

UNIVERSIDAD PRIVADA ANTENOR ORREGO
FACULTAD DE INGENIERÍA
ESCUELA PROFESIONAL DE INGENIERÍA DE COMPUTACIÓN
Y DE SISTEMAS



TRABAJO DE TESIS PARA OBTENER EL TITULO PROFESIONAL DE
INGENIERO DE COMPUTACION Y SISTEMAS

“SOLUCION DE INTELIGENCIA DE NEGOCIOS PARA LA
GENERACIÓN DE INDICADORES QUE AYUDE EN EL ANALISIS
DE INFORMACION EN EL AREA DE VENTAS DE LA
FERRETERIA SOLANO BASADA EN LA METODOLOGÍA
HEFESTO UTILIZANDO PENTAHO BI.”

Línea de Investigación:

Gestión de Datos y de Información.

Autor:

BR. DIEGO ALONSO CORILLA AMAYA.

Asesor:

ING. AGUSTIN EDUARDO ULLON RAMIREZ

2017

**SOLUCION DE INTELIGENCIA DE NEGOCIOS PARA LA
GENERACIÓN DE INDICADORES QUE AYUDE EN EL ANALISIS
DE INFORMACION EN EL AREA DE VENTAS DE LA
FERRETERIA SOLANO BASADA EN LA METODOLOGÍA
HEFESTO UTILIZANDO PENTAHO BI**

Elaborado por:

Br. Diego Alonso Corilla Amaya

Aprobada por:

Ing. Carlos Alberto Gaytán Toledo
Presidente
CIP: 84519

Ing. Freddy Henry Infantes Quiroz
Secretario
CIP: 139578

Mg. Percy Lucio Carranza Medina
Vocal
CIP: 149877

Ing. Agustín Eduardo Ullón Ramírez
Asesor
CIP: 137602

PRESENTACIÓN

Señores Miembros del Jurado:

De conformidad y en cumplimiento de los requisitos estipulados en el reglamento de grados y Títulos de la Universidad Privada Antenor Orrego y el Reglamento Interno de la Escuela Profesional de Ingeniería de Computación y Sistemas, ponemos a vuestra disposición el presente Trabajo de Tesis: **“SOLUCION DE INTELIGENCIA DE NEGOCIOS PARA LA GENERACIÓN DE INDICADORES QUE AYUDE EN EL ANALISIS DE INFORMACION EN EL AREA DE VENTAS DE LA FERRETERIA SOLANO BASADA EN LA METODOLOGÍA HEFESTO UTILIZANDO PENTAHO BI”** para obtener el Título Profesional de Ingeniero de Computación y Sistemas.

El contenido de la presente tesis ha sido desarrollado tomando como marco de referencia los lineamientos establecidos por la Facultad de Ingeniería, la Escuela Profesional de Ingeniería de Computación y Sistemas y los conocimientos adquiridos durante nuestra formación profesional, además de consulta de fuentes bibliográficas.

El Autor.

DEDICATORIA

A Dios

Por haberme permitido llegar hasta este punto de mi vida y haberme dado salud y conocimiento para poder lograr mis objetivos y seguir continuando cumpliendo mis metas.

A mi mama Gladys

Por haberme dado la vida por haberme apoyado en todo el lapso de mi carrera estudiantil y profesional sin ella no hubiera podido estar realizando este trabajo agradecerle por los valores y la educación que me inculco desde muy pequeño.

A mi papa Saúl

Por haber estado en todo momento pendiente de mí en el transcurso de mi carrera profesional y por haberme dado mucha motivación para seguir adelante.

AGRADECIMIENTO

Quiero expresar mi agradecimiento al personal del área ventas de la corporación Solano E.I.R.L y al señor Edgar Richard Solano Ramírez Gerente y dueño de la empresa quienes me brindaron todas las facilidades para conocer más sobre la problemática, dándonos acceso a la información necesaria para el desarrollo de la presente tesis.

También agradezco al el Ing. Agustín Ullón, por su apoyo y asesoramiento en el desarrollo de la presente Tesis.

Y a todos aquellos que prefirieron permanecer en el anonimato pero que estuvieron siempre con nosotros durante todo el camino de este trabajo de tesis.

Muchas Gracias.

El Autor.

RESUMEN
SOLUCION DE INTELIGENCIA DE NEGOCIOS PARA LA
GENERACIÓN DE INDICADORES QUE AYUDE EN EL ANALISIS
DE INFORMACION EN EL AREA DE VENTAS DE LA
FERRETERIA SOLANO BASADA EN LA METODOLOGÍA
HEFESTO UTILIZANDO PENTAHO BI

Por:

Br. Diego Alonso Corilla Amaya

El uso de una herramienta de inteligencia de negocios (BI) permite a una empresa analizar y organizar los datos transaccionales de sus sistemas de información así su data se encuentre en distintas fuentes, para obtener un conocimiento más profundo que facilite la interpretación y la comprensión de la información que servirá para la toma de decisiones generando una ventaja competitiva ante las otras empresas del mismo rubro. El presente trabajo de tesis realizó la implementación de una solución de inteligencia de negocios para la generación de indicadores que ayude en el análisis de información en el área de ventas de la ferretería Solano basada en la metodología Hefesto utilizando Pentaho BI se eligió Hefesto porque es una metodología ágil para la elaboración de soluciones BI, para la realización de este proyecto se hizo uso de la herramienta de software libre Pentaho BI la cual permitió realizar el tratamiento de los datos así como su transformación y carga al nuevo modelo que se implementó como solución, los resultados obtenidos de la aplicación permitieron obtener un nuevo conocimiento a la empresa que los gerentes ignoraban, se obtuvieron menores tiempos de respuesta con un alto nivel de granularidad que es analizada por los tomadores de decisiones y por los jefes de área, apreciada mediante gráficos dinámicos diseñados para un buen entendimiento de los indicadores y perspectivas generando nuevas decisiones que serán tomadas por los administrativos para la compra de productos, medición de eficiencia de personal, entrada y salida de productos y otros.

ABSTRACT

BUSINESS INTELLIGENCE SOLUTION FOR THE GENERATION OF INDICATORS TO HELP IN THE ANALYSIS OF INFORMATION IN THE SALES AREA OF THE SOLE HARDWARE STORE BASED ON THE METHODOLOGY HEFESTO USING PENTAHO BI

By:

Br.Diego Alonso Corilla Amaya

The use of a business intelligence (BI) tool allows a company to analyze and organize the transactional data of its information systems so that its data is in different sources, to obtain a deeper knowledge that facilitates the interpretation and understanding of the information that will serve to make decisions generating a competitive advantage over other companies in the same field. The present work of thesis realized the implementation of a business intelligence solution for the generation of indicators that help in the analysis of information in the sales area of Solano hardware based on the methodology Hephaestus using Pentaho BI Hephaestus was chosen because it is a an agile methodology for the development of BI solutions, for the realization of this project was made use of the free software tool Pentaho BI which allowed to perform the processing of the data as well as its transformation and load to the new model that was implemented as a solution, the results obtained from the application allowed to obtain a new knowledge to the company that the managers ignored, they obtained smaller times of response with a high level of granularity that is analyzed by the decision makers and by the heads of area, appreciated by means of dynamic graphs designed for a good understanding of the indicators and perspectives you are generating new decisions that will be taken by the administrative staff for the purchase of products, measurement of personnel efficiency, entry and exit of products and others.

Índice de Contenido

PRESENTACIÓN	ii
DEDICATORIA	iii
AGRADECIMIENTO	iv
RESUMEN	v
ABSTRACT	vi
Índice de Contenido.....	vii
INDICE DE IMÁGENES.....	xi
INDICE DE TABLAS	xiv
1. INTRODUCCION	1
1.1. Planteamiento del problema.....	2
1.2. Delimitación del problema.....	4
1.3. Formulación del problema	4
1.4. Formulación del hipótesis	4
1.5. Objetivos del estudio	5
1.6. Justificación del estudio	6
2. MARCO TEÓRICO	8
2.1. Antecedentes	8
2.2. DEFINICIONES	14
2.2.1. INTELIGENCIA DE NEGOCIOS (BI).....	14
2.2.2. VENTAJAS DE LA INTELIGENCIA DE NEGOCIOS.....	15
2.2.3. OLAP	16
2.2.4. TIPOS DE SISTEMAS OLAP.....	17
2.2.5. DATAMART	18
2.2.6. DATAWAREHOUSE.....	18
2.2.7. PROCESOS DE EXTRACION TRANSFORMACION Y CARGAR (ETL)....	20
2.2.8. ESQUEMA ESTRELLA PARA DATAWAREHOUSE.....	22
2.2.9. ESQUEMA COPO DE NIEVE PARA DATAWAREHOUSE.....	23
2.2.10. SISTEMAS DE SOPORTE A LAS DECISIONES.....	24
2.2.11. SISTEMAS DE INFORMACION.....	26
2.2.11.1. FACTORES DE UN SISTEMA DE INFORMACION.....	26
2.2.11.2. TIPOS DE SISTEMAS DE INFORMACION.....	27
2.2.12. PROCESAMIENTO DE TRANSACCIONES EN LINEA (OLTP).....	32
2.2.13. CUBOS OLAP	34
2.2.13.1. FUNCIONES	34
2.2.13.2. VENTAJAS DE LOS CUBOS OLAP	35
2.2.13.3. DESVENTAJAS DE LOS CUBOS OLAP.....	35

2.2.13.4.	DIMENSIONES	36
2.2.13.5.	MEDIDAS.....	37
2.2.13.6.	JERARQUIAS	38
2.2.14.	MICROSOFT SQL SERVER 2012	39
2.2.15.	MYSQL SERVER	39
2.2.16.	PENTAHO BI.....	39
2.2.16.1.	CARACTERISTICAS DE PENTAHO BI.....	40
2.2.17.	QLICKVIEW	41
2.3.	METODOLOGIA PARA EL DESARROLLO DEL PROYECTO	43
2.3.1.	Metodología Inmon	43
2.3.2.	Metodología de kimball	44
2.3.3.	Tipos de Arquitectura.....	46
2.3.4.	Comparación de metodologías	48
2.3.5.	Selección de la metodología.....	49
2.4.	METODOLOGIA HEFESTO.....	49
2.4.1.	Fases de la metodología Hefesto	50
2.4.1.1.	Análisis de los requerimientos	50
2.4.1.2.	Análisis de los OLTP	51
2.4.1.3.	Elaboración del Modelo lógico DW.....	52
2.4.1.4.	Procesos ETL	53
3.	MATERIALES Y METODOS	54
3.1.	Material.....	54
3.1.1.	Población.....	54
3.1.2.	Muestra.....	54
3.1.3.	Unidad de análisis	54
3.2.	Método	54
3.2.1.	Tipo de investigación	54
	Aplicada.....	54
3.2.2.	Diseño de Investigación	54
3.2.3.	Variables de estudio y Operacionalización	55
	<input type="checkbox"/> Independiente (VI): Solución de Inteligencia de Negocios usando la metodología Hefesto y Pentaho BI.....	55
	<input type="checkbox"/> Dependiente(VD): Indicadores que dan soporte a la toma de decisiones en el Área de ventas de la Ferretería Solano	55
3.2.4.	Instrumentos de recolección de Datos.....	56
3.2.5.	Procedimientos y análisis de datos.....	56
3.2.6.	Modelos estadísticos de análisis de datos.....	57
4.	RESULTADOS.....	62

4.1.	Definición del proyecto	62
4.2.	Justificación , Objetivos y alcance.....	62
4.3.	El negocio	63
4.4.	Procesos del área del negocio	65
4.4.1.	Descripción del proceso del área de ventas.....	65
4.4.2.	Descripción del proceso de almacén	66
4.4.3.	Descripción del proceso de compras	67
4.5.	Estrategia y método de la implementación	68
4.5.1.	Situación actual del proceso de generación de reportes	68
4.6.	Orígenes de datos.....	69
4.7.	Análisis de los requerimientos	70
4.7.1.	Requerimientos.....	71
4.8.	Identificar Indicadores y Perspectivas de Análisis	73
4.8.1.	Indicadores	73
4.8.2.	Perspectivas.....	73
4.9.	Modelo conceptual.....	74
4.9.1.	Diagrama de perspectivas por requerimiento	74
4.9.2.	Modelo conceptual del DataWareHouse por áreas	80
4.10.	Análisis de las fuentes de datos.....	82
4.10.1.	Fuente de información ferretería solano.....	83
4.11.	Análisis de los OLTP.....	84
4.11.1.	Determinar indicadores	84
4.12.	Modelo Dimensional	85
4.12.1.	Tipo de modelo lógico del Datawarehouse	86
4.12.2.	Tablas de dimensiones	89
4.12.3.	Tablas de hechos	94
4.12.4.	Modelo físico del Data Warehouse	97
4.13.	Establecer correspondencias.....	98
4.14.	Nivel de granularidad.....	103
4.15.	Modelo Conceptual ampliado	106
4.16.	Selección de las herramientas	109
4.16.1.	Instalación de las herramientas.....	109
4.17.	Integración de datos	113
4.18.	Proceso ETL para las dimensiones y hechos.....	113
4.19.	Creación del cubo OLAP	137
4.20.	Reportes.....	147
5.	DISCUSION DE RESULTADOS	159

Formulación de la hipótesis	159
5.1. MANERA PRESENCIAL	159
5.1.1. DISEÑO PREEXPERIMENTAL PRE-PRUEBA Y POST-PRUEBA	160
5.1.2. CÁLCULO DE LOS INDICADORES DE LA HIPÓTESIS.....	161
5.1.3. APLICACIÓN DEL RANGO DE SATISFACCIÓN A LOS INDICADORES DE LA HIPÓTESIS	171
5.1.4. ANÁLISIS ESTADÍSTICO PARA LA PRUEBA PRESENCIAL DE LA HIPÓTESIS	172
5.2. MANERA NO PRESENCIAL	174
6. CONCLUSIONES	177
7. RECOMENDACIONES	178
8. REFERENCIAS BIBLIOGRAFICAS	180
9. ANEXOS.....	182

INDICE DE IMÁGENES

<i>Ilustración 1.Ciclo del business Intelligence</i>	15
Ilustración 2 Proceso ETL	21
Ilustración 3.Esquema estrella.....	22
Ilustración 4 Esquema copo de nieve	23
Ilustración 5.Factores de un sistemas de información.....	26
Ilustración 6.Piramide de los sistemas de información	28
Ilustración 7.Cubo OLAP	36
Ilustración 8.Cubo OLAP y sus dimensiones	37
Ilustración 9.Jerarquias de dimensiones	38
Ilustración 10.Arquitectura Inmon	44
Ilustración 11.Arquitectura de kimball	45
Ilustración 12.Arquitectura Top Down	46
Ilustración 13.Arquitectura Bottom up.....	46
Ilustración 14. Arquitectura Hibrida.....	47
Ilustración 15.Modelo conceptual	51
Ilustración 16.Fases de la metodología Hefesto	53
Ilustración 17.Region de aceptación de la hipótesis.....	58
Ilustración 18.ubicacion geográfica de la ferretería solano	63
Ilustración 19.ubicacion geográfica de la ferretería solano sucursal 02.....	64
Ilustración 20.Pictograma de ventas de ferretería solano	65
Ilustración 21.Pictograma del proceso de almacén	66
Ilustración 22.Pictograma del proceso de compras	67
Ilustración 23.Proceso actual de generación de reportes de todas las áreas de la ferretería solano.....	68
Ilustración 24.Estructura de la base de datos origen BDFerreteriaSolano	69
Ilustración 25.requerimiento N°1	74
Ilustración 26.Requerimiento N°2.....	74
Ilustración 27.Requerimiento N°3	74
Ilustración 28.Requerimiento N°4	75
Ilustración 29.Requerimiento N°5	75
Ilustración 30.Requerimiento N°6.....	75
Ilustración 31.Requerimiento N°7	76
Ilustración 32.Requerimiento N°8.....	76
Ilustración 33.Requerimiento N°9.....	76
Ilustración 34.Requerimiento N°10.....	77
Ilustración 35.Requerimiento N°11	77
Ilustración 36.Requerimiento N°12.....	77
Ilustración 37.Requerimiento N°13	78
Ilustración 38.Requerimiento N°14.....	78
Ilustración 39.Requerimiento N°15	78
Ilustración 40.Requerimiento N°16.....	79
Ilustración 41.Requerimiento N°17	79
Ilustración 42.Requerimiento N°18	79
Ilustración 43.Modelo conceptual del área de compras	80
Ilustración 44.Modelo conceptual del proceso de almacén.....	81
Ilustración 45.Modelo conceptual del proceso de ventas	81
Ilustración 46.Procesos diarios en BD transaccional.....	82
Ilustración 47.Diagrama entidad relación de la base de datos transaccional	

BDFerreteriaSolano	83
Ilustración 48.Modelo lógico de datos del proceso de compras.....	86
Ilustración 49,Modelo lógico del proceso de almacén	87
Ilustración 50.Modelo lógico del proceso de ventas	88
Ilustración 51.Tabla Dim_producto	89
Ilustración 52.Tabla Dim_material	90
Ilustración 53.Tabla Dim_cliente	90
Ilustración 54.Tabla Dim_sucursal	91
Ilustración 55.Tabla Dim_ubigeo	91
Ilustración 56.Tabla Dim_empleado.....	92
Ilustración 57.Tabla Dim_movimiento.....	93
Ilustración 58.Tabla Dim_proveedor	93
Ilustración 59.Tabla Dim_tiempo	94
Ilustración 60.Tabla ventas_hechos	95
Ilustración 61.Tabla Compras_hechos.....	95
Ilustración 62.Tabla Almacen_hechos	96
Ilustración 63.Modelo entidad relación de la base de datos dimensional	97
Ilustración 64.Nivel de granularidad del proceso de ventas.....	103
Ilustración 65.Nivel de granularidad del proceso de compras.....	104
Ilustración 66.Nivel de granularidad proceso de almacén.....	105
Ilustración 67.Modelo conceptual ampliado del proceso de ventas	106
Ilustración 68.Modelo conceptual del proceso de compras.....	107
Ilustración 69.Modelo conceptual ampliado del proceso de almacén	108
Ilustración 70.Archivos de Pentaho Data Integration.....	110
Ilustración 71.utilitarios del schema workbench	111
Ilustración 72.pentaho data Integration	112
Ilustración 73.Inicio de sesión para el repositorio de Pentaho	112
Ilustración 74.conexión de la base de datos origen	113
Ilustración 75.Consulta SQL para la elección de campos	114
Ilustración 76.Mapeo de datos dim_producto	115
Ilustración 77.Proceso ETL culminado Dim_producto	115
Ilustración 78.consulta para poblamiento Dim_empleado	116
Ilustración 79.Mapeo Dim_empleado	117
Ilustración 80.Carga de datos Dim_empleado.....	117
Ilustración 81, consulta para Dim_sucursal.....	118
Ilustración 82.mapeo dim_producto	118
Ilustración 83, carga de datos Dim_sucursal	119
Ilustración 84.Consulta SQL para Dim_cliente.....	120
Ilustración 85.Mapeo Dim_cliente	120
Ilustración 86.carga de datos Dim_cliente	121
Ilustración 87.Consulta SQL para Dim_tiempo_ventas	122
Ilustración 88.Mapeo tabla Dim_tiempo_ventas.....	122
Ilustración 89.Carga de datos para Dim_tiempo_ventas	123
Ilustración 90.Carga de datos Dim_tiempo_almacen.....	124
Ilustración 91. Carga de tabla Dim_tiempo_compras	124
Ilustración 92.consulta SQL para la tabla Dim_proveedor	125
Ilustración 93.mapeo tabla Dim_proveedor	125
Ilustración 94.carga de datos Dim_proveedor.....	126
Ilustración 95.Carga de datos de tabla Dim_material.....	126
Ilustración 96.carga de datos de la tabla Dim_ubigeo.....	127

Ilustración 97.consulta para la tabla de compra_ hechos	128
Ilustración 98.ordenar filas y cálculos compra_ hechos	129
Ilustración 99.mapeo compra_ hechos	129
Ilustración 100.carga de datos compra_ hechos	130
Ilustración 101.consulta para almacen_ hechos	131
Ilustración 102. Ordenar campos Almacen_ hechos	132
Ilustración 103.mapeo de la tabla Almacen_ hechos.....	132
Ilustración 104.carga de datos de la tabla Almacen_ hechos	133
Ilustración 105.consulta SQL para ventas_ hechos	134
Ilustración 106.ordenar campos ventas_ hechos	135
Ilustración 107.mapeo ventas_ hechos	135
Ilustración 108.carga de datos ventas_ hechos	136
Ilustración 109.conexion a la base de datos para el cubo	137
Ilustración 110.vista previa de creación del cubo ventas_ hechos	138
Ilustración 111.esquema del cubo ventas_ hechos	139
Ilustración 112.estructura del cubo almacen_ hechos	140
Ilustración 113.estructura del cubo Compras_ hechos	141
Ilustración 114.publicación del cubo	141
Ilustración 115.página principal de Pentaho BI server	142
Ilustración 116.conexion de la base de datos en BI SERVER.....	143
Ilustración 117, cubo ventas_ hechos	144
Ilustración 118.cubo Almacen_ hechos	145
Ilustración 119.cubo Compras_ hechos	146
Ilustración 120.reporte requerimiento N°1	147
Ilustración 121.reporte de requerimiento N°2	148
Ilustración 122.reporte de requerimiento N°3	149
Ilustración 123.reporte de Requerimiento N°4.....	149
Ilustración 124.reporte de requerimiento N°5	150
Ilustración 125.reporte de requerimiento N°6	150
Ilustración 126, reporte de Requerimiento N°7.....	151
Ilustración 127.reporte de Requerimiento N°8.....	151
Ilustración 128.reporte de Requerimiento N°9.....	152
Ilustración 129.reporte de requerimiento N°10	152
Ilustración 130.reporte de requerimiento n°11	153
Ilustración 131.Reporte de requerimiento N°12.....	153
Ilustración 132.reporte de requerimiento N°13	154
Ilustración 133.reporte de requerimiento N°14	155
Ilustración 134.reporte de requerimiento N°15	155
Ilustración 135.reporte de requerimiento N°16	156
Ilustración 136.Reporte de requerimiento N°17.....	157
Ilustración 137.reporte de requerimiento 18.....	158
Ilustración 138. Carta de aceptación por la empresa	182
Ilustración 139. Modelo de encuesta para hipótesis	183
Ilustración 140.Modelo de encuesta para hipótesis 2	184

INDICE DE TABLAS

Tabla 1.Diferencia entre Datawarehouse y bases de datos transaccionales	19
Tabla 2.Comparacion de metodologías	48
Tabla 3.Diagrama de investigación	55
Tabla 4.variables de estudio y Operacionalización	55
Tabla 5.Hipotesis nula y alternativas	57
Tabla 6.Tablas de datos.....	58
Tabla 7.tablas Muestrales	59
Tabla 8.tabla datos ponderado	60
Tabla 9.Tabla resumen	60
Tabla 10.tabla resultados	61
Tabla 11.tabla de requerimientos funcionales de la ferretería solano.....	72
Tabla 12.Indicadores y su descripción.....	84
Tabla 13.Mapeo Dim_producto.....	98
Tabla 14.Mapeo Dim_cliente	98
Tabla 15.Mapeo Dim_proveedor.....	99
Tabla 16.Mapeo Dim_material.....	99
Tabla 17.Mapeo Dim_movimiento.....	99
Tabla 18.Mapeo Dim_sucursal.....	100
Tabla 19.Mapeo Dim_ubigeo	100
Tabla 20.Mapeo Dim_tiempo.....	100
Tabla 21.Mapeo Dim_tiempo_ventas.....	101
Tabla 22.Mapeo Dim_tiempo_almacen.....	101
Tabla 23.Mapeo Ventas_hechos.....	101
Tabla 24.Mapeo Compras_hechos.....	102
Tabla 25.Mapeo Almacen_hechos	102
Tabla 26.tablas de rangos de satisfacción.....	161
Tabla 27.Tabla de indicadores promedios y valores.....	172
Tabla 28.Tabla diferencia de las medias	172
Tabla 29.calculo prueba de la hipótesis	173

1. INTRODUCCION

La presente tesis es una investigación que tiene como objetivo mejorar el proceso de toma de decisiones de la ferretería Solano utilizando la inteligencia de negocios como herramienta tecnológica, aumentando la capacidad de explotar y de organizar la información que se encuentra alojada en sus bases de datos de distintas fuentes, las cuales mediante distintos tipos de procesos permitirán la creación de una nueva información que permitirá realizar un análisis de la información a nivel estratégico, facilitando la interpretación y la comprensión de los usuarios finales, encargados de la toma de decisiones en la empresa.

La herramienta propuesta para el presente trabajo permite reunir, depurar y transformar la información que se encuentra desestructurada y sin normalizar, que provienen de distintos tipos de bases de datos, eliminando las redundancias y los datos irrelevantes.

Se llevaran a cabo distintos procesos o fases establecidas por la metodología Hefesto la cual es un metodología para el desarrollo de aplicaciones BI el cual empezara por el análisis de los datos continuara con la extracción y carga y finalizara con la explotación de la información la cual generara los resultados finales mostrando reportes estadísticos con información precisa y exacta para poder llevar a cabo la toma de decisiones.

1.1. Planteamiento del problema

Actualmente la Tecnología de la Información se ha convertido en una herramienta clave en el proceso de desarrollo continuo dentro de las empresas. La competitividad del mercado y la globalización de la industria plantean un reto mayor dentro de toda organización pues se hace necesaria la innovación y el planeamiento estratégico que permita a la empresa trascender con un producto, proceso o servicio diferenciado.

El manejo eficiente y eficaz de los datos es un componente indispensable para la toma de decisiones, que genera un conocimiento importante para cualquier organización en este contexto es ahí donde aparecen conceptos como el de inteligencia de negocios , que busca facilitar a las empresas realizar la recolección de su información corporativa, la presentación de sus resultados, en resumen, ayudar a la empresa a comprender el significado de los datos que diariamente almacena en sus sistemas logísticos para que éstas puedan tomar decisiones que ayuden a crecer su rentabilidad .

Una empresa que no tome la inteligencia de negocios como una solución a sus problemas y mejora de procesos es una empresa que quedaría varada frente a un mercado competitivo, que aprovecha de manera inteligente los recursos tecnológicos y las tecnologías de información que existen hoy en día. La Ferretería Solano S.A.C, es una empresa que tiene más de 30 años en el mercado, liderando en ventas de materiales de Construcción así como también en Herramientas de Carpintería, Productos para pintores entre otros. Trabajando en el abastecimiento de Constructoras y de otras Ferreterías.

Ferretería Solano S.A.C cuenta con una Gerencia que se encarga de dirigir los diferentes Procesos que se llevan a cabo así como también de la Toma de Decisiones, de una Área Contable que se encarga de llevar el control de los Ingresos y Egresos que tiene la Ferretería, Área de Ventas que se encarga de ver el inventario diario de los productos vendidos así como también las ganancias, Área de Compras que se encarga de los Pedidos y Abastecimiento de Productos provenientes de los diferentes Proveedores Locales y Nacionales y de una Área de Almacén que lleva un registro del Movimiento de los Productos así como también como el Stock de estos.

La empresa requiere remojar el análisis de la información para generar indicadores que le den un mayor conocimiento del record de sus ventas por temporadas y así

tener un mejor panorama del estado de su estado actual para poder realizar una mejor toma de decisiones de su área de ventas utilizando la inteligencia de negocios como herramienta tecnológica ya que esta carece de conocimiento para mejorar sus procesos y aumentar su rentabilidad. La prioridad de esta organización es mejorar sus procesos internos con el fin de mejorar la gestión del área del problema y que la sub gerencia tenga indicadores con resultados que les permita tomar mejores decisiones y así mejorar la gestión y los procesos del área interesada. hay un problema que existe en las diferentes grandes empresas, que es la falta de análisis en sus datos , que son requeridos y necesarios para cada línea de negocio , esto se genera por las distintas formas en que se integran sus sistemas de información generando problemas de información imprecisa o sin valor que no ayude al crecimiento de la empresa y a una buena toma de decisión debido a datos dispersos falta de un almacenamiento integro de la información que unifique los datos en un repositorio central .Estas limitaciones hacen que la información que se obtiene a través de los sistemas transaccionales sean solo datos en un repositorio sin ningún valor, Los análisis van perdiendo valor por el tiempo que toma en organizarlos manualmente, por la manipulación de los datos y posibles errores humanos que ello ocasiona. Al momento de extraer los datos que son necesarios para crear el conocimiento extraídos de los repositorios de bases de datos de la empresa son considerados como información irrelevante para poder realizar un análisis de datos. El área de ventas al solo tener datos en tablas y no tener un reporte con datos contemporáneos e indicadores que indique el buen funcionamiento de los diferentes procesos de la empresa carece de información para poder tomar decisiones acerca de las distintas áreas y poder realizar un plan para poder tener ventaja competitiva tanto interna como externamente. Ya que el área de ventas es una de las áreas más importantes que es la encargada de generar las ganancias y la seguridad de la ventaja competitiva de la ferretería solano.

1.2. Delimitación del problema

En el desarrollo del presente proyecto de investigación, puede presentarse posibles riesgos que al presentarse podrían perjudicar esta investigación a continuación se listaran unos posibles problemas:

- Falta de interés en el trabajo de investigación por parte del autor.
- La empresa no desee proporcionar la información requerida.
- Algún problema de salud del autor o del asesor.
- El tiempo establecido en el cronograma no es suficiente para poder concluir el trabajo.
- Falta de conocimiento de parte del equipo investigador y del asesor para poder realizar el proyecto.
- Problemas económicos del equipo investigador que impida el avance del proyecto.

1.3. Formulación del problema

Dada la problemática de la investigación se formula lo siguiente:

¿Cómo mejorar el análisis de la información para generar indicadores que den soporte a la toma de decisiones al área de ventas de la ferretería “solano”?

1.4. Formulación del hipótesis

El desarrollo de una solución de Inteligencia de Negocios usando la metodología Hefesto y Pentaho BI en el Área de Ventas de la Ferretería Solano permitirá mejorar el análisis de la información generando indicadores que den soporte a la toma de decisiones.

1.5. Objetivos del estudio

El Objetivo general es:

Realizar una Aplicación de inteligencia de negocios para la generación de indicadores en el área de ventas de la ferretería Solano basada en la metodología Hefesto y utilizando Pentaho BI.

