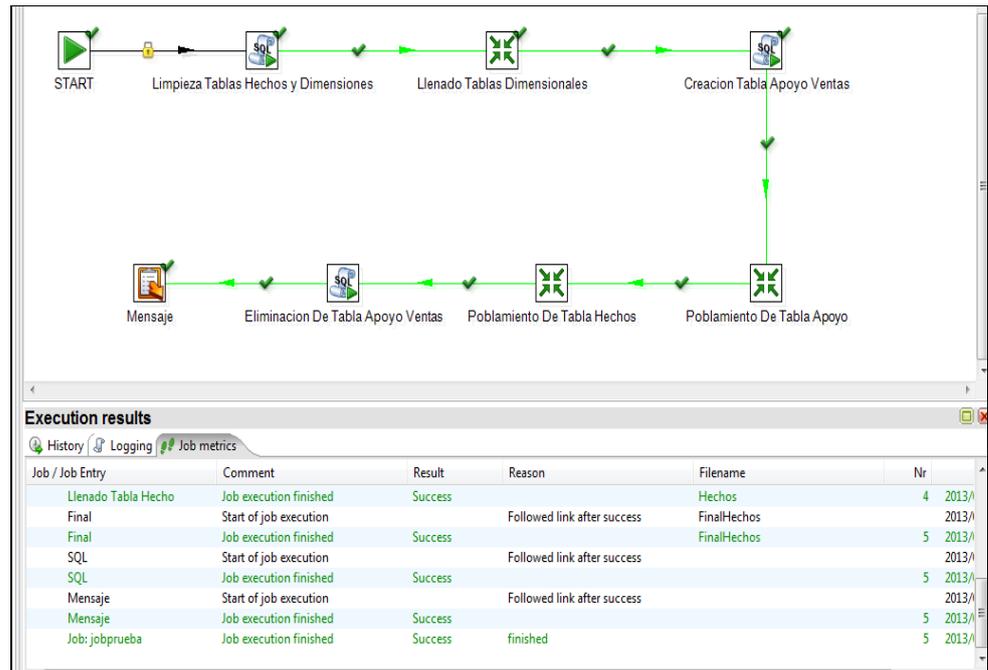


Figura 105: Proceso de carga satisfactoria.

Luego de la ejecución del proceso de carga, se verifica que las tablas dimensionales y hechos OLAP contengan los datos que se encuentran en la base de datos OLTP:

Se puede observar en la siguiente figura el Data Mart “ventasmart” totalmente poblado.

MySQL Workbench

SQL Editor (Mysql@127.0.0.1...)

File Edit View Query Database Plugins Scripting Help

Object Browser

SCHEMAS

Search objects

sakila

test

ventasmart

Tables

dbo_dimbuscadores

dbo_dimcliente

dbo_dimfrases

dbo_dimhost

dbo_dimpaginas

dbo_dimpaises

dbo_dimproducto

dbo_dimpunto

dbo_dimtiempo

dbo_dimtiempoweblog

dbo_dimurlprevia

dbo_hechosventa

dbo_hechosweblog

Views

Routines

weblog

dbo_dimproducto

dbo_hechosventa

dbo_hechosweblog

dbo_dimpaises

dbo_dimhost

dbo_dimfrases

Filter:

KeyProducto	CodProducto	Producto	Tipo	Grupo
1	0101-010001	M3D - KING KONG MEDIANO 3 DULCES X 900GR SR	NO PROMOCION	AZUL
2	0101-010006	P3D - KING KONG PEQUENO 3 DULCES X 450GR SR	NO PROMOCION	AZUL
3	0101-010011	PROMOC. MINIPEQUE 03 DUL 225G SR	PROMOCION	AZUL
7	0101-020001	PROMOCION BARRA 02 DULCES	PROMOCION	NARANJAS
8	0101-020006	P2D - KING KONG PEQUENO 2 DULCES X 450GR SR	NO PROMOCION	NARANJAS
9	0101-020016	POR2D - PORCION 2 DULCES X 75GR SR	NO PROMOCION	NARANJAS
14	0101-020041	BMP BARRA DE MANJAR-PINA X 250GR SR	NO PROMOCION	NARANJAS
18	0101-030001	M1D - KING KONG MEDIANO 1 DULCE X 900GR SR	NO PROMOCION	ROJAS
19	0101-030006	P1D - KING KONG PEQUENO 1 DULCE X 450GR SR	NO PROMOCION	ROJAS
20	0101-030011	B1D - BARRA DE MANJARBLANCO X 250GR SR	NO PROMOCION	ROJAS
21	0101-030016	POR1D - PORCION 1 DULCE X 75GR SR	NO PROMOCION	ROJAS
27	0101-030046	KKMINI - KING KONG MINI X UNID - 30GR SR	NO PROMOCION	ROJAS
28	0101-040001	PROMOCION KLM - KING KONG LATA SR	PROMOCION	LATA
33	0101-050001	BPL - BARRA PURA LUCUMA X 250GR SR	NO PROMOCION	VERDE

Figura 106: Data Mart Ventas Poblado.

b) Pruebas de Interfaces

Como resultado final se obtienen las interfaces en sí. A continuación se muestra a manera de ejemplo las pruebas a uno de los tipos de interface desarrollados, los tableros de mando. Se realizara la prueba con dos tableros de mando desarrollados.

En las siguientes figuras se muestra para el tablero de mando “Análisis de Ventas” y para el tablero “Análisis de Ventas por Productos”, los resultados esperados y los resultados del tablero.

- **Los resultados esperados**, son los resultados que se obtienen después de trabajar la data manualmente y a los cuales se les propuso una mejora en la presentación de estos, en la etapa de diseño de la explotación de la fase de diseño.
- **Los resultados del tablero**, son los que se obtienen a través del datamart a través de la interface desarrollada.

Tablero Análisis de Ventas

Resultado Esperado:

CUADRO DE MANDO ANALISIS DE VENTAS				
Analisis de Ventas			Indicador Cantidad de ventas	
		Año		13 500
		2011	2012	
Sucursal	Subfamilia			
	Bebidas	11540,00	13400,000	
	Dulceria	9540,000	9750,000	
Lambayeque	Helados	8400,00	4320,00	
	King Kong	9500,00	7850,000	
	Potes	12640,00	9560,000	
	Miel de Abeja Ecologica	11300,00	11200,00	
	Total Soles	62920,000	56080,000	Indicador Total de Ventas
	Dulceria	9780,00	11340,00	125 650
Distribucion	King Kong	10500,00	9540,00	
	Potes	12740,00	13256,00	
	Total Soles	33020,00	34136,00	
	Bebidas	12640,00	12900,00	
	Dulceria	11230,00	11340,00	
Lima	Helados	14320,00	9560,00	
	King Kong	12450,00	10420,00	
	Total Soles	50640,00	44220,00	
Filtro	Año : 2010			

Figura 107: Ejecución Análisis de Ventas - Muestra.

Resultado del Tablero:

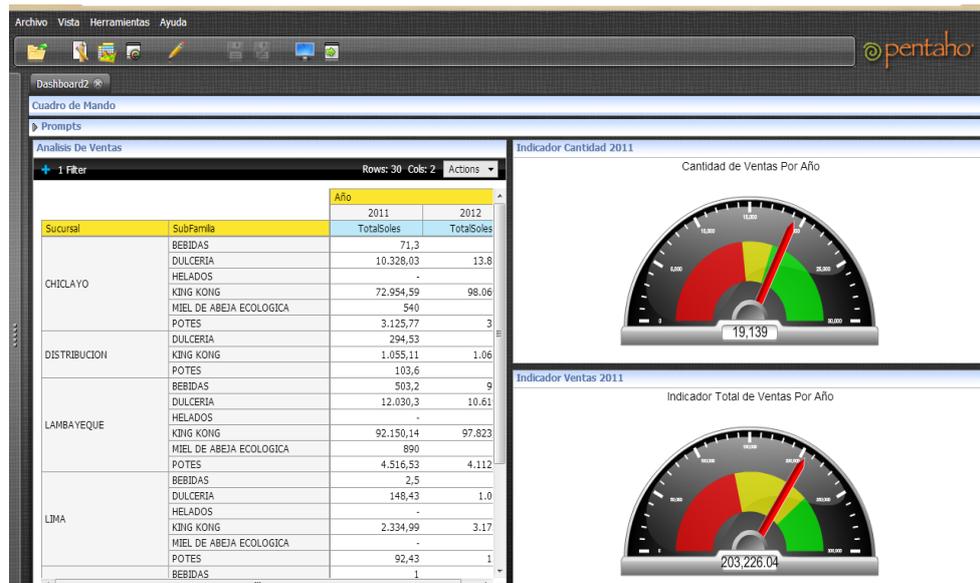


Figura 108: Ejecución Análisis de Ventas - Tablero.

Resultados:

Se verifica que el tablero muestra los datos según los resultados esperados.

Tablero Análisis de Ventas por Productos

Resultado Esperado:

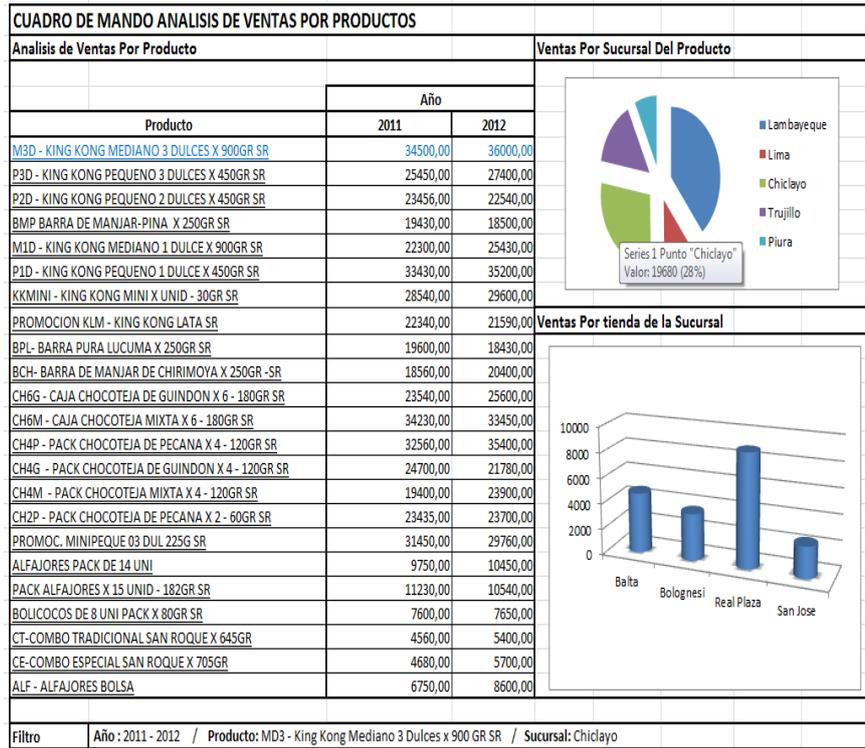


Figura 109: Ejecución Análisis de Ventas por Productos - Muestra.

Resultado del Tablero:

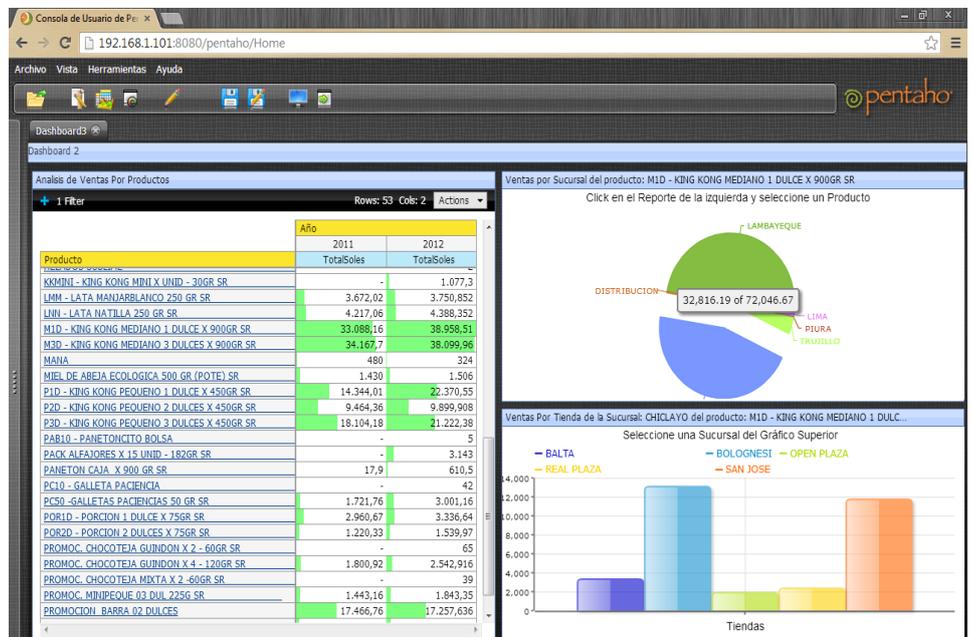


Figura 110: Ejecución Análisis de Ventas por Productos - Tablero.

Resultados:

Se verifica que el tablero muestra los datos según los resultados esperados.

5.1.6. FASE 6: IMPLANTACIÓN

5.1.6.1. IMPLANTACIÓN DE LA SOLUCIÓN

Para esta etapa algunas de las actividades a realizar son:

- Preparar ambiente para la implantación.
- Instalación de la aplicación BI.
- Administración de usuarios.
- Constante capacitación a los usuarios.
- Establecer procedimientos de operación y mantenimiento.
- Administración de procesos ETL.
- Administración de aplicaciones de explotación.

Monitoreo

Se realizaran las funciones de soporte, que incluye mantener la base de datos, tráfico, programar y correr el job (ETL), mantenimiento de metadata, tuning (puesta a punto de consulta) de rendimiento sobre las consultas, monitorear el comportamiento del sistema.

Se verificara que las cargas de datos y las consultas se estén ejecutando correctamente.

CAPITULO VI: RESULTADOS

En este capítulo, se muestran los resultados de la evaluación del modelo de inteligencia de negocio propuesto, a través de la opinión de expertos y también se muestran los resultados obtenidos de la aplicación del modelo propuesto en el área de ventas de la empresa San Roque S.A.

6.1. Proceso de Toma de Decisiones

Como resultado al análisis del proceso de toma de decisiones actual en el área de ventas de la empresa San Roque, se obtuvo el siguiente modelo de proceso de Toma de Decisiones, el cual describe la forma en cómo se realizaba este proceso y se muestra en la siguiente figura:

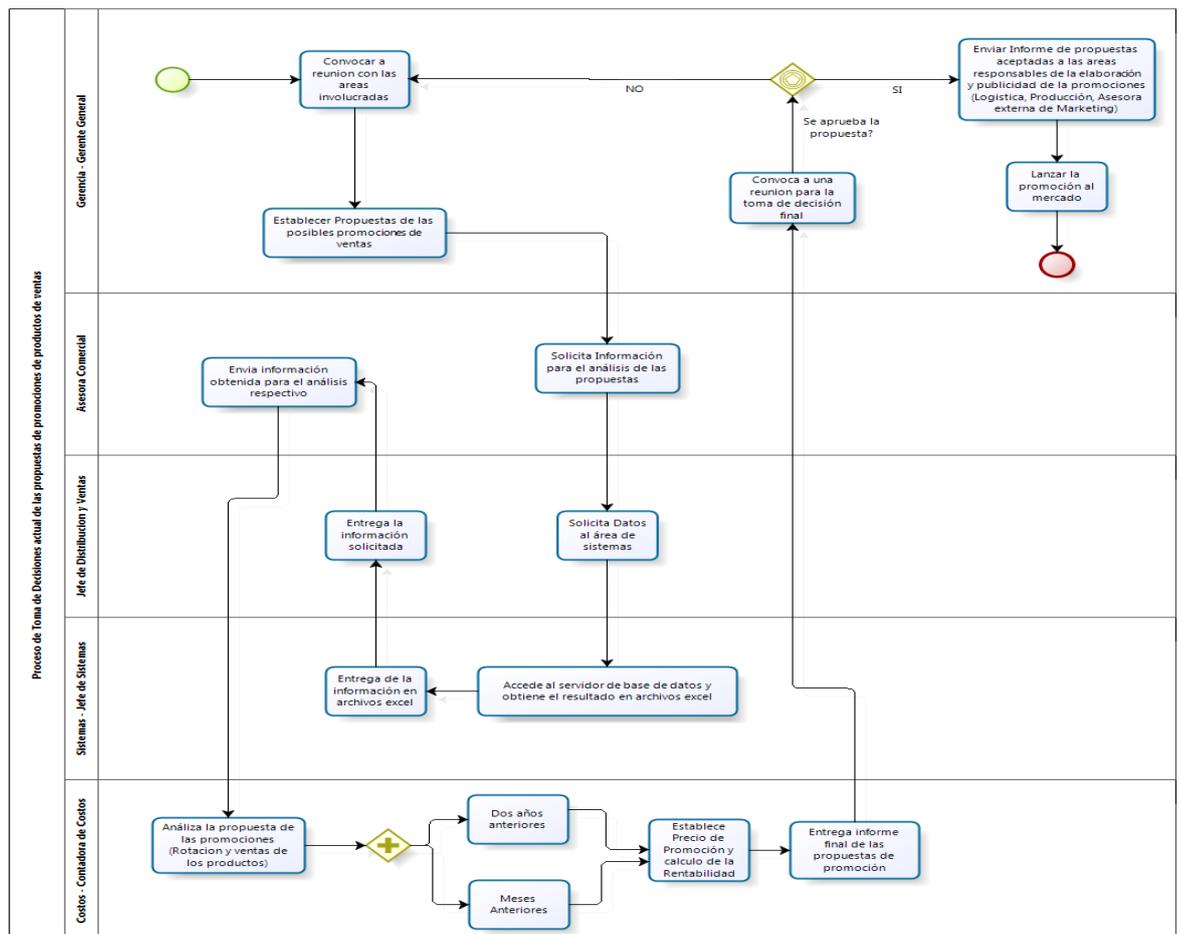


Figura 111: Modelo del proceso de toma de decisiones actual.

6.2. Propuesta del Modelo de Inteligencia de Negocio

Como resultado al estudio y comparación realizados a las metodologías de desarrollo de inteligencia de negocios, tales como la metodología de Bill Inmon, Ralph Kimball, Josep Curto y Big Data, se propuso un modelo de inteligencia de negocio para el presente proyecto el cual se muestra en la siguiente figura:

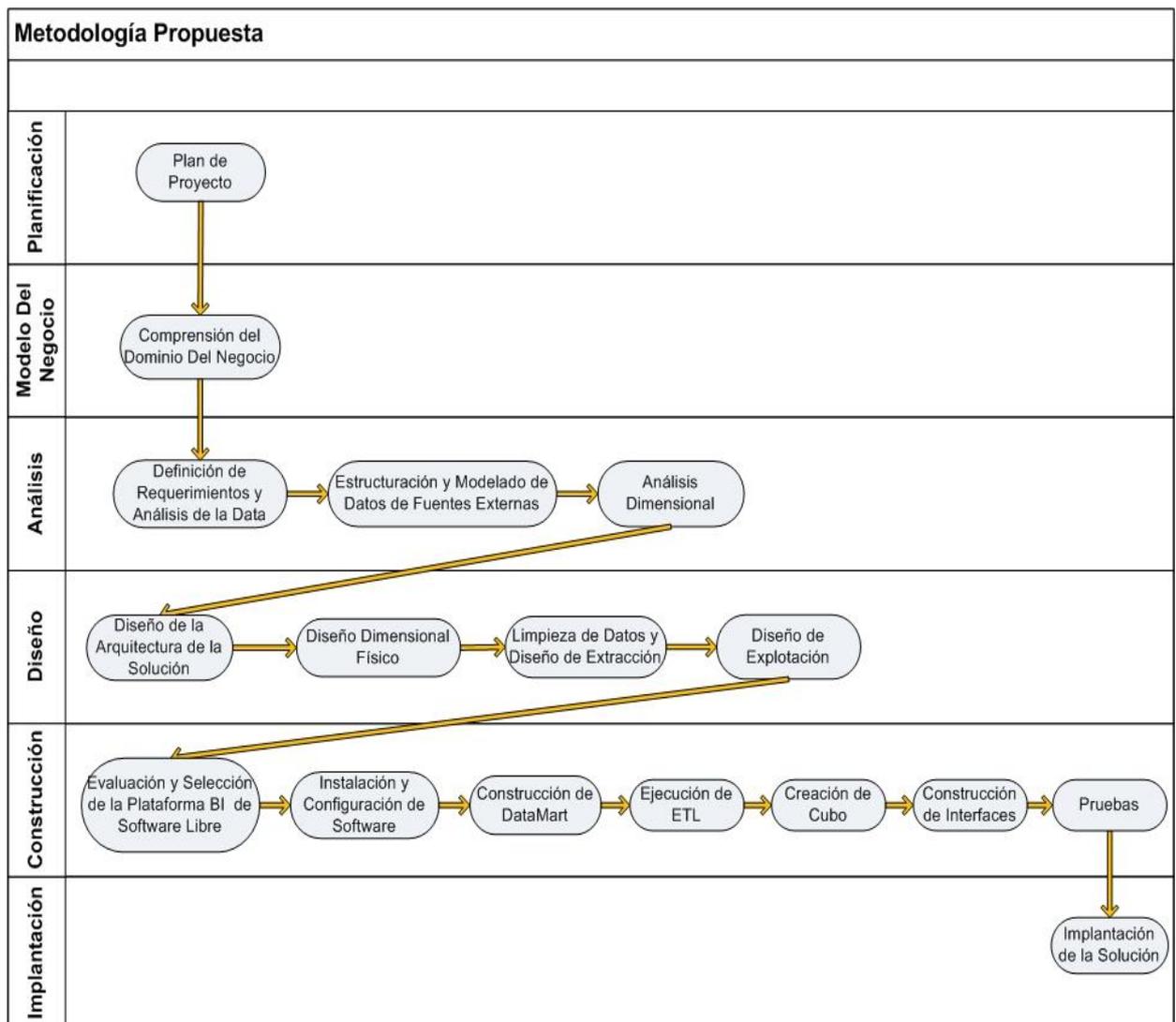


Figura 112: Modelo de Inteligencia de Negocio Propuesta.

6.3. Despliegue del Modelo de Inteligencia de Negocio Propuesto

En la siguiente figura se muestra el proceso de toma de decisiones aplicando el Modelo de inteligencia de negocio propuesto y la herramienta BI implementada con la plataforma Pentaho:

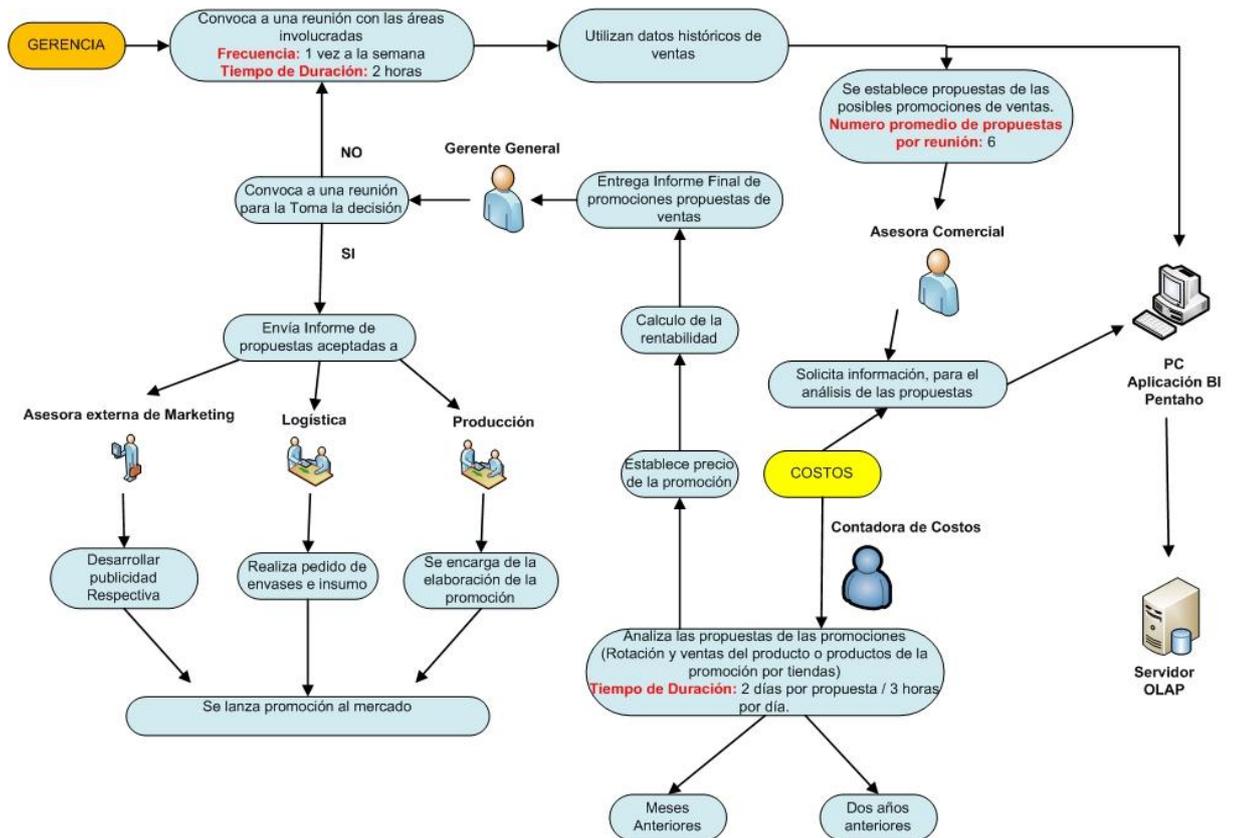


Figura 113: Proceso de Toma de Decisiones aplicando el Modelo BI propuesto.

6.4. Validación del Modelo de Inteligencia de Negocio Propuesto

VARIABLE INDEPENDIENTE: El Modelo de Inteligencia de Negocios.

Los resultados de la evaluación del modelo de inteligencia de negocios a través de la opinión de expertos son las siguientes:

6.4.1. Cualidades evaluadas por expertos en el Modelo de Inteligencia de Negocio Propuesta.

Para obtener las siguientes puntuaciones de los atributos del Modelo de Inteligencia de Negocio propuesta, se implementó una solución de inteligencia de negocio y se expuso las fases (actividades, características y componentes) a cinco expertos en el desarrollo de soluciones de Inteligencia de Negocios de la empresa AC SOFT LIVE E.I.R.L., los cuales respondieron a la encuesta mostrada en el ANEXO 19, obteniendo los resultados mostrados en el ANEXO 20, en la siguiente tabla se muestra el promedio de la puntuación por atributo otorgada por los expertos del modelo propuesto.