Los objetivos específicos son los siguientes:

- ✓ Realizar un análisis bibliográfico de las metodologías en el desarrollo de soluciones de Inteligencia de Negocios.
- ✓ Obtener y analizar los requerimientos de las áreas críticas del negocio.
- ✓ Desarrollar el proceso ETL utilizando como herramienta Pentaho BI
- ✓ Diseñar la estructura de que tendrán las dimensiones y cubos pertinentes para crear el diseño de vista de análisis y la navegabilidad que tendrá la información utilizando Pentaho BI Server.
- ✓ Utilizar la herramienta Qlikview para crear y diseñar los reportes correspondientes al Modelo implementado.
- ✓ Evaluar los resultados de los informes que muestra la solución BI implementada.
- ✓ Utilizar como ciclo de vida de desarrollo del proyecto la metodología Hefesto

1.6. Justificación del estudio

Importancia de la investigación

- El modelo de inteligencia de negocios que se realizara en La ferretería solano permitirá recolectar, depurar y transformar los datos que se encuentren desestructurados de la base de datos transaccional de la empresa evitando la redundancia de datos.
- La importancia de tener en la empresa un modelo de inteligencia de negocios es que este englobara los sistemas donde se pueda obtener información de distintas áreas de donde se requiera indicadores para tomar una decisión y así crear una ventaja competitiva en el mercado.
- La solución de inteligencia de negocios representa un apoyo de gran importancia al momento de realizar informes y reportes ya que el personal encargado de realizar la tarea de brindar los reportes prioriza realizar esta tarea a tiempo para poder brindar la información a la gerencia para poder definir las estrategias y las decisiones que tomaran los gerentes.
- Tener una buena consistencia en los datos es un factor muy importante hoy en día para cualquier empresa sea empresa grande o una pyme ya que de acuerdo a esto se podrá tener una mejor fiabilidad de la información que se procesara para la toma de decisiones.
- Los datos estadísticos que se generaran de la herramienta de inteligencia de negocios permitirá determinar medidas de cambio y mejora provenientes de los datos procesados que podrá brindar una realidad más clara a la situación actual del negocio.

El presente trabajo de tesis está organizado en diferentes puntos que facilitarán el uso y entendimiento del mismo dando a continuación una breve descripción del mismo:

Marco teórico: Fundamento teórico y metodología, en este punto damos el conocimiento formal de los temas y el modelo de referencia a utilizar para la solución del problema planteado. En esta parte damos los conocimientos básicos de que es un Datamart, Data Warehouse, indicadores, Soporte de tomas de decisiones y Cubos OLAP.

También en este capítulo referimos la metodología que empleamos para el desarrollo del proyecto, presentando el Enfoque, Esquema y Herramientas que intervendrán en la misma.

Resultados: Desarrollo del trabajo, en éste punto se muestra el desarrollo de los pasos enunciados en el Esquema de la Metodología. Mostramos los resultados obtenidos con relación a los objetivos planteados al inicio de este proyecto.

Discusión de resultados: En este capítulo se verifica si la Hipótesis es aceptada a esto se le llama Contrastación de la Hipótesis.

Conclusiones y Recomendaciones, en éste último capítulo se encuentra las conclusiones a las que se llegó después de haber culminado el proyecto y las recomendaciones derivadas de la experiencia de la misma.

Además en la parte final del documento se encuentran los anexos con información adicional, referidos a los capítulos enunciados.

2. MARCO TEÓRICO

2.1. Antecedentes

- **Autores:** Manuel Javier Valdiviezo Basauri, Iván Zico Segundo Herrera Geldres, Gerardo Daniel Jáuregui Puertas, título de la investigación: ANALISIS Y DISEÑO DE UNA HERRAMIENTA DE INTELIGENCIA DE NEGOCIOS–ANALISIS DIMENSIONAL” (2011).

Descripción: Hoy en día las empresas necesitan depositar mucha confianza en la toma de decisiones sobre sus negocios, basándose en la información que poseen sobre sus operaciones y lo que ocurre en el mercado. Estas decisiones, que son tomadas en un contexto de alta y creciente competencia, se toman cada vez más utilizando la mayor cantidad de información que la empresa posee. Una mala decisión puede afectar significativamente a la organización. La toma de decisiones implica incertidumbre y por lo tanto riesgo. Para minimizar estos riesgos es necesario disponer de información. La información en la empresa debe ser clara, precisa, oportuna, completa, de fácil acceso y sobre todo necesaria y no superflua. Para aquellos que desarrollan y mantienen los sistemas de información de soporte de decisiones, se hace necesario contar con una herramienta que ofrezca versatilidad y facilidad en la implementación de soluciones. Esta herramienta debe permitir generar reportes de diversos tipos e integrar información y datos desde diversos contextos. La arquitectura de datos e información debe ser diseñada del modo más simple posible, pero a la vez efectivo, entendible y eficiente. Es necesario tener un modelo que represente el negocio de la manera más adecuada y permita diseñar la obtención de la información de ésta del modo más natural. Además, otro punto clave en este tipo de sistemas es la población y la calidad de los datos con los que se cuenta. En base a un modelo, o metadata, se puede determinar el tipo de información que va a ser contenida en éste. Una tarea importante es probar los datos a ser puestos en este esquema, por lo que la utilidad de una herramienta que basada en reglas las pueble con datos es significativa. Por lo tanto, el presente proyecto de tesis consistió en preparar

el análisis y diseño para el desarrollo de una aplicación que asista en el diseño de modelos de datos dimensionales, parte fundamental de la datamart. Estos modelos podrán ser enviados a motores de base de datos y generados en ellos, automáticamente. Otra característica importante de este tema de tesis es dar a los usuarios la posibilidad de, sobre estos modelos creados, generar datos en base a reglas definidas y con éstos poblarla base de datos. En este caso la herramienta trabajará por medio de una interfaz con un Sistema Generador de Datos ya existente. Este trabajo de tesis presenta también el análisis y diseño necesarios para este caso. El producto se aplica para los principales manejadores de base de datos relacionales, tales como Oracle y MS SQL Server. Los datos podrán ser almacenados en archivos de texto y XML. Este tema de tesis propone desarrollar las etapas de análisis y diseño del proceso de desarrollo de productos de software para la herramienta propuesta. Las etapas de construcción y las pruebas de desempeño de dicha herramienta son presentadas en la tesis “Construcción y Pruebas de una herramienta de desarrollo de soluciones para inteligencia de negocios –Análisis dimensional”.

- **Autores:** Gastañadui Caballero Yajaira Ross Yta, Tamayo Avalos, Jhon Pier, título de la investigación: “INTELIGENCIA DE NEGOCIOS PARA LA DINAMIZACIÓN EN LA TOMA DE DECISIONES EN LA GESTIÓN CONTABLE Y PRESUPUESTAL DE LA MUNICIPALIDAD DE VÍCTOR LARCO HERRERA” (2016)

Descripción: La Municipalidad Distrital de Víctor Larco Herrera es una institución pública que brinda servicios públicos de alta calidad, promoviendo una fuerte gobernabilidad democrática y asegurando la mayor participación ciudadana en la formulación de las políticas locales logrando que los vecinos Victorlarquenses vivan en un lugar saludable, ordenado, seguro, turístico y atractivo. El accionar de la municipalidad es auditado cada cierto tiempo por organismos gubernamentales, especializados en cada área de la municipalidad; el manejo de los ingresos, egresos y presupuesto tiene particular importancia en caso de auditorías es debido a esto que el control de las áreas de contabilidad y manejo presupuestario tienen una mayor necesidad de acceso a la información específica de manera ágil. Al aprobar una partida presupuestal para la municipalidad se inicia con el proceso de distribución de

la misma para cumplir con las obligaciones propias de las municipalidades, además de los proyectos a ejecutar; antes de la aprobación de los proyectos, durante y posteriormente a dicha ejecución se necesita tomar decisiones para planear el proyecto controlar el gasto y evaluar los resultados a nivel económico; los reportes cumplen un papel importante pues en ellos se sustenta la toma de decisiones. Se tomó una medición de los tiempos de emisión de reportes, usando el sistema de información transaccional actual dando como resultados los tiempos indicados en la tabla N°1, de esta medición logramos identificar tiempos de emisión de reportes muy elevados los cuales generan inconvenientes al momento de tomar decisiones.

- **Autores:** Arratia Vinatea, Ricardo José; Pastor Quispe, Carolina Aloise
Titulo de la investigación: PROPUESTA DE IMPLEMENTACIÓN DE UNA SOLUCIÓN BUSINESS INTELLIGENCE PARA EL PROYECTO DE CONTROL Y ERRADICACIÓN DE LA MOSCA DE LA FRUTA DEL SENASA (2013).

Descripción: El objetivo principal del proyecto es la elaboración de una propuesta de implementación de una solución Business Intelligence para el proyecto de control y erradicación de la mosca de la fruta del SENASA. Para el SENASA el proyecto de control y erradicación de la mosca de la fruta es un proyecto estratégico que está orientado a la erradicación de la plaga que causa daños en los cultivos a nivel nacional, esta plaga causa un impacto negativo a los agricultores que se ven afectados debido a que sus cultivos son invadidos por la plaga. Además, el Perú como país se ve afectado porque los mercados internacionales exigen que los frutos que van a ser exportados estén libres de esta plaga. El SENASA mantiene sistemas de información que registran de manera transaccional la información del estado de la plaga a nivel nacional, pero el análisis posterior de esta información no es explotado de manera eficiente, se requieren de horas hombre para obtener reportes de análisis de información del estado de la plaga. Además de esto, el análisis de información solo se realiza considerando algunas variables de datos, y no se aprovecha la información de la base de datos del SENASA. Por ser una plaga de rápida propagación, el tiempo necesario para el análisis de información es crítico para la toma de decisiones. El proyecto plantea una propuesta de implementación de una solución BI que apoye al proyecto de control y

erradicación de la mosca de la fruta del SENASA a la explotación de la información existente en los sistemas transaccionales usando una herramienta BI como IBM Cognos. En el proyecto se propone una metodología para la implementación para el SENASA, en esta metodología se identificarán los requerimientos del proyecto de Control y Erradicación de la mosca de la fruta del SENASA, luego se desarrollará el diagrama starnet para identificar la información a ser considerada en el modelo de datos de la solución, el diagrama de arquitectura tecnológica para considerar las características técnicas de tecnologías de información existentes, se definirán los principales procesos de carga ETL y finalmente los prototipos de análisis de información, los mismos que fueron validados por el área usuaria y la Unidad de Informática y Estadística del SENASA. La evaluación financiera demuestra que el proyecto es un proyecto viable que tiene un retorno de inversión de 10 meses. Este proyecto puede ser el primer hito para que los otros procesos del proyecto de control y erradicación de la mosca de la fruta del SENASA como son los procesos de producción y distribución de moscas estériles en laboratorio se implementen posteriormente al datamart de la solución planteada.

- **Autores:** Horna Gutiérrez Diana Carolina, Rodríguez Marquina Fabián Lorenzo, Título de la investigación: “IMPLEMENTACIÓN DE UN S.I. DE ANÁLISIS PREDICTIVO PARA LA TOMA DE DECISIONES EN EL PROCESO DE ATENCIÓN MÉDICA DEL HOSPITAL VÍCTOR LAZARTE USANDO LA HERRAMIENTA DE COGNOS BI DE IBM”(2016).

Descripción: El Hospital “Víctor Lazarte Echegaray” fue inaugurado el lunes 29 de enero de 1951, para la atención de la clase obrera denominándose en un principio Hospital Obrero de Trujillo. Luego en 1973 Hospital Zonal y desde 1980 Hospital Víctor Lazarte Echegaray. Los orígenes del Hospital Víctor Lazarte se remontan a 1912, durante el gobierno de Augusto B. Leguía quien promulgo la ley sobre accidentes de trabajo para obreros y empleados, formulándose en esta, una serie de dispositivos que reconocían el derecho al bienestar del trabajador el mismo que incluía promoción, recuperación y subsidios, para el cuidado de su salud. Los objetivos de este hospital, se orientan a aumentar la atención preventiva promocional mediante programas,

dando una atención recuperativa de calidad, buscando la satisfacción plena del usuario. Para ello privilegia el uso de protocolos en la atención de los asegurados, usándose una nueva y única historia clínica a nivel nacional. Otro aspecto al interior del Hospital es Auditoria Medica, la misma que ya se practica en todos los servicios y para todos los procedimientos, especialmente la historia clínica, el manejo terapéutico y la morbilidad. La principal función del Hospital Víctor Lazarte es dar cobertura a los asegurados y sus derechos habientes, a través del otorgamiento de prestaciones de prevención, promoción, recuperación, rehabilitación, prestaciones económicas, y prestaciones sociales que corresponden al régimen contributivo de la Seguridad Social en Salud, así como otros seguros de riesgos humanos. Sin embargo, existen diversos problemas en dicho Hospital con respecto a las necesidades de asignación de algunos consultorios, doctores, personal de apoyo; que es insuficiente en algunas áreas del sector de una institución de salud, lo cual el Director del Hospital no cuenta con un sistema que apoye a la toma de decisiones (modelo de análisis predictivo) que le ayude a mejorar el proceso de toma de decisiones en el Hospital Víctor Lazarte.

- **Autores** : Alejandro Rojas Zaldívar ,Titulo de la investigación: “IMPLEMENTACIÓN DE UN DATA MART COMO SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS, BAJO LA METODOLOGÍA DE RALPH KIMBALL PARA OPTIMIZAR LA TOMA DE DECISIONES EN EL DEPARTAMENTO DE FINANZAS DE LA CONTRALORÍA GENERAL DE LA REPÚBLICA (2014)”

Descripción: La Contraloría General de la República, es una entidad pública cuya misión es: “Trabajar para promover el desarrollo de una gestión eficaz y moderna de los recursos públicos en beneficio de todos los peruanos.” La Gerencia Central de Administración y Finanzas es el órgano dependiente del Contralor General de la República, encargado de diseñar e implementar la organización, estructura, políticas, controles y sistemas referidos a la gestión de recursos humanos, gestión financiera y logística; las políticas y procesos para la asignación presupuestal por unidad orgánica, incluyendo las responsabilidades de gasto por partida presupuestal y destino del gasto; el sistema de información de gestión administrativo y financiera y la implementación de la organización para soportar el desarrollo y

mantenimiento de los sistemas requeridos por el plan de transformación. De igual forma, conduce la Escuela Nacional de Control que está a cargo de un gerente central. Dentro de la Gerencia Central de Administración y Finanzas, podemos encontrar el Departamento de Logística. El Departamento de Finanzas es el órgano dependiente de la Gerencia Central de Administración y Finanzas, encargado de efectuar la programación, ejecución y evaluación de las actividades relacionadas con los procesos de presupuesto, contabilidad y tesorería, así como elaborar e interpretar los estados financieros y presupuestales de la Contraloría General de la República. Dentro de las funciones del Departamento de Finanzas, las que están más ligadas a la Alta Dirección son: “Organizar, programar y ejecutar los procesos que permitan presentar a la alta dirección la información económica y financiera oportuna para la toma de decisiones, formular el presupuesto institucional en coordinación con las unidades orgánicas, emitir los informes de disponibilidad presupuestaria para la adquisición o contratación de bienes, servicios u obras que requiera la entidad, formular la consolidación mensual de gastos y efectuar la consolidación con los saldos de balance y presentar a la alta dirección los estados financieros y presupuestales del pliego”. La Alta dirección y el Departamento de Finanzas, realizan reportes mediante su sistema transaccional actual: Sistema de Gestión Administrativa (SIGA), donde cada usuario del área obtiene sus reportes de manera propia. Cada gerente encuentra los reportes en un menú en el sistema actual. Está información económica y financiera es solicitada por la Alta Dirección mensualmente para la adecuada toma de decisiones.

2.2. DEFINICIONES

2.2.1. INTELIGENCIA DE NEGOCIOS (BI)

Según el autor (Cano, 2007) El objetivo básico de la Business Intelligence es apoyar de forma sostenible y continuada a las organizaciones para mejorar su competitividad, facilitando la información necesaria para la toma de decisiones. El primero que acuñó el término fue Howard Dresner que, cuando era consultor de Gartner, popularizó Business Intelligence o BI como un término paraguas para describir un conjunto de conceptos y métodos que mejoraran la toma de decisiones, utilizando información sobre qué había sucedido (hechos). Mediante el uso de tecnologías y las metodologías de Business Intelligence pretendemos convertir datos en información y a partir de la información ser capaces de descubrir conocimiento. (Cano, 2007)

Los procesos de inteligencia de negocios se han convertido en una parte esencial de la gestión empresarial, ya que nos proporcionan información de mucha importancia para una organización:

Nos permiten saber con exactitud ¿quiénes son nuestros clientes?

- ¿dónde están los clientes?
- ¿cómo son los clientes?
- ¿qué pautas de comportamiento tienen?
- ¿qué es lo que quieren los clientes?

Así mismo Permiten conocer el funcionamiento de los procesos internos de una organización. Dan respuesta a preguntas importantes como:

- ¿cuánto cuesta cada uno de los productos o servicios que ofrece la empresa?
- ¿cuáles de esos productos o servicios son realmente rentables?
- ¿dónde está perdiendo la empresa?
- ¿cuánto cuesta conseguir cada cliente?
- ¿cuáles departamentos de la empresa funcionan correctamente?
- ¿qué servicios puede externalizar la empresa?

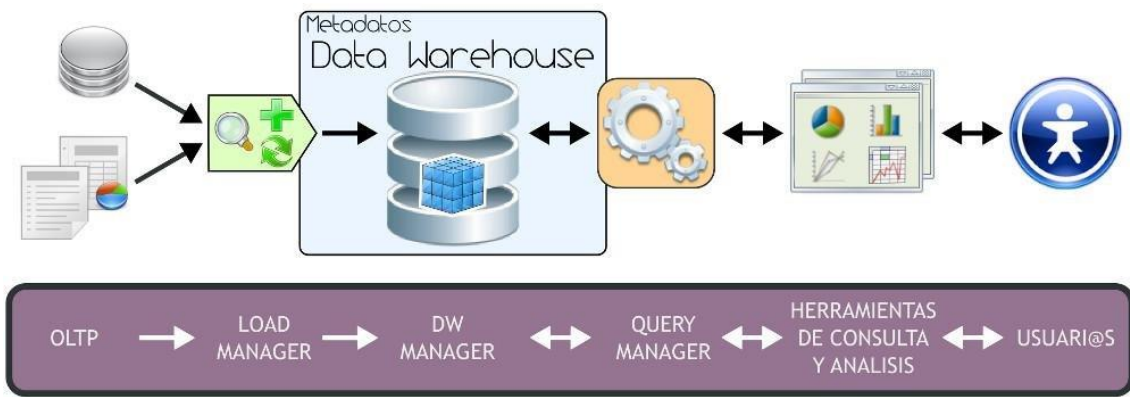


Ilustración 1. Ciclo del business Intelligence

Fuente: (bernabeu, 2010)

2.2.2. VENTAJAS DE LA INTELIGENCIA DE NEGOCIOS

La adquisición de un sistema de BI debe proporcionar las siguientes ventajas: (rio, 2006)

- Una plataforma de tecnología integrada que se añadiría a las inversiones ya realizadas por una organización, para proporcionar información de alta calidad a cada computadora o servidor de cada departamento de una empresa, añadiendo valor en cada paso del proceso y proporcionando una versión única de la realidad.
- Acceso amplio y potenciado a las capacidades de análisis ya conocidas que ayudan a conocer el pasado de una organización para controlar y comunicar el presente y predecir el futuro con fiabilidad.
- Interfaces de usuario personalizadas que “se adapten a cada tarea”, concebida y diseñadas para todos los niveles de experiencia y patrones de uso de los usuarios de la información (como expertos, ejecutivos, gerentes, tecnólogos, etc.).
- Una gama de soluciones para satisfacer la demanda de información y generación de informes de diferentes sectores de actividad (servicios financieros, fabricación, telecomunicaciones, salud, etc.) y de toda la empresa (gestión del rendimiento empresarial, inteligencia de clientes, inteligencia financiera, gestión del capital humano, inteligencia de la cadena de suministro, etc.).

2.2.3. OLAP

El procesamiento analítico en línea permite obtener acceso a datos organizados y agregados de orígenes de datos empresariales, organiza subconjuntos de datos con una estructura multidimensional de manera que represente un significado especial o responda a una pregunta en particular (rousseau, 2006) Estas herramientas soportan el análisis interactivo de la información de resumen, soportando muchas tareas de agrupación de datos que no pueden realizarse empleando las facilidades básicas de agregación y agrupamiento .

La tecnología de Procesamiento Analítico en Línea –OLAP- (Online Analytical Processing) permite un uso más eficaz de los datawarehouses para el análisis de datos en línea, lo que proporciona respuestas rápidas a consultas analíticas complejas e iterativas utilizada generalmente para sistemas de ayuda para la toma de decisiones.

Primero y más importante, el OLAP presenta los datos a los usuarios a través de un modelo de datos intuitivo y natural. Con este estilo de navegación, los usuarios finales pueden ver y entender más efectivamente la información de sus bases de datos, permitiendo así a las organizaciones reconocer mejor el valor de sus datos. Las aplicaciones OLAP deberían proporcionar análisis rápidos de información multidimensional compartida. Las características principales del OLAP son:

- **Rápido:** proporciona la información al usuario a una velocidad constante. La mayoría de las peticiones se deben de responder al usuario en cinco segundos o menos.
- **Análisis:** realiza análisis estadísticos y numéricos básicos de los datos, predefinidos por el desarrollador de la aplicación o definido “ad hoc” por el usuario.
- **Compartida:** implementa los requerimientos de seguridad necesarios para compartir datos potencialmente confidenciales a través de una gran población de usuarios.
- **Multidimensional:** llena la característica esencial del OLAP, que es ver la información en determinadas vistas o dimensiones.

- **Información:** acceden a todos los datos y a la información necesaria y relevante para la aplicación, donde sea que ésta resida y no esté limitada por el volumen.

El OLAP es un componente clave en el proceso de almacenamiento de datos (data warehousing) y los servicios OLAP proporcionan la funcionalidad esencial para una gran variedad de aplicaciones que van desde reportes corporativos hasta soporte avanzado de decisiones.

Dentro de cada dimensión de un modelo de datos OLAP, los datos se pueden organizar en una jerarquía que represente niveles de detalle de los datos. Por ejemplo, dentro de la dimensión de tiempo, se puede tener estos niveles: años, meses y días; de manera similar, dentro de la dimensión geografía, Se puede tener estos niveles: país, región, estado/provincia y ciudad. Una instancia particular del modelo de datos OLAP tendrá valores para cada nivel en la jerarquía. Un usuario que vea datos OLAP se moverá entre estos niveles para ver información con mayor o menor detalle. (Kimball, 2013).

2.2.4. TIPOS DE SISTEMAS OLAP

Tradicionalmente, este sistema se clasifica según las Siguietes categorías:

- **ROLAP.** Implementación que almacena los datos en un motor relacional. Típicamente, los datos son detallados, evitando las agregaciones y las tablas se encuentran normalizadas.
- **MOLAP.** Esta implementación almacena los datos en una base de datos multidimensional. Para optimizar los tiempos de respuesta, el resumen de la información es usualmente calculado por adelantado.
- **HOLAP.** Almacena algunos datos en un motor relacional y otros en una base de datos multidimensional

2.2.5. DATAMART

Un Datamart es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. Un datamart puede ser alimentado desde los datos de un Datawarehouse, o integrar por sí mismo un compendio de distintas fuentes de información. Por tanto, para crear el datamart de un área funcional de la empresa es preciso encontrar la estructura óptima para el análisis de su información, estructura que puede estar montada sobre una base de datos OLTP, como el propio Datawarehouse, o sobre una base de datos OLAP. La designación de una u otra dependerá de los datos, los requisitos y las características específicas de cada departamento.

2.2.6. DATAWAREHOUSE

Es el proceso de extraer datos de distintas aplicaciones (internas y externas), para que una vez depurados y especialmente estructurados sean almacenados en un depósito de datos consolidado para el análisis del negocio. Requiere una combinación de metodologías, técnicas, hardware y los componentes de software que proporcionan en conjunto la infraestructura para soportar el proceso de información (robert stackowiak & greenwald, 2007) La estructura que se defina debe reflejar las necesidades y características del negocio, sus departamentos, equipos de trabajo y directivos, esto permitirá responder a interrogantes generados al tratar de tomar las decisiones (Ian H. witten, 2000) y con el tiempo se va convirtiendo en la memoria corporativa (Wang, 2009) ; describiendo el pasado y el presente de la empresa. Data Warehouse desglosa, resume, ordena y compara, pero no descubre, ni predice. (Flores, 2004) .

Para la construcción de un Data Warehouse se establecen tres etapas; la primera está dedicada a examinar el esquema Entidad Relación de la base de datos operacional, generando los esquemas multidimensionales candidatos.

La segunda etapa, consiste en recoger los requisitos de usuario por medio de entrevistas, para obtener información acerca de las necesidades de análisis de estos, y la tercera etapa, contrasta la información obtenida en la segunda etapa, con los esquemas multidimensional candidatos formados en la primera etapa

generando así, una solución que refleja los requisitos de usuario (sanchez, 2008)

Por otra parte implementar una solución de este tipo, ocasiona un costo que no todas las organizaciones están dispuestas a pagar (debido a sus capacidades de inversión), es por eso que los promotores del proyecto dentro de la empresa deben persuadir a los directivos y compañeros de trabajo, una buena alternativa de hacerlo es mediante el uso de técnicas administrativas, que permitan conocer a los directivos como se puede establecer el retorno de la inversión del proyecto equiparando inversión contra beneficios. (Arturo L, 2001)

DATAWAREHOUSE	BD TRANSACCIONALES
Está orientado a la toma de decisiones	Está orientado a los procesos relacionados con la operativa de la empresa
Almacena datos categorizándolos o estructurándolos de forma que favorezcan el análisis de datos y pueda proporcionar análisis históricos	Automatizan tareas operativas de la organización
Suelen ser intensivos en cálculos y escasos en entradas y salidas de información	Sus cálculos y procesos suelen ser simples y poco sofisticados
Explotación de los datos con herramientas data mining	Tienen la propiedad de ser recolectores de información
Sistemas de información interactivos y amigables , con altos estándares de diseño gráficos y visual	

Tabla 1. Diferencia entre Datawarehouse y bases de datos transaccionales

(Ecured, 2017)

2.2.7. PROCESOS DE EXTRACION TRANSFORMACION Y CARGAR (ETL)

(Power Data, 2016) Podemos definir la integración de datos como un proceso de transformación y conciliación de datos que permita una mayor agilidad en la gestión, proporcionando datos conectados, seguros y de calidad.

Integrar significa combinar datos que se encuentran en diferentes fuentes para permitirle al usuario final tener una vista unificada de los mismos para una accesibilidad idónea, que sirva a las necesidades de negocio.

Actualmente, la aparición de nuevas tecnologías y la explosión de datos plantean un gran desafío en este aspecto, y ello se traduce en una mayor complejidad técnica a la hora de implementar un plan de integración de datos, si bien el objetivo sigue siendo el mismo: evitar su fragmentación mediante el desarrollo de soluciones ad hoc. Las tecnologías de integración de datos permiten asumir proyectos que impliquen la transferencia de datos, transformaciones complejas de datos, el acceso a fuentes de datos múltiples, sistemas heterogéneos, con tiempos de latencia apropiados (batch, tiempo real) y minimizando los riesgos más frecuentes vinculados a este tipo de proyectos, entre otros:

- Tiempos de desarrollo excesivos
- Costes de mantenimiento altos
- Dificultades a la hora de responder a las necesidades empresariales en continuo cambio

Los proyectos de Data Integration comprenden desde el aprovisionamiento de datos para proyectos de Business Intelligence (BI) migraciones que requieren la transformación de los modelos de datos; hasta la sincronización de bases de datos o la consolidación de sistemas. En la actualidad, las empresas también se enfrentan al desafío de integrar datos desestructurados (Big Data) o datos en la nube las etapas del proceso ETL son las siguientes:

- **Extracción:** En este proceso se extraerán los datos desde múltiples fuentes y se los integra. Primero se identificarán las fuentes, luego se seleccionará las fuentes a usar, y finalmente se unen los datos de las fuentes para ser extraídos.
- **Transformación:** Conjunto de técnicas y herramientas para cambiar los datos extraídos de las diferentes fuentes, es decir quitar y añadir información, corregir si hay errores y uniformizar los datos (asignar un formato estándar); en otras palabras, refinarlo y agregarles calidad
- **Carga:** En este proceso se Suben los datos transformados al Data Warehouse o almacén de datos a través de métodos, ya que esto controlará la Actualización de datos de acuerdo al tiempo programado para Su elaboración. Finalizado el proceso de Carga se debería tener en el Data Warehouse datos de calidad, los cuales servirán de soporte para la toma de decisiones (sinnexus, 2014).

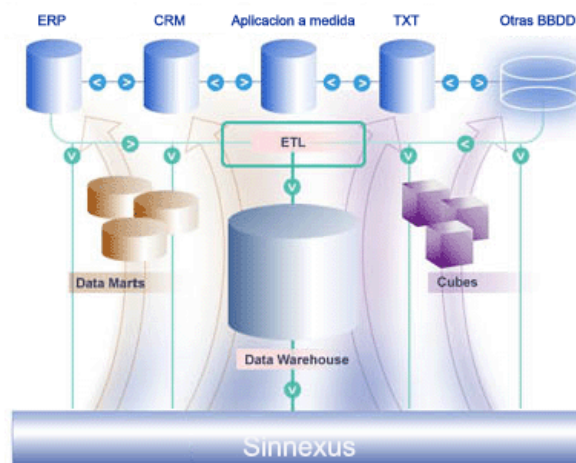


Ilustración 2 Proceso ETL

Fuente: (sinnexus, 2014)

2.2.8. ESQUEMA ESTRELLA PARA DATAWAREHOUSE

Consiste en estructurar la información en procesos, vistas y métricas recordando a una estrella. Es decir, tendremos una visión multidimensional de un proceso que medimos a través de unas métricas. A nivel de diseño, consiste en una tabla de hechos en el centro para el hecho objeto de análisis y una o varias tablas de dimensión (dimensión table) por cada dimensión de análisis que participa de la descripción de ese hecho. En la tabla de hecho encontramos los atributos destinados a medir (cuantificar) el hecho: sus métricas. Mientras, en las tablas de dimensión, los atributos se destinan a elementos de nivel (que representan los distintos niveles de las jerarquías de dimensión) y a atributos de dimensión (encargados de la descripción de estos elementos de nivel). En el esquema en estrella la tabla de hechos es la única tabla del esquema que tiene múltiples join que la conectan con otras tablas (foreign keys hacia otras tablas). El resto de tablas del esquema (tablas de dimensión) únicamente hacen join con esta tabla de hechos. Las tablas de dimensión se encuentran además totalmente desnormalizadas, es decir, toda la información referente a una dimensión se almacena en la misma tabla. (Curto, 2007)



Ilustración 3. Esquema estrella

Fuente: (IBM, 2017)

2.2.9. ESQUEMA COPO DE NIEVE PARA DATAWAREHOUSE

El esquema en copo de nieve es un esquema de representación derivado del esquema en estrella, en el que las tablas de dimensión se normalizan en múltiples tablas. Por esta razón, la tabla de hechos deja de ser la única tabla del esquema que se relaciona con otras tablas, y aparecen nuevas join gracias a que las dimensiones de análisis se representan ahora en tablas de dimensión normalizadas. En la estructura dimensional normalizada, la tabla que representa el nivel base de la dimensión es la que hace join directamente con la tabla de hechos. La diferencia entre ambos esquemas (star y snowflake) reside entonces en la estructura de las tablas de dimensión. Para conseguir un esquema en copo de nieve se ha de tomar un esquema en estrella y conservar la tabla de hechos, centrándose únicamente en el modelado de las tablas de dimensión, que, si bien en el esquema en estrella se encontraban totalmente de normalizadas, ahora se dividen en subtablas tras un proceso de normalización. Es posible distinguir dos tipos de esquemas en copo de nieve, un snowflake completo (en el que todas las tablas de dimensión en el esquema en estrella aparecen ahora normalizadas en el snowflake) o un snowflake parcial (sólo se lleva a cabo la normalización de algunas de ellas).