Puntuación del Atributo por Expertos (Máxima Puntuación 100)						
Atributo	Experto 1	Experto 2	Experto 3	Experto 4	Experto 5	Promedio del Modelo Propuesto
Adaptación del Modelo	64	67	63	66	69	65,73
Costo de Desarrollo	70	71	73	75	67	71,10
Tiempo de Desarrollo	82	83	86	81	84	83,27
Nivel del Calidad del Modelo	82	83	85	83	79	82,4
Promedio de Atributos del Modelo						75,63

Tabla 65: Puntuación de Atributos por Expertos, en el Modelo de Inteligencia de Negocios Propuesto.

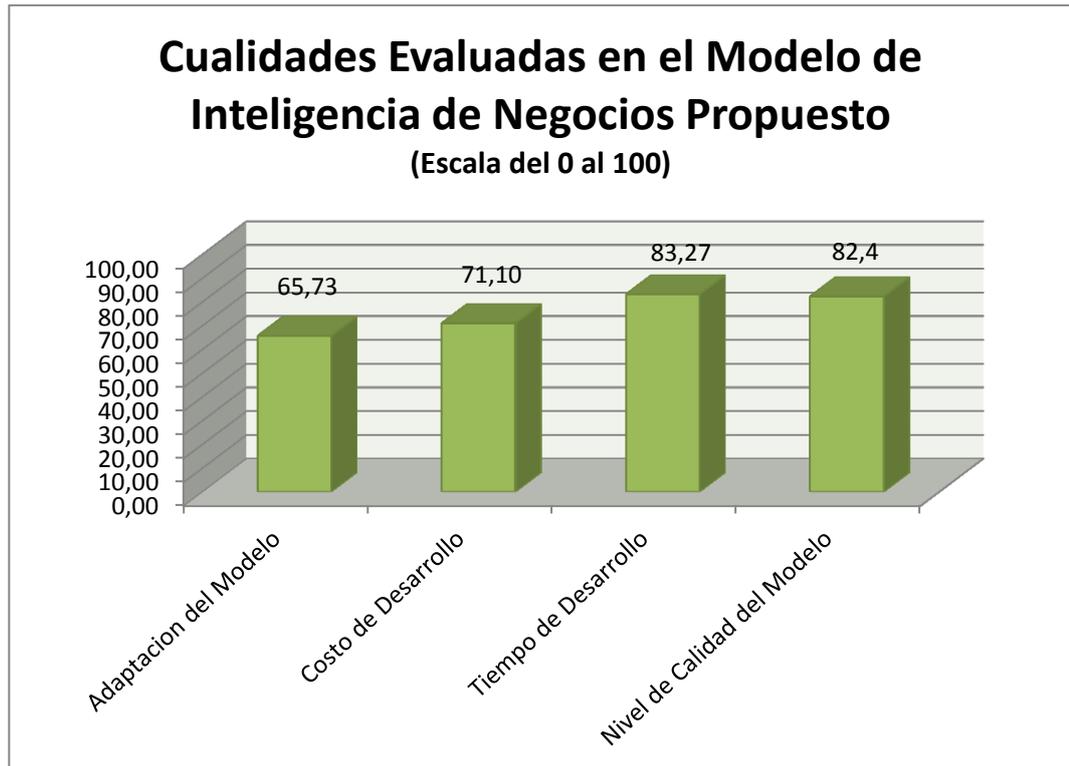


Figura 114: Puntuación promedio de Cualidades Evaluadas por expertos en el Modelo de Inteligencia de Negocios.

Se puede aplicar el Alfa de Crombach a todos los resultados obtenidos y validarlos.

Para identificar que nuestros datos RESULTADO son fiables el Alfa de Cronbach debe tener un valor cercano a 1, caso contrario no serán datos útiles para aplicar análisis estadístico y en tal sentido se sugeriría una normalización de datos.

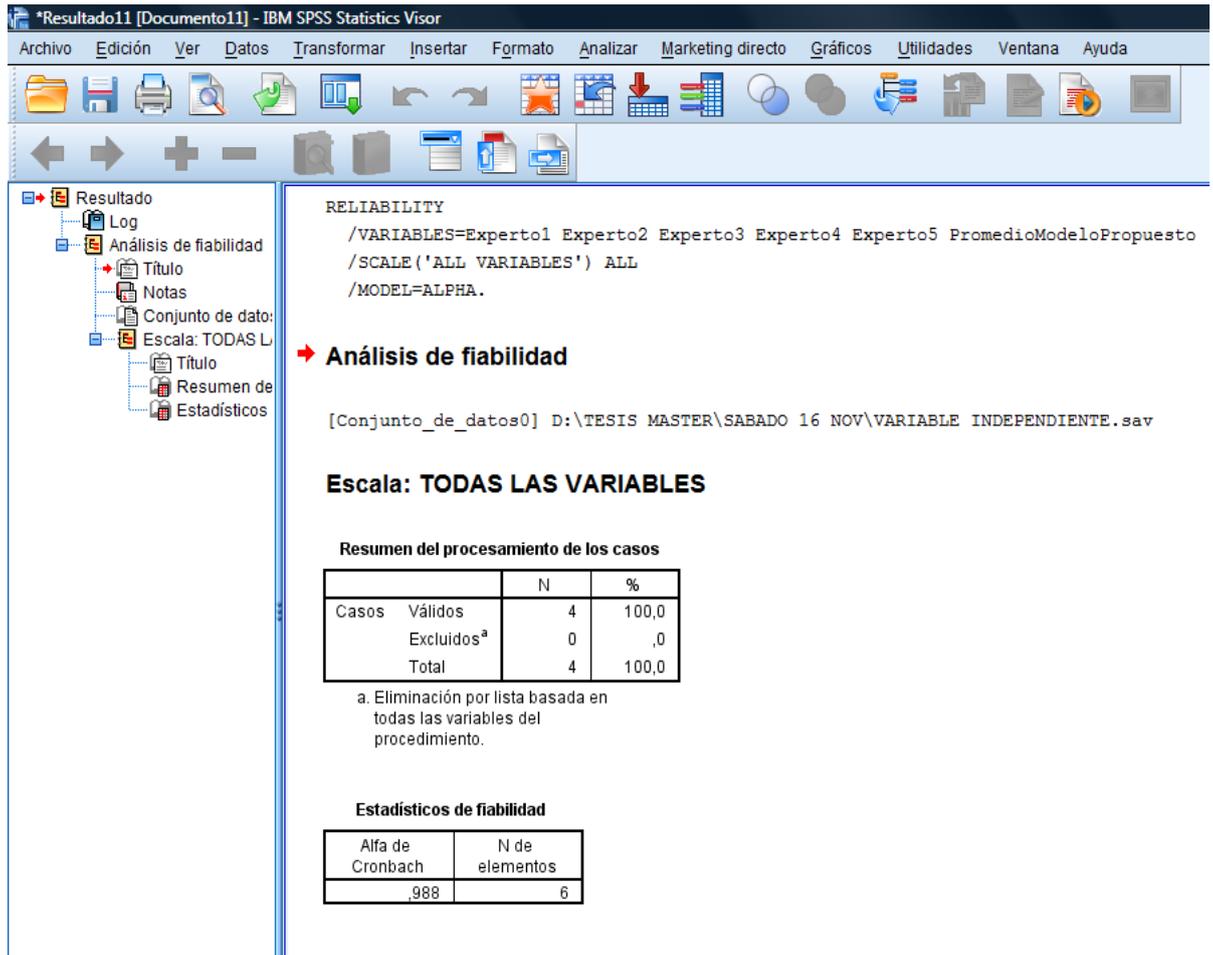


Figura 115: Obtención del Alfa de Cronbach con el Software SPSS.

En el presente caso el Alfa de Cronbach es de 0.988 en las seis columnas o variables de SPSS.

El Alfa de Cronbach, permite medir la fiabilidad (Si los datos son escalares es factible aplicar el Alfa de Cronbach), de una escala de medida que muestra indirectamente una cualidad observable en una población.

VARIABLE DEPENDIENTE: La toma de decisiones en el proceso de venta de la empresa SAN ROQUE S.A.

Los resultados de esta variable comúnmente tienen que ser tomados antes de la aplicación de la variable independiente y luego de la aplicación de la variable independiente (pre test y post test). Así, se

tienen en las siguientes tablas los resultados obtenidos sin la aplicación del modelo de inteligencia de negocios versus los resultados obtenidos de la aplicación del modelo de inteligencia de negocios propuesto al área de ventas de la empresa SAN ROQUE S.A., cuya duración fue de tres meses para ambos casos, durante el mes de marzo a mayo para los datos obtenidos sin el modelo y en el mes de octubre a diciembre para los datos obtenidos con el modelo, para este proceso se tuvo en cuenta la evaluación de los siguientes indicadores:

1. Tiempo de respuesta en formulación de propuestas por reunión.

Nro de Semanas	Sin el Modelo Propuesto.	Con el Modelo Propuesto.
1	3	2
2	2,5	1,5
3	4	2
4	3	1,5
5	2,5	1,5
6	4	2
7	3	1,5
8	3	2
9	2,5	1,5
10	3	2
11	3	1,5
12	2,5	1,5
Promedio:	3	2

Tabla 66: Resultados de la evaluación del indicador - Tiempo de respuesta en formulación de propuestas por reunión.

2. Índice de tiempo de respuesta en el proceso de análisis por propuesta formulada.

Nro de Semanas	Sin el Modelo Propuesto.	Con el Modelo Propuesto.
1	0,25	0,13
2	0,21	0,13
3	0,29	0,08
4	0,25	0,13
5	0,21	0,08
6	0,25	0,13
7	0,25	0,13
8	0,29	0,08
9	0,21	0,13
10	0,29	0,13
11	0,21	0,08
12	0,25	0,13
Promedio:	0,25	0,11

Tabla 67: Resultados de la evaluación del indicador – Índice de Tiempo de respuesta en el proceso de análisis de propuesta formulada.

El detalle de cómo se obtuvieron los resultados presentados de estos dos primeros indicadores se encuentran en el ANEXO 21.

3. Numero de propuestas formuladas por reunión.

Nro de Semanas	Sin el Modelo Propuesto.	Con el Modelo Propuesto.
1	2	4
2	3	5
3	3	6
4	2	4
5	2	5
6	3	4
7	3	4
8	2	4
9	3	5
10	2	6
11	3	6
12	3	6
Promedio	3	5

Tabla 68: Resultados de la evaluación del indicador – Numero de propuestas formuladas por reunión.

CAPITULO VII: DISCUSIÓN DE RESULTADOS

En este capítulo, se somete a discusión, mediante estadística descriptiva, los resultados obtenidos de la evaluación de los atributos del Modelo de Inteligencia de Negocio propuesto para el presente proyecto.

También se somete a discusión, mediante estadística descriptiva, los resultados obtenidos de la evaluación de los indicadores para la toma de decisiones, entre la aplicación del modelo de inteligencia de negocios propuesto versus los resultados obtenidos sin la aplicación del modelo de inteligencia de negocios propuesto, para exponer que la hipótesis: “El desarrollo de un modelo de Inteligencia de Negocios, permitirá una mejora en la toma de decisiones en el área de ventas en la empresa SAN ROQUE S.A.” ha sido demostrada

7.1. Análisis del Proceso de Toma de Decisiones.

Del estudio del proceso de toma de decisiones en el área de ventas de la empresa San Roque S.A., se identificaron los actores y actividades o tareas más relevantes que permitan tomar una buena decisión:

Actores y Actividades

- Gerente General: Establecer las propuestas de las posibles promociones de ventas.
- Asesora Comercial: Solicitar información para el análisis de las propuestas.
- Contadora de Costos: Analizar las propuestas de las promociones en base a la información de la rotación y venta de los productos.
- Contadora de Costos: Establecer el precio de promoción y cálculo de la rentabilidad.

- Gerente General: Convocar a una reunión para evaluar y aprobar las propuestas formuladas.

Propuesta

Luego de identificar las actividades más representativas del proceso, se propone las siguientes actividades para optimizar o mejorar el proceso:

- Gerente General: Establecer las propuestas de las posibles promociones de ventas en base a información histórica y consolidada de las fuentes internas y externas de la organización haciendo uso de una herramienta BI.
- Asesora Comercial: Analizar las propuestas de las promociones como por ejemplo la posibilidad de asociaciones de productos, en base a la información de las ventas de los productos lanzados en promociones anteriormente y que han tenido éxito, y a la información obtenida de mis fuentes externas (Redes Sociales, WebSite) para conocer el impacto de mis productos de las propuestas de las posibles promociones a lanzar, haciendo uso de una herramienta BI.
- Contadora de Costo: Establecer el precio de promoción y cálculo de la rentabilidad en base a la información de la rotación y venta de los productos haciendo uso de una herramienta BI.
- Gerente General: Evalúa y aprueba las propuestas formuladas.

7.2. Análisis del Modelo de Inteligencia de Negocio Propuesto.

El Modelo de Inteligencia de Negocio propuesto proporciona un marco de trabajo consolidado y eficiente que permite un análisis profundo de nuestros datos tanto internos como externos y convertirlo en información valiosa para su posterior análisis que permita tomar decisiones eficientes y acertadas en base a

información sólida, a diferencia de las metodologías estudiadas las cuales presentan ciertas deficiencias y a la vez no se encuentran consolidados los beneficios que ofrecen cada una de ellas en un solo modelo, las características más representativas y ventajosas y destacadas del modelo propuesto son:

- Se está tomando las mejores prácticas de dos metodologías de BI tradicional como son Kimball y Curto y también de la metodología Big Data la cual es considerada la innovación del BI tradicional que proporciona grandes ventajas como el análisis de gran cantidad de información externa no estructurada y cruzarla con nuestra información interna estructurada de fuentes tradicionales como por ejemplo bases de datos transaccionales, esto es destacable ya que va más allá de las metodologías de BI tradicionales propuestas por Kimball y Curto.
- Se está basando en el uso de software libre el cual en la actualidad presenta gran impacto en diversos sectores de las organizaciones.
- Se está cubriendo deficiencias de las metodologías estudiadas como por el ejemplo en el caso de kimball: la limpieza de datos.
- Los atributos o características del Modelo propuesto han sido analizados por expertos de los cuales se han obtenido resultados favorables hacia el modelo propuesto.
- El modelo propuesto ha sido validado a través de su aplicación al área de ventas de la empresa San Roque S.A., del cual se obtuvo resultados positivos que garantizan que el modelo propuesto es de calidad.

7.3. Análisis del Despliegue del Modelo de Inteligencia de Negocio Propuesto.

De los resultados obtenidos en el despliegue del Modelo de Inteligencia de Negocio, se puede observar una reducción del tiempo para generar y analizar las propuestas, así como también el número de personas involucradas en el proceso de toma de decisiones lo cual indica que el proceso de toma de decisiones en el área de ventas de la empresa San Roque S.A., tuvo una mejora significativa.

7.4. Análisis de la validación del Modelo de Inteligencia de Negocio Propuesto.

ANALISIS DE LOS ATRIBUTOS DEL MODELO DE INTELIGENCIA DE NEGOCIOS PROPUESTA.

Según la tabla de resultados obtenidos de la evaluación por expertos de los atributos del modelo propuesto, expuestos en el ANEXO 20, se puede apreciar claramente que el Modelo propuesto resulta mucho más ventajoso en cada uno de sus atributos:

- a) Adaptación del Modelo:** en casi todas las fases presentadas, exceptuando la fase de Planificación y Diseño.
- b) Costos de desarrollo:** es menor en casi todas las fases presentadas, exceptuando la fase de análisis.
- c) Tiempo de desarrollo:** es menor en casi todas las fases presentadas, exceptuando la fase de diseño.
- d) Nivel de calidad del Modelo:** en casi todas las fases presentadas a excepción de la fase modelo del negocio, lo cual resalta a las fases presentadas como fortalezas de la propuesta.

Dando un ponderado a cada criterio para un proyecto de este tipo se encontró que los atributos más resaltantes favorable y

desfavorablemente son: los atributos denominados Nivel de Calidad del Modelo, Tiempo de Desarrollo como favorables y el atributo denominado Adaptación del Modelo y Costo de Desarrollo como desfavorables.

Atributos	Puntaje Final	Ponderado Acorde al Proyecto	Relevancia Final
Adaptación del Modelo	65,73	2	131,46
Costo de Desarrollo	71,10	2	142,2
Tiempo de Desarrollo	83,27	3	249,81
Nivel de Calidad del Modelo	82,4	3	247,2

Tabla 69: Relevancia Final de cada atributo del Modelo de Inteligencia de Negocio Propuesta.

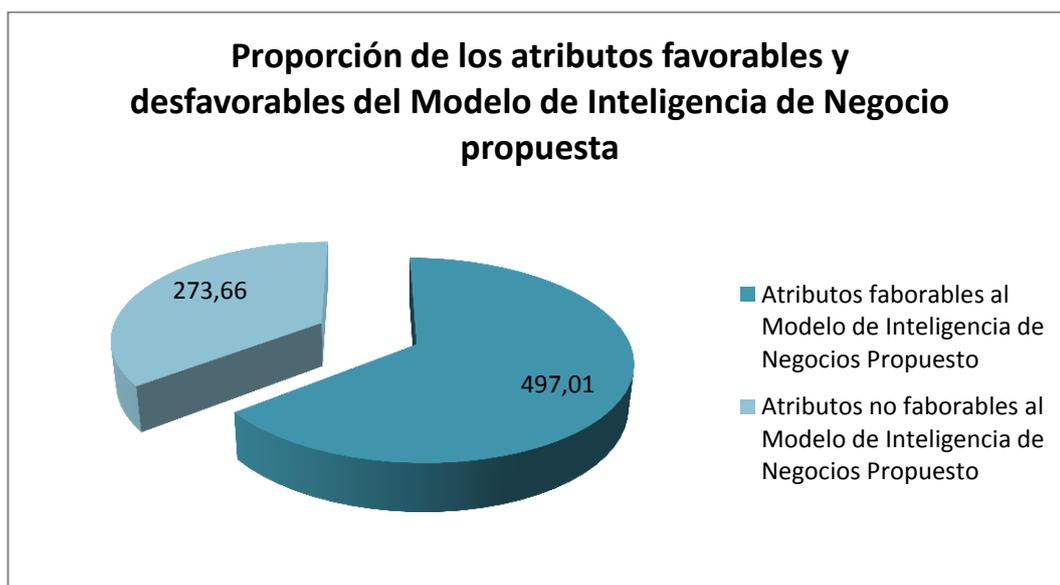


Figura 116: Proporción de los atributos favorables y desfavorables del Modelo de Inteligencia de Negocios Propuesta.

ANALISIS DE LOS RESULTADOS OBTENIDOS DE LA EVALUACION DE LOS INDICADORES DE LA TOMA DE DECISIONES CON EL MODELO DE INTELIGENCIA DE NEGOCIOS PROPUESTO EN COMPARACION CON LOS

RESULTADOS OBTENIDOS SIN EL MODELO DE INTELIGENCIA DE NEGOCIOS PROPUESTO.

En la tabla de Resultados que muestran los datos obtenidos a través de sus respectivos indicadores de la toma de decisiones con la aplicación del modelo de inteligencia de negocios propuesto al área de ventas de la empresa SAN ROQUE S.A., se puede medir a través de la comparación si los resultados indican que el modelo produce una elevación del rendimiento, indicando si los resultados antes y después son significativamente diferentes.

Para el presente proyecto no hace falta realizar Test estadísticos muy sofisticados como la prueba de PRUEBA DE T-STUDENT, Chi-cuadrado, o la prueba de Fisher, para demostrar que realmente mejora el proceso de toma de decisiones en el área de ventas de la empresa San Roque S.A. con la aplicación del Modelo de Inteligencia de Negocios Propuesto, ya que los resultados obtenidos están bastante claros, en la siguiente tabla y figura se puede apreciar a través de los resultados de la evaluación de los indicadores, que si existe una mejora significativa.

Indicadores	Unidad de Medida	Promedio	
		Sin el Modelo Propuesto	Con el Modelo Propuesto
Tiempo de respuesta en formulación de propuestas por reunión.	Horas	3	2
Índice de Tiempo de respuesta en el proceso de análisis por propuesta formulada.	Porcentaje	0,25	0,11
Numero de propuestas formuladas por reunión.	Escalar	3	5

Tabla 70: Promedio Final de cada indicador de la toma de decisiones con el Modelo de Inteligencia de Negocio Propuesto en comparación sin el Modelo Propuesto.

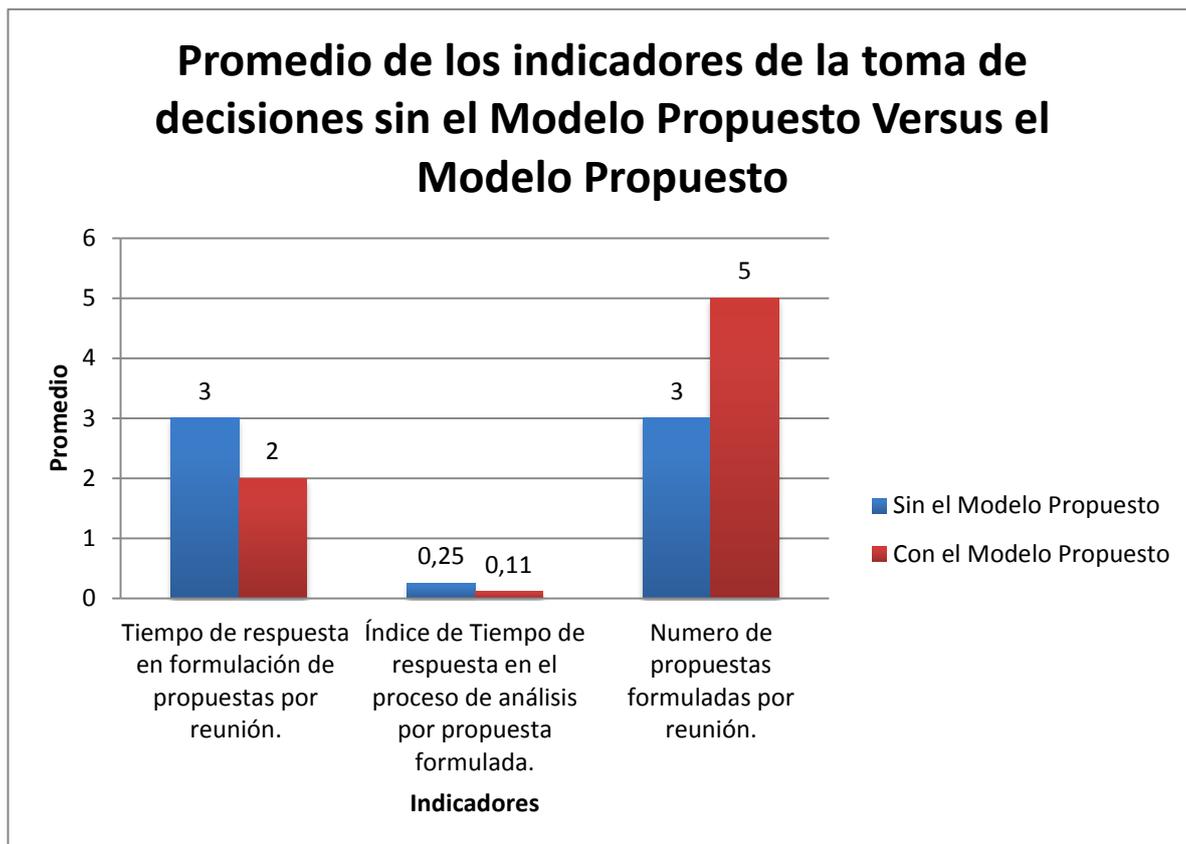


Figura 117: Promedio de los indicadores de la toma de decisiones sin el Modelo Propuesto versus el Modelo Propuesto.

De los resultados obtenidos de la tabla y figura anterior se deduce, que los valores mostrados en ambos casos, son significativamente diferentes. Lo cual nos indica que el Modelo Propuesto es significativamente más relevante que el Modelo de Inteligencia de Negocio que han estado utilizando en la organización, ya que se puede apreciar claramente una disminución tanto en el tiempo en formular propuestas como en el proceso de análisis de propuestas formuladas, también se puede apreciar que el modelo propuesto muestra un incremento del número de propuestas formuladas por reunión.

Por tanto, aceptaríamos la hipótesis de que “El desarrollo de un Modelo de Inteligencia de Negocios, permitirá una mejora en la toma

de decisiones en el área de ventas en la empresa SAN ROQUE S.A.”, por lo que podemos concluir que existe evidencia significativa como para decir, que el Modelo de Inteligencia de Negocio propuesto tiene incidencia en la toma de decisiones en el área de ventas de la empresa San Roque S.A.

CONCLUSIONES

- El proceso de toma de decisiones en el área de ventas en la empresa San Roque S.A presenta una serie de actividades: convocar a una reunión con las áreas involucradas, establecer propuestas de promoción, solicitar información de las propuestas, solicitar datos al área de sistemas, acceder al servidor de base de datos, enviar la información obtenida para el análisis, analizar la propuesta de promociones, establecer precio de promoción y cálculo de rentabilidad, entregar informe final de las propuestas, convocar a una reunión para la toma de decisión final, enviar informe de propuestas aceptadas a las áreas responsables de la elaboración y publicidad de las promociones, y por ultimo lanzar la promoción al mercado, de las cuales las más relevantes son: establecer las propuestas de las posibles promociones de ventas, solicitar información para el análisis de las propuestas, analizar las propuestas de las promociones en base a la información de la rotación y venta de los productos, establecer el precio de promoción y cálculo de la rentabilidad, convocar a una reunión para evaluar y aprobar las propuestas formuladas.
- Teniendo en cuenta que la metodología de Kimball tiene las siguientes fases: Planeación y administración del Proyecto, Definición de los Requerimientos del Negocio, Modelado Dimensional, Diseño Físico, Diseño y Desarrollo de la Presentación de Datos, Diseño de la Arquitectura Técnica, Selección de Productos e Instalación, Especificación de Aplicaciones para Usuarios Finales, Desarrollo de Aplicaciones para Usuarios Finales, Implementación, Mantenimiento y crecimiento, Gestión del Proyecto de la cual destaca la fase de Mantenimiento y crecimiento y la metodología de Curto por su parte tiene las siguientes fases: Análisis y Requerimientos, modelización, desarrollo, producción, formación y documentación la cual destaca por ser más

tecnológica y por su orientación al software libre y la metodología Big Data por su parte tiene las siguientes fases: adquisición de datos y grabación, extracción y pre procesamiento de la información, representación, agregación e integración de datos, procesamiento de peticiones, modelado de datos y análisis, interpretación de los datos la cual destaca por ser una tendencia del BI tradicional hacia el manejo de grandes cantidades de información no estructurada, se propuso un marco cuyas fases son: Planificación, Modelo del Negocio, Análisis, Diseño, Construcción e Implantación para el diseño del modelo de inteligencia de negocios propuesto, el cual demuestra que la analítica avanzada que proporciona Big Data y el BI tradicional son perfectamente complementarios.