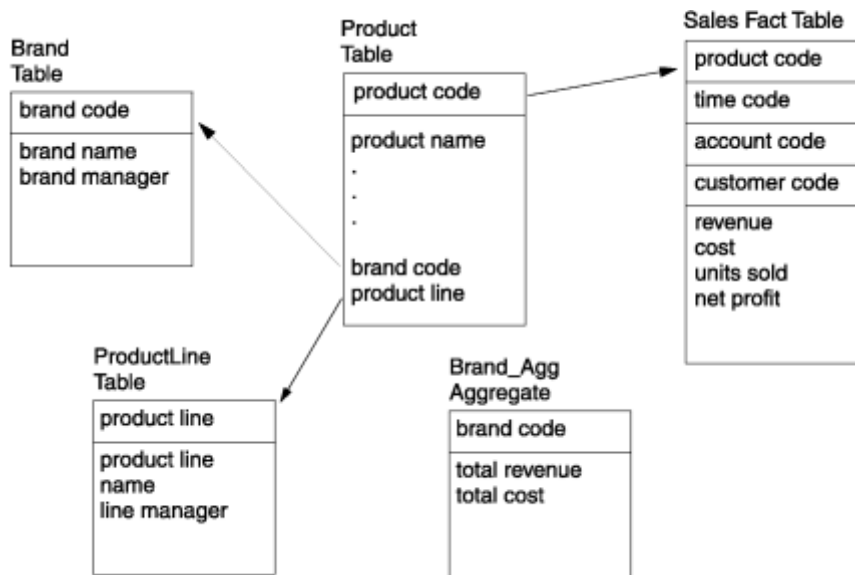


Ilustración 4 Esquema copo de nieve

Fuente: (IBM, 2017)

2.2.10. SISTEMAS DE SOPORTE A LAS DECISIONES

En la empresa no todas las decisiones son de carácter recurrente, sino que algunas se presentan muy pocas veces o incluso una sola vez. Los dss son instrumentos para abordar problemas de definición o estructuración menos precisa, y de carácter más esporádico. Los sistemas de apoyo a la decisión ayudan a los directivos que deben tomar decisiones no estructuradas. Una decisión se considera no estructurada si no existen procedimientos claros para tomarla y tampoco es posible identificar, con antelación, todos los factores que deben considerarse en la decisión. Hay que decir que todos los sistemas de información sirven de apoyo a la toma de decisiones, aunque sea de forma indirecta. Los dss son el tipo de sistemas de información desarrollados expresamente para servir de apoyo en el proceso de toma de decisiones. Estos sistemas facilitan un diálogo con el usuario que está considerando soluciones alternativas a un problema, y el sistema proporciona modelos construidos para la presentación de la información y acceso a bases de datos. Los sistemas de apoyo a la decisión son interactivos y su objetivo es la ampliación del razonamiento humano en la resolución de problemas particulares de toma de decisiones no estructuradas. Este tipo de sistemas se centra en los procesos de decisión y deberá proporcionar de forma fácil, rápida y exacta hechos importantes relacionados con la decisión a tomar y facilitando el acceso interactivo a medios de tratamiento que se utilizan creativamente y que permiten explorar las distintas posibilidades, suministrando las informaciones necesarias para responder a los problemas planteados. Al utilizar un dss, un directivo considerará un número posible de escenarios con el razonamiento: ¿Qué pasaría si? Por ejemplo, un directivo que estuviera estudiando fijar el precio de un nuevo producto puede utilizar un sistema de apoyo a la decisión en el área marketing. El sistema contendrá un modelo que relacione varios factores, tales como el precio del producto, el coste de los materiales, los gastos de promoción, que afecten a la estimación de beneficios de las ventas del producto para un período de unos cinco años. Variando el precio del producto en el modelo, el directivo puede comparar resultados previstos y, entonces, seleccionar un precio (alcami & Carlos Devece carañana, 2009).

El DSS es una de las herramientas más emblemáticas del Business Intelligence ya que, entre otras propiedades, permiten resolver gran parte de las limitaciones de

los programas de gestión. Estas son algunas de sus características principales:

- Informes dinámicos, flexibles e interactivos, de manera que el usuario no tenga que ceñirse a los listados predefinidos que se configuraron en el momento de la implantación, y que no siempre responden a sus dudas reales.
- No requiere conocimientos técnicos. Un usuario no técnico puede crear nuevos gráficos e informes y navegar entre ellos, haciendo drag&drop o drill through. Por tanto, para examinar la información disponible o crear nuevas métricas no es imprescindible buscar auxilio en el departamento de informática.
- Rapidez en el tiempo de respuesta, ya que la base de datos subyacente suele ser un Datawarehouse corporativo o un datamart, con modelos de datos en estrella o copo de nieve. Este tipo de bases de datos están optimizadas para el análisis de grandes volúmenes de información.
- Integración entre todos los sistemas/departamentos de la compañía. El proceso de ETL previo a la implantación de un Sistema de Soporte a la Decisión garantiza la calidad y la integración de los datos entre las diferentes unidades de la empresa. Existe lo que se llama: integridad referencial absoluta.
- Cada usuario dispone de información adecuada a su perfil. No se trata de que todo el mundo tenga acceso a toda la información, sino de que tenga acceso a la información que necesita para que su trabajo sea lo más eficiente posible.
- Disponibilidad de información histórica. En estos sistemas está a la orden del día comparar los datos actuales con información de otros períodos históricos de la compañía, con el fin de analizar tendencias, fijar la evolución de parámetros de negocio (Sinnexus, 2016).

2.2.11. SISTEMAS DE INFORMACION

Un conjunto de componentes interrelacionados que reúne (u obtiene), procesa, almacena y distribuye información para apoyar la toma de decisiones y el control en una organización (Laudon & Laudon, 2012).

2.2.11.1. FACTORES DE UN SISTEMA DE INFORMACION

Un sistema de información contiene información sobre una organización y su entorno. Dentro de este entorno figuran clientes, proveedores, competidores, accionistas y agencias reguladoras que interactúan con la organización y sus sistemas de información.



Ilustración 5. Factores de un sistemas de información

Fuente: (Laudon & Laudon, 2012)

2.2.11.2. TIPOS DE SISTEMAS DE INFORMACION

Los sistemas de información se desarrollan con diversos propósitos, según las necesidades de la empresa. Los sistemas de procesamiento de transacciones (TPS, Transaction Processing Systems) funcionan al nivel operativo de una organización, los sistemas de automatización de la oficina (OAS, Office Automation Systems) y los sistemas de trabajo del conocimiento (KWS, Knowledge Work Systems) apoyan el trabajo al nivel del conocimiento. Los sistemas de información gerencial (MIS, Management Information Systems) y los sistemas de apoyo a la toma de decisiones (DSS, Decisión Support Systems) se encuentran entre los sistemas de alto nivel. Los sistemas expertos aplican el conocimiento de los encargados de la toma de decisiones para solucionar problemas estructurados específicos. Los sistemas de apoyo a ejecutivos (ESS, Executive Support Systems) se encuentran en el nivel estratégico de la administración. Los sistemas de apoyo a la toma de para solucionar problemas estructurados específicos. Los sistemas de apoyo a ejecutivos (ESS, Executive Support Systems) se encuentran en el nivel estratégico de la administración. Los sistemas de apoyo a la toma de decisiones en grupo (GDSS, Group Decisión Support Systems) y los sistemas de trabajo corporativo apoyado por computadora (CSCWS, Computer-Supported Collaborative Work Systems), descrito de manera más general, auxilian la toma de decisiones semiestructuradas o no estructuradas a nivel de grupo. (Kendall & Kendall, 2005)

Los Tipos de sistemas de información:

- SISTEMAS DE PROCESAMIENTO DE TRANSACCIONES
- SISTEMAS DE AUTOMATIZACIÓN DE LA OFICINA Y SISTEMAS DE TRABAJO DEL CONOCIMIENTO
- SISTEMAS DE INFORMACIÓN GERENCIAL
- SISTEMAS DE APOYO A LA TOMA DE DECISIONES
- SISTEMAS EXPERTOS E INTELIGENCIA ARTIFICIAL
- SISTEMAS DE APOYO A LA TOMA DE DECISIONES EN GRUPO Y SISTEMAS DE TRABAJO COLABORATIVO APOYADOS POR COMPUTADORA
- SISTEMAS DE APOYO A EJECUTIVOS

Tipos de Sistemas de Información

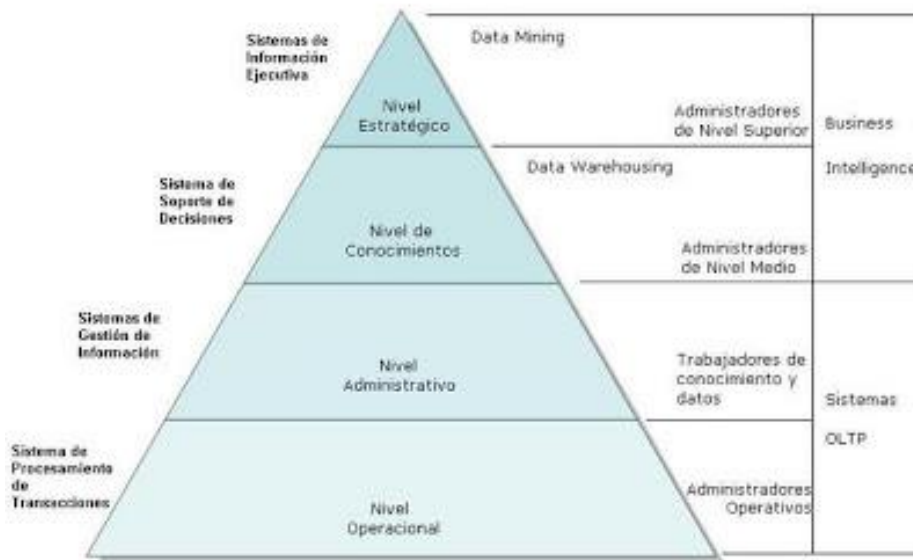


Ilustración 6. Piramide de los sistemas de información

Fuente: (Kendall & Kendall, 2005)

a. Sistemas de Procesamiento de Transacciones

Los sistemas de procesamiento de transacciones son sistemas de información encargados de procesar gran cantidad de transacciones rutinarias, es decir son todas aquellas que se realizan rutinariamente en la empresa entre estas tenemos el pago de nómina, facturación, entrega de mercancía y depósito de cheques. Estas transacciones varían de acuerdo al tipo de empresa.

Los sistemas de procesamiento de transacción o TPS (Transaction Procesation System) por sus siglas en inglés, eliminan el trabajo tedioso de las transacciones operacionales y como resultado reducen el tiempo que se empleaba en ejecutarlas actualmente, aunque los usuarios todavía deben alimentar de datos a los TPS.

"Los sistemas de procesamiento de transacciones son sistemas que traspasan sistemas y que permiten que la organización interactúe con ambientes externos. Debido a que los administradores consultan los datos generados por el TPS para información al minuto acerca de lo que está pasando en sus compañías, es esencial para las operaciones diarias que estos sistemas funcionen lentamente y sin interrupción". (Laudon & Laudon, 2012)

b. Sistemas de información de Gerencial

Los sistemas de información gerencial (MIS por sus siglas en ingles) no reemplazan a los sistemas de procesamiento de transacciones ni tampoco son los mismos, sino que estos sistemas incluyen procesamiento de transacciones. Los sistemas de información gerencial son sistemas de información computarizada que trabajan con la interacción entre usuarios y computadoras. Requieren que los usuarios, el software (programas de computadora) y el hardware (computadoras, impresoras, etc.) trabajen a un mismo ritmo. Los sistemas de información gerencial dan soporte a un aspecto más amplio de tareas organizacionales, a comparación de los sistemas de procesamiento de transacciones, los sistemas de información gerencial incluyen el análisis de decisiones y la toma decisiones.

"Para poder ligar la información, los usuarios de un sistema de información gerencial comparten una base de datos común. La base de datos guarda modelos que ayudan a los usuarios a interpretar y aplicar esos mismos datos. Los sistemas de información gerencial producen información que es usada en la toma de decisiones. Un sistema de información gerencial también puede llegar a unificar algunas de las funciones de información computarizada, aunque no exista como una estructura singular en ningún lugar del negocio". (Laudon & Laudon, 2012)

c. Sistemas de Apoyo a Decisiones

Los sistemas de apoyo a decisiones (DSS) ayudan a los gestores a tomar decisiones que son únicas, que cambia rápidamente, y no es fácil definirse de antemano. Ellos tratan de problemas que el procedimiento para llegar a una solución no puede ser plenamente predefinidos de antemano. Aunque el Departamento de Servicios Sociales uso de información interna de TPS y MIS, que a menudo traen consigo la información de fuentes externas, tales como los precios de las acciones o los precios de los productos de los competidores. Es evidente que, de diseño, DSS tienen más poder analítico que otros sistemas. Utilizan una gran variedad de modelos para analizar los datos, o se condensan grandes cantidades de datos en un formulario en el que puedan ser analizados por los encargados de adoptar decisiones. DSS están diseñados para que los usuarios puedan trabajar con ellos directamente, estos sistemas incluyen explícitamente el software de uso fácil. DSS son interactivos, el usuario puede cambiar las hipótesis, pedir nuevas preguntas, e incluir nuevos datos. (Laudon & Laudon, 2012)

d. Sistemas de Apoyo a Ejecutivos

Los Sistemas de Apoyo a Ejecutivos (ESS) son utilizados en el nivel estratégico de la organización. Los ESS no solo están diseñados para incorporar información sobre eventos externos, como las nuevas leyes fiscales o de los competidores, sino que también sacar un resumen de la información interna de los sistemas MIS y DSS. Estos sistemas pueden filtrar, comprimir, y realizar un seguimiento de datos críticos, mostrando los datos de la mayor importancia para los altos ejecutivos. Por ejemplo, el CEO de Leiner Health Products, el mayor fabricante de etiqueta privada vitaminas y suplementos en los Estados Unidos, tiene un ESS que ofrece en su escritorio, minuto a minuto la vista de la empresa: estados financieros, medido por capital de trabajo, cuentas por cobrar, cuentas por pagar, flujo de caja, e inventario.

Los Sistemas de Apoyo a Ejecutivos, emplean los más avanzados software de gráficos y puede presentar gráficos y datos de muchas fuentes. A menudo la información se entrega a los altos ejecutivos a través de un portal, que utiliza una interfaz Web integrada para presentar contenido personalizado de negocio de una variedad de fuentes.

A diferencia de los otros tipos de sistemas de información, los ESS no están diseñados principalmente para resolver problemas específicos. Aunque muchos DSS están diseñados para ser muy analítico, los ESS tienden a hacer menos uso de modelos analíticos. Los Sistemas de Apoyo a Ejecutivos ayudan a responder a las siguientes preguntas: ¿En qué negocios deberíamos estar? ¿Cuáles son nuestros competidores? ¿Qué nuevas adquisiciones deberíamos realizar? ¿Qué unidades debemos vender más para recaudar más utilidades? (Laudon & Laudon, 2012)

2.2.12. PROCESAMIENTO DE TRANSACCIONES EN LINEA (OLTP)

Es un tipo de sistema diseñado para admitir aplicaciones orientadas a transacciones. Los sistemas de OLTP se diseñan para responder inmediatamente a las solicitudes de usuario. Cada solicitud se considera una transacción única. Las solicitudes pueden implicar procesos de adición, recuperación, actualización o eliminación de datos”. (Symantec, 2016)

(Nima Ramos , 2014) OLTP es la sigla en inglés de Procesamiento de Transacciones En Línea (Online Transaction Processing) es un tipo de procesamiento que facilita y administra aplicaciones transaccionales, usualmente para entrada de datos y recuperación y procesamiento de transacciones (gestor transaccional). Los paquetes de software para OLTP se basan en la arquitectura cliente-servidor ya que suelen ser utilizados por empresas con una red informática distribuida. En muchas ocasiones, un almacén de datos se utiliza como el fundamento de un sistema de ayuda para la toma de decisiones. Los almacenes de datos se han diseñado para superar algunos de los problemas que una organización encuentra cuando intenta realizar un análisis estratégico mediante la misma base de datos que utiliza para realizar el proceso de transacciones en línea (OLTP).

Normalmente, los sistemas OLTP:

- ✓ Admiten el acceso simultáneo de muchos usuarios que agregan y modifican datos.
- ✓ Representan el estado, en cambio constante, de una organización, pero no guardan su historial.
- ✓ Contienen grandes cantidades de datos, incluidos los datos extensivos utilizados para comprobar transacciones.
- ✓ Tienen estructuras complejas.
- ✓ Se ajustan para dar respuesta a la actividad transaccional.
- ✓ Proporcionan la infraestructura tecnológica necesaria para admitir las operaciones diarias de la empresa.

A continuación se enumeran algunas de las dificultades que se suelen encontrar al utilizar bases de datos OLTP para realizar análisis en línea:

- ✓ Los analistas carecen de la experiencia técnica necesaria para crear consultas "ad hoc" contra la compleja estructura de datos.
- ✓ Las consultas analíticas que resumen grandes volúmenes de datos afectan negativamente a la capacidad del sistema para responder a las transacciones en línea.
- ✓ El rendimiento del sistema cuando está respondiendo a consultas analíticas complejas puede ser lento o impredecible, lo que causa un servicio poco eficiente a los usuarios del proceso analítico en línea.
- ✓ Los datos que se modifican con frecuencia interfieren en la coherencia de la información analítica.
- ✓ La seguridad se complica cuando se combina el análisis en línea con el proceso de transacciones en línea.

El almacén de datos proporciona una de las claves para resolver estos problemas al organizar los datos para analizarlos. Los almacenes de datos:

- ✓ Pueden combinar datos de orígenes heterogéneos en una única estructura homogénea.
- ✓ Organizan los datos en estructuras simplificadas buscando la eficiencia de las consultas analíticas más que del proceso de transacciones.
- ✓ Contienen datos transformados que son válidos, coherentes, consolidados y con el formato adecuado para realizar el análisis.
- ✓ Proporcionan datos estables que representan el historial de la empresa.
- ✓ Se actualizan periódicamente con datos adicionales, no con transacciones frecuentes.
- ✓ Simplifican los requisitos de seguridad.
- ✓ Proporcionan una base de datos organizada para OLAP, no para OLTP.

Un puesto de datos es un formato especial de almacén de datos que suele contener un subconjunto de datos de la empresa orientado a un tema determinado y apropiado para una función empresarial específica. (Nima Ramos , 2014)

2.2.13. CUBOS OLAP

En el mundo de las soluciones para Business Intelligence, una de las herramientas más utilizadas por las empresas son las aplicaciones OLAP (OLAP significa ‘On-Line Analytical Processing’), ya que las mismas han sido creadas en función a bases de datos multidimensionales, que permiten procesar grandes volúmenes de información, en campos bien definidos, y con un acceso inmediato a los datos para su consulta y posterior análisis. (Ecured, 2017).

El término Cubo OLAP, que nació de la mano de Edgar F. Codd, de la compañía EF Codd & Associates, sólo se utilizaban bases de datos relacionales para el proceso de la información, con sistemas tales como el ROLAP.

Gracias a la incorporación de las bases de datos de tipo multidimensional, y el nacimiento del nuevo concepto Cubo OLAP, las herramientas de soluciones para sistemas Business Intelligence han avanzado notablemente en cuanto a las prestaciones que estas aplicaciones brindan a las empresas, donde la información confiable, precisa y en el momento oportuno, son uno de los bienes más preciados (Ecured, 2017).

2.2.13.1. FUNCIONES

Las herramientas OLAP proporcionan a las compañías un sistema confiable para procesar datos que luego serán utilizados para llevar a cabo análisis e informes que permitan mejorar las operaciones productivas, tomar decisiones inteligentes y optimizar la competitividad en el mercado.

Para funcionar, las aplicaciones OLAP utilizan un tipo de base de datos que posee la peculiaridad de ser multidimensional, denominada comúnmente Cubo OLAP.

Básicamente, el Cubo OLAP, es una base de datos que posee diversas dimensiones, ampliando las posibilidades que hasta el momento ofrecían las conocidas hojas de cálculo.

Mediante la incorporación de estos vectores o cubos, se han ampliado las posibilidades de las bases de datos relacionales, permitiendo el procesamiento de importantes volúmenes de información, que de lo contrario sería imposible realizar.

Cada una de las dimensiones que posee la base de datos incorpora un campo determinado para un tipo de dato específico, que luego podrá ser comparado con la información contenida en el resto de dimensiones, para hacer posible

la evaluación y posteriores informes de la información realmente relevante para una compañía.

Una base de datos multidimensional puede contener varios cubos o vectores que extenderán las posibilidades del sistema OLAP con el cual se trabaja.

Por ello, si bien en general los sistemas OLAP suelen estar compuestos por tres dimensiones, lo cierto es que existe la posibilidad de que el sistema OLAP albergue más de tres dimensiones mediante la utilización de estos Cubos OLAP.

Para tener una idea más simple de la función de los Cubos OLAP dentro de una base de datos multidimensional, cabe destacar que cada una de las dimensiones o escalas del cubo corresponde básicamente a una jerarquía de datos (Ecured, 2017).

2.2.13.2. VENTAJAS DE LOS CUBOS OLAP

- Tiene acceso a grandes cantidades de información.
- Analiza las relaciones entre muchos tipos de elementos empresariales.
- Involucra datos agregados.
- Compara datos agregados a través de periodos jerárquicos.
- Presentan los datos en diferentes perspectivas.
- Involucran cálculos complejos entre elementos de datos.
- Pueden responder con rapidez a consultas de usuarios

2.2.13.3. DESVENTAJAS DE LOS CUBOS OLAP

A pesar de las grandes ventajas que presenta este tipo de base de datos multidimensional que incluye Cubos OLAP, la cual permite obtener mayor rapidez en las consultas y en el procesamiento de la información, lo cierto es que su gran falla reside en la imposibilidad de realizar cambios en su estructura.

Debido a su forma de funcionamiento y almacenamiento de la información, cuando los usuarios requieren realizar modificaciones en la estructura de este tipo de base de datos, deben rediseñar el Cubo OLAP, sin posibilidades de poder utilizar la estructura en la que se trabajó hasta el momento (Ecured, 2017).

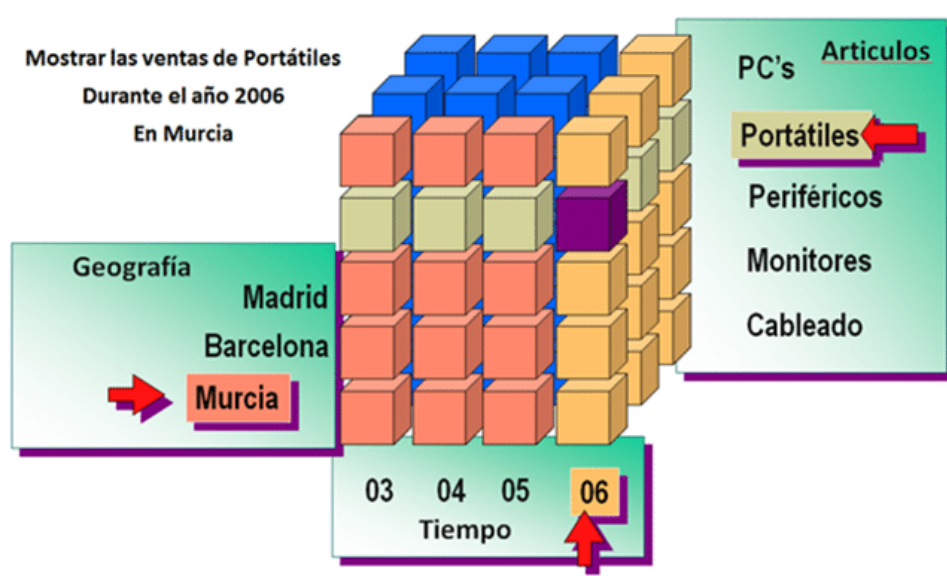


Ilustración 7. Cubo OLAP

Fuente: (Ramos, 2012)

2.2.13.4. DIMENSIONES

Una dimensión en SSAS hace referencia a una dimensión del almacenamiento de datos de Service Manager. En System Center 2012 – Service Manager, una dimensión es prácticamente equivalente a una clase de módulo de administración. Cada clase de módulo de administración tiene una lista de propiedades, mientras que cada dimensión contiene una lista de atributos, con cada atributo asignado a una propiedad en una clase. Las dimensiones permiten el filtrado, la agrupación y el etiquetado de datos. Por ejemplo, puede filtrar los equipos por el sistema operativo instalado y agrupar a los usuarios en categorías por sexo o edad. A continuación, los datos se presentan en un formato donde se clasifican de forma natural en estas jerarquías y categorías para permitir un análisis más exhaustivo. Las dimensiones también pueden tener jerarquías naturales para permitir a los usuarios “profundizar” hasta niveles más precisos de detalle. Por ejemplo, la dimensión Fecha tiene una jerarquía de la que se pueden obtener más detalles por Año, a continuación Trimestre, a continuación, Mes, a continuación, Semana y a continuación Día (microsoft, 2016).



Ilustración 8. Cubo OLAP y sus dimensiones

Fuente: (microsoft, 2016)

2.2.13.5. MEDIDAS

Las medidas son los valores numéricos que los usuarios desean reorganizar, agregar y analizar; son uno de los motivos fundamentales por los que desearía crear cubos OPAL mediante la infraestructura de almacenamiento de datos. Mediante el uso de SSAS, puede crear cubos OLAP que aplicarán reglas de negocio y cálculos para aplicar formato a las medidas y mostrarlas en un formato personalizable. Gran parte del tiempo empleado en el desarrollo de un cubo OLAP sirve para la determinación y la definición de las medidas que se mostrarán y de cómo se deben calcular.

Las medidas son valores que normalmente se asignan a columnas numéricas en una tabla de hechos del almacenamiento de datos, pero también se pueden crear en atributos de dimensión y de dimensión degenerada. Estas medidas son los valores más importantes de un cubo OLAP que se analizan y el interés principal de los usuarios finales que exploran el cubo OLAP. Las funciones de agregación se calculan en medidas para habilitar el análisis más profundo de los datos. La función de agregación más común es Suma. Una consulta común de cubo OLAP, por ejemplo, resume el tiempo total para todas las actividades In Progress. Otras funciones comunes de agregación son Mín, Máx y Cuenta.

Una vez procesados los datos sin procesar en un cubo OLAP, los usuarios pueden realizar cálculos y consultas más complejos mediante expresiones multidimensionales (MDX) para definir sus expresiones de medida o

miembros calculados propios. MDX es el estándar de la industria para consultar y obtener acceso a datos almacenados en sistemas OLAP. SQL Server no se diseñó para funcionar con el modelo de datos compatible con bases de datos multidimensionales (microsoft, 2016).

2.2.13.6. JERARQUIAS

Una jerarquía de cubos es un subconjunto de una jerarquía y se utiliza en un cubo. Una jerarquía de cubos hace referencia a la jerarquía de la que deriva (jerarquía padre) y puede tener un conjunto de niveles de cubo que son un subconjunto de los niveles padre de la jerarquía padre.

Una dimensión de cubo sólo puede tener una jerarquía de cubos. En general, una jerarquía de cubos tiene el mismo tipo de jerarquía y mecanismo de despliegue que la jerarquía de la que deriva. Si la jerarquía es de tipo de red, la jerarquía de cubos está equilibrada si no falta ningún miembro y es desigual si faltan miembros. Una jerarquía de cubos únicamente puede incluir el nivel Todos si la jerarquía padre incluye el nivel Todos.

En el Centro de OLAP, se crea un cubo en el contexto de un modelo de cubo utilizando el asistente para Cubo. Debe seleccionar cuál de las dimensiones del modelo del cubo desea tener en el cubo. Para cada dimensión que incluya como dimensión de cubo, puede seleccionar los niveles de cubo que se deben incluir en la jerarquía de cubos (IBM, 2017).

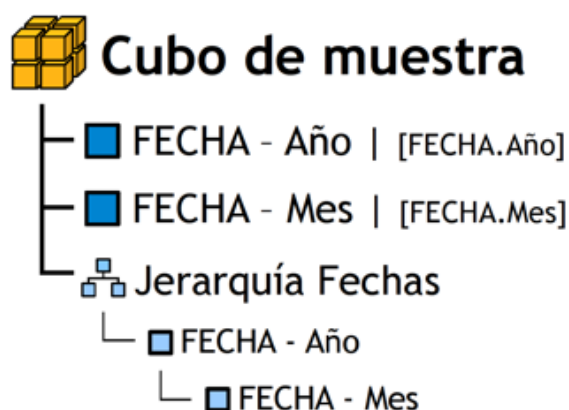


Ilustración 9. Jerarquías de dimensiones

Fuente: (Dataprix, 2009)

2.2.14. MICROSOFT SQL SERVER 2012

Con la aparición de **SQL Server 2012** el mundo de las Bases de datos está cambiando. Los desarrolladores ahora pueden ubicar su código apropiadamente en relación a su funcionalidad, acceder a datos nativos como XML, y construir sistemas complejos que sean manejados por el servidor de Bases de Datos. Estos puntos hacen que el desarrollo de Bases de Datos esté encaminado hacia una integración. **SQL Server 2012** es más que un sistema gestor de Bases de Datos ya que incluye múltiples componentes y servicios que la convierten en una plataforma de aplicaciones corporativas.

2.2.15. MYSQL SERVER

Es una herramienta visual unificada para los arquitectos de base de datos, desarrolladores y DBA'S. MySQL Workbench proporciona el modelado de datos, desarrollo de SQL y herramientas completas de administración de servidor de administración de configuración de usuario. Incluye el conjunto más completo de funciones avanzadas, herramientas de gestión y soporte técnico para alcanzar los niveles más altos de escalabilidad, seguridad, fiabilidad y tiempo de actividad de MySQL. Reduce el riesgo, el costo y la complejidad en el desarrollo, implementación y administración de aplicaciones MySQL críticas para el negocio.

2.2.16. PENTAHO BI

Pentaho es una herramienta de Business Intelligence desarrollada bajo la filosofía del software libre para la gestión y toma de decisiones empresariales. Es una plataforma compuesta de diferentes programas que satisfacen los requisitos de BI. Ofreciendo soluciones para la gestión y análisis de la información, incluyendo el análisis multidimensional OLAP, presentación de informes, minería de datos y creación de cuadros de mando para el usuario. La plataforma ha sido desarrollada bajo el lenguaje de programación Java y tiene un ambiente de implementación también basado en Java, haciendo así que Pentaho sea una solución muy flexible al cubrir una alta gama de necesidades empresariales (gravitar, 2017).

2.2.16.1. CARACTERISTICAS DE PENTAHO BI

Pentaho Business Intelligence abarca las siguientes áreas de reporte:

- **Pentaho reporting:** Pentaho Reporting es un potente generador de informes: Permite la distribución de los resultados del análisis en múltiples formatos.
- **Pentaho Data Integration:** Muchas organizaciones tienen información disponible en aplicaciones y base de datos separados. Pentaho Data Integration abre, limpia e integra esta valiosa información y la pone en manos del usuario. Provee una consistencia, una sola versión de todos los recursos de información, que es uno de los más grandes desafíos para las organizaciones TI hoy en día. Pentaho Data Integration permite una poderosa ETL (Extracción, Transformación y Carga). El uso de kettle permite evitar grandes cargas de trabajo manual frecuentemente difícil de mantener y de desplegar.
- **Pentaho BI Plataforma:** La plataforma de Pentaho BI provee de servicios críticos incluyendo programación, seguridad, integración automatización y flujo de trabajo. Proporcionando habilidades a los usuarios finales de Pentaho y proveniente un lugar central para administrar y mantener el despliegue de la empresa BI (pentaho, 2017)
 - **Integración con procesos de negocio.**
 - **Administra y programa reportes.**
 - **Administra seguridad de usuarios.**

- **Componentes de Pentaho**

- ✓ Plataforma 100% J2EE, asegurando la escalabilidad, integración y portabilidad.
- ✓ Servidor: puede correr en servidores compatibles con J2EE como JBOSS AS, WebSphere, Tomcat, WebLogic y Oracle AS.
- ✓ Base de datos: vía JDBC, IBM DB2, Microsoft SQL Server, MySQL, Oracle, PostgreSQL, NCR Teradata, Firebird.
- ✓ Sistema operativo: no hay dependencia. Lenguaje interpretado.
- ✓ Lenguaje de programación: Java, JavaScript, JSP, XSL (XSLT/XPath/XSL-FO).
- ✓ Interfaz de desarrollo: Java SWT, Eclipse, Web-based.
- ✓ Repositorio de datos basado en XML.
- ✓ Todos los componentes están expuestos vía Web Services para facilitar la integración con Arquitecturas Orientadas a Servicios (SOA). (pentaho, 2017)

2.2.17. QCLICKVIEW

El producto que distribuye QlikTech se llama Qlikview y destaca por su sencillez de uso y por ser muy visual. A diferencia de otros proveedores, dispone de un único producto principal que no aspira a cubrir todas las necesidades de Business Intelligence. No compiten en el mercado del reporting, por ejemplo. Sin embargo, se trata de un producto muy interesante que puede cubrir muchas necesidades empresariales (o departamentales), desde un cuadro de mando, hasta una solución analítica más general. Las aplicaciones de Qlikview suelen ser muy rápidas, ofrecen una experiencia de usuario muy positiva y tienen un interfaz de usuario atractivo. Para conseguir estos tiempos de respuesta tan buenos, emplean una "nube de datos" residente en memoria. No emplean un modelo relacionan tradicional ni los clásicos cubos. Su tecnología, dicen, es diferente. Le llaman "tecnología asociativa" y es una especie de base de datos basadas en columnas donde cada "dato" se almacena una única vez. Sea lo que sea, obtienen un buen rendimiento para un volumen de datos moderado. (facil, 2009)

- ✓ **Servicio de datos y Plataformas** : Qlikview está disponible en varias versiones desde la Desktop para mono puestos, hasta la versión server que

permite centralizar y servir datos a usuarios que acceden a estos desde diferentes plataformas Windows, Mac, Linux incluso desde dispositivos móviles como Ipad, iPhone o Smartphone con Android. (Martínez, 2012)

- ✓ **Modelado de datos e integración:** Qlikview dispone de su propio lenguaje de modelado de datos. No voy a negar que tenga su complejidad, pero con sus asistentes, integrar datos es cuestión de minutos. Los datos a diferencia de otras plataformas BI se integran con el propio cuadro de mandos, permitiendo que este trabaje desconectado de los orígenes de datos. (Martínez, 2012)
- ✓ **Búsqueda Asociativa:** El punto fuerte de Qlikview sin duda es la presentación de datos. Posee gran cantidad de objetos: Velocímetros, Gráficos de barras, de Tarta... Además los gráficos son interactivos y permiten hacer Drill Down, para profundizar en el análisis y concretar en la selección de datos. (Martínez, 2012)

2.3.METODOLOGIA PARA EL DESARROLLO DEL PROYECTO

Existen diversas tecnologías y metodologías para el desarrollo de un BI entre las que encontramos:

2.3.1. Metodología Inmon

Esta metodología fue desarrollada por Bill Inmon en el año 1992 en su libro "Building the data Warehouse". Para esta metodología un Data Warehouse es considerado como un conjunto de datos orientados por temas, integrados, variantes en el tiempo y no volátiles, cuyo objetivo es ser un apoyo importante para la toma de decisiones estratégicas. El Data Warehouse es una parte de un sistema de BI y los Data Mart obtienen su información a partir de este Data Warehouse y para ser almacenada la información es necesario normalizarla antes de almacenarse.

Debido a las grandes cargas de datos que debe realizarse, Inmon considera que el ambiente de origen de los datos y el ambiente de acceso de datos deben estar físicamente separados en diferentes bases de datos y en equipos separados, y los datos deben ser cuidadosamente gestionados y condensados. En esta metodología los Data Mart son considerados como capas del Data Warehouse y los Data Mart son dependientes del depósito central de datos de Data Warehouse Corporativo y por lo tanto se construyen después de él (Inmon, 2002).

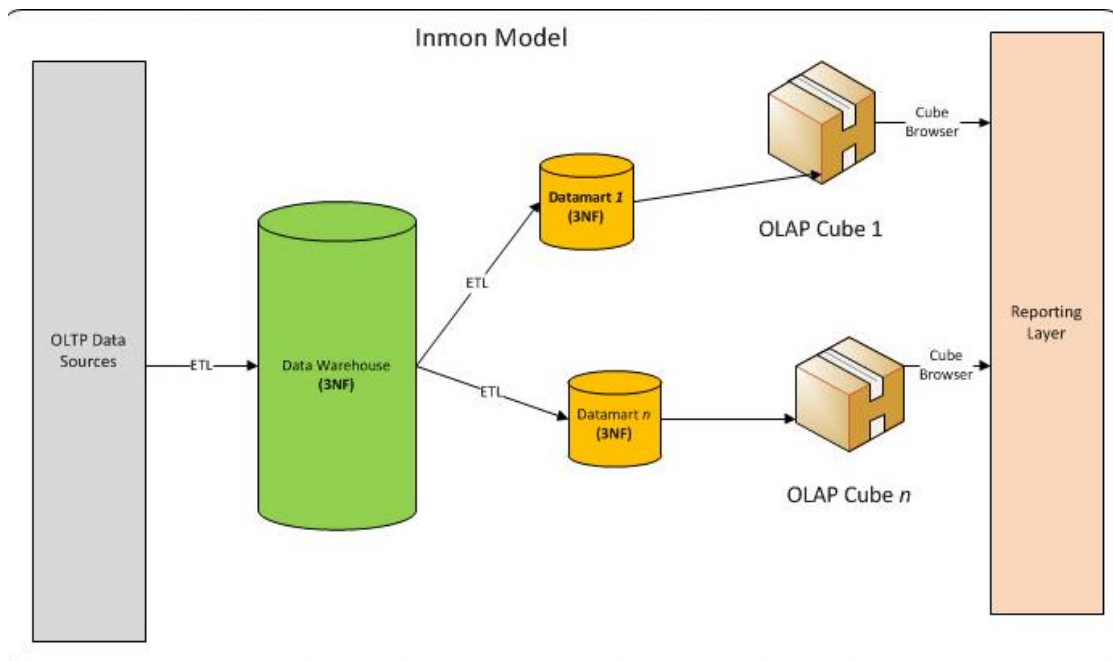


Ilustración 10.Arquitectura Inmon

Fuente: (austin, 2010)

2.3.2. Metodología de kimball

En esta metodología del DW se forma del conjunto de todos los Datamart generados existentes en una empresa. La información siempre se almacena en un modelo dimensional. Otra forma de denominar esta aproximación es como “Bottom-up”.

En la metodología de kimball un Data Warehouse se considera como una copia de los datos transaccionales para realizar la consulta y el análisis de los mismos. La metodología de Kimball, también conocida como modelo dimensional, está basado en el ciclo de vida dimensional del negocio.

Para almacenar los datos. Las estructuras de estos requieren una presentación dimensional. En el modelo Dimensional se constituyen modelos de tablas y relaciones con el propósito de optimizar la toma de decisiones , basándose en la sentencia SQL realizadas en una base de datos relacional que están ligadas con la medición o un conjunto de mediciones de los resultados de los procesos del negocio.

Este modelo dimensional, tiene como objetivo presentar los datos dentro de un marco de trabajo estándar e intuitivo, para permitir su acceso con un alto rendimiento, cada modelo dimensional está compuesta por una tabla llave

combinada, llamada tabla de hechos, y un conjunto de tablas más pequeñas denominadas dimensiones, los elementos de estas tablas se pueden definir de la siguiente manera

- **Hechos:** es una colección de piezas de datos y datos de contexto. cada hecho representa una parte del negocio , una transacción o un evento,
- **Dimensiones:** es una colección de miembros, unidades o individuos del mismo tipo.
- **Medidas:** son atributos numéricos de un hecho que representan el comportamiento del negocio relativo a una dimensión.

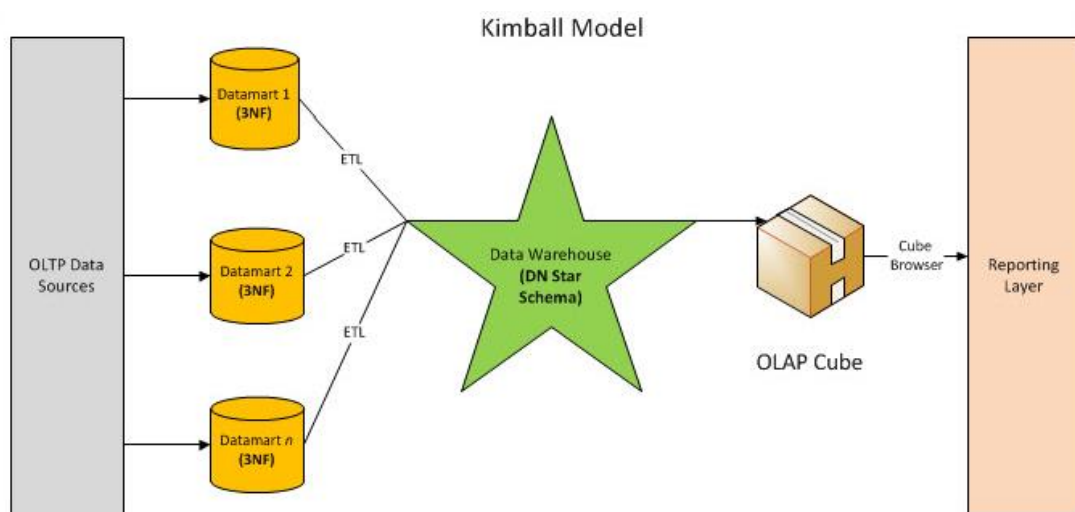


Ilustración 11.Arquitectura de kimball

Fuente: (austin, 2010)

2.3.3. Tipos de Arquitectura

Top-Down: con esta arquitectura se define inicialmente DW y después de realiza el desarrollo, y se construyen y cargan los DM.

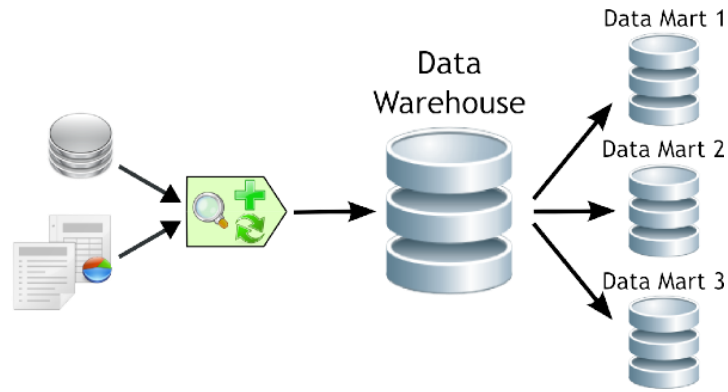


Ilustración 12.Arquitectura Top Down

Fuente: (Dataprix, 2009)

Bottom-up: con este tipo de arquitectura, se definen previamente los datamart o un Dw, de acuerdo al alcance empresarial y la magnitud de datos y fuentes de análisis.

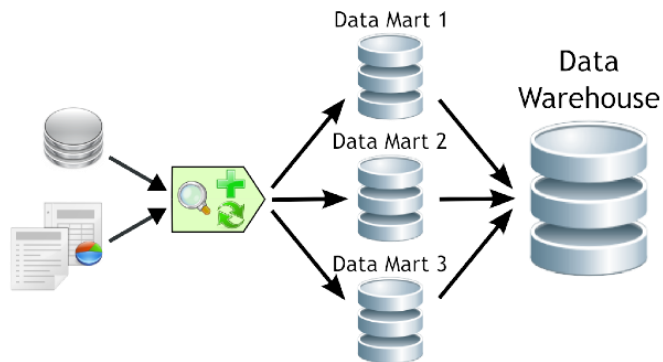


Ilustración 13.Arquitectura Bottom up

Fuente: (Dataprix, 2009)

Híbrida: con este tipo de arquitectura se puede realizar el desarrollo de un DM o un DW, de acuerdo al alcance empresarial y la magnitud de datos y fuentes de análisis.

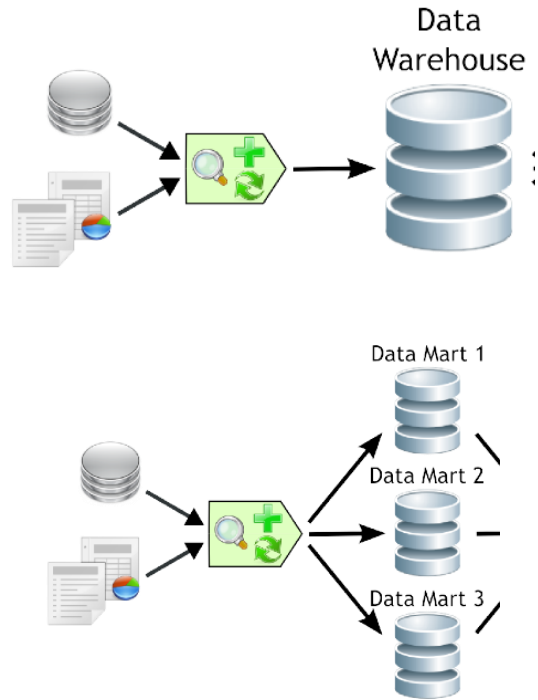


Ilustración 14. Arquitectura Híbrida

Fuente: (Dataprix, 2009)

2.3.4. Comparación de metodologías

NOMBRE	INMON	KIMBALL	HEFESTO
METODOLOGÍA			
AUTOR	BILL INMON	RALPH KIMBALL	ING. BERNABEU .RICARDO DARIO
ARQUITECTURA	TOP DOWN	BOTTOM UP	HIBRIDA
ENFASIS	DATA WAREHOUSE	DATA MART	DATA WAREHOUSE Y DATAMART
	LOS PROBLEMAS QUE SE VAN A RESOLVER YA SON CONOCIDOS DE ANTEMANO , TIENEN UN MAYOR COSTO YA QUE CADA FASE ALCANZA UN GRAN NIVEL DE DETALLE	ESTA BASADA EN EXPERIMENTOS Y PROTOTIPOS , ES MAS FLEXIBLE Y MENOS COSTOSA	SE ENFOCA EN EL ANALISIS DE REQUERIMIENTOS PARA IDENTIFICAR INDICADORES , PERSPECTIVAS Y REALIZAR EL ANALISIS DE LOS DATOS

Tabla 2. Comparación de metodologías

Fuente: Propia

2.3.5. Selección de la metodología

Se ha seleccionado la Metodología Hefesto, debido a que las fases que esta comprende , se acopla al desarrollo y está enfocada al análisis de los requerimientos , identificación de indicadores y perspectivas , que permitan realizar el análisis de datos respectivos y obtener resultados que apoyen a la toma de decisiones de la organización.

2.4.METODOLOGIA HEFESTO

HEFESTO es una metodología propia, cuya propuesta está fundamentada en una muy amplia investigación, comparación de metodologías existentes, experiencias propias en procesos de confección de almacenes de datos. Cabe destacar que HEFESTO está en continua evolución, y se han tenido en cuenta, como gran valor agregado, todos los feedbacks que han aportado quienes han utilizado esta metodología en diversos países y con diversos fines. La idea principal, es comprender cada paso que se realizará, para no caer en el tedio de tener que seguir un método al pie de la letra sin saber exactamente qué se está haciendo, ni por qué. La construcción e implementación de un DW puede adaptarse muy bien a cualquier ciclo de vida de desarrollo de software, con la salvedad de que, para algunas fases en particular, las acciones que se han de realizar serán muy diferentes. Lo que se debe tener muy en cuenta, es no entrar en la utilización de metodologías que requieran fases extensas de reunión de requerimientos y análisis, fases de desarrollo monolítico que conlleve demasiado tiempo y fases de despliegue muy largas. Lo que se busca, es entregar una primera implementación que satisfaga una parte de las necesidades, para demostrar las ventajas del DW y motivar a los usuarios.

La metodología HEFESTO, puede ser embebida en cualquier ciclo de vida que cumpla con la condición antes declarada. (bernabeu, 2010)

2.4.1. Fases de la metodología Hefesto

2.4.1.1. Análisis de los requerimientos

- **Identificar preguntas :** El primer paso comienza con el acopio de las necesidades de información, el cual puede llevarse a cabo a través de muy variadas y diferentes técnicas, cada una de las cuales poseen características inherentes y específicas, como por ejemplo entrevistas, cuestionarios, observaciones, etc. El análisis de los requerimientos de los diferentes usuarios, es el punto de partida de esta metodología, ya que ellos son los que deben, en cierto modo, guiar la investigación hacia un desarrollo que refleje claramente lo que se espera del depósito de datos, en relación a sus funciones y cualidades. El objetivo principal de esta fase, es la de obtener e identificar las necesidades de información clave de alto nivel, que es esencial para llevar a cabo las metas y estrategias de la empresa, y que facilitará una eficaz y eficiente toma de decisiones (Ecured, 2017).
- **Identificar indicadores y perspectivas de análisis:** Una vez que se han establecido las preguntas claves, se debe proceder a su descomposición para descubrir los indicadores que se utilizarán y las perspectivas de análisis que intervendrán. Para ello, se debe tener en cuenta que los indicadores, para que sean realmente efectivos son, en general, valores numéricos y representan lo que se desea analizar concretamente, por ejemplo: saldos, promedios, cantidades, sumatorias, fórmulas, etc. En cambio, las perspectivas se refieren a los objetos mediante los cuales se quiere examinar los indicadores, con el fin de responder a las preguntas planteadas, por ejemplo: clientes, proveedores, sucursales, países, productos, rubros, etc. Cabe destacar, que el Tiempo es muy comúnmente una perspectiva.
- **Modelo Conceptual:** En esta etapa, se construirá un modelo conceptual a partir de los indicadores y perspectivas obtenidas en el paso anterior. A través de este modelo, se podrá observar con claridad cuáles son los alcances del proyecto, para luego poder trabajar sobre ellos, además al poseer un alto nivel de definición de los datos, permite que pueda ser presentado ante los usuarios y explicado con facilidad.

2.4.1.2. Análisis de los OLTP

- **Establecer correspondencias con los requerimientos:** Establecer correspondencias con los requerimientos. El objetivo de este análisis, es el de examinar los OLTP disponibles que contengan la información requerida, como así también sus características, para poder identificar las correspondencias entre el modelo conceptual y las fuentes de datos. En el caso de los indicadores, deben explicitarse como se calcularán, y más aún si son fórmulas u operaciones complejas. La idea es, que todos los elementos del modelo conceptual estén correspondidos en los OLTP.
- **Seleccionar los campos que integrarán cada perspectiva. Nivel de granularidad:** Una vez que se han establecido las relaciones con los OLTP, se examinarán y seleccionarán los campos que contendrá cada perspectiva, ya que será a través de estos por los que se manipularán y filtrarán los indicadores. Para ello, basándose en las correspondencias establecidas en el paso anterior, se debe presentar al usuario los datos de análisis disponibles para cada perspectiva. Es muy importante conocer en detalle que significa cada campo y/o valor de los datos encontrados en los OLTP, por lo cual, es conveniente investigar su sentido, ya sea a través de diccionarios de datos, reuniones con los encargados del sistema, análisis de los datos propiamente dichos, etc.

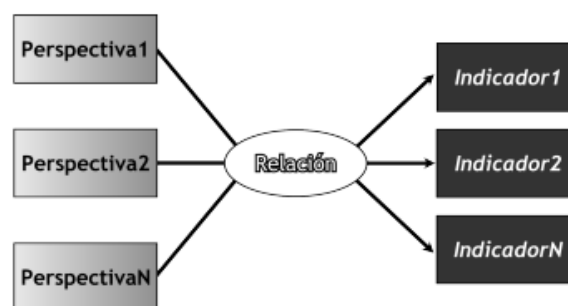


Ilustración 15. Modelo conceptual

Fuente: (bernabeu, 2010)

2.4.1.3.Elaboración del Modelo lógico DW

- **Diseñar tablas de dimensiones:** (Ecured, 2017) Este paso, se aplicará por igual a todos los tipos de esquemas lógicos. Lo primero que se hará será crear las dimensiones del mismo, para ello se tomará cada perspectiva con sus atributos relacionados y se les realizará el siguiente proceso:
 - ✓ Se elegirá un nombre que identifique la dimensión.
 - ✓ Se añadirá un campo que represente su clave principal.
 - ✓ Se redefinirán los nombres de los atributos si es que no son lo bastante explicativos.
- **Diseñar tablas de hechos:** (Ecured, 2017)En este paso, se definirán las tablas de hechos, que son las que contendrán los indicadores de estudio. Para los esquemas en estrella y copo de nieve, se realizará lo siguiente:
 - ✓ Al igual que las dimensiones, se le deberá asignar un nombre a la tabla de hechos que en este caso represente la información analizada, área de investigación, negocio enfocado, etc.
 - ✓ Se definirá su clave primaria, que se compone de la combinación de las claves primarias de cada dimensión que se utilizará para generar las consultas.
 - ✓ Se renombrarán los hechos o indicadores si es que no llegasen a ser lo suficientemente explícitos.
- **Realizar uniones:** Para los tres tipos de esquemas, se realizarán las uniones correspondientes entre sus tablas de dimensiones y sus tablas de hechos.
- **Determinar jerarquías :** Para los esquemas en estrella y constelación, se deberán especificar las jerarquías que existirán dentro de cada tabla de dimensión, teniendo siempre presente cual es el objetivo de las mismas. Para representar las jerarquías en el modelo lógico, se deberán colocar los atributos pertenecientes a las jerarquías en sus respectivas tablas, en orden descendente y acompañado con un número ordinal encerrado entre corchetes (Ecured, 2017).

2.4.1.4. Procesos ETL

Una vez construido el modelo lógico, se deberá proceder a probarlo con datos, a través de procesos ETL. Para realizar la compleja actividad de extraer datos de diferentes fuentes, para luego integrarlos, filtrarlos y depurarlos, existen varios software que facilitan estas tareas, por lo cual este paso se centrará solo en la generación de las sentencias SQL que contendrán los datos que serán de interés. Antes de realizar la carga de datos, es conveniente efectuar una limpieza de los mismos, para evitar valores faltantes y anómalos. En la cláusula ORDER BY de las sentencias SQL, que se efectuarán para cargar cada tabla, deben figurar los atributos, medidas y claves en orden de aparición de sus respectivas tablas. Al realizar esta acción, se logrará aportar mayor eficiencia cuando se realizan búsquedas de datos. (bernabeu, 2010)

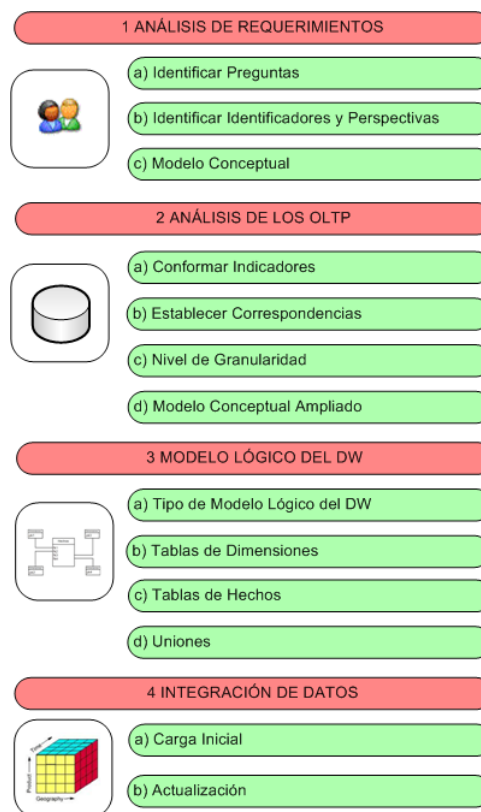


Ilustración 16. Fases de la metodología Hefesto

Fuente: (Ecured, 2017)

3. MATERIALES Y METODOS

3.1. Material

3.1.1. Población

Todo Personal involucrado en el proceso de ventas y logística de la ferretería “Solano”

3.1.2. Muestra

Al tratarse de una investigación cuasi experimental, se seleccionará solo a los tomadores de decisiones.

3.1.3. Unidad de análisis

Área de ventas de la ferretería “Solano”.

3.2. Método

3.2.1. Tipo de investigación

Aplicada.

3.2.2. Diseño de Investigación

Se realizará un diseño cuasi-experimental en el cual se evaluará la eficiencia de la nueva solución (BI) para el soporte en la toma de decisiones del área de ventas.

Se realizarán observaciones antes y después del tratamiento.

Diagrama de investigación cuasi-experimental	G -> O1 -> X -> O2
G (Grupo a investigar)	Tomadores de decisión en la Empresa (02 personas)
X (Tratamiento)	Aplicación del BI
O (Observación)	O1: Observación pre-test

Diagrama de investigación cuasi-experimental	G -> O1 -> X -> O2
	O2: Observación post-test

Tabla 3. Diagrama de investigación

3.2.3. Variables de estudio y Operacionalización

- ✓ Independiente (VI): Solución de Inteligencia de Negocios usando la metodología Hefesto y Pentaho BI
- ✓ Dependiente (VD): Indicadores que dan soporte a la toma de decisiones en el Área de ventas de la Ferretería Solano

Variable	Dimensión	Indicador	Unidad de medida	Instrumento de Investigación
VI	Tiempo	Tiempo en obtener información para tomar decisiones	Horas	Hoja de captura de datos
	Costo	Costo en la toma de decisiones	Soles	Hoja de costos
VD	Oportunidad	Tiempo para analizar información a través de indicadores	Horas	Hoja de captura de datos
	Ahorro	Costo de Servicio Futuro	Soles	Hoja de costos

Tabla 4. variables de estudio y Operacionalización

3.2.4. Instrumentos de recolección de Datos

- Búsqueda de información:
 - Ficha bibliográfica.
 - Hoja resumen.
- Entrevista:
 - Guion de entrevista.
 - Grabador de audio.
 - Hoja resumen.
- Observación:
 - Guion de Observación.

3.2.5. Procedimientos y análisis de datos

3.2.5.1. Procesamiento de datos

Una vez la puesta en marcha el cuestionario los datos serán ingresados en el Ms Excel para la realización de análisis descriptivo. Las respuestas que se obtengan se presentarán en tablas y gráficos dinámicos, dando a conocer como es los resultados numéricos y porcentuales de las variables en estudio según los objetivos planteados anteriormente.

La presentación de reportes se hace a través de la herramienta de Qlikview.

3.2.5.2. Análisis de datos

El análisis de los datos se llevará a cabo por medio de cuadros estadísticos descriptivos (Pruebas hipótesis nula y alternativa y las Pruebas t de Student), ya que la información obtenida será analizada y mostrada por medio de cuadros y gráficos.

3.2.6. Modelos estadísticos de análisis de datos

Como la cantidad de registros para la prueba es *menor a 30*, se aplicará la *Distribución t de Student*, en caso contrario se aplicará la *Distribución Z (normal)*. En ambos casos, se seguirá el siguiente proceso:

1. Establecer la Hipótesis Nula y la Hipótesis Alternativa

Hipótesis Nula (H_0)	Hipótesis Alternativa (H_A)	Expresión
La solución de Inteligencia de Negocios usando la metodología Hefesto y Pentaho BI en el Área de Ventas de la Ferretería Solano permite mejorar el tiempo en el análisis de la información generando indicadores que den soporte a la toma de decisiones.	La solución de Inteligencia de Negocios usando la metodología Hefesto y Pentaho BI en el Área de Ventas de la Ferretería Solano no permite mejorar el tiempo en el análisis de la información generando indicadores que den soporte a la toma de decisiones.	$H_0: \mu_A - \mu_D \geq 0$ $H_0: \mu_A - \mu_D < 0$
La solución de Inteligencia de Negocios usando la metodología Hefesto y Pentaho BI en el Área de Ventas de la Ferretería Solano mejora el costo en el análisis de la información generando indicadores que den soporte a la toma de decisiones.	La solución de Inteligencia de Negocios usando la metodología Hefesto y Pentaho BI en el Área de Ventas de la Ferretería Solano no mejora el costo en el análisis de la información generando indicadores que den soporte a la toma de decisiones.	

Tabla 5. Hipotesis nula y alternativas

2. Establecer el Nivel de Significancia:

Es la posibilidad de aceptar H_0 cuando en realidad es falsa y ha sido definido en 95%, cuando se calculó la muestra.

Por lo anterior, el Nivel de Significancia (α) para la prueba de la hipótesis es del 5% y el Nivel de Confianza ($1 - \alpha$) será del 95%

3. Determinar el valor crítico de las tablas:

t_t : Valor Crítico z_t : Valor Crítico

El valor crítico de T de Student se obtiene buscando en Tablas $t(1 - \alpha)(n-1)$, siendo n el valor de la muestra y (n-1) los grados de libertad.

4. Determinar la Región de Aceptación:

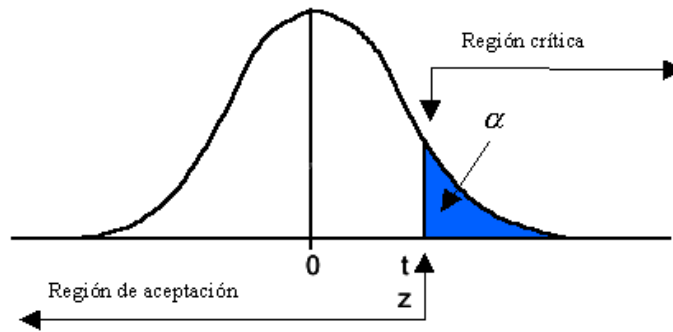


Ilustración 17. Región de aceptación de la hipótesis

Fuente: Propia

5. Calcular el valor estadístico de la prueba:

Se seguirá el siguiente procedimiento:

- Tabular los valores obtenidos antes y después en la siguiente tabla:

N°	Pregunta	Antes			Después		
		X_{Ai}	$X_{Ai} - \bar{X}_A$	$(X_{Ai} - \bar{X}_A)^2$	X_{Di}	$X_{Di} - \bar{X}_D$	$(X_{Di} - \bar{X}_D)^2$
1							
2							
...							
n-1							
n							
Suma							
Promedio							

Tabla 6. Tablas de datos

Dónde:

- n: Cantidad de Preguntas de la Encuesta.
- Pregunta: Enunciado de la misma.
- i: Número de Pregunta.
- Suma: Es la suma de los valores de cada columna.
- Promedio: Es el promedio de los valores de cada columna.
- x_{Ai} : Respuestas de Pregunta i antes del experimento.