- Se realizó el despliegue del modelo utilizando la herramienta de software libre Pentaho en su versión 4.8 Enterprise, el cual permitió implementar todo lo necesario para desarrollar la solución de Inteligencia de negocios que proporcione información interna y externa de la organización de forma integrada para su respectivo análisis en el proceso de toma de decisiones, el cual consta de procesos ETL (para la información interna y externa), reportes, tablas de análisis y tableros de mando, por otra parte para el procesamiento de la información externa se utilizó el framework apache Hadoop 2.2 y Hive 0.12 herramientas de Big Data. Se usó el software libre debido a su principal ventaja: el costo económico. El uso de la interfaz BI de Pentaho permitió un manejo intuitivo y sencillo a los usuarios finales para generar sus reportes y análisis acorde a sus necesidades en comparación al uso de hojas de cálculo.
- Los resultados de la evaluación a través de la opinión de expertos de los atributos del Modelo de Inteligencia de Negocios propuesta, nos muestran una mejora significativa en la mayoría de las fases del modelo propuesto, por otro lado también se clasifico los

atributos desfavorables (Adaptación del Modelo y Costo de Desarrollo) con una relevancia final de 273,66 puntos y favorables (Tiempo de Desarrollo y Nivel de Calidad del Modelo) con una relevancia final de 497,01 puntos para el Modelo Propuesto.

- Se analizaron los datos obtenidos a través de la implementación del modelo propuesta en el área de ventas de la empresa San Roque S.A, para luego con esta información contrastar la hipótesis, con lo que se demostró mejoras en los indicadores, en donde el promedio de tiempo de respuesta en formulación de propuestas por reunión fue de 3 horas y el promedio de tiempo con el modelo fue de 2 horas, el promedio de Índice de Tiempo de respuesta en el proceso de análisis por propuesta formulada fue de 0,25 y el promedio de índice de tiempo de respuesta con el modelo fue de 0,11, el promedio del número de propuestas formuladas por reunión fue de 3 y el promedio del número de propuestas con el modelo fue de 5, por otro lado el número de personas involucradas en el proceso de toma de decisiones disminuyo de 5 a 3, estos resultados hacen adecuada, y significativamente relevante, la aplicación del Modelo de Inteligencia de Negocio propuesta en el área de ventas de la empresa San Roque S.A.

RECOMENDACIONES

- Para mejorar el proceso se recomienda que las siguientes actividades: establecer las propuestas de las posibles promociones de ventas, analizar las propuestas de las promociones, establecer el precio de promoción y cálculo de la rentabilidad en base a la información de la rotación y venta de los productos, evaluar y aprobar las propuestas formuladas, se realicen con el análisis de información obtenida de un mayor número de fuentes externas que posea la organización para el proceso de toma de decisiones, como por ejemplo las cuentas de las redes sociales con las que cuenta la empresa, no solo de su WebSite como se ha realizado en el presente proyecto, esto permitirá conocer a mayor detalle el impacto de sus productos de las propuestas de las posibles promociones a lanzar.
- Se recomienda para trabajos futuros usar otras nuevas tendencias que están surgiendo e irrumpiendo con fuerza en el mundo del Business Intelligence que no se están tomando en cuenta o considerando en el modelo propuesto, entre ellas cabe destacar Real Time BI, Mobile BI, BI en la nube e integración Social Media, son algunos de los representantes de las grandes revoluciones que se están produciendo en el campo del Business Intelligence y que viene de la mano de los profundos cambios que ha establecido la llamada Web 2.0, una evolución tecnológica y de modelo de negocio del Web tradicional, con grandes difusores como las redes sociales, Google, YouTube, etc. En el presente modelo propuesto se ha considerado solo la evaluación de plataformas BI de software libre para el despliegue del modelo, sin embargo el modelo se puede ampliar para la evaluación de cualquier tipo de plataforma BI, por lo tanto se recomienda para futuros trabajos aplicarlo a otros tipos de plataforma de BI que no necesariamente sean de software libre.

- Para la versión 4.8 de la Suite de Pentaho Enterprise se recomienda que el componente para poder conectarse y analizar fuentes externas de gran cantidad de datos como los archivos log del servidor web o redes sociales, sea mejorado, ya que por el momento el proceso para trabajar con Big Data es un poco engorroso para esta versión, por lo tanto para posteriores versiones del Pentaho se recomienda incluir un wizard o asistente que permita manejar estos procesos de manera sencilla y en unos pocos pasos, en especial Pentaho debe incrementar el esfuerzo por potenciar todo lo que se refiere al Big Data ya que es una tendencia de BI que presenta un gran impacto en la actualidad, por otro lado se recomienda también incluir en una nueva versión una plataforma de informes para MongoDB, ya que es el sistema de bases de datos NoSQL más relevante de la industria, agregar la opción de diseño personalizado de Dashboards o tableros de mando, mejorar la interfaz de usuario, mejorar la experiencia del administrador, agregar nuevas capacidades de visualización, una amplia y profunda integración Big Data es decir Integraciones actualizadas y certificaciones para los data warehouse que garanticen mantener el ritmo de los cambios, entre ellas podemos mencionar las integraciones con Splunk, Amazon Redshift, Cloudera Impala, y certificaciones en MongoDB, Cassandra, DataStax, Cloudera , Intel, Hortonworks y MapR.
- Se recomienda para futuros trabajos que el tiempo para aplicar el Modelo de Inteligencia de Negocio propuesto a una organización sea de un estudio anual, en lugar de que sea solo de 12 semanas como se realizó en el presente proyecto, es decir de un mayor tiempo, para que permita evaluar el modelo de manera más rigurosa y así obtener una mayor confiabilidad y validez.

REFERENCIAS BIBLIOGRAFICAS

- *DatosIntensos*. (05 de Junio de 2012). Recuperado el 25 de Noviembre de 2013, de <http://datosintensos.blogspot.com/2012/06/herramientas-open-source-para-big-data.html>
- *ZooskDev*. (19 de Octubre de 2012). Recuperado el 12 de Noviembre de 2013, de <http://zooskdev.wordpress.com/2012/10/19/marriage-of-hadoop-and-olap-best-of-both-worlds-to-make-sense-of-200-terabytes-of-data/>
- *Acis*. (11 de Julio de 2013). Recuperado el 21 de Noviembre de 2013, de <http://www.acis.org.co/fileadmin/Conferencias/CBigData.pdf>
- *DataPrix*. (23 de Mayo de 2013). Recuperado el 19 de Noviembre de 2013, de <http://www.dataprix.com/empresa/recurso-it/business-intelligence/nuevas-tendencias-business-intelligence-big-data-social-intelligence>
- *DataPrix*. (03 de Abril de 2013). Recuperado el 12 de Noviembre de 2013, de <http://www.dataprix.com/blog-it/tendencias-tecnologicas/big-data/introduccion-hadoop-su-ecosistema>
- *DataPrix*. (15 de Agosto de 2013). Recuperado el 23 de Noviembre de 2013, de <http://www.dataprix.com/empresa/recursos/recursos-pentaho-su-integracion-apache-hadoop>
- *Iosutron*. (27 de Febrero de 2013). Recuperado el 26 de Noviembre de 2013, de <http://iosutron.wordpress.com/tag/hdfs/>
- *Lynko*. (21 de Agosto de 2013). Recuperado el 14 de Noviembre de 2013, de <http://www.lynkoo.com/analitica-web/que-es-el-marketing-de-big-data/>
- *Microsoft*. (21 de Marzo de 2013). Recuperado el 13 de Agosto de 2013, de <http://www.microsoft.com/sqlserver/2008/en/us/pricing.aspx>

- *Oracle*. (16 de Enero de 2013). Recuperado el 13 de Agosto de 2013, de <http://www.oracleparamiempresa.com/productos>
- *Palentino*. (29 de Julio de 2013). Recuperado el 12 de Noviembre de 2013, de <http://www.palentino.es/blog/big-data-herramientas-para-el-analisis-de-datos-y-aplicaciones-distribuidas/>
- *Pentaho*. (24 de Febrero de 2013). Recuperado el 13 de Agosto de 2013, de <http://www.pentaho.com>
- *Pentaho*. (06 de Noviembre de 2013). Recuperado el 27 de Noviembre de 2013, de <http://blog.pentaho.com/tag/data-integration/>
- *Pentaho Community*. (14 de Enero de 2013). Recuperado el 13 de Agosto de 2013, de <http://community.pentaho.com>
- *PowerData*. (05 de Septiembre de 2013). Recuperado el 23 de Noviembre de 2013, de <http://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/328879/Cómo-se-relacionan-Big-Data-y-Hadoop>
- *PowerData*. (01 de Octubre de 2013). Recuperado el 20 de Noviembre de 2013, de <http://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/337030/Big-world-Big-Data>
- Aaker, D. A. (2010). En *Estrategia de cartera de marcas*. (pág. 103). España: Ediciones Gestión 2000. Base de datos en línea: Google Book.
- Amaya Amaya, J. (2010). En *Toma de decisiones gerenciales: métodos cuantitativos para la administración* (2da ed., págs. 4-10). Ecuador: Ecoe.
- Anthony, S., Sinfield, J., Johnson, M., & Altman, E. (2010). En *Guía del innovador para crecer* (pág. 190). España: Editorial Deusto.
- Association, C. R. (19 de Febrero de 2012). *Challenges and Opportunities with Big Data*. Recuperado el 25 de Noviembre de 2013, de <http://www.cra.org/ccc/files/docs/init/bigdatawhitepaper.pdf>

- Bouman, R., & Dongen, J. (2010). En *Pentaho Solutions: Business Intelligence and Data Warehousing with Pentaho and mysql* (1ra ed., págs. 19-26). Indiana, United States Of America: Wiley Publishing. Base de datos en línea: Google Book.
- Clark, C. (2011). En *Brainstorming - The Dynamic New Way to Create Successful Ideas* (págs. 12 – 19). United States Of America: Editorial Classic Business Bookshelf. Base de datos E-Libro.
- Curto Díaz, J., & Conesa Caralt, J. (2010). En *Introducción al Business Intelligence* (págs. 13-16). España: UOC. Base de datos en línea: E-Libro.
- Harvard Business Essentials. (2010). En *Como Crear Una Empresa Exitosa* (pág. 51). España: Ediciones Deusto. Base de datos en línea: Google Book.
- Howson, C. (2009). En *Business intelligence: estrategias para una implementación exitosa* (págs. 14-18). México: McGraw-Hill. Base de datos en línea: E-Libro.
- Hurwitz, J., Nugent, A., Halper, F., & Kaufman, M. (2013). *Big Data For Dummies*. New Jersey: Editorial Wiley. Base de Datos en Línea: Google Books.
- Inmon, W., Strauss, D., & Neushloss, G. (2010). En *DW 2.0: The Architecture for the Next Generation of Data Warehousing*. (págs. 78-85). USA: Editorial Morgan Kaufmann. Base de datos en línea: Google Book.
- Jara, J. (2012). *Big Data & Web Intelligence*. Asunción - Paraguay: Universidad Católica.
- Kimball, R., & Ross, M. (2010). En *The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence*. (págs. 35-45). United States Of America, Indiana: Wiley Publishing. Base de datos en línea: Google Book.
- Kommineni Sivaganesh, S., & Suresh Chandra, S. (2012). En *Optimization of ETL Work Flow in Data Warehouse* (Vol. III, págs. 1579-1586). India: Engg Journals Publications. Revista:

International Journal on Computer Science and Engineering.
Catalogo de Revistas Especializadas en Línea: DOAJ.

- Laundon, J., & Laundon, K. (2012). En *Sistemas de Información Gerencial* (págs. 24-30). México: Editorial ADDISON-WESLEY. Base de datos en línea: E-Libro.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: La revolución de los datos masivos*. España: Ediciones Deusto.
- Sainz de Vicuña, J. M. (2010). En *El Plan del Marketing en la PYME*. (págs. 110-112). España: Editorial ESIC. Base de datos en línea: Google Book.
- Salcedo Parra, O., Milena Galeano, R., & Rodriguez B, L. (2010). En *Modelamiento Dimensional de Datos* (págs. 18-30). España: Editorial NETBIBLO. Revista: Red de Revistas Científicas de América Latina, el Caribe, España y Portugal. Base de datos en línea: E-Libro.
- Stanton, W., Etzel, M., & Walker, B. (2009). En *Fundamentos de Marketing*. (págs. 604-607). México: Editorial McGraw-Hill. Base de datos en línea: E-Libro.
- Vega Torres, L., Rojas Díaz, L., & Placeres Villar, C. (2009). En *La Inteligencia De Negocio. Su Implementación mediante la Plataforma PENTAHO*. (págs. 8-9). Cuba: Editorial CITMATEL. Base de datos en línea: E-Libro.
- White, T. (2012). *Hadoop: The Definitive Guide, 3rd Edition*. Cambridge: O'Reilly Media.
- Zikopoulos, P., Eaton, C., Deroos, D., Lapis, G., & Deutsch, T. (2012). *Understanding Big Data*. Silicon Valley: McGraw-Hill.