- \bar{x}_A : Promedio de Respuestas antes del experimento.
 - $x_{Ai} - \bar{x}_A$: Diferencia entre la Respuesta de Pregunta i menos Promedio de Respuestas de Preguntas antes del experimento.
 - x_{Di} : Respuestas de Pregunta i después del experimento.
 - \bar{x}_D : Promedio de Respuestas después del experimento.
 - $x_{Di} - \bar{x}_D$: Respuesta de Pregunta i menos Promedio de Respuestas de Preguntas después del experimento.
- Calcular las Varianzas Muestrales usando las expresiones de la tabla siguiente:

	Antes	Después
Varianza Muestral	$\sigma_A^2 = \frac{\sum_{i=1}^n (x_{Ai} - \bar{x}_A)^2}{n - 1}$	$\sigma_D^2 = \frac{\sum_{i=1}^n (x_{Di} - \bar{x}_D)^2}{n - 1}$

Tabla 7. tablas Muestrales

- Calcular el Valor Estadístico de la Prueba usando la expresión:

$$Z_c = \frac{\bar{x}_A - \bar{x}_D}{\sqrt{\left(\frac{\sigma_A^2}{n_A} + \frac{\sigma_D^2}{n_D}\right)}} \quad (1)$$

- Comparar ambos valores (de tablas y calculado) para redactar la Conclusión de la Prueba Estadística.
- Tabular los valores obtenidos antes y después en la siguiente

N°	Pregunta	Respuestas				Puntaje Pregunta (PP)	Puntaje Promedio Pregunta (PPP)
		MI	I	PI	NI		
		3	2	1	0		
1							
2							
...							
N							
	Suma						
	Promedio						

Tabla 8. tabla datos ponderado

Dónde:

- n: Cantidad de Preguntas de la Encuesta.
- i: Número de Pregunta.
- Respuestas: Escala de Licker de cualificación y peso (P_j), que será determinada durante la investigación.
- R_{ij} : Respuesta j de Pregunta i.
- Puntaje de la Pregunta i (PP_i): $PP_i = \sum_{j=1}^4 (P_j * R_{ij})$
- Puntaje Promedio de la Pregunta i: $PPP_i = \frac{PP_i}{n}$

- Luego se llena la tabla siguiente:

Pregunta	Puntaje Promedio		DP_i	DP_i^2
	Pre-Prueba	Post-Prueba		
	PPP_{Ai}	PPP_{Di}		
1				
2				
...				
N				
Suma				
Promedio				

Tabla 9. Tabla resumen

Dónde:

- n: Cantidad de Preguntas.
- i: Número de Pregunta.
- Suma: Es la suma de los valores de cada columna.

- Promedio: Es el promedio de los valores de cada columna.
 - PPP_{Ai} : Puntaje Promedio de la pregunta i antes del experimento.
 - PPP_{Di} : Puntaje Promedio de la pregunta i después del experimento.
 - DP_i : Diferencia de Puntajes Promedio $i = PPP_{Ai} - PPP_{Di}$
 - DP_i^2 : Cuadrado de la Diferencia de Puntajes Promedio $i = (PPP_{Ai} - PPP_{Di})^2$
- Calcular el Valor Estadístico de la Prueba usando la expresión

$$S^2 = \frac{n \sum_{i=1}^n D_i^2 - (\sum_{i=1}^n D_i)^2}{n(n-1)} \quad (2)$$

- Calcular el Valor Estadístico de la Prueba usando la expresión:

$$t_c = \frac{\bar{D}\sqrt{n}}{\sqrt{S^2}} \quad (3)$$

- Comparar ambos valores (de tablas y calculado) para redactar la Conclusión de la Prueba Estadística.

6. Decisión:

Si $t_0, z_0 \in$ Región crítica \rightarrow Rechazamos $H_0: \mu_A - \mu_D \geq 0$

Y Aceptamos $H_0: \mu_A - \mu_D < 0$

- a. **Interpretación de los datos:** Se analizan e interpretan los datos para corroborar o rechazar la investigación.

<i>Indicador</i>	<i>Antes</i>		<i>Después</i>		<i>Valor tabla</i>	<i>Valor cálculo</i>	<i>Conclusión</i>
	\bar{X}_A	σ_A^2	\bar{X}_D	σ_D^2			
							H ₀ se acepta
							H ₀ se acepta

Tabla 10.tabla resultados

4. RESULTADOS

4.1. Definición del proyecto

Para el desarrollo del proyecto se determinó desarrollar una solución BI para dar soporte a las necesidades del negocio y a sus requerimientos, con el fin de facilitar el control y la toma de decisiones por parte del gerente administrador de la empresa que requieren mantener monitoreadas las actividades de compras, ventas y almacén del negocio verificando el funcionamiento y los cambios o soluciones que la empresa requiere como tal , de esta manera se generara independencia y facilidad en la manipulación de información para que los usuarios finales no tengan inconvenientes ni pérdidas de tiempo al querer realizar algún reporte requeridos por ellos .

4.2. Justificación , Objetivos y alcance

La definición, los objetivos y el alcance están definidos en la parte de introducción del presente proyecto, mismo que enfatiza en la solución del problema que ya se definió.

La solución se focaliza en la generación de indicadores que permitan a través de esquemas de reportes visualizar: tiempos, cantidades, totales, proveedores etc. Lo cual facilitara tener una mejor perspectiva del negocio y de las decisiones y planes que se pueden tomar a partir del nuevo conocimiento generado por los indicadores, reduciendo la carga de trabajo para el usuario final y optimizando la nueva información que se va a obtener.

4.3. El negocio

Ferretería Solano S.A.C, es una empresa que tiene más de 30 años en el mercado, liderando en ventas de materiales de Construcción así como también en Herramientas de Carpintería, Productos para pintores entre otros. Trabajando en el abastecimiento de Constructoras y de otras Ferreterías. Ferretería Solano S.A.C cuenta con dos sucursales las cuales la especificaremos en la ilustración 17 y 18 Ferretería Solano S.A.C cuenta con una Gerencia que se encarga de dirigir los diferentes Procesos que se llevan a cabo así como también de la Toma de Decisiones, de una Área Contable que se encarga de llevar el control de los Ingresos y Egresos que tiene la Ferretería, Área de Ventas que se encarga de ver el inventario diario de los productos vendidos así como también las ganancias, Área de Compras que se encarga de los Pedidos y Abastecimiento de Productos provenientes de los diferentes Proveedores Locales y Nacionales y de una Área de Almacén que lleva un registro del Movimiento de los Productos así como también como el Stock de estos.

MISION: Dar servicio de calidad y satisfacer las necesidades del mercado.

VISION: Consolidarse como una Empresa Líder en la Región Norte.

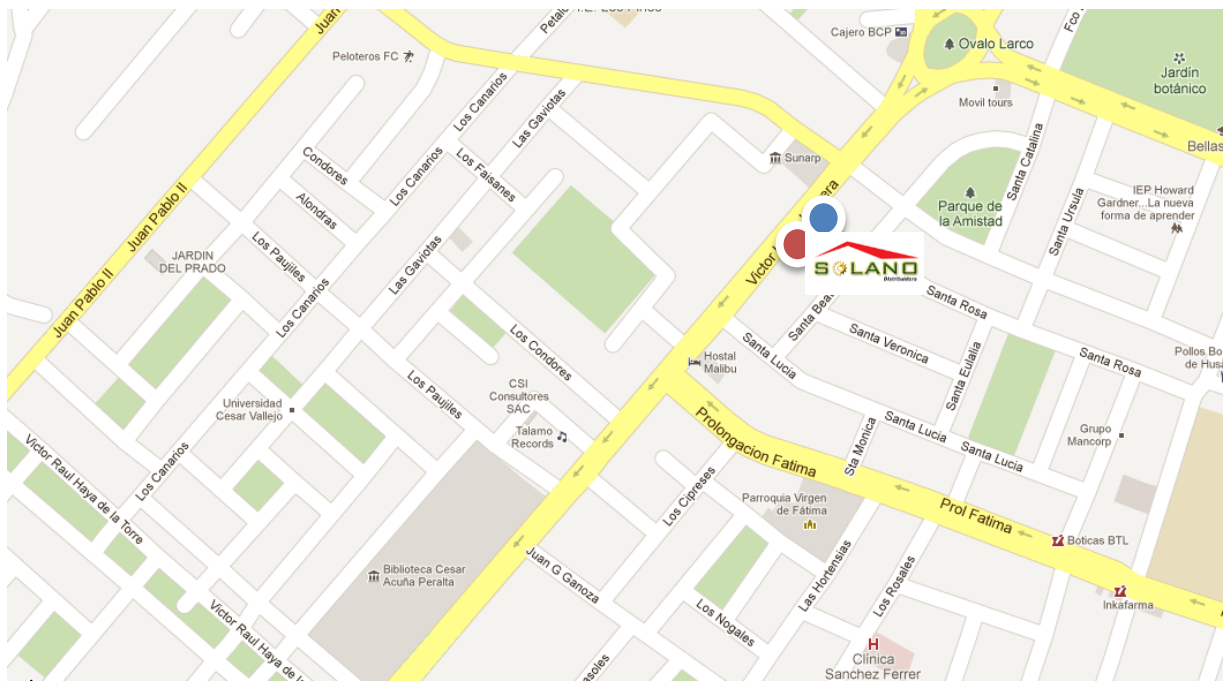


Ilustración 18.ubicacion geográfica de la ferretería solano

Fuente: google maps

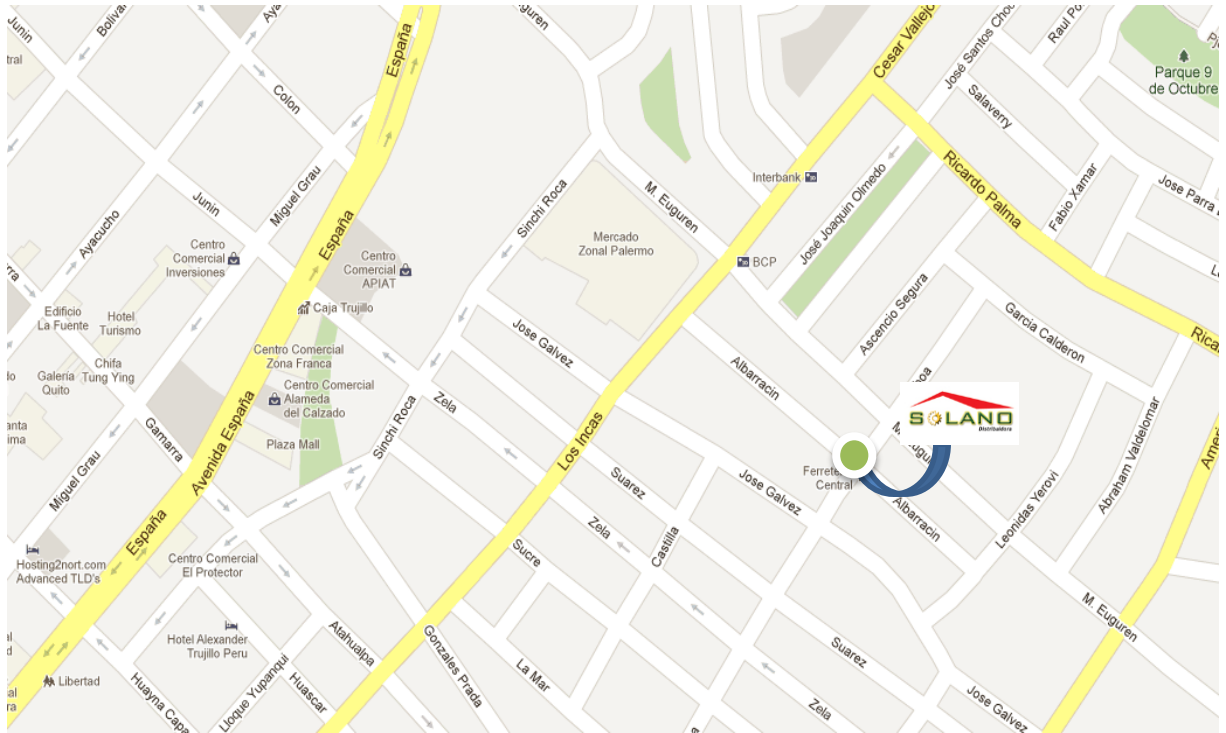


Ilustración 19.ubicacion geográfica de la ferreteria solano sucursal 02

Fuente: google maps

4.4. Procesos del área del negocio

4.4.1. Descripción del proceso del área de ventas

El cliente se dirige a la ferretería, pide los productos que necesita al vendedor, se verifica en el sistema los productos solicitados en el sistema con su stock la lista de productos con los datos del cliente son mandados a caja , el cajero pregunta qué tipo de pago es si en efectivo o crédito si es en efectivo el cliente paga y se le entrega la boleta con su producto y su cambio y si es en crédito el cajero le solicita las cuotas en que quiere pagarlo y le brinda su boleta y su producto.

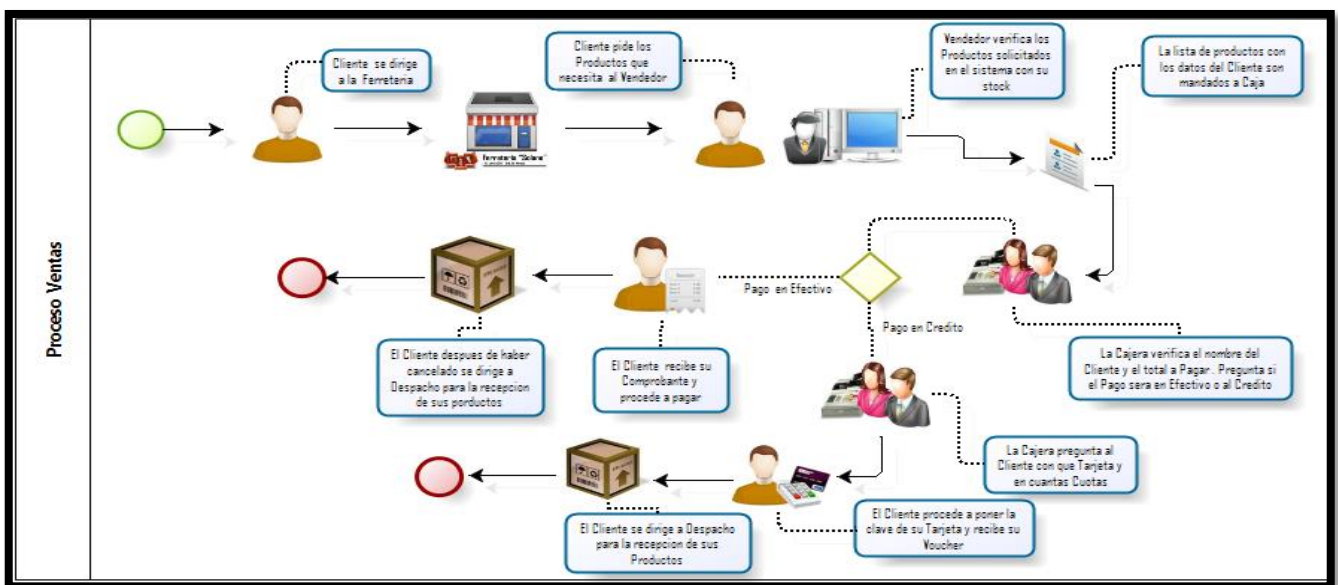


Ilustración 20. Pictograma de ventas de ferretería solano

Fuente: propia

4.4.2. Descripción del proceso de almacén

El proveedor llega a la ferretería y hace entrega de los productos que se la han solicitados hace la entrega de la guía de remisión de los productos a los almaceneros los almaceneros ingresan los nuevos productos al sistema al mismo tiempo es actualizado el sistema de ventas automáticamente por la base de datos actualizada.

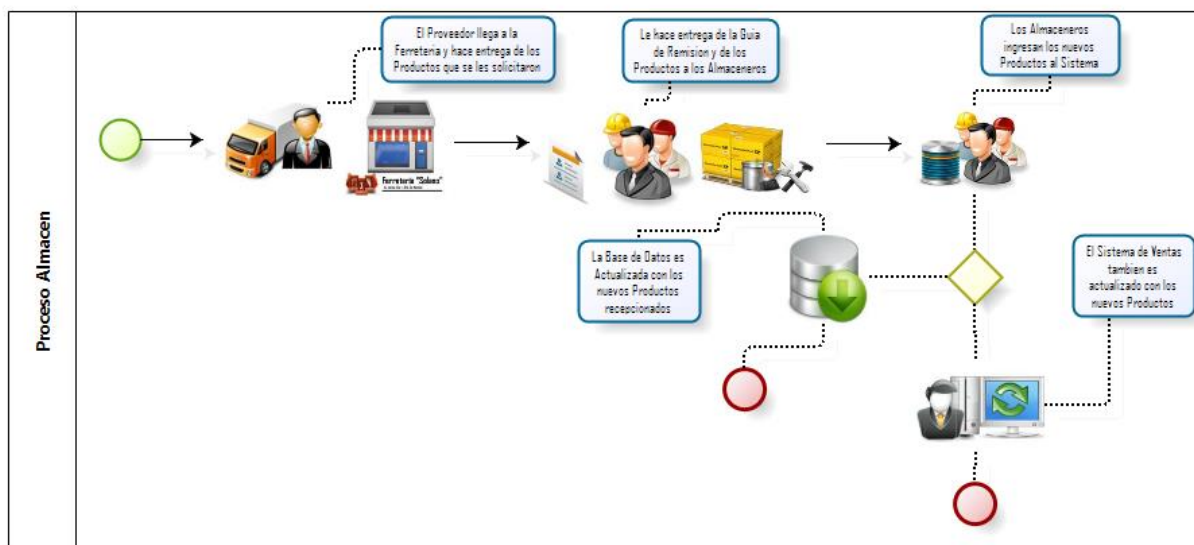


Ilustración 21. Pictograma del proceso de almacén

Fuente: propia

4.4.3. Descripción del proceso de compras

Los compradores de la ferretería solicitan la visita de los proveedores, llegan los proveedores con la diferentes cotizaciones de los productos solicitados por los compradores, después del análisis de los precios se llega a un acuerdo con los proveedores los compradores elaboran una guía de los productos que se decepcionaran y se la entregan al proveedor, los almaceneros reciben el producto y los compradores verifican los productos recepcionados y firman la guía de remisión.

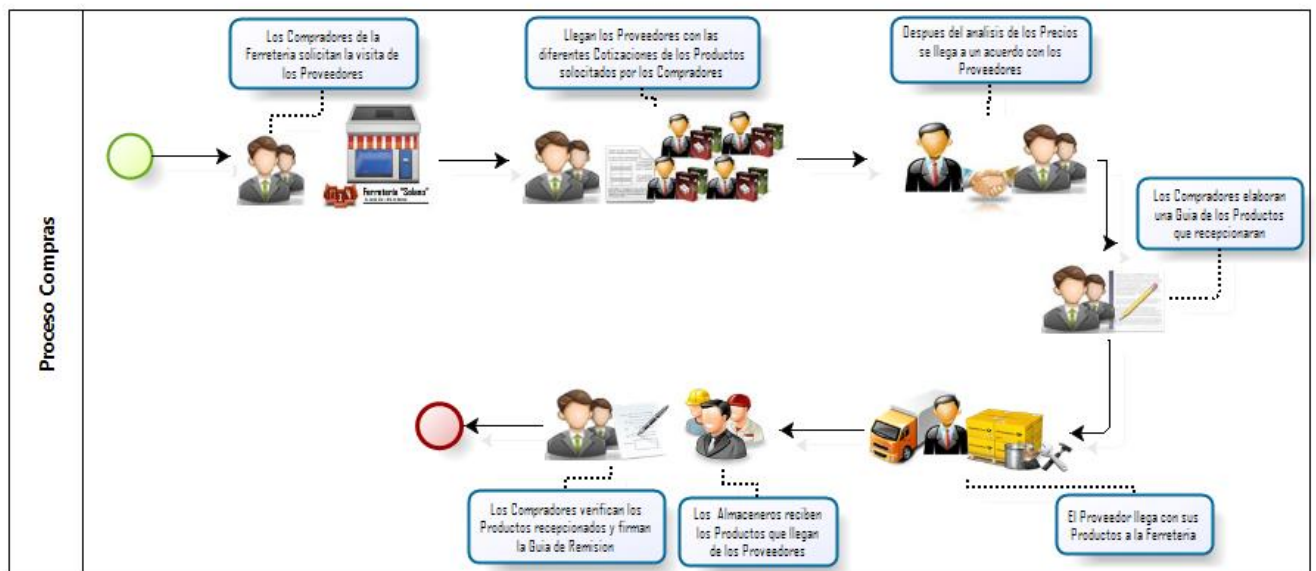


Ilustración 22. Pictograma del proceso de compras

Fuente: propia

4.5. Estrategia y método de la implementación

Se realizara la implementación de una aplicación de Business Intelligence, para la generación de indicadores que apoye a la toma de decisiones del área de ventas de la ferretería solano S.A.C.

La implementación de la aplicación BI con la herramienta Pentaho Community permitirá realizar la integración de datos así como también los cubos dimensionales para poder generar los orígenes de datos que permitirán el desarrollo de los reportes que se realizaran con la herramienta QCLICKVIEW.

Se implementara un cubo para poder realizar varias combinaciones y otorgar la información que el usuario requiera.

Se utilizara la metodología Hefesto V2.0, para realizar la implementación de la aplicación BI, que es una metodología, propia para el trabajo con soluciones bajo la suite de Pentaho, se tomó la decisión de utilizar Hefesto con el fin de que no se requieran fases extensas de reunión de requerimientos y análisis sin afectar la calidad del producto.

4.5.1. Situación actual del proceso de generación de reportes

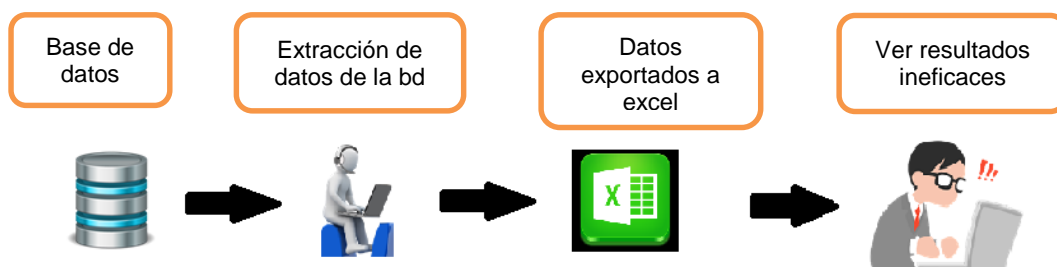


Ilustración 23. Proceso actual de generación de reportes de todas las áreas de la ferretería solano

Fuente: propia

4.6. Orígenes de datos

Para poder realizar la aplicación BI se utilizara la base de datos transaccional de los sistemas de la ferretería “solano” que engloba todas las áreas necesarias para poder realizar el trabajo de estudio, la base de datos de la ferretería solano denominada “BDFerreteriaSolano” encuentra implementada en el gestor de base de datos SQL server la cual cuenta con más de 40 tablas normalizadas que sirven de repositorio para las transacciones que se realizan en el negocio diariamente, de donde se extraerán los datos necesarios para la elaboración de la solución BI.

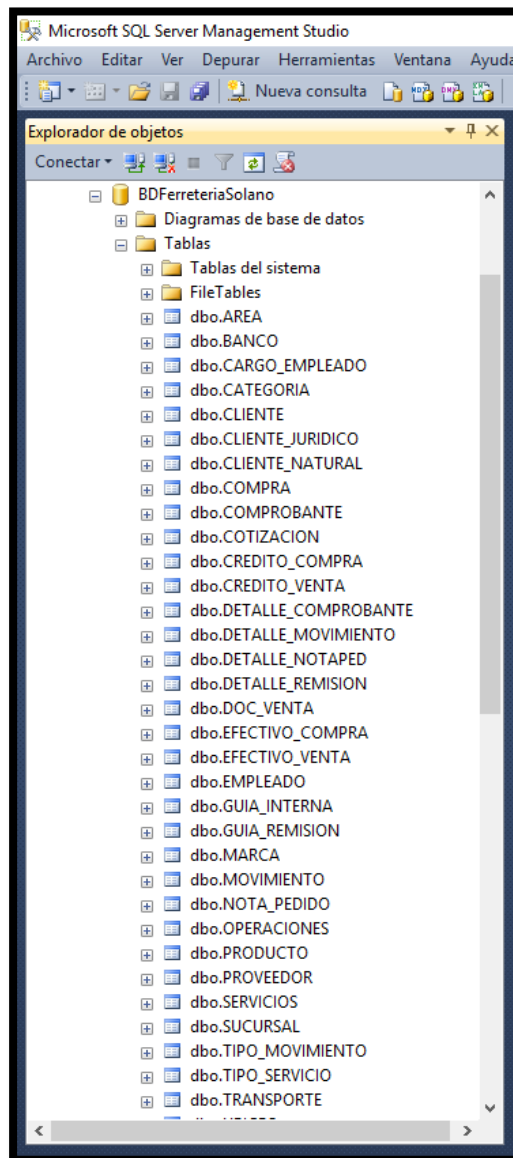


Ilustración 24. Estructura de la base de datos origen BDFerreteriaSolano

4.7. Análisis de los requerimientos

Se realizaron reuniones con el personal de ventas de la ferretería con el jefe de almacén y los compradores para poder determinar las necesidades y requerimientos para poder definir qué información es la que se necesita obtener en el tiempo real una vez realizado el proyecto .

Después de haber analizado la problemática y definido la información que se necesita obtener se decidió realizar las siguientes tareas:

- Analizar los Procesos que se llevan a cabo dentro de la Empresa, así como la documentación que utilizan para estos.
- Analizar el Sistema Transaccional de la Empresa.
- Reunir los Requerimientos de las diferentes Áreas de la Empresa.
- Analizar los Requerimientos.
- Realizar el diseño y la arquitectura para el DataWareHouse.
- Implementar el DataWareHouse.
- Implementar la Interfaz de Usuario.

4.7.1. Requerimientos

Con las entrevistas realizadas y las reuniones que se realizaron con los trabajadores de la empresa e implicados en los distintos procesos se establecieron los siguientes requerimientos que involucraran el área de ventas, almacén y compras.

NRO	REQUERIMIENTOS	INDICADOR	DIMENSIONES CANDIDATAS
1	Cuantos Productos han sido comprados y sus Proveedores en un Tiempo determinado	Cant productos	Proveedor Tiempo
2	Conocer los Precio total de pedidos por <u>Proveedores</u> en un tiempo determinado	Monto productos	Material Producto Tiempo
3	Resumen de cantidad de Compras en las Diferentes Sucursales en un mes determinado	Cant Compras	Sucursal Tiempo
4	Resumen de los Proveedores frecuentes en un mes determinado	Cant proveedores	Proveedor tiempo
5	Conocer cuántos Pedidos de productos se hacen un mes determinado	Cant Productos	Pedido Producto tiempo
6	Conocer los egresos en las diferentes Sucursales en una determinado semestre	Monto Egresos	Sucursal tiempo
7	Conocer los Productos más vendidos en un tiempo determinado	Productos	Movimiento Productos tiempo
8	Conocer los Productos Recibidos en distintas sucursales en un tiempo determinado	Cant productos	Producto sucursal Tiempo
9	Conocer la Categoría de producto y su demanda en un trimestre determinado	Cant Productos	Producto tiempo

10	Conocer las condiciones de los Productos en un mes determinado	Cant Productos	Producto tiempo
11	Conocer la cantidad de Productos en que entraron en las diferentes sucursales en un tiempo determinado	Cant productos	Sucursal Producto tiempo
12	Resumen de los Productos en buen estado en un tiempo determinado	Productos	Producto tiempo
13	Conocer el record de Ventas de cada Empleado en las diferentes Sucursales	Monto Total ventas	Empleado Sucursal
14	Conocer cantidad de productos vendidos en cada sucursal por trimestre y año	Cant productos	Tiempo
15	Conocer la cantidad de productos compradas por clientes más frecuentes en un trimestre determinado	Cant productos	Cliente productos tiempo
16	Conocer los Productos comprados por Cliente en un mes determinado	Cant Productos	Producto Cliente tiempo
17	Conocer el Resumen de Ventas en un determinado mes en las diferentes sucursales	Monto total ventas	Tiempo sucursal
18	Reporte del promedio de ventas por sucursal en tiempos determinados	Promedio ventas	Productos Sucursal tiempo

Tabla 11. tabla de requerimientos funcionales de la ferretería solano

Fuente: propia

4.8. Identificar Indicadores y Perspectivas de Análisis

4.8.1. Indicadores

- Cantidad de productos vendidos.
- Cantidad de movimientos de productos de entrada y salida realizados
- Sumatoria de las ventas realizadas.
- Record de ventas de los empleados.
- Totales de egresos e ingresos por cada sucursal.
- Resumen de ventas por sucursales.
- Resumen de productos vendidos en un semestre determinado.
- Total de compras realizadas por sucursal y tiempo.
- Cantidad de existencias de productos en almacén.
- Total de productos en buen estado y mal estado.

4.8.2. Perspectivas

- Año
- Mes
- Trimestre
- Productos
- Sucursales
- Movimientos entrada
- Movimientos salida
- Compras.
- Ventas.
- Clientes

4.9. Modelo conceptual

4.9.1. Diagrama de perspectivas por requerimiento

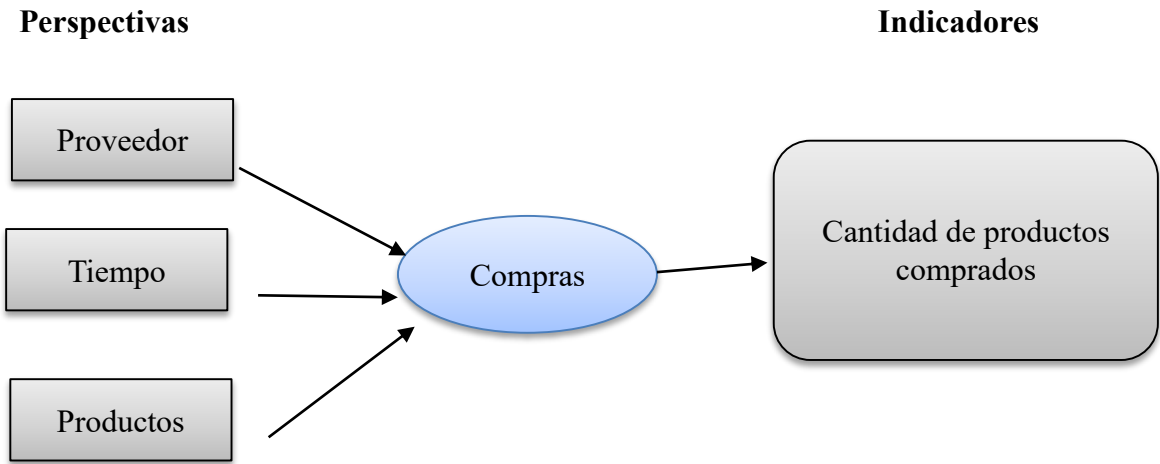


Ilustración 25. requerimiento N°1

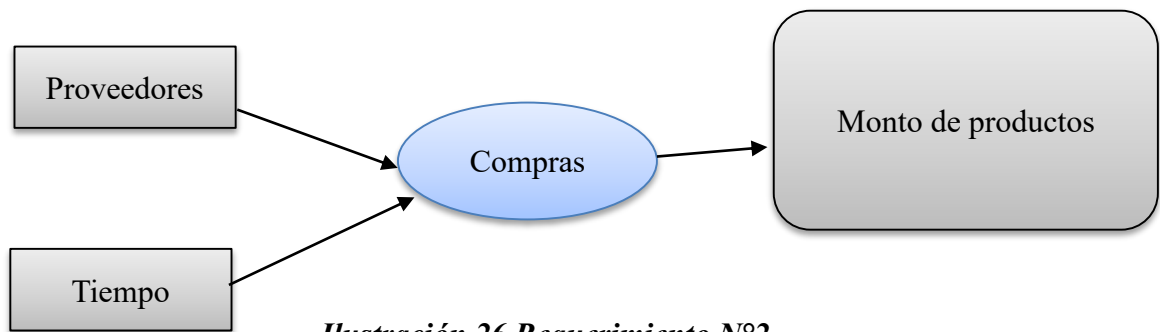


Ilustración 26. Requerimiento N°2

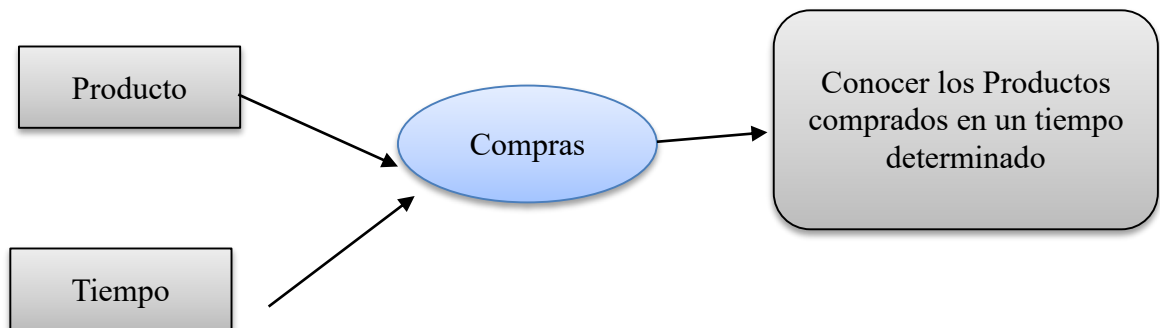


Ilustración 27. Requerimiento N°3

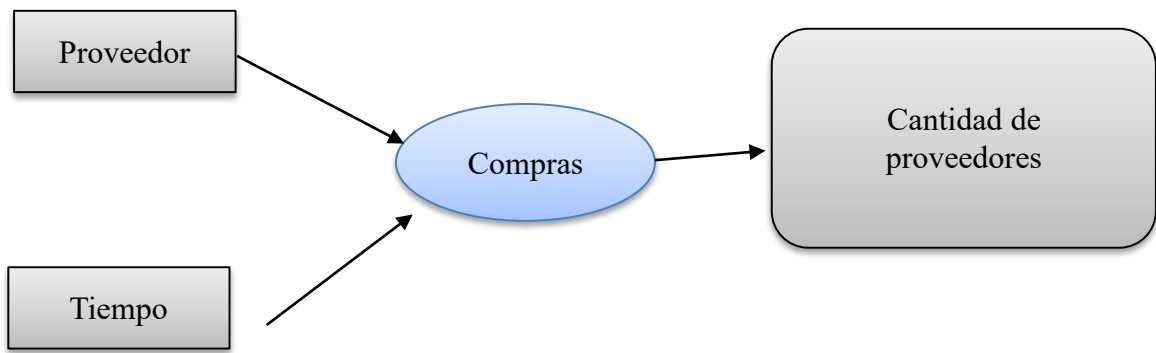


Ilustración 28. Requerimiento N°4

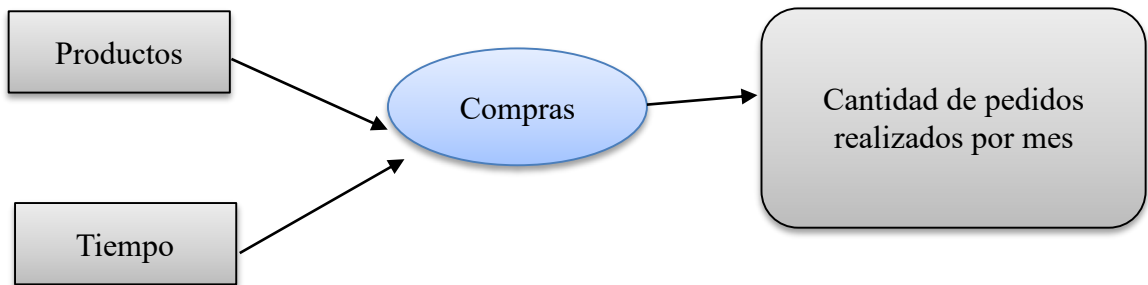


Ilustración 29. Requerimiento N°5

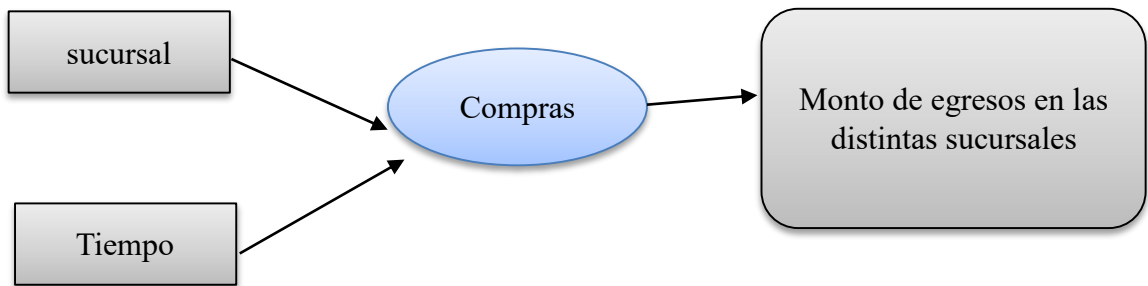


Ilustración 30. Requerimiento N°6

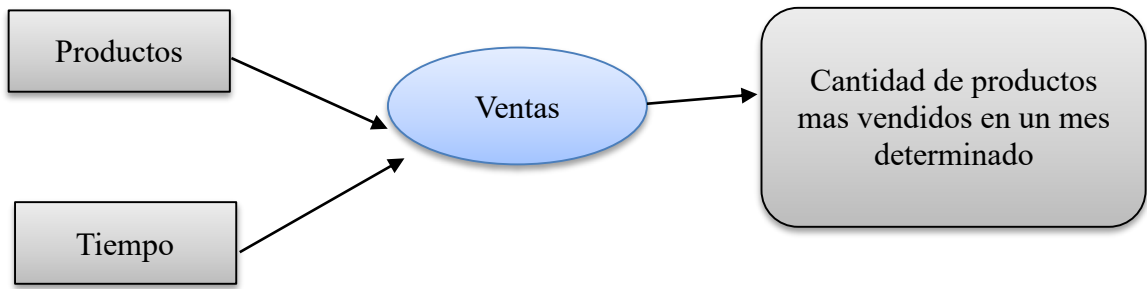


Ilustración 31.Requerimiento N°7

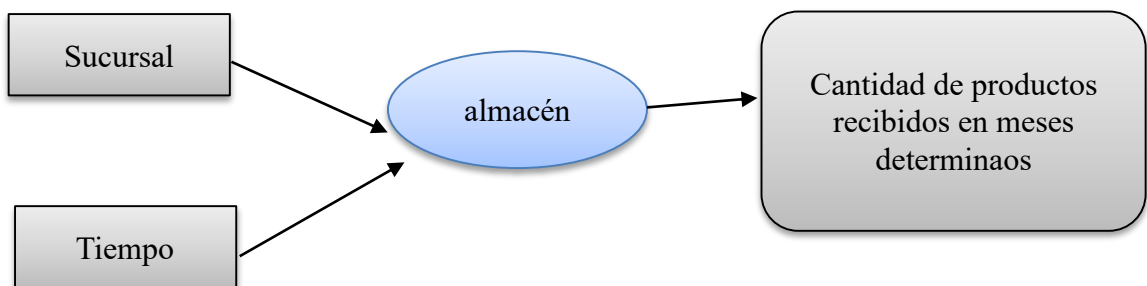


Ilustración 32.Requerimiento N°8

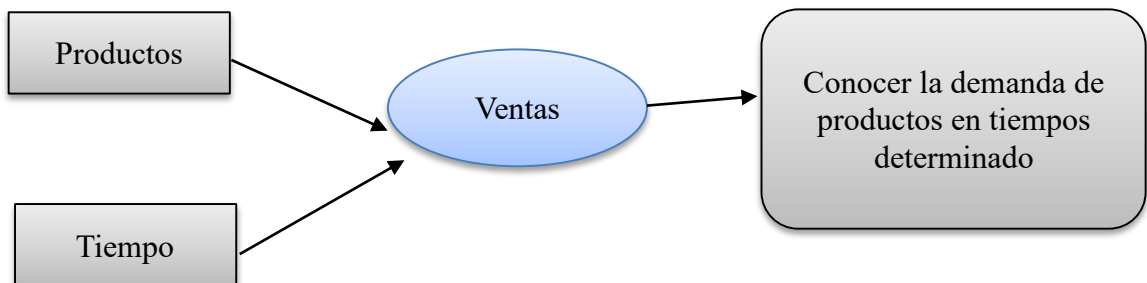


Ilustración 33.Requerimiento N°9

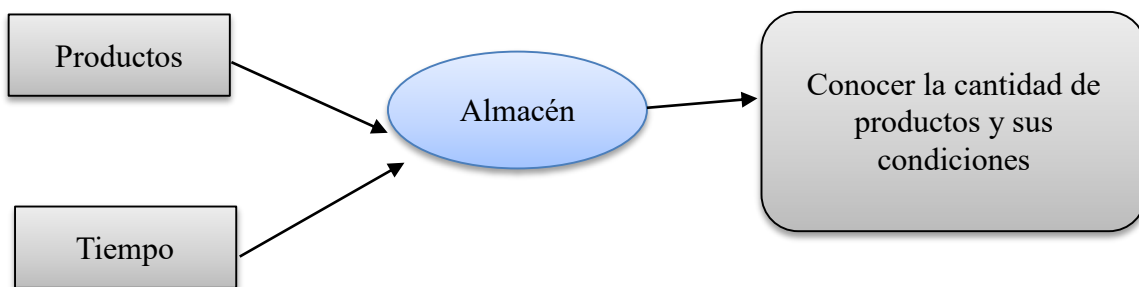


Ilustración 34.Requerimiento N°10

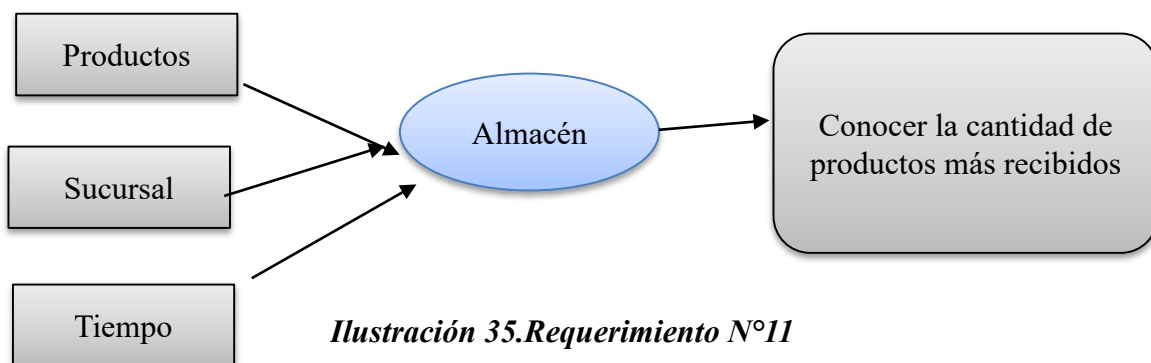


Ilustración 35.Requerimiento N°11

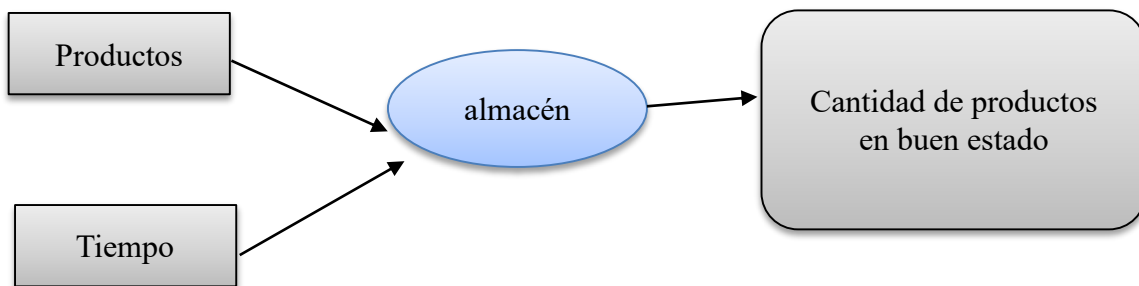


Ilustración 36.Requerimiento N°12

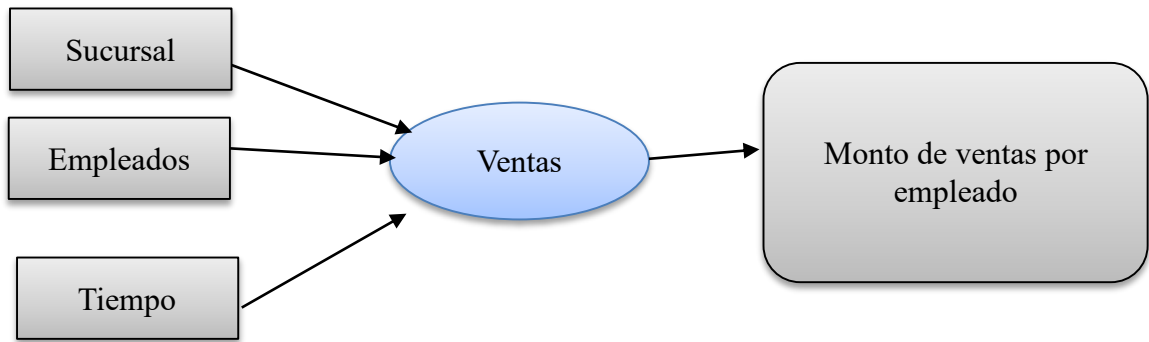


Ilustración 37.Requerimiento N°13

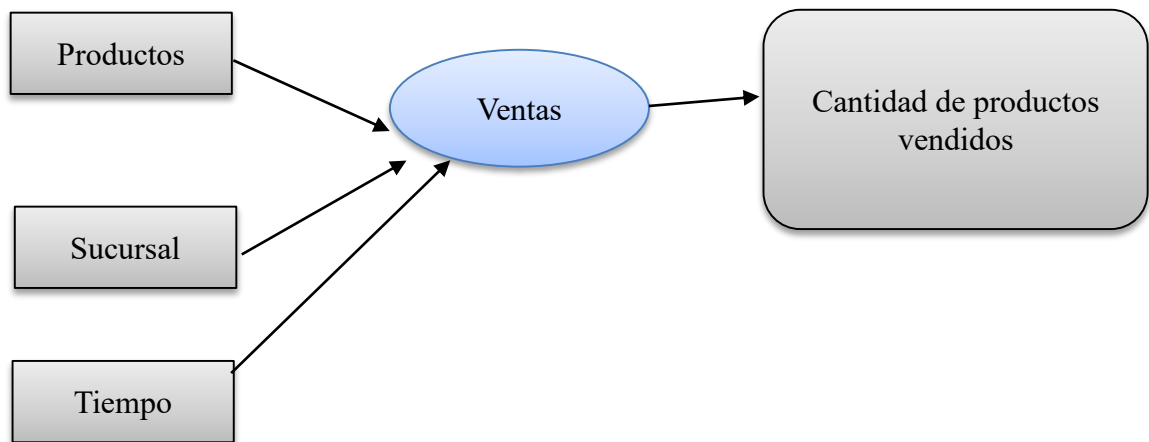


Ilustración 38.Requerimiento N°14

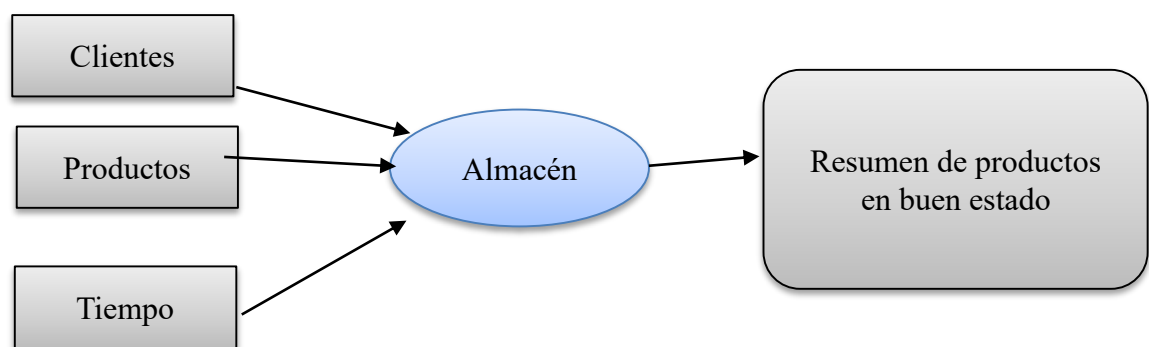


Ilustración 39.Requerimiento N°15

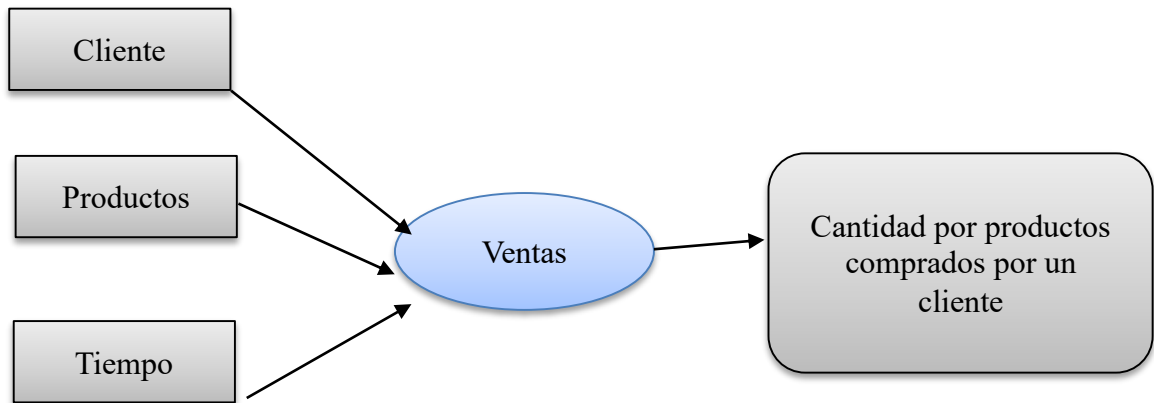


Ilustración 40.Requerimiento N°16

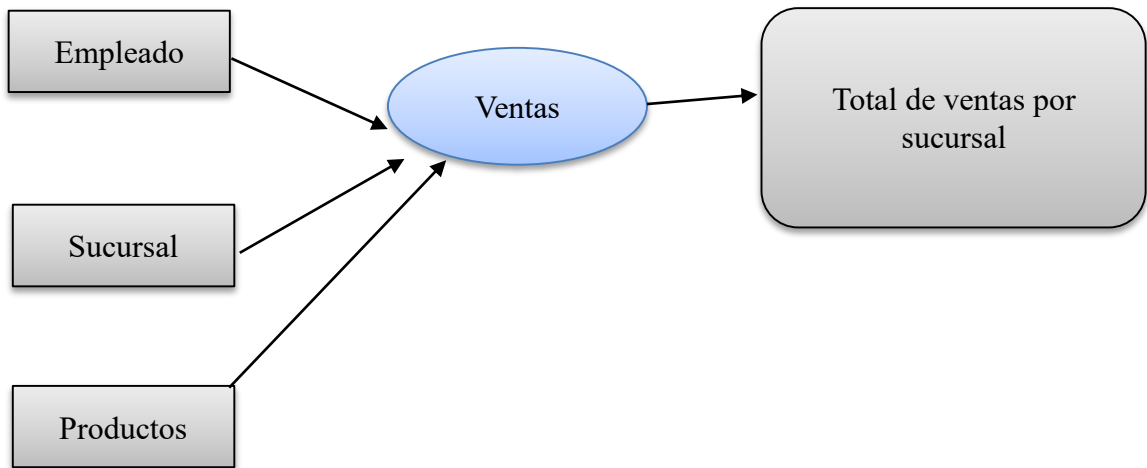


Ilustración 41.Requerimiento N°17

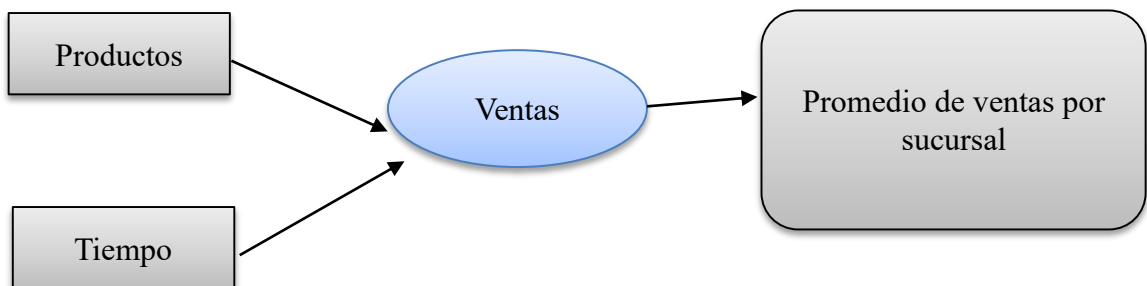


Ilustración 42.Requerimiento N°18

4.9.2. Modelo conceptual del DataWareHouse por áreas

Considerando las perspectivas e indicadores analizadas y mostradas en el paso anterior por los requerimientos funcionales del DW que permite demostrar con mayor facilidad el alcance del proyecto a continuación se presentara el modelo de las 3 áreas que se trabajaran en el proyecto.

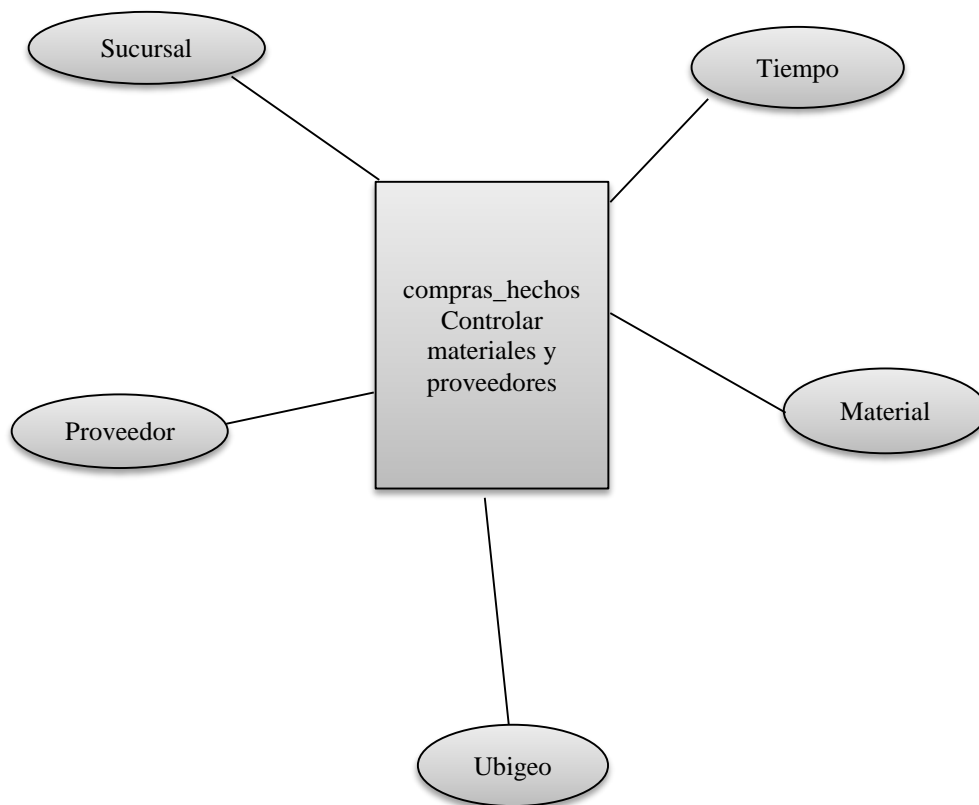


Ilustración 43. Modelo conceptual del área de compras

Fuente: propia

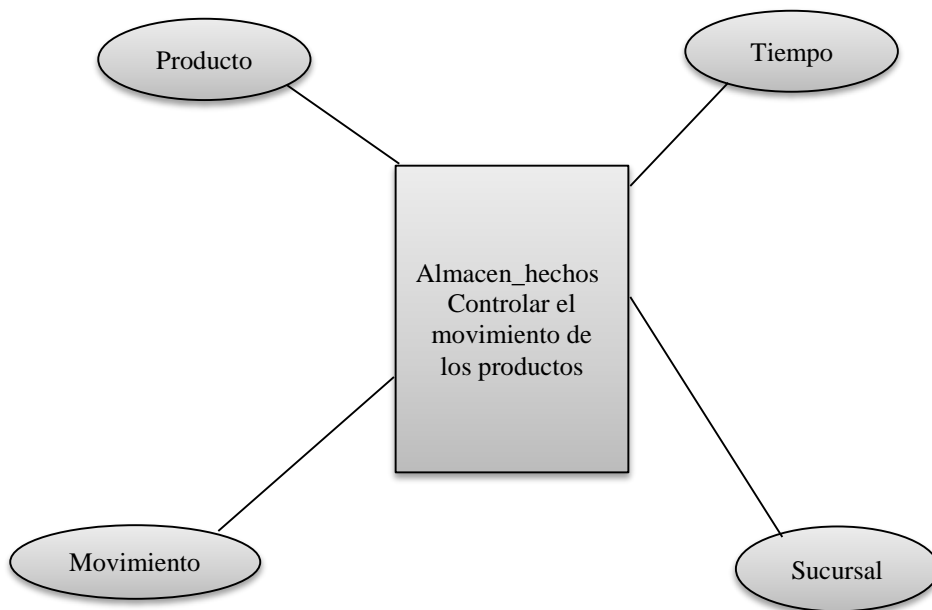


Ilustración 44. Modelo conceptual del proceso de almacén

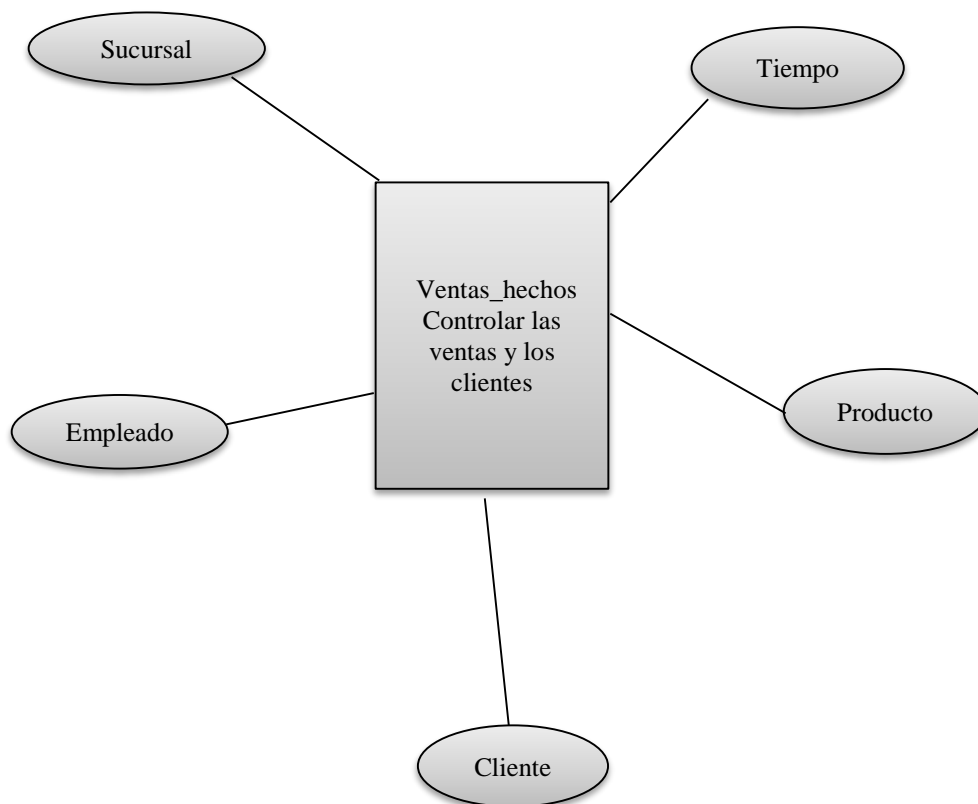


Ilustración 45. Modelo conceptual del proceso de ventas

4.10. Análisis de las fuentes de datos

Una vez analizados los requerimientos del proyecto, el segundo paso según la metodología hefesto es el análisis de los OLTP. Los OLTP son las fuentes de información orientadas al proceso de transacciones , donde se involucran los ingresos los egresos las actualizaciones de data que se realizan día a día en un sistema de información mediante las transacciones . Entre las cantidades de transacciones se encuentran: registro de ventas, registro de compras, entrada de productos, salida de productos, actualización de almacén, adquisiciones

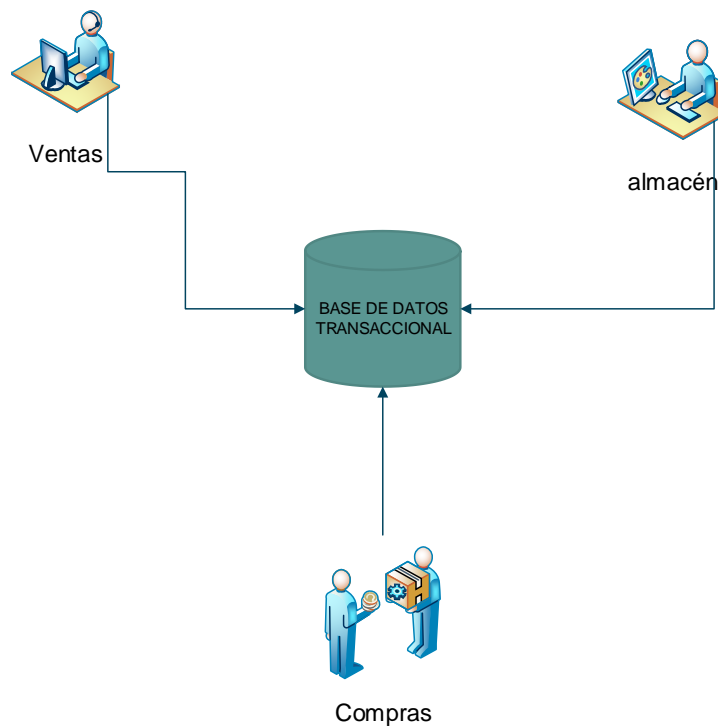


Ilustración 46. Procesos diarios en BD transaccional

Fuente: propia

4.10.1. Fuente de información ferretería solano

El sistema de la ferretería solano trabaja con gestor de base de datos SQL server 2008 el que contiene toda la información de los distintos procesos y transacciones comerciales de esta entidad que contiene las siguientes funciones:

- Registro de ventas
- Registro de compras
- Actualización de movimientos de entrada y salida
- Registro de productos de almacén
- Registro de guías de remisiones
- Registro y consulta de kardex

La base de datos transaccional cuenta con distintas tablas que están normalizadas para realizar las transacciones diarias que requiere la tienda a continuación se mostrara el diagrama de base de datos de la ferretería solano

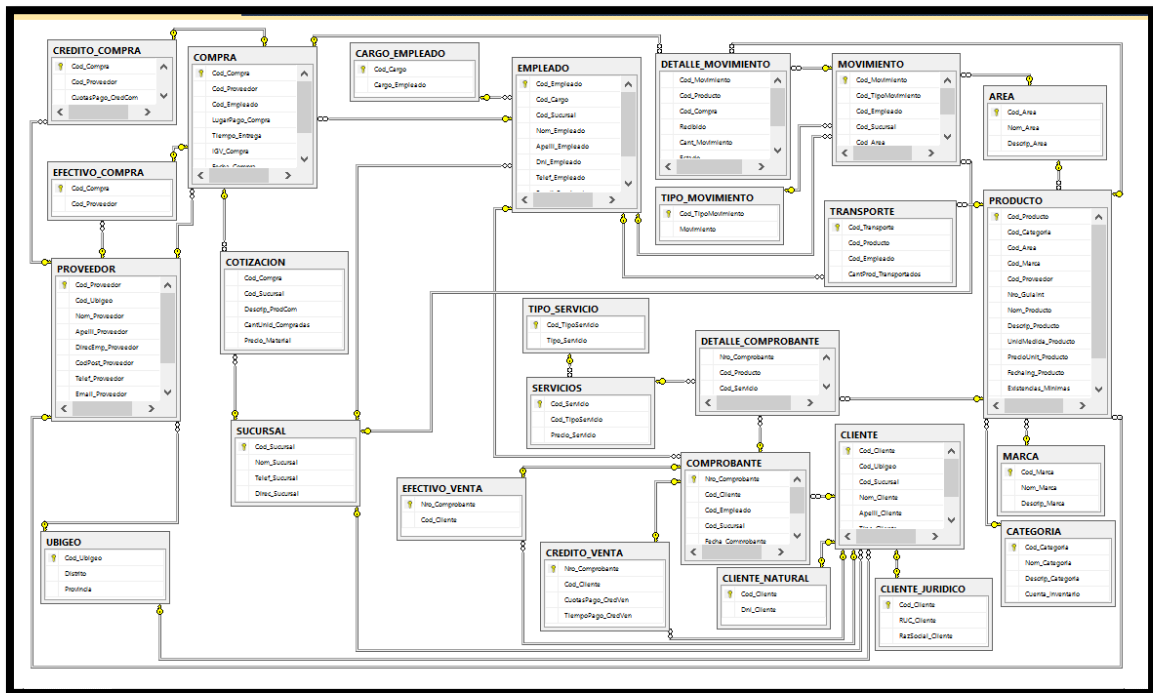


Ilustración 47. Diagrama entidad relación de la base de datos transaccional BDFerreteriaSolano

4.11. Análisis de los OLTP

En la sección pasada **Análisis de las fuentes de datos** se describió el origen de las fuentes de datos para la realización del proyecto, en esta sección se continúa con las fases de la metodología hefesto.