ANEXOS

Anexo N° 01: Guía de Entrevista N° 01

Entrevistado: Fabiola Díaz Orbegoso	Fecha: 04-Abril-2013
Entrevistador: Omar Sánchez Guevara	Área o Departamento: Gerencia Comercial
Objetivo: Conocer la situación actual de la empresa, con respecto a la información general de la Empresa.	
Dirigido a: Jefe de Comercialización y Ventas	
Preguntas: <ol style="list-style-type: none"> 1. ¿Cuál es su función principal dentro del área? 2. ¿Cuánto tiempo de creación tiene la empresa? 3. ¿Cuáles son los productos que mayormente sobresalen? 4. ¿Con cuántos puntos de venta cuenta actualmente San Roque S.A? 5. ¿Cada cuánto tiempo se realizan las promociones de ventas en la empresa? 6. En cuanto a la exportación de los productos, ¿A qué países actualmente exporta San Roque? 	
Resumen: San Roque S.A. es una empresa perteneciente al Grupo San Roque que se encarga de la elaboración de dulces tradicionales desde hace más de 90 años, se ubica en la carretera Panamericana Norte Km. 780 en la ciudad de Lambayeque. Cuentan con una gran variedad de productos entre los cuales destaca el King Kong, alfajor de gran tamaño y de diversos sabores. La comercialización de estos productos se realiza a través de sus puntos de ventas; de manera nacional cuentan con seis tiendas en Chiclayo y dos tiendas en las ciudades de Piura, Trujillo y Lambayeque respectivamente, además se distribuyen los productos en supermercados y autoservicios en las ciudades de Lima, Chimbote, Tacna y Arequipa. Y de manera internacional en los países de Estados Unidos, Suiza, Italia, Canadá y Costa Rica	

Anexo N° 02: Guía de Entrevista N° 02

Entrevistado: Víctor Seclén Esquivés Entrevistador: Omar Sánchez Guevara	Fecha: 05-Abril-2013 Área o Departamento: Sistemas
Objetivo: Conocer la problemática de la empresa San Roque S.A. con respecto al uso de los sistemas de información.	
Dirigido a: Responsable del Área de Sistemas	
Preguntas: <ol style="list-style-type: none"> 1. ¿Cuál es su función dentro de la empresa? 2. ¿El área de sistemas cuenta con sus propios desarrolladores? 3. ¿Cuál es la problemática actual con respecto a los sistemas de información que se utilizan? 4. ¿Cuáles son los sistemas de información con los que cuenta actualmente la empresa San Roque? 	
Resumen: Desde setiembre del 2004 se implanto un sistema de información, el cual incluía la gestión de procesos de las áreas de Contabilidad, Compras, Almacén, Tesorería, Recursos Humanos y Comercial. El sistema funcionaba de manera correcta sobre todo en la parte contable, pero el rápido crecimiento de la empresa, volvió limitado al sistema en cuanto a escalabilidad, presentando inconsistencia de datos, por lo que se vieron en la necesidad de realizar una licitación para la implantación de un Sistema ERP desde enero del año 2011.	

Anexo N° 03: Guía de Entrevista N° 03

Entrevistado: Jorge Piscoya Madueño	Fecha: 04-Abril-2013
Entrevistador: Omar Sánchez Guevara	Área o Departamento: Gerencia General
Objetivo: Conocer la problemática de la empresa San Roque S.A. con respecto a la toma de decisiones en el proceso de ventas.	
Dirigido a: Gerente General.	
Preguntas: <ol style="list-style-type: none"> 1. ¿Cuál es su función dentro de la empresa? 2. ¿De qué manera se realiza la toma de decisiones en la empresa? 3. ¿Cada cuánto tiempo se realizan las reuniones y cuál es su duración? 4. ¿Qué tipo de decisiones se toman en las reuniones realizadas? 5. ¿Utiliza alguna estrategia para la toma de decisiones? 6. ¿Cuánto demora en lanzarse al mercado una promoción propuesta? 	
Resumen: La toma de decisiones es de vital importancia en la empresa, por eso se realizan reuniones todos los días martes que tienen por duración de 3 horas aproximadamente. Ahí se definen uno de los puntos estratégicos como es el de establecer promociones de ventas a través de la técnica "Brainstorming", con opiniones de los involucrados pero sin basarse en datos históricos de ventas, el número promedio de propuestas establecidas por reunión es de tres. Desde que se inicia la idea hasta que se lanza el producto al mercado demora entre dos a tres meses, eso depende del tiempo que le dediquen durante las reuniones ya definidas.	

Anexo N° 04: Guía de Entrevista N° 04

Entrevistado:	Ana Mendoza Saavedra	Fecha:	08-Abril-2013
Entrevistador:	Omar Sánchez Guevara	Área o Departamento:	Costos
Objetivo:			
Saber la problemática de la empresa San Roque S.A. con respecto al análisis de los costos de las propuestas emitidas en las reuniones de Gerencia.			
Dirigido a:			
Contadora de costos.			
Preguntas:			
<ol style="list-style-type: none"> 1. ¿Cuál es su función principal dentro del área? 2. ¿Cada cuánto tiempo se realizan las promociones de ventas en la empresa? 3. ¿Cuántos tipos de promociones de ventas realizan? 4. ¿Qué productos han intervenido mayormente en promociones? 5. ¿Cuánto se demora en promedio en elaborar una promoción de venta? 6. ¿Podría explicar cómo realiza el proceso de elaboración de promociones de ventas? 7. ¿Qué recursos utiliza en este proceso? 			
Resumen:			
<p>Las propuestas son analizadas en primer lugar según la rotación y ventas del producto o los productos de la promoción por tiendas en los dos últimos años y los meses anteriores. Esta información la consigue utilizando los reportes del sistema y los exporta a una hoja de cálculo, analizando los datos obtenidos de la rotación y venta del producto para establecer de esta manera el posible descuento de la promoción calculando si la rentabilidad satisface al indicador ya determinado, el tiempo utilizado es de 4 días en promedio.</p> <p>Luego se realiza un informe general para ser presentada en la siguiente reunión de Gerencia para tomar la decisión si la promoción se lanza al mercado o si se cambia a otra propuesta de promoción, si es la primera opción se enviará el informe a la asesora externa de marketing para la publicidad respectiva, a logística para el pedido de envases e insumos y a producción para que se encargue de la elaboración de la promoción, caso contrario se repite el proceso de análisis de la contadora de costos, hasta que la promoción sea la aceptada por Gerencia.</p>			

Anexo N° 05: Números de Promociones periodo 2007-2012.

Periodos	2007	2008	2009	2010	2011	2012
Enero	-	-	-	-	0	0
Febrero	-	-	-	-	0	0
Marzo	-	-	-	-	0	0
Abril	-	-	-	-	1	1
Mayo	-	-	-	-	1	1
Junio	-	-	-	-	0	0
Julio	-	-	-	-	1	1
Agosto	-	-	-	-	0	1
Septiembre	-	-	-	-	0	1
Octubre	-	-	-	-	0	0
Noviembre	-	-	-	-	0	0
Diciembre	-	-	-	-	1	1
Total	-	-	-	-	4	6

Tabla 71: Promociones de Ventas Mensuales. Periodo 2007 – 2012.

Fuente: Jefe de Distribución y Ventas. Fabiola Díaz Orbegoso.

Anexo Nº 06: Ventas Mensuales periodo 2007-2012.

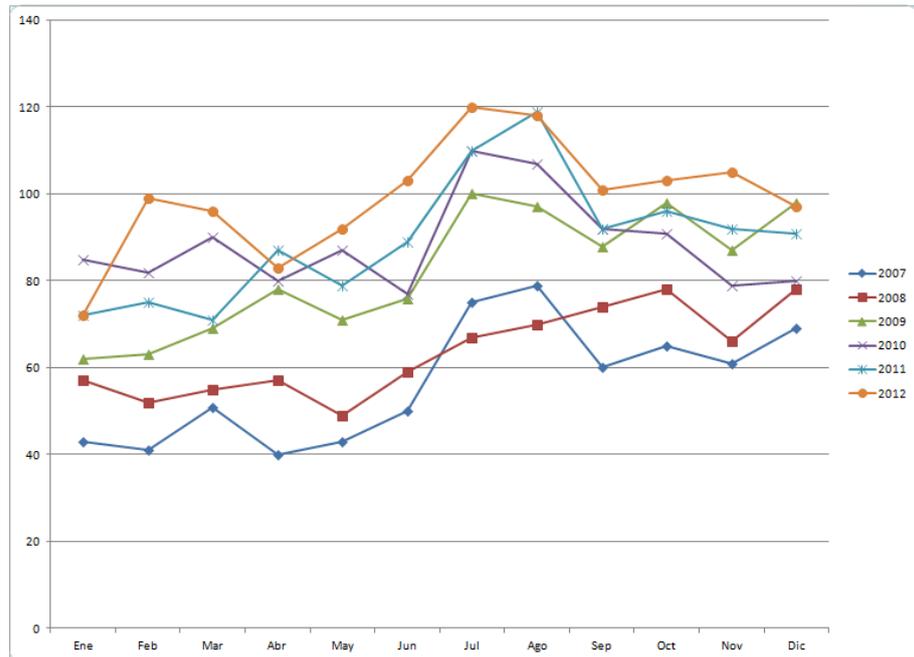


Figura 118: Ventas Mensuales en Miles de Unidades. 2007 – 2012

Fuente: Contadora de Costos. Ana Mendoza Saavedra.

Anexo N° 07: Ficha de Entrevista - Reunión N°...

Entrevistado:	Fecha:
Entrevistador:	Área o Departamento:
<p>Objetivo General: Conocer cuánto tiempo se invierte en la toma de decisiones en los proceso de ventas.</p> <p>Objetivos específicos:</p> <ul style="list-style-type: none"> ➤ Identificar al personal implicado en la toma de decisiones ➤ Conocer el tiempo invertido en la formulación de estrategias. ➤ Percibir el tiempo de demora del proceso de análisis de las propuestas formuladas. ➤ Lograr obtener el índice de propuestas aceptadas por Gerencia. 	
<p>Dirigido a:</p> <p>Personal involucrado en la toma de decisiones</p>	
<p>Preguntas:</p> <ol style="list-style-type: none"> 1. ¿Cuántas personas se involucran en la toma de decisiones? 2. ¿Cuánto tiempo se tardan en comunicar y debatir las ideas de cada persona involucrada en la toma de decisiones? 3. ¿Cuánto tiempo se invierte en decidir la propuesta idónea? 4. ¿Cuántas reuniones se realizan mensualmente? 5. ¿Cuántas propuestas se designan para el proceso de ventas? 6. ¿Cuántas horas se emplean en el proceso de análisis de la(s) propuesta(s) concertada(s) en reunión? 7. ¿Cuántas propuestas concertadas se analizan mensualmente? 8. ¿Cuántas de las propuestas formuladas al mes, son aceptadas para el proceso de ventas? 9. Después de obtener mensualmente propuestas aceptadas ¿Cuánto demora en decidir la que se lanzara para el proceso de las ventas? 10. Luego de haber decidido la propuesta favorable. ¿Cuánto tiempo tarda en lanzarla al mercado? 	
<p>¿Qué se logró?</p>	
<p>Conclusiones:</p>	

Anexo N° 8: Ficha de Observación - Reunión N°...

Personal: Evaluador:		Lugar y Fecha de aplicación: Campos de aplicación:			
N°	Acciones a Evaluar	Registro de cumplimiento			Observaciones
		SI	NO	NA	
1	Se realizan reuniones semanales				
2	Utiliza el Brainstorming para debatir propuestas				
3	Utiliza equipos multimedia en las reuniones				
4	Invierte mucho tiempo en el debate de las ideas				
5	Invierte tiempo considerable para determinar las propuestas				
6	Se basan en datos históricos para determinar las propuestas				
7	Invierte el % de horas laborables para el proceso de análisis				
8	Respetar el tiempo determinado para la realización del análisis de la propuesta				
9	Intervienen muchas personas en el proceso de análisis de la propuesta formulada				
10	Las propuestas analizadas siempre son aceptadas por gerencia				

Anexo Nº 9: Ficha de Análisis Documental

Datos de la Aplicación

a) Nombre de la Institución:

b) Fechas:

Fecha de aplicación 1: __/__/__

Fecha de aplicación 2: __/__/__

Fecha de aplicación 3: __/__/__

Fecha de aplicación 4: __/__/__

c) Miembros que participaron:

d) Documentos de la organización revisados:

Documentos	Tiene		Se revisó	
	Si	No	Si	No
Plan Estratégico				
Reportes de Gestión				
...				

Anexo N° 10: Documento Visión del Proyecto

Título del Proyecto

(Incluir que se va a hacer y Para qué se va a hacer)

Introducción del Proyecto

(Descripción breve de la intención del documento)

Objetivos del Proyecto

(Basado en los requerimientos obtenidos del negocio). . .

Alcance del Proyecto

(Definir los beneficiarios –áreas, departamentos- directos)

Descripción de los Stakeholders

(Son las personas, áreas o instituciones –internas o externas- que intervienen directa o indirectamente en la preparación del producto)

Stakeholder	Cargo	Función
		Función que desempeña en el proyecto

Requisitos del Proyecto

(Son los requisitos humanos, materiales y de equipamiento necesario para desarrollar el proyecto y en el caso del Equipamiento para su operatividad.)

a) Equipo del Proyecto

(Son los recursos humanos que directamente para la consecución del producto – tanto del lado de Gestión como de Tecnologías Información).