4.11.1. Determinar indicadores

En este paso se indicara como se van a realizar los cálculos para generar los indicadores de cada tabla de hechos que contiene el DW a continuación se describirán los hechos por proceso.

Nombre de la tabla	Nombre de la columna	Descripción del indicador	Métrica SQL
Compras_hechos	total_compras	El total del dinero invertido en las compras de los materiales para la tienda	Función SUM
Compras_hechos	Cantidad	La cantidad de los productos que se han comprado por pedido	Función COUNT
Almacen_hechos	cant_pro_mov	Cantidad de productos por movimiento	Función SUM
Almacen_hechos	Cant_mov	Cantidad de movimientos de entrada y salida por producto	Función SUM
ventas_hechos	Cantidad	Cantidad de ventas realizadas por producto	Función SUM
ventas_hechos	Total_ventas	Total de dinero ganado por las ventas de los productos	Función SUM
ventas_hechos	Promedio_ventas	Promedio de las ventas por cada producto.	Función AVERAGE

Tabla 12.Indicadores y su descripción

Fuente: propia

4.12. Modelo Dimensional

Estándar Modelo Dimensional

El estándar que se utilizara para dar nombre a las dimensiones creadas y para el caso de las nuevas dimensiones serán las siguientes:

- ✓ **Dim:** Para representar una dimensión.
- ✓ **Nombre tabla:** representara el nombre de la tabla a la que hará referencia la dimensión.

Lista de dimensiones

- ✓ Dim_cliente
- ✓ Dim_empleado
- ✓ Dim_sucursal
- ✓ Dim_material
- ✓ Dim_movimiento
- ✓ Dim_producto
- ✓ Dim_proveedor
- ✓ Dim_tiempo
- ✓ Dim_tiempo_almacen
- ✓ Dim_tiempo_ventas
- ✓ Dim_ubigeo

4.12.1. Tipo de modelo lógico del Datawarehouse

Se selecciona el tipo de esquema que más se adapte a los requerimientos a utilizar para la estructura del Data Warehouse. El esquema seleccionado para el desarrollo del Data Warehouse será el modelo estrella debido a la compatibilidad con la herramienta seleccionada y la simplicidad del diseño el modelo lógico de dividirá de acuerdo a su proceso correspondiente a continuación los esquemas de los distintos procesos del negocio.

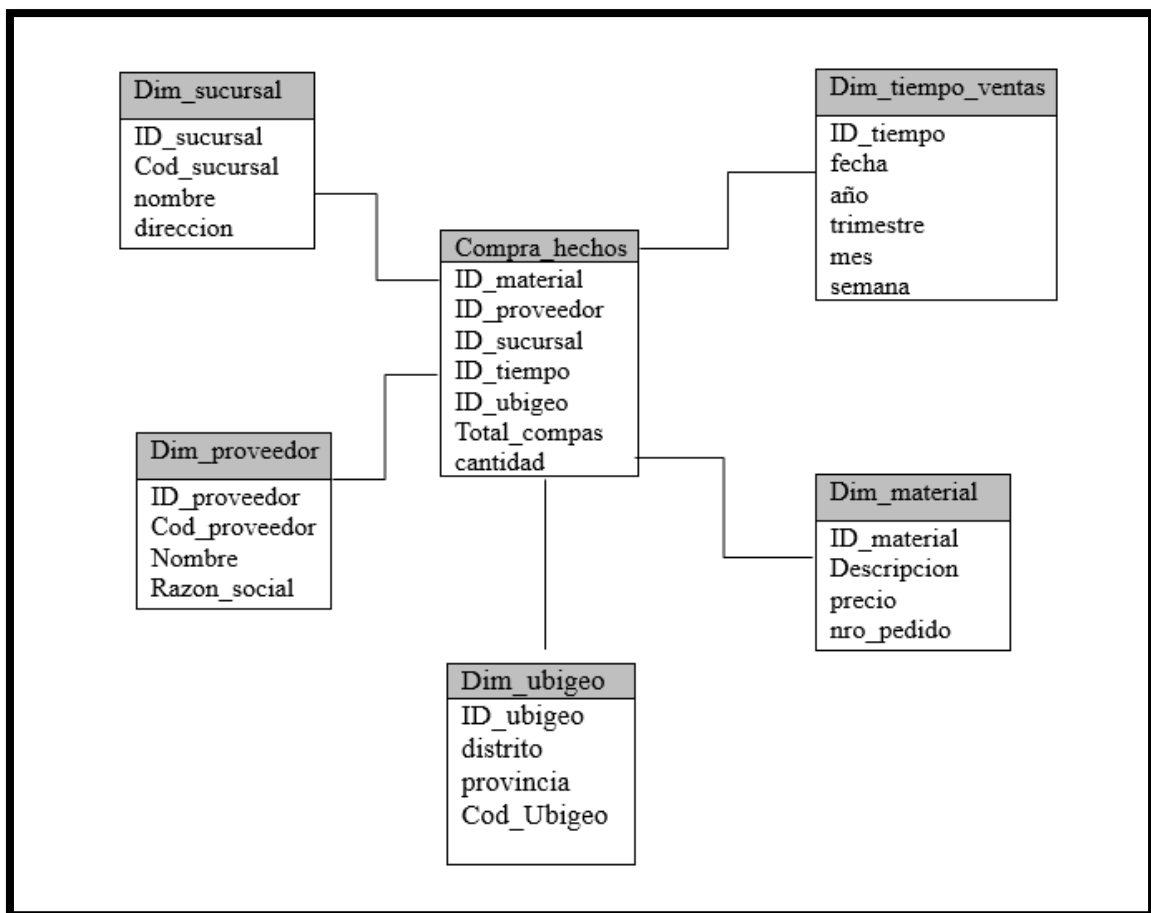


Ilustración 48. Modelo lógico de datos del proceso de compras

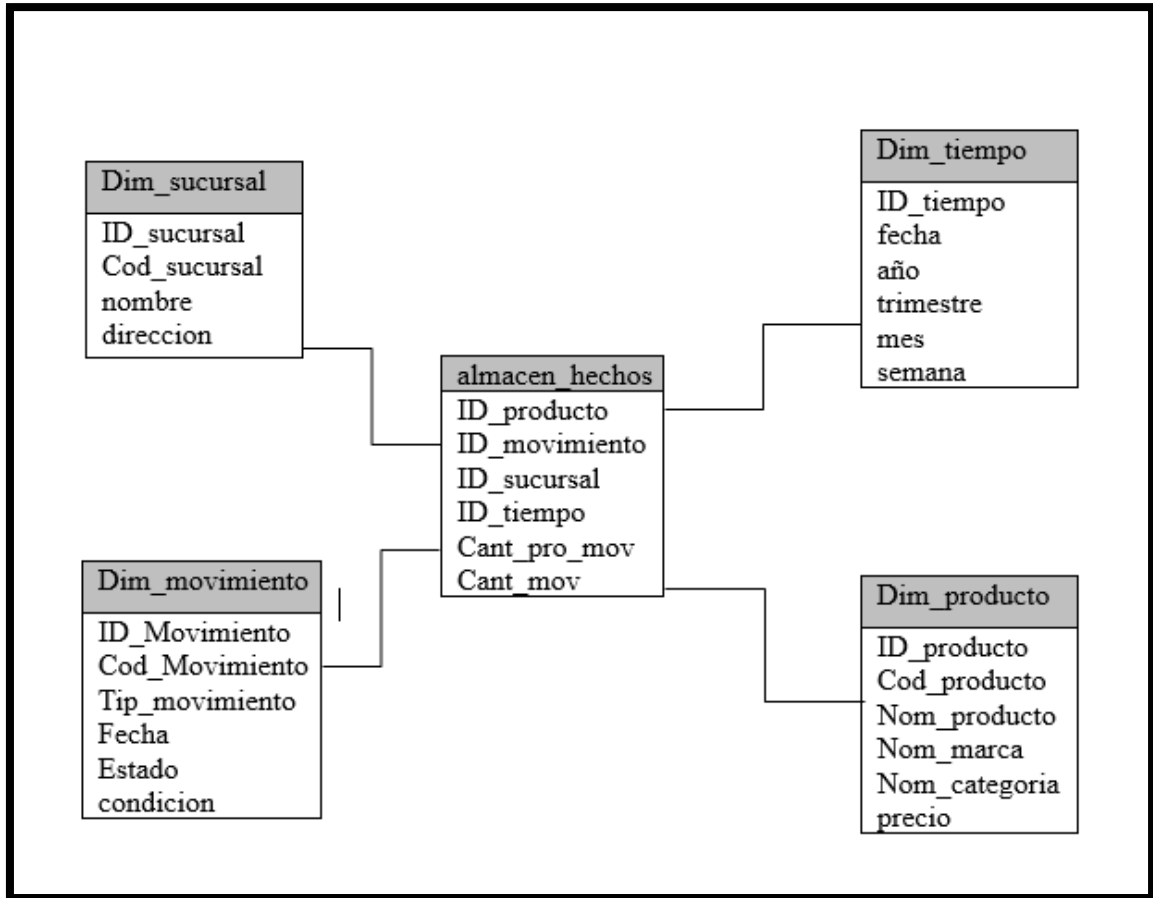


Ilustración 49, Modelo lógico del proceso de almacén

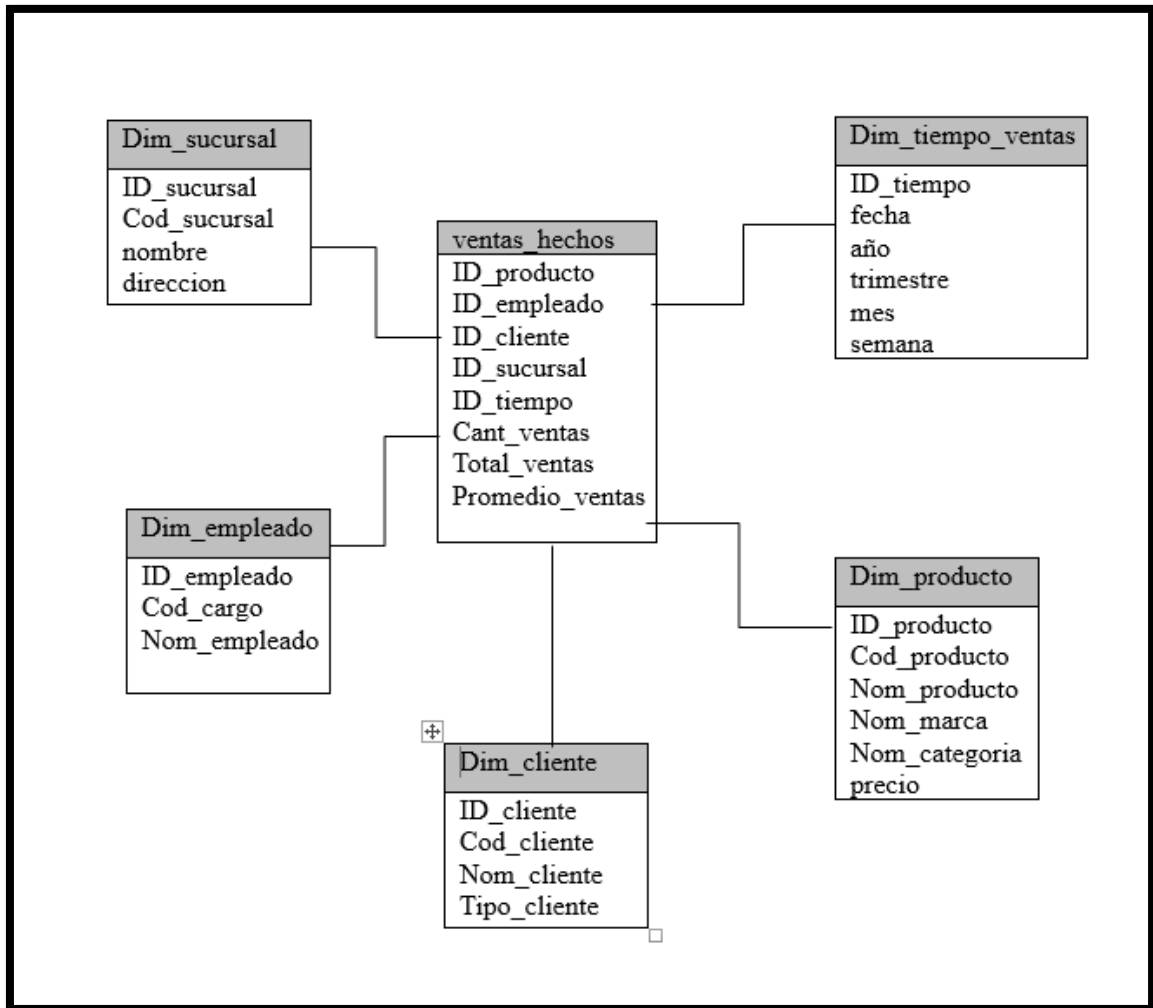


Ilustración 50. Modelo lógico del proceso de ventas

4.12.2. Tablas de dimensiones

En este paso se diseñaran las dimensiones que formaran parte del modelo de Data warehouse conglomerando todos los procesos involucrados en una sola base de datos dimensional , para la creacion de las dimensiones se debe tener en cuenta cual sera el nombre de la tabla , cada tabla dimensional contiene una clave primaria y sus respectivos nombres de los campos.

Perspectiva producto : la tabla dimension producto hara referencia a los productos en venta que tiene la ferreteria la cual tiene los siguientes atributos :

- Se le agregara la clave principal llamada ID_producto
- Contendra el atributo Cod_Producto que referencia al codigo de producto de la tabla PRODUCTOS base de datos transaccional .
- Contendra el atributo Nom_producto referente al nombre de los productos de la tabla PRODUCTOS
- El atributo Nom_marca referente a la marca del producto.
- El atributo Nom_categoria referente a la categoria a la que pertenece el producto
- Finalmente el atributo precio de los productos .

Dim_producto
ID_producto
Cod_producto
Nom_producto
Nom_marca
Nom_categoria
Precio

Ilustración 51. Tabla Dim_producto

Perspectiva Material : la dimension material hara referencia a los productos que se encuentran en almacen que recién han sido comprados a los proveedores la cual contiene los siguientes atributos .

- La clave primaria sera el atributo ID_material
- El atributos descripcion del material referenciada en la tabla

COTIZACION de la bd transaccional origen.

- El atributo precio
- El atributo Nro_pedido que sera referenciado con el codigo de pedido de material

Dim_material
ID_material
Descripcion
precio
nro_pedido

Ilustración 52. Tabla Dim_material

Perspectiva cliente : la tabla dimension cliente contendra los clientes que se registran a diario cuando se realiza una venta en la ferreteria solano la cual contendra los siguientes atributos :

- El atributo ID_cliente como clave principal de la tabla
- El atributo Cod_cliente referenciado al codigo del cliente de la tabla Clientes de la base de datos transaccional.
- El atributo Nom_cliente que sera transformado de los atributos nombre y apellido de cliente de la base de dato transaccional.
- Finalmente el atributo Tipo_cliente

Dim_cliente
ID_cliente
Cod_cliente
Nom_cliente
Tipo_cliente

Ilustración 53. Tabla Dim_cliente

Perspectiva sucursal: la tabla dimension sucursal contendrá las sucursales de la ferretería en estudio factor importante ya que se obtienen datos distintos de cada sucursal la cual tendrá los siguientes atributos :

- El Atributo ID_sucursal como clave principal de la tabla
- Cod_sucursal el código de la sucursal obtenido de la tabla SUCURSAL de la bd transaccional
- Nombre de la sucursal
- Y la dirección

Dim_sucursal
ID_sucursal
Cod_sucursal
nombre
direccion

Ilustración 54. Tabla Dim_sucursal

Perspectiva ubigeo: esta perspectiva contiene los distritos donde se encuentran las sucursales de las tiendas la cual tendrá los siguientes atributos:

- El atributo Id_ubigeo como clave principal a la tabla
- El atributo distrito obtenido de la tabla UBIGEO de la base de datos transaccional
- El atributo provincia
- Y el código de ubigeo

Dim_ubigeo
ID_ubigeo
distrito
provincia
Cod_Ubigeo

Ilustración 55. Tabla Dim_ubigeo

Perspectiva Empleado: la tabla dimensión empleado contendrá los empleados solo que realicen transacciones de ventas dado los requerimientos se estableció de esa manera, contendrá los siguientes atributos:

- El Atributo ID_empleados como clave principal de la tabla
- Cod_cargo el código del cargo del empleado extraído de la tabla Empleado de la base de datos transaccional
- Nom_empleado que será una fusión de nombre y apellido en un solo campo extraído de la bd transaccional

Dim_empleado
ID_empleado
Cod_cargo
Nom_empleado

Ilustración 56. Tabla Dim_empleado

Perspectiva movimiento: en esta dimensión contendrá todos los datos relacionados a los movimientos de almacén de la ferretería la cual contiene los siguientes atributos

- Id_movimiento como clave principal.
- Cod_movimiento referente a los movimientos
- Tip_movimiento referente al tipo de movimiento
- Atributo fecha donde se realizado el movimiento
- Atributo estado se refiere a como se encuentra el producto puede ser buen o mal estado
- Condición atributo que refiere en donde está el producto si en almacén o en ventas o devolución.

Dim_movimiento
ID_Movimiento
Cod_Movimiento
Tip_movimiento
Fecha
Estado
condicion

Ilustración 57. Tabla Dim_movimiento

Perspectiva proveedor: en la dimensión proveedor se guardan los datos relevantes de los proveedores de la empresa la cual se decidió obtener los siguientes atributos:

- Id_proveedor
- Cod_proveedor
- Nombre
- Razón social

Dim_proveedor
ID_proveedor
Cod_proveedor
Nombre
Razon_social

Ilustración 58. Tabla Dim_proveedor

Perspectiva tiempo: En esta dimensión se utilizaran 3 tablas que utilizaran los mismos atributos pero el tiempo será distinto en esta tabla las fechas realizadas por compras por ventas y las fechas de almacén serán divididas en distintas tablas para obtener una visión más clara del tiempo en que se realizaron cada proceso las tablas tendrán los siguientes atributos

- Id_tiempo
- Fecha
- Año
- trimestre
- mes

- semana

Estos atributos se obtendrán después de una transformación derivada de la fecha correspondiente de la transacción que se realizó en cada proceso.

Dim_tiempo_compras
ID_tiempo
fecha
año
trimestre
mes
semana

Dim_tiempo_almacén
ID_tiempo
fecha
año
trimestre
mes
semana

Dim_tiempo_ventas
ID_tiempo
fecha
año
trimestre
mes
semana

Ilustración 59. Tabla Dim_tiempo

4.12.3. Tablas de hechos

En este paso se definen las tablas de hechos que son las que contendrán las métricas o hechos a través de las cuales se construirán los indicadores de estudios. Al igual que las tablas dimensionales se debe tomar en cuenta el nombre de la tabla, pero en estas tablas la diferencia es que contiene todas las claves primarias de las dimensiones involucradas en el proceso que serán parte de los atributos al igual que los atributos calculados.

Tabla ventas_hechos: esta tabla es la que contiene los cálculos del proceso del área de ventas, tendrá como claves principales todas las claves de las dimensiones antes definidas para esta tabla se utilizarán las siguientes claves: ID_producto, ID_employado, ID_cliente, ID_sucursal, ID_tiempo. Esta tabla contendrá los siguientes hechos que fueron

detallados anteriormente los cuales serán;

- Cant_ventas
- Total_ventas
- Promedio de ventas

ventas hechos
ID_producto
ID_empleado
ID_cliente
ID_sucursal
ID_tiempo
Cant_ventas
Total_ventas
Promedio_ventas

Ilustración 60. Tabla ventas_hechos

Tabla Compras_hechos: en esta tabla se encuentran las métricas que se realizaran para el proceso de ventas al igual que la tabla anterior esta tabla contendrá las claves primarias de todas las dimensiones involucradas en esta área las cuales serán: ID_material, ID_proveedor, ID_sucursal, ID_tiempo, ID_ubigeo. Esta tabla contendrá 2 hechos los cuales son los siguientes:

- Total_compras
- Cantidad

Compra hechos
ID_material
ID_proveedor
ID_sucursal
ID_tiempo
ID_ubigeo
Total_compras
cantidad

Ilustración 61. Tabla Compras_hechos

Tabla Almacen_hechos: esta tabla contiene las métricas del área de almacén la cual está conformada por las claves de las dimensiones

involucradas que son: ID_producto, ID_movimiento, ID_sucursal, ID_tiempo y como hechos los siguientes atributos

- Cant_pro_mov
- Cant_mov

almacen_hechos
ID_producto
ID_movimiento
ID_sucursal
ID_tiempo
Cant_pro_mov
Cant_mov

Ilustración 62. Tabla Almacen_hechos

Relaciones de las tablas de hechos: las uniones de las tablas de hechos se harán con las dimensiones involucradas mediante clave primaria de cada una de estas, se relacionaran con la tabla de hechos con el fin de involucra la data que contiene cada una y así poder crear una unión entre la data de las dimensiones y las métricas de la tabla de hechos las imágenes de la unión de las tablas de este proyectos se vieron en la **ilustración 50, 51 y 52.**

4.12.4. Modelo físico del Data Warehouse

Después de haber analizados los requerimientos, diseñados las dimensiones, establecidos los hechos se pasara a construir el Data Warehouse en el gestor de base de datos MySQL a continuación se mostrara en la imagen el diagrama físico del Data Warehouse.

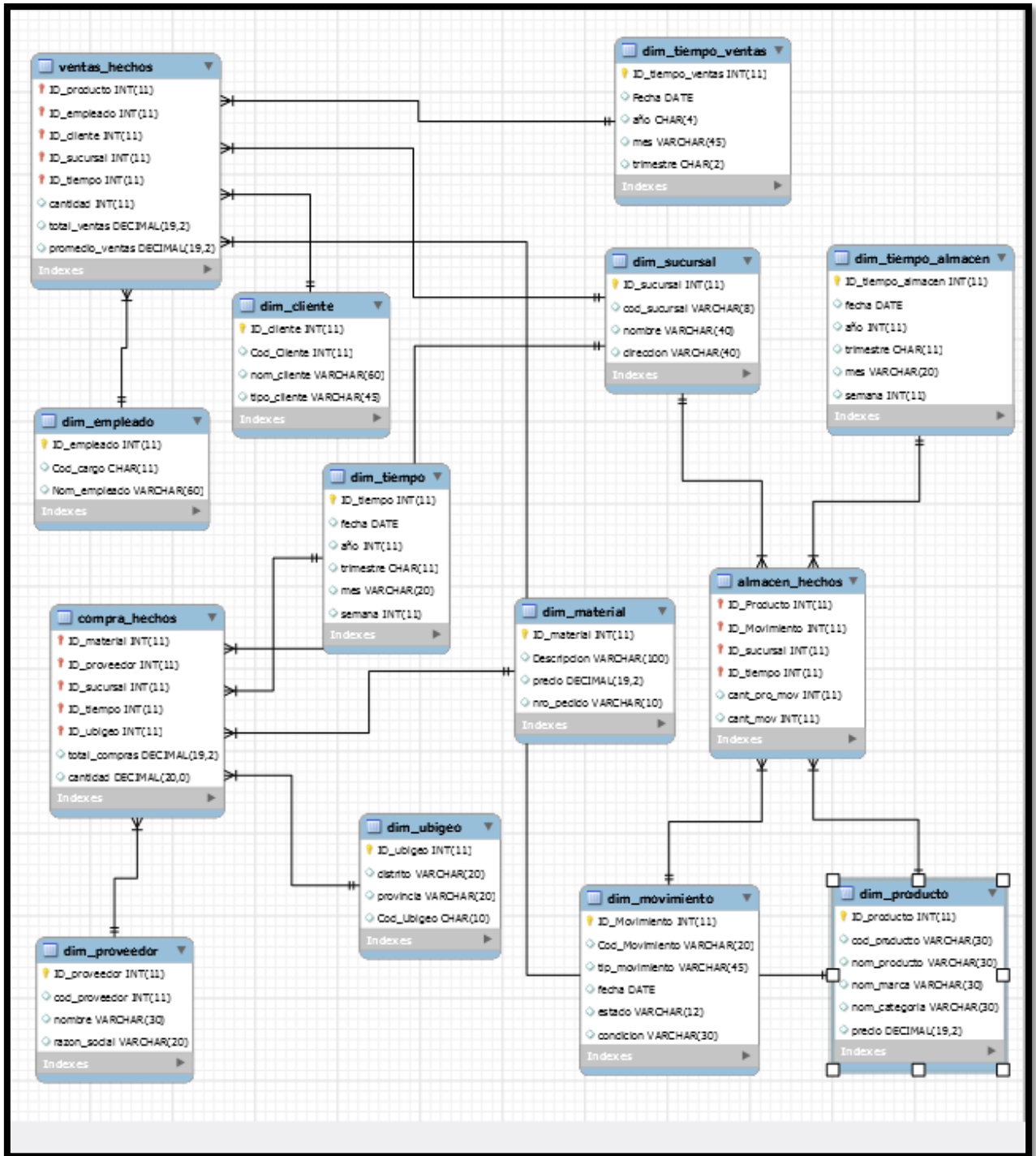


Ilustración 63. Modelo entidad relación de la base de datos dimensional

4.13. Establecer correspondencias

El mapeo de los datos se realiza para explicar de dónde se han extraído los datos para realizar el modelo dimensional y como se relacionan con los campos que serán su destino en la solución BI, a continuación se muestran los siguientes cuadros por dimensión y por hechos.

tabla	Dim_producto							
tipo de tabla	dimensión							
Destino DW_Solano					Origen BDFerreteriaSolano			
columna	tipo de dato	tamaño	clave	fk	Tabla	campo	tipo de dato	longitud
ID_producto	int	11	pk					
cod_producto	varchar	30			Producto	Cod_producto	char	8
nom_producto	varchar	30			Producto	Nom_producto	varchar	30
nom_marca	varchar	30			Marca	Nom_marca	varchar	20
nom_categoria	varchar	30			categoria	Nom_categoria	varchar	30
precio	decimal	19,2						

Tabla 13.Mapeo Dim_producto

tabla	Dim_cliente							
tipo de tabla	dimensión							
Destino DW_Solano					Origen BDFerreteriaSolano			
columna	tipo de dato	tamaño	clave	fk	Tabla	campo	tipo de dato	longitud
ID_cliente	int	11	pk					
cod_cliente	int	11			Cliente	Cod_Cliente	int	
nom_cliente	varchar	60			Cliente	Nom_cliente	varchar	20
					Cliente	Apelli_Cliente	varchar	30
tipo_cliente	varchar	45			Cliente	Tipo_Cliente	char	20

Tabla 14.Mapeo Dim_cliente

tabla	Dim_proveedor							
tipo de tabla	dimensión							
Destino DW_Solano					Origen BDFerreteriaSolano			
columna	tipo de dato	tamaño	clave	fk	Tabla	campo	tipo de dato	longitud
ID_proveedor	int	11	pk					
cod_proveedor	int	11			proveedor	Cod_Proveedor	int	
Razón_social	varchar	30			proveedor	Nom_Proveedor	varchar	20
nombre	varchar	30			proveedor	Apelli_Proveedor	varchar	30

Tabla 15.Mapeo Dim_proveedor

tabla	Dim_material							
tipo de tabla	Dimensión							
Destino DW_Solano					Origen BDFerreteriaSolano			
columna	tipo de dato	tamaño	clave	fk	Tabla	campo	tipo de dato	longitud
ID_material	int	11	pk					
Descripción	varchar	100			Cotización	Descrip_ProdCom	varchar	150
precio	decimal	19,2			Proveedor	Precio_Material	decimal	19,2
Nro_pedido	varchar	10			Cotización	Cod_compra	varchar	10