✓ Gestión

Persona	Cargo	Rol	Función
			Función que desempeña en el Proyecto

✓ Tecnologías Información

Persona	Rol	Función
		Función que desempeña en el proyecto

b) Oportunidad de Negocio

(Factores que permitan obtener una ventaja competitiva acerca de la situación actual del negocio)

c) Restricciones

(Factores que limitan el desarrollo del proyecto propuesto)

d) Riesgos del Proyecto

(Factores que podrían incidir en la ejecución normal del proyecto que impidan que se realice en el tiempo previsto o determinen que un sobre costo al presupuesto asignado)

e) Objetivos no considerados

(Factores que no son tomados en cuenta o no abarca el presente proyecto)

Anexo N° 11: Cuestionario para el Personal de Gestión

Nombre y Cargo que desempeña

(Personal de Gestión)

Introducción

(Discutir los objetivos del proyecto y las metas de las entrevistas)

Conocimiento del Negocio

(Entender el negocio desde el punto de vista de gestión, identificando: objetivos, estrategias, indicadores y medidas)

- ✓ ¿Cuáles son los objetivos de la organización o del proceso de negocios que representa?
- ✓ ¿Cómo está tratando de cumplirlos?
- ✓ ¿Cuáles son las metas prioritarias?
- ✓ ¿Cuáles son las medidas de éxito de su gestión?
- ✓ ¿Cómo sabe si está haciendo bien su gestión?

Analizar Requerimientos

(Definir las necesidades directas basadas en los requerimientos analíticos y la técnica de modelamiento dimensional. Es altamente recomendable haber definido: Que? , Quien? , Como?, Donde?, De qué forma?, etc)

- ✓ ¿Qué capacidad de análisis quisiera tener con respecto al ..**Qué?**
- ✓ ¿Qué capacidad de análisis quisiera tener con respecto al ..**Quién?**
- ✓ ¿Qué reportes utiliza generalmente?,
- ✓ ¿Qué datos se presentan en el reporte?
- ✓ ¿Podría obtener una muestra de ese reporte?
- ✓ ¿Qué oportunidades existen en el negocio basado en el análisis de la información. Influyen y en qué medida en las finanzas?

Anexo N° 12: Cuestionario para el Personal de TI

Nombre y Cargo que desempeña

(Personal de IT)

Introducción

(Discutir los objetivos del proyecto y las metas de las entrevistas)

Analisis y Requerimientos de Datos

(Definir la disponibilidad de la data para obtener los reportes de gestión)

- ✓ Cuál es el proceso actualmente usado para analizar información
- ✓ Que herramientas se usan para analizar información y quienes las usan?
- ✓ Dependen en la elaboración de información de análisis de su Area?. Ha creado reportes estandarizados

Disponibilidad y Calidad de la Data

(Definir el origen de los datos, tipos de SGBD donde actualmente reside la data y la integridad de la misma)

- ✓ ¿Qué manejador de BD utiliza?
- ✓ ¿Con qué frecuencia se actualiza la data y en qué momento de producen los cierres respectivos de operaciones?
- ✓ ¿Qué tanto la data histórica se encuentra disponible y desde cuando se tiene datos en la BD?
- ✓ ¿Cuáles son los datos obligados y cuales los opcionales?
- ✓ ¿Existe documentación del modelo de datos?

Anexo N° 13: Tablas involucradas en el Modelo de Datos Actual

Nombre de Tabla	Descripción Tabla

Anexo N° 14: Hoja de Gestión

Proceso

(Definir el nombre del Proceso de Negocios a Implementar)

Objetivos

(Objetivos encontrados del proceso de negocios)

Estrategias

(Estrategias Implementar asociadas a la consecución de los objetivos planteados)

Indicadores

(Permiten realizar un seguimiento de la consecución de los objetivos planteados)

Indicador	Medidas	Estado	Valores		
Nombre	Fórmula de Cálculo	Bueno			
		Regular			
		Malo			

Anexo N° 15: Hoja de Análisis

Proceso

(Definir el nombre del Proceso de Negocios a Implementar)

Medidas

(Definir las medidas encontradas en los indicadores y otras de ser requeridas)

Dimensiones y Niveles

(Incluir las dimensiones encontradas y los puntos de análisis para dimensiones: niveles)

Dimensiones	Niveles
Nombre	Definir Niveles

Anexo N° 16: Cuadro de Dimensiones y Jerarquías

Proceso

(Definir el nombre del Proceso de Negocios a Implementar)

Dimensiones	JERARQUIAS (Organizar Niveles para permitir análisis de lo genérico a lo detallado o viceversa)				
	Nivel 1	Nivel 2	Nivel 3	Nivel 4	Nivel 5
Nombre Dimensión	Nombre Nivel (empezar desde el nivel más detallado)				

Anexo Nº 17: Cuadro de Dimensiones y Medidas

Proceso

(Definir el nombre del Proceso de Negocios a Implementar)

Medidas	DIMENSIONES <i>(Especificar las dimensiones encontradas)</i>				
	Dimensión 1	Dimensión 2	Dimensión 3	...	Dimensión n
<i>Nombre Medida 1</i>	<i>Marcar con X si la Medida puede ser analizada esta dimensión</i>				
<i>Nombre Medida 2</i>					
...					
<i>Nombre Medida n</i>					

Anexo N° 18: Diseño de Extracción

DIMENSION TIEMPO

1. Descripción de las tablas fuentes

Tipo de Fuente	Nombre de Fuente	Descripción
Base de datos transaccional	mst01fac	Ventas: Esta tabla contiene el listado de todas las ventas realizadas.

Tabla 72: Tabla Fuente de la dimensión Tiempo.

2. Fuente de Datos

Tabla	Nombre	Descripción	Llave	Tipo	Formato	Consideración Importante
mst01fac	fecha	Fecha de venta		DateTime	Fecha	No Nulo

Tabla 73: Fuente de Datos de la Dimensión Tiempo.

3. Estandarización y Limpieza de datos

Nombre	Llave	Tipo	Formato	Limpieza	Valor Por Defecto
KeyTiempo	PK	Int(11)	Numero entero	No Nulo	0
Fecha		DateTime	Texto	No Nulo	
DiaSemana		Varchar(30)	Texto	Formato: "1 -Lunes"	

Mes		Varchar(30)	Texto	Formato: "01 -Enero"	
Anio		Int(11)	Numero entero	Formato: "2007"	
Semestre		Varchar(30)	Texto	Formato: "1 - Semestre"	
Trimestre		Varchar(30)	Texto	Formato: "1 - Trimestre"	
Bimestre		Varchar(30)	Texto	Formato: "1 - Bimestre"	
FinSemana		Char(2)	Texto	Solo puede ser "SI" o "No"	

Tabla 74: Limpieza de Datos de la Dimensión Tiempo.

4. Mapeo de datos

Tabla:	dbo_dimtiempo	
Campo	Tipo	Mapeo
KeyTiempo	Int(11)	Identificador Numérico Autogenerado
Fecha	DateTime	mst01fac.fecha
DiaSemana	Varchar(30)	cast(datepart(dw, mst01fac.fecha) as varchar(1)) + '-' + upper(datename(dw, mst01fac.fecha))
Mes	Varchar(30)	case len(datepart(mm, mst01fac.fecha)) when (1) then '0' + cast(datepart(mm, mst01fac.fecha) as char(2)) + '-' + upper(datename(month, mst01fac.fecha)) else cast(datepart(mm, mst01fac.fecha) as char(2))+ '-' + upper(datename(month, mst01fac.fecha)) end

Anio	Int(11)	datepart(yy, mst01fac.fecha)
Semestre	Varchar(30)	convert(varchar, floor((datepart(mm, mst01fac.fecha)- 1) / 6) + 1)+ '- semestre'
Trimestre	Varchar(30)	convert(varchar, datepart(qq, mst01fac.fecha))+ '-trimestre'
Bimestre	Varchar(30)	convert(varchar, floor((datepart(mm, mst01fac.fecha)- 1) / 2) + 1)+ '- bimestre'
FinSemana	Char(2)	case datepart(dw, mst01fac.fecha) when (1) then 'si' when (7) then 'si' else 'no' end

Tabla 75: Mapeo de Datos de la dimensión Tiempo.

5. Proceso

- **Dimensión Tiempo**

El proceso que se realiza para la extracción, transformación y carga de la dimensión tiempo es la siguiente:

- ✓ **Extracción de los datos de las tablas mst01fac:** Se toma los datos de las tablas OLTP mencionadas y se realiza la transformación de acuerdo al mapeo indicado.
- ✓ **Carga de la Dimensión Tiempo:** Se transforma e ingresan los datos en la dimensión tiempo. En el caso que sean nuevas fechas se insertan, en el caso que sean antiguas fechas sólo se actualizan los datos.

TABLA HECHO

1. Descripción de las tablas fuentes

Tipo de Fuente	Nombre de Fuente	Descripción
Base de datos transaccional	mst01fac	Ventas: Esta tabla contiene el listado

		de todas las ventas realizadas.
Base de datos transaccional	dtl01fac	Detalle de Venta: Esta tabla contiene el detalle de todas las ventas realizadas.

Tabla 76: Tablas Fuentes para la tabla hechos.

2. Fuente de Datos

Tabla	Nombre	Descripción	Llave	Tipo	Formato	Consideración Importante
mst01fac	ndocu	Numero de documento de venta, puede estar en moneda local o en dólares.	PK	Char(12)	Texto	No Nulo
dtl01fac	preu	Precio unitario del producto vendido, puede estar en moneda local o en dólares.		Decimal(19,4)	Número Real	Requiere conversión a moneda local en caso este en dólares.
mst01fac	tcam	Valor del Tipo de cambio del dólar.		Decimal(19,4)	Número Real	No Nulo
mst01fac	mone	Tipo de		Char(1)	Carácter	Sólo puede

		Moneda (Soles ó Dólares)				contener "S" o "D"
dtl01fac	cost	Costo de producción del producto vendido, puede estar en moneda local o en dólares.		Decimal(19,4)	Número Real	Requiere conversión a moneda local o a dólares, dependiendo de la elección del tipo de moneda.
dtl01fac	tota	Monto total bruto de la venta, puede estar en moneda local o en dólares.		Decimal(19,4)	Número Real	CAMPO PRE CALCULADO Este dato resulta de la operación interna de la BD transaccional, multiplicando el precio unitario por la cantidad: $preu * cant$
dtl01fac	totn	Monto total neto de la venta aplicando el IGV (19%), puede estar en moneda local o en dólares.		Decimal(19,4)	Número Real	CAMPO PRE CALCULADO Este dato resulta de la operación interna de la BD transaccional, multiplicando monto total bruto por el IGV (19%): $tota * 19\%$
dtl01fac	cant	Cantidad de productos vendidos		Int(11)	Numero entero	No Nulo

dtl01fac	codi	Código de producto en la tabla detalle venta		Char(12)	Texto	No Nulo
mst01fac	Codcli	Código de cliente en la tabla venta		Char(6)	Texto	No Nulo
mst01fac	Codpto	Código de punto de venta en la tabla venta		Char(8)	Texto	No Nulo
mst01fac	Fecha	Fecha en la tabla venta		DateTime	Texto	No Nulo

Tabla 77: Fuente de Datos de la Tabla Hecho.

3. Estandarización y Limpieza de datos

Nombre	Llave	Tipo	Formato	Limpieza	Valor Por Defecto
KeyProducto	FK	Int(11)	Numero entero	No Nulo	
KeyTiempo	FK	Int(11)	Numero entero	No Nulo	
KeyCliente	FK	Int(11)	Numero entero	No Nulo	
KeyPunto	FK	Int(11)	Numero entero	No Nulo	
NroDocumento		Char(12)	Texto	No Nulo	
PrecioUnitario		Decimal(19,4)	Número Real	>0	
Cantidad		Int(11)	Numero	>0	

			entero		
CostoNS		Decimal(19,4)	Número Real	>0	
CostoUS		Decimal(19,4)	Número Real	>0	
VVentaNS		Decimal(19,4)	Número Real	>0	
VVentaUS		Decimal(19,4)	Número Real	>0	
TotalNS		Decimal(19,4)	Número Real	>0	
TotalUS		Decimal(19,4)	Número Real	>0	

Tabla 78: Limpieza de Datos de la Tabla Hecho.

4. Mapeo de datos

Tabla:	dbo_hechosventa	
Campo	Tipo	Mapeo
KeyProducto	Int(11)	Se obtiene de realizar un inner join entre los campos código de producto de las tablas dbo_dimproducto.CodProducto y dtl01fac.codi
KeyTiempo	Int(11)	Se obtiene de realizar un inner join entre los campos fecha de las tablas dbo_dimtiempo.Fecha y mst01fac.Fecha
KeyCliente	Int(11)	Se obtiene de realizar un inner join entre los campos código de cliente de las tablas dbo_dimcliente.CodCliente y mst01fac.codcli

KeyPunto	Int(11)	Se obtiene de realizar un inner join entre los campos código de punto de las tablas dbo_dimpunto.CodPunto y mst01fac.codpto
NroDocumento	Char(12)	mst01fac.ndocu
PrecioUnitario	Decimal(19,4)	case when mst01fac .mone='S' then cast(dtl01fac.preu as money) else cast(dtl01fac.preu* mst01fac.tcam as money) end
Cantidad	Int(11)	dtl01fac.cant
CostoNS	Decimal(19,4)	case when mst01fac.mone='S' then cast(dtl01fac.cost* dtl01fac.cant as money) else cast((dtl01fac.cost* mst01fac.tcam)* dtl01fac.cant as money) end
CostoUS	Decimal(19,4)	case when mst01fac.mone='D' then cast(dtl01fac.cost* dtl01fac.cant as money) else cast((dtl01fac.cost/ mst01fac.tcam)* dtl01fac.cant as money) end
VVentaNS	Decimal(19,4)	case when mst01fac.mone='S' then cast(dtl01fac.tota as money) else cast(dtl01fac.tota* mst01fac.tcam as money) end
VVentaUS	Decimal(19,4)	case when mst01fac.mone='D' then cast(dtl01fac.tota as money) else cast(dtl01fac.tota/ mst01fac.tcam as money) end
TotalNS	Decimal(19,4)	case when mst01fac.mone='S' then cast(dtl01fac.totn as money) else cast(dtl01fac.totn* mst01fac.tcam as money) end
TotalUS	Decimal(19,4)	case when mst01fac.mone='d' then cast(dtl01fac.totn as money) else cast(dtl01fac.totn/ mst01fac.tcam as money)

		end
--	--	-----

Tabla 79: Mapeo de Datos de la Tabla Hecho.

5. Proceso

- **Tabla Hecho**

El proceso que se realiza para la extracción, transformación y carga de la tabla hecho es la siguiente:

- ✓ **Extracción de los datos de las tablas mst01fac, dtl01fac y de las dimensiones dimtiempo, dimcliente, dimpunto, dimproducto:** Se toma los datos de las tablas OLTP y de las dimensiones mencionadas y se realiza la transformación de acuerdo al mapeo indicado.
- ✓ **Carga de la Tabla Hecho:** Se transforma e ingresan los datos a la tabla hecho de ventas. En el caso que sean nuevos registros se insertan, en el caso que sean antiguos registros sólo se actualizan los datos.

Anexo N° 19

ENCUESTA PARA EVALUAR EL MODELO DE INTELIGENCIA DE NEGOCIO PROPUESTO

Antes de la aplicación de la presente encuesta sobre los atributos del Modelo de Inteligencia de Negocio propuesta, se implementó una solución de inteligencia de negocio y se expuso las fases (actividades, características y componentes) a expertos en el desarrollo de soluciones de Inteligencia de Negocios de la empresa AC SOFT LIVE E.I.R.L.

PREGUNTAS

Su puntuación puede ir de 0 a 100.

- a) Adaptación del Modelo:** evalúe las fases planteadas del modelo propuesto con respecto al nivel de adaptación, se debe considerar el mayor puntaje a las fases del modelo que mejor se adaptan.

	Item	Valor
1	Evalué la adaptación del modelo en la Fase de Planificación.	
2	Evalué la adaptación del modelo en la Fase de Modelo del Negocio.	
3	Evalué la adaptación del modelo en la Fase de Análisis.	
4	Evalué la adaptación del modelo en la Fase de Diseño.	
5	Evalué la adaptación del modelo en la Fase de Construcción.	
6	Evalué la adaptación del modelo en la Fase de Implantación.	

Tabla 80: Formato de encuesta al atributo Adaptación del modelo.

- b) Costo de Desarrollo:** evalúe las fases planteadas del modelo propuesto con respecto al costo de desarrollo, se debe considerar el mayor puntaje a las fases del modelo que generan un menor costo de desarrollo.

	Item	Valor
1	Evalué el costo de desarrollo en la Fase de Planificación.	
2	Evalué el costo de desarrollo en la Fase de Modelo del Negocio.	
3	Evalué el costo de desarrollo en la Fase de Análisis.	
4	Evalué el costo de desarrollo en la Fase de Diseño.	
5	Evalué el costo de desarrollo en la Fase de Construcción.	
6	Evalué el costo de desarrollo en la Fase de Implantación.	

Tabla 81: Formato de encuesta al atributo costo de desarrollo.

c) Tiempo de Desarrollo: evalué las fases planteadas del modelo propuesto con respecto al tiempo de desarrollo, se debe considerar el mayor puntaje a las fases del modelo que generan un menor tiempo de desarrollo.

	Item	Valor
1	Evalué el tiempo de desarrollo en la Fase de Planificación.	
2	Evalué el tiempo de desarrollo en la Fase de Modelo del Negocio.	
3	Evalué el tiempo de desarrollo en la Fase de Análisis.	
4	Evalué el tiempo de desarrollo en la Fase de Diseño.	
5	Evalué el tiempo de desarrollo en la Fase de Construcción.	
6	Evalué el tiempo de desarrollo en la Fase de Implantación.	

Tabla 82: Formato de encuesta al atributo tiempo de desarrollo.

d) Nivel de Calidad del Modelo: evalué las fases planteadas del modelo propuesto con respecto al nivel de calidad, se debe considerar el mayor puntaje a las fases del modelo que consideran tienen un mayor nivel de calidad.

	Ítem	Valor
1	Evalué el nivel de calidad del modelo en la Fase de Planificación.	
2	Evalué el nivel de calidad del modelo en la Fase de Modelo del Negocio.	
3	Evalué el nivel de calidad del modelo en la Fase de Análisis.	
4	Evalué el nivel de calidad del modelo en la Fase de Diseño.	
5	Evalué el nivel de calidad del modelo en la Fase de Construcción.	
6	Evalué el nivel de calidad del modelo en la Fase de Implantación.	

Tabla 83: Formato de encuesta al atributo nivel de calidad del modelo.

Anexo N° 20

RESULTADOS DE LA ENCUESTA PARA EVALUAR EL MODELO PROPUESTO

RESUMEN DE RESULTADOS DE LAS ENCUESTAS EFECTUADAS							
	ATRIBUTOS DEL MODELO	EXPERTOS					
I	Adaptación del Modelo	E1	E2	E3	E4	E5	PROMEDIO
	1. Fase de Planificación	50	55	45	45	65	52
	2. Fase de Modelo del Negocio	60	65	63	60	55	61
	3. Fase de Análisis	75	70	72	80	75	74
	4. Fase de Diseño	55	60	50	65	65	59
	5. Fase de Construcción	75	75	73	72	75	74
	6. Fase de Implementación	70	75	72	75	80	74
	PROMEDIO POR EXPERTO:	64	67	63	66	69	
	PROMEDIO TOTAL POR EXPERTO:	65,73					
II	Costo de Desarrollo	E1	E2	E3	E4	E5	PROMEDIO
	1. Fase de Planificación	82	74	80	85	65	77
	2. Fase de Modelo del Negocio	65	50	60	75	70	64
	3. Fase de Análisis	60	65	70	55	50	60
	4. Fase de Diseño	65	70	75	70	66	69
	5. Fase de Construcción	70	80	85	85	75	79
	6. Fase de Implementación	75	85	70	80	76	77
	PROMEDIO POR EXPERTO:	70	71	73	75	67	
	PROMEDIO TOTAL POR EXPERTO:	71,1					
III	Tiempo de Desarrollo	E1	E2	E3	E4	E5	PROMEDIO
	1. Fase de Planificación	75	70	85	74	90	79
	2. Fase de Modelo del Negocio	85	90	90	85	80	86
	3. Fase de Análisis	80	85	80	70	84	80
	4. Fase de Diseño	75	70	80	85	70	76
	5. Fase de Construcción	90	95	93	90	88	91
	6. Fase de Implementación	85	90	88	84	92	88
	PROMEDIO POR EXPERTO:	82	83	86	81	84	
	PROMEDIO TOTAL POR EXPERTO:	83,27					
IV	Nivel de Calidad del Modelo	E1	E2	E3	E4	E5	PROMEDIO
	1. Fase de Planificación	75	80	90	78	82	81
	2. Fase de Modelo del Negocio	80	75	88	70	74	77
	3. Fase de Análisis	74	82	85	90	76	81
	4. Fase de Diseño	90	84	80	88	82	85
	5. Fase de Construcción	90	95	88	80	83	87
	6. Fase de Implementación	85	80	80	90	78	83
	PROMEDIO POR EXPERTO:	82	83	85	83	79	
	PROMEDIO TOTAL POR EXPERTO:	82,4					

Tabla 84: Resultado de encuestas de evaluación del modelo propuesto.

Los expertos que participaron en la evaluación del modelo reúnen los requisitos suficientes y necesarios para ser considerados válidos y confiables, y por tanto, aptos para ser aplicados en el logro de los objetivos que se plantean en la investigación. Se les considera expertos por las siguientes características:

a) Experto 1

- ✓ Ha liderado proyectos de Business Intelligence y Gestión del Conocimiento.
- ✓ Desarrollo de soluciones de Inteligencia de Negocios utilizando tanto software propietario como de software libre para diversas entidades entre ellas: HYUNDAI filial Chiclayo, CHICLAYO MOTORS S.A.C, y HORIZONTES S.R.L entre otras.

b) Experto 2

- ✓ Experto tecnológico en plataformas de Business Intelligence Pentaho, brindando soporte.
- ✓ Líder de Proyecto BI en la empresa DISTRIBUCIONES AMERICA S.A.C.
- ✓ Implementación de modelos de minería de datos para diversas organizaciones.

c) Experto 3

- ✓ Jefe de Proyectos BI.
- ✓ Creación de cubos OLAP para clientes y soporte tanto interno como externo a la empresa GUSTOSITO E.I.R.L sobre la suite Pentaho y diseño e implementación de indicadores de gestión (KPI).
- ✓ Desarrollo de una solución de DATA MINING para el área de marketing del instituto IDAT filial Chiclayo.

d) Experto 4

- ✓ Jefe de Proyectos de BI en la empresa AC SOFT LIVE E.I.R.L.
- ✓ Implementación de un Data Mart utilizando la suite BI Pentaho en la empresa VIAJERO NIYU S.A.C, empleando metodología de Ralph Kimball, modelamiento estrella.

e) Experto 5

- ✓ Consultor de Inteligencia de Negocios en el supermercado el SUPER S.A.C.
- ✓ Creador de un sistema de Business Intelligence para ayuda a la toma de decisiones en el instituto Instituto Superior de Administración Gerencial - ISAG, a través de la programación de OLAP por medio del software Analysis Services y Database Engine de SQL Server 2008, para el uso de toda la institución y para la obtención de estadísticas unificadas y confiables, en todos los ámbitos, académicos, administrativos y financieros (presupuestos, ejecuciones, etc)

Anexo N° 21

RESULTADOS DE LA EVALUACION DE LOS INDICADORES

Para obtener los resultados de estos dos indicadores se tuvo que aplicar la fórmula establecida en el capítulo Material y Métodos, para los indicadores de la variable dependiente:

a) Tiempo de respuesta en formulación de propuestas por reunión.

La fórmula a aplicar es la siguiente:

$$RFP = (HIL + HID) / NR$$

RFP: Tiempo de respuesta en formulación de propuestas por reunión.

HIL: Horas invertidas en lanzamiento de ideas.

HID: Horas invertidas en discusiones de las ideas.

NR: Número de reuniones.

La unidad de medida es la Hora.

Nro de Semanas	Sin el Modelo		Con el Modelo	
	Aplicando Formula	Total	Aplicando Formula	Total
1	$(2 + 1) / 1$	3	$(1,5 + 0,5) / 1$	2
2	$(1,5 + 1) / 1$	2,5	$(1 + 0,5) / 1$	1,5
3	$(2 + 2) / 1$	4	$(1,5 + 0,5) / 1$	2
4	$(2 + 1) / 1$	3	$(1 + 0,5) / 1$	1,5
5	$(1,5 + 1) / 1$	2,5	$(1 + 0,5) / 1$	1,5
6	$(2 + 2) / 1$	4	$(1,5 + 0,5) / 1$	2
7	$(2 + 1) / 1$	3	$(1 + 0,5) / 1$	1,5
8	$(2 + 1) / 1$	3	$(1,5 + 0,5) / 1$	2
9	$(1,5 + 1) / 1$	2,5	$(1 + 0,5) / 1$	1,5
10	$(2 + 1) / 1$	3	$(1,5 + 0,5) / 1$	2
11	$(2 + 1) / 1$	3	$(0,5 + 1) / 1$	1,5

12	$(1,5 + 1) / 1$	2,5	$(1 + 0,5) / 1$	1,5
----	-----------------	------------	-----------------	------------

Tabla 85: Resultado de aplicar la fórmula al indicador tiempo de respuesta en formulación de propuestas por reunión.

b) Índice de Tiempo de respuesta en el proceso de análisis por propuesta formulada.

La fórmula a aplicar es la siguiente:

$$PAP = (HIP * DT / JL)$$

PAP: Índice de tiempo de respuesta en el proceso de análisis por propuesta formulada.

HIP: Horas invertidas en el proceso de análisis.

JL: Jornada laboral.

DT: Días trabajados

La unidad de medida es el porcentaje.

Nro de Semanas	Sin el Modelo		Con el Modelo	
	Aplicando Formula	Total	Aplicando Formula	Total
1	$(3 * 4) / 48$	0,25	$(3 * 2) / 48$	0,13
2	$(2,5 * 4) / 48$	0,21	$(3 * 2) / 48$	0,13
3	$(3,5 * 4) / 48$	0,29	$(2 * 2) / 48$	0,08
4	$(3 * 4) / 48$	0,25	$(3 * 2) / 48$	0,13
5	$(2,5 * 4) / 48$	0,21	$(2 * 2) / 48$	0,08
6	$(3 * 4) / 48$	0,25	$(3 * 2) / 48$	0,13
7	$(3 * 4) / 48$	0,25	$(3 * 2) / 48$	0,13
8	$(3,5 * 4) / 48$	0,29	$(2 * 2) / 48$	0,08
9	$(2,5 * 4) / 48$	0,21	$(3 * 2) / 48$	0,13
10	$(3,5 * 4) / 48$	0,29	$(3 * 2) / 48$	0,13
11	$(2,5 * 4) / 48$	0,21	$(2 * 2) / 48$	0,08
12	$(3 * 4) / 48$	0,25	$(3 * 2) / 48$	0,13

Tabla 86: Resultado de aplicar la formula al indicador tiempo de respuesta en el proceso de análisis por propuesta formulada.

c) Numero de propuestas formuladas por reunión

Para este indicador no se realizó ninguna fórmula, por lo tanto estos datos se muestran directamente en el capítulo de resultados.