Tabla 16.Mapeo Dim_material

Tabla	Dim_movimiento							
tipo de tabla	Dimensión							
Destino DW_Solano					Origen BDFerreteriaSolano			
Columna	tipo de dato	tamaño	clave	fk	Tabla	campo	tipo de dato	longitud
ID_Movimiento	int	11	pk					
Cod_Movimiento	varchar	20			Detalle_Movimiento	Descrip_ProdCom	varchar	150
Tip_movimiento	varchar	45			Tipo_Movimiento	Precio_Material	decimal	19,2
Fecha	date	10			Movimiento	Fecha_Movimiento	date	10
Estado	varchar	12			Detalle_Movimiento	Estado	Varchar	12
Condición	varchar	30			Detalle_Movimiento	Condicion	Varchar	30

Tabla 17.Mapeo Dim_movimiento

tabla	Dim_sucursal							
tipo de tabla	dimensión							
Destino DW_Solano					Origen BDFerreteriaSolano			
columna	tipo de dato	tamaño	clave	fk	Tabla	campo	tipo de dato	longitud
ID_sucursal	int	11	pk					
cod_sucursal	varchar	8			Sucursal	Cod_Sucursal	char	8
nombre	varchar	40			Sucursal	Nom_Sucursal	varchar	40
dirección	varchar	40			sucursal	Direc_Sucursal	varchar	40

Tabla 18.Mapeo Dim_sucursal

tabla	Dim_ubigeo							
tipo de tabla	dimensión							
Destino DW_Solano					Origen BDFerreteriaSolano			
columna	tipo de dato	tamaño	clave	fk	Tabla	campo	tipo de dato	longitud
ID_ubigeo	int	11	pk					
distrito	varchar	20			Ubigeo	Distrito	varchar	20
provincia	varchar	29			Ubigeo	provincia	varchar	20
Cod_ubigeo	char	10			ubigeo	Cod_Ubigeo	char	8

Tabla 19.Mapeo Dim_ubigeo

tabla	Dim_tiempo							
tipo de tabla	Dimensión							
Destino DW_Solano					Origen BDFerreteriaSolano			
columna	tipo de dato	tamaño	clave	fk	Tabla	campo	tipo de dato	longitud
ID_tiempo	int	11	pk					
Fecha	Date	20			Detalle_Movimiento	Fecha_compra	Date	
Año	int	11						
Trimestre	char	11						
Mes	varchar	20						
Semana	int	11						

Tabla 20.Mapeo Dim_tiempo

Tabla	Dim_tiempo_ventas							
tipo de tabla	Dimensión							
Destino DW_Solano					Origen BDFerreteriaSolano			
columna	tipo de dato	tamaño	clave	fk	Tabla	campo	tipo de dato	longitud
ID_tiempo	int	11	pk					
Fecha	Date	20			comprobante	Fecha_comprobante	Date	
Año	int	11						
Trimestre	char	11						
Mes	varchar	20						
Semana	int	11						

Tabla 21.Mapeo Dim_tiempo_ventas

Tabla	Dim_tiempo_almacen							
tipo de tabla	Dimensión							
Destino DW_Solano					Origen BDFerreteriaSolano			
Columna	tipo de dato	tamaño	clave	fk	Tabla	campo	tipo de dato	longitud
ID_tiempo	int	11	pk					
Fecha	Date	20			comprobante	Fecha_Movimiento	Date	
Año	int	11						
Trimestre	char	11						
Mes	varchar	20						
Semana	int	11						

Tabla 22.Mapeo Dim_tiempo_almacen

Tabla	Ventas_hechos							
tipo de tabla	hechos							
Destino DW_Solano					Origen BDFerreteriaSolano			
columna	tipo de dato	tamaño	clave	fk	Tabla	campo	tipo de dato	funcion
ID_producto	int	11	pk					
ID_employado	int	11	Pk					
ID_cliente	int	11	Pk					
ID_sucursal	int	11	Pk					
ID_tiempo	int	11	pk					
cantidad	int	11			Detalle_comprobante	Cantunid_ventas	int	Count
Total_ventas	Decimal	19,2			Producto	PrecioUnit_producto	money	sum
Promedio_ventas	Decimal	19,2						

Tabla 23.Mapeo Ventas_hechos

Tabla	Compras_hechos								
tipo de tabla	hechos								
Destino DW_Solano					Origen BDFerreteriaSolano				
columna	tipo de dato	tamaño	clave	fk	Tabla	campo	tipo de dato	función	
ID_material	Int	11	pk						
ID_proveedor	Int	11	Pk						
ID_sucursal	Int	11	Pk						
ID_tiempo	Int	11	Pk						
ID_ubigeo	Int	11	pk						
Total_compras	Int	11			Cotización	Precio_Material	Decimal(19,2)	sum	
cantidad	Decimal	19,2			cotización	CantUnid_compradas	int	count	

Tabla 24.Mapeo Compras_hechos

tabla	Almacen_hechos								
tipo de tabla	hechos								
Destino DW_Solano					Origen BDFerreteriaSolano				
columna	tipo de dato	tamaño	clave	fk	Tabla	campo	tipo de dato	función	
ID_producto	int	11	pk						
ID_Movimiento	int	11	Pk						
ID_sucursal	int	11	Pk						
ID_tiempo	int	11	Pk						
Cant_pro_mov	int	11			Detalle_Movimiento	Cant_movimiento	int	sum	
Cant_mov	int	11			Detalle_Movimiento	Cod_compra	varchar	count	

Tabla 25.Mapeo Almacen_hechos

4.14. Nivel de granularidad

De acuerdo con las correspondencias establecidas anteriormente, se analizaron los campos que serán utilizados para el desarrollo de los reportes de los cuales se determinaron las perspectivas y los indicadores.

Primero se examinaron las fuentes de datos entregadas para el análisis, para entender y definir los significados de cada campos, después de consulto con el personal de ventas, almacén y compras que manejan los datos para comprender los campos que nos permitirán obtener los indicadores requeridos para el negocio a continuación se realizara un esquema del nivel de granularidad por proceso.

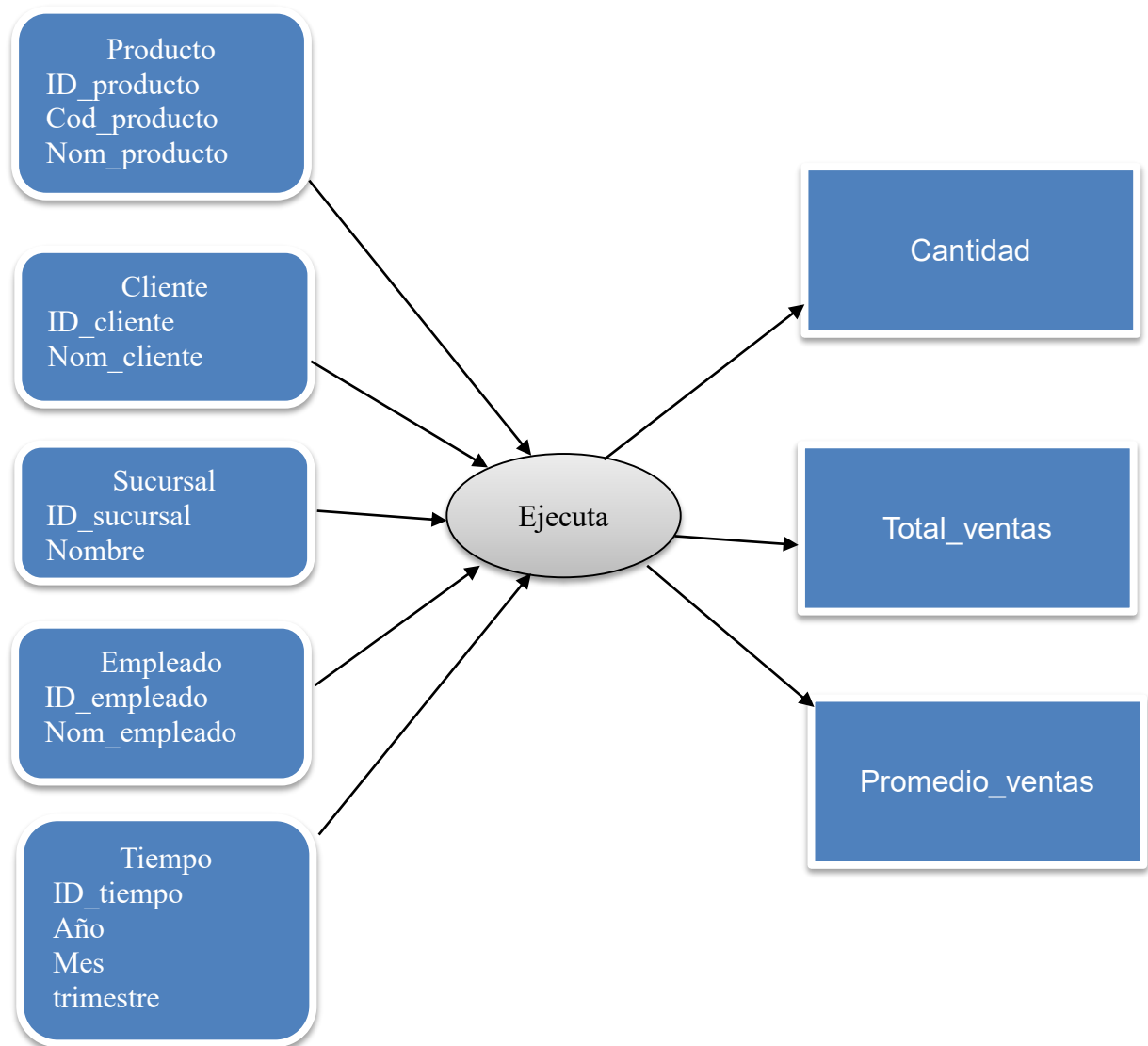


Ilustración 64. Nivel de granularidad del proceso de ventas

Fuente: propia

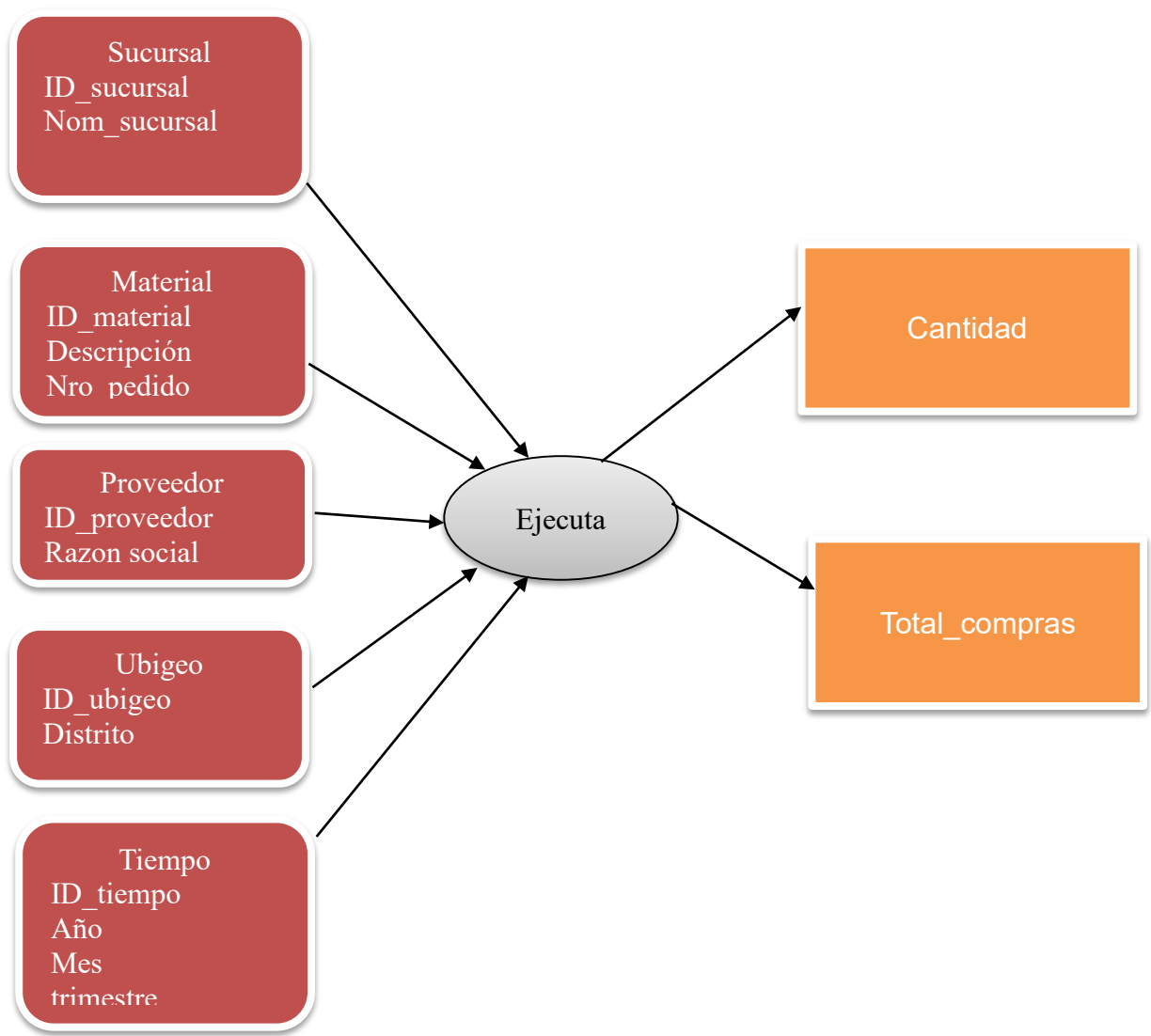


Ilustración 65. Nivel de granularidad del proceso de compras

Fuente: propia

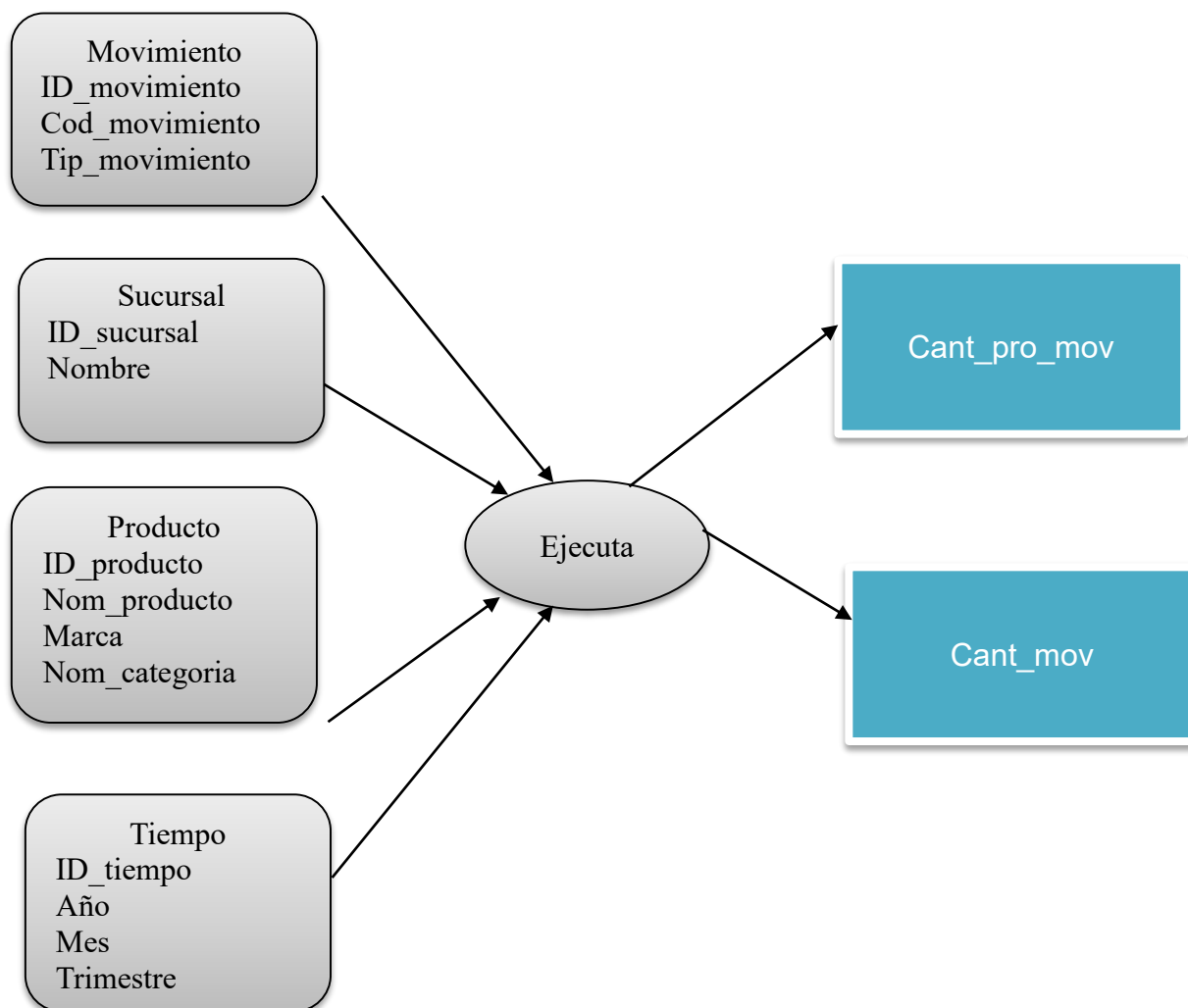


Ilustración 66. Nivel de granularidad proceso de almacén

Fuente: propia

4.15. Modelo Conceptual ampliado

En este modelo se mostrara las dimensiones y los hechos que se han implementado en el proyecto enfocándose en sus perspectivas con un nivel de granularidad completos utilizando todos los atributos que tiene cada dimensión y los atributos que se necesitaran para realizar los cálculos.

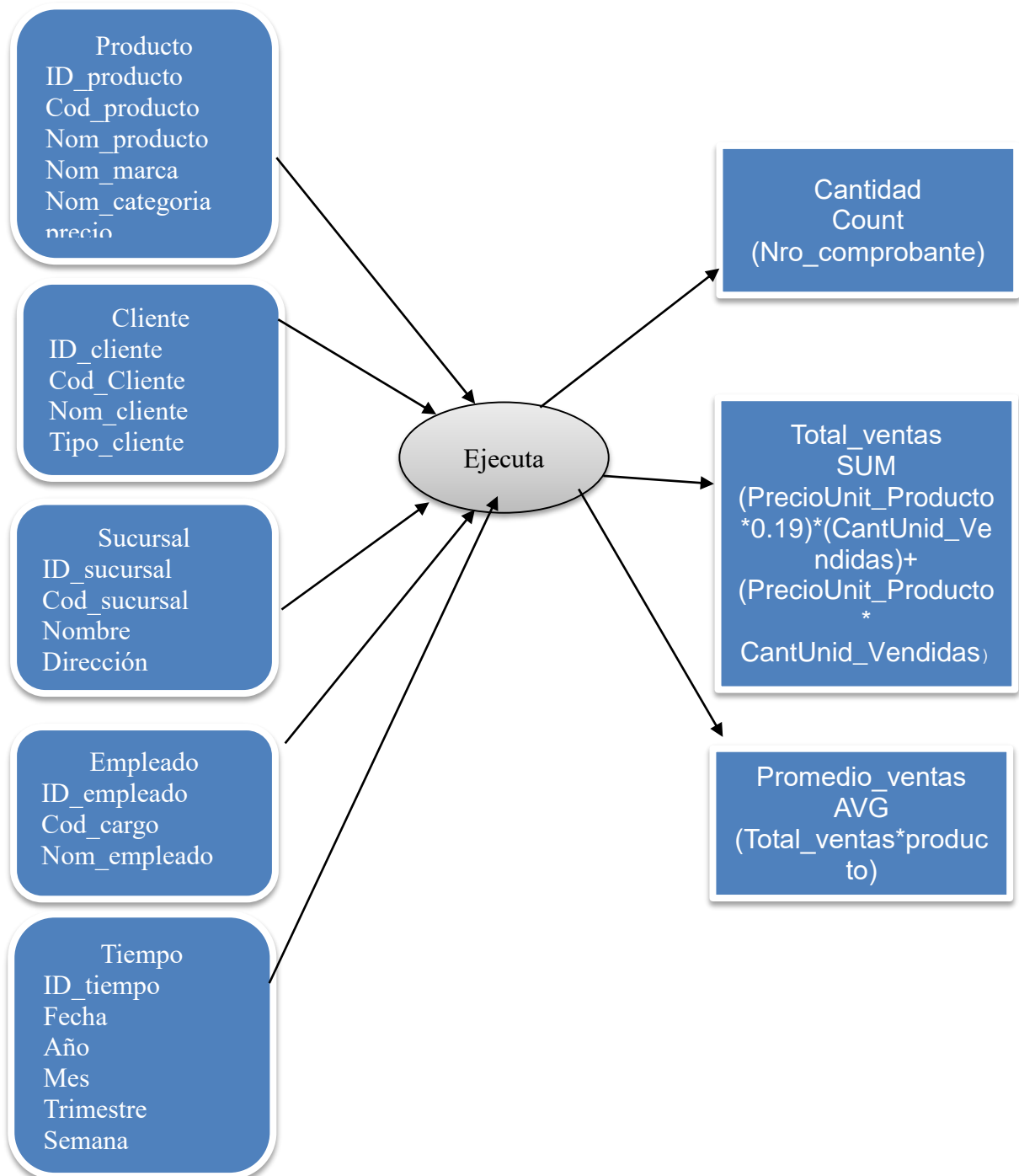


Ilustración 67. Modelo conceptual ampliado del proceso de ventas

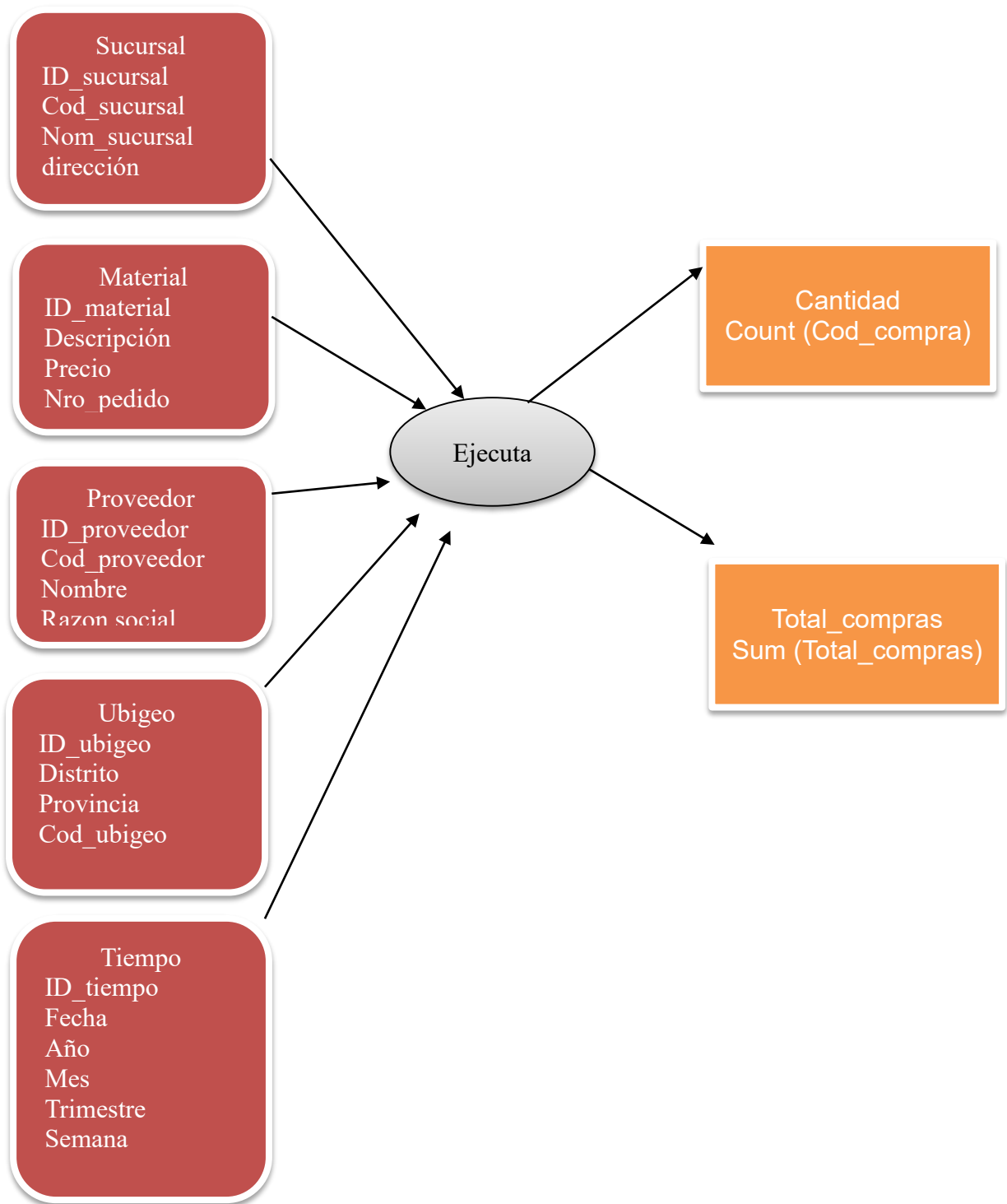


Ilustración 68. Modelo conceptual del proceso de compras

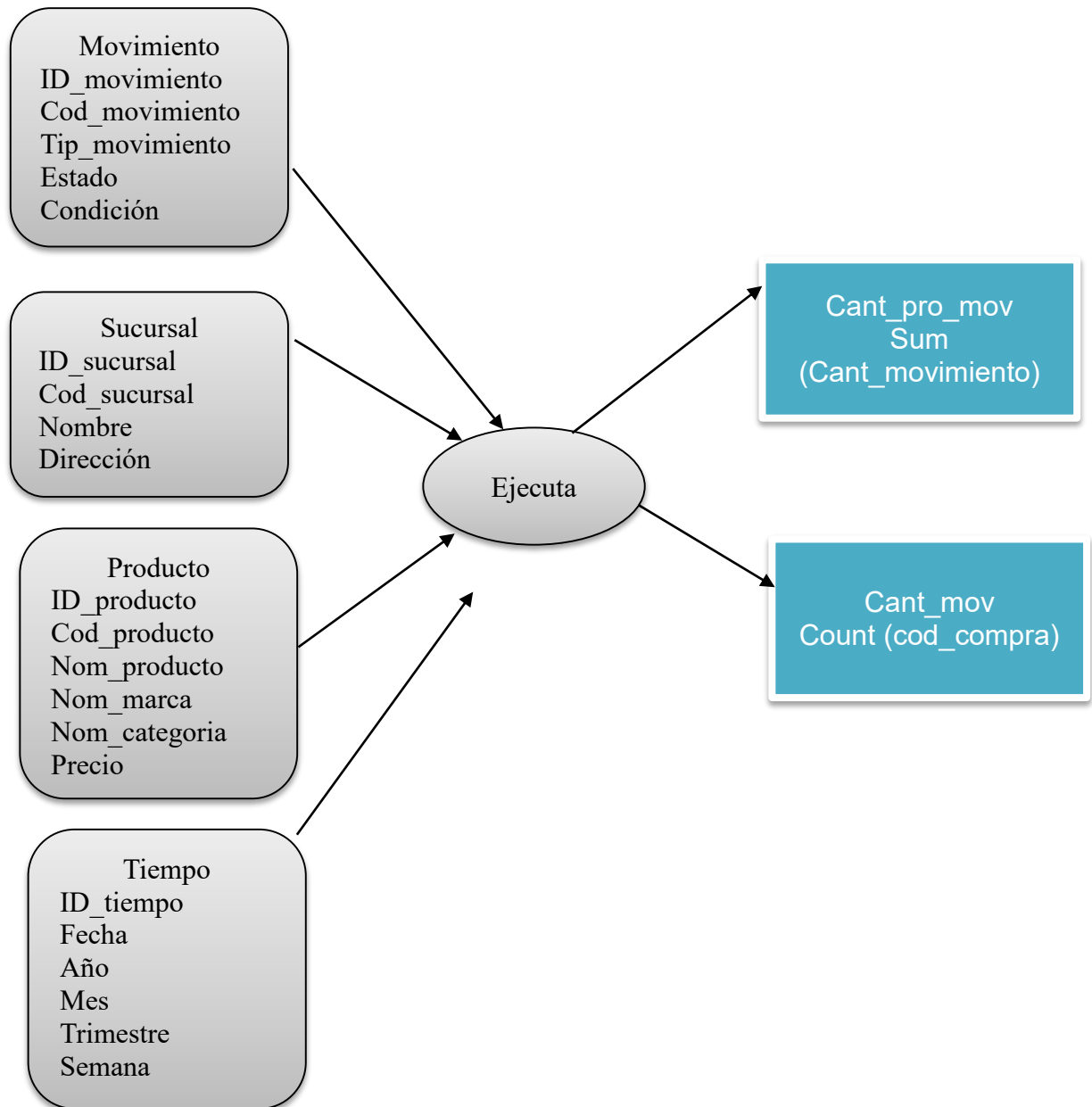


Ilustración 69. Modelo conceptual ampliado del proceso de almacén

4.16. Selección de las herramientas

Para mantener la integridad de los datos de selecciono el gestor de base de datos MySQL para la creación del BI en donde se almacenaran los datos correspondientes a los distintos procesos en estudio.

Pentaho Community Edition es una herramienta de software libre con la que se implementara la etapa de integración de datos, la creación de cubos dimensionales para la generación de los cálculos necesarios para los requerimientos.

Las características y funcionalidades del presente software están descritas en el marco teórico de este documento.

4.16.1. Instalación de las herramientas

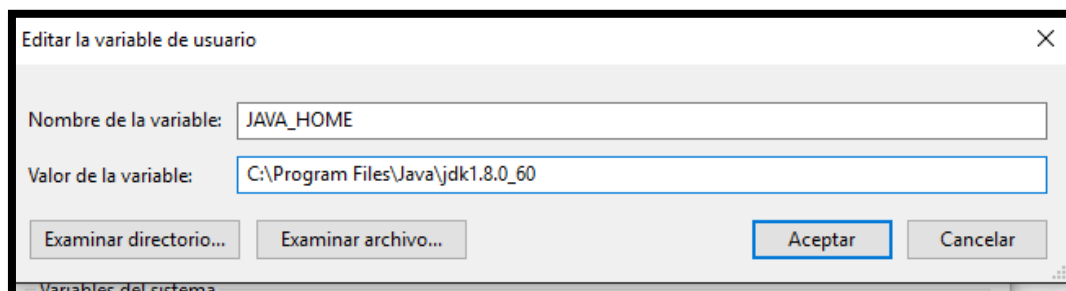
La instalación del gestor de base de datos MySQL ya se había realizado con anterioridad en el computador donde se realizara el proyecto

El Pentaho Integration se descargó desde su página principal:

<http://www.pentaho.com/product/data-integration>

- **Pentaho Data Integration 7.0 Kettle**
- **Pentaho schema-workbench 3.14**
- **Pentaho BI server 7**
- **Instalación del bd**

Configuración de las variables de entorno: classpath y JAVA_HOME con el jdk actual para el correcto funcionamiento del Pentaho ya que está diseñado en JAVA .



Pentaho Community no tiene un asistente de instalación como la mayoría de programas, solo se descarga un archivo que contiene carpetas con más archivos y su ejecutable es un archivo .bat llamado spoon y su ubicación será de acuerdo al usuario donde descomprima el archivo se creara su carpeta y componentes.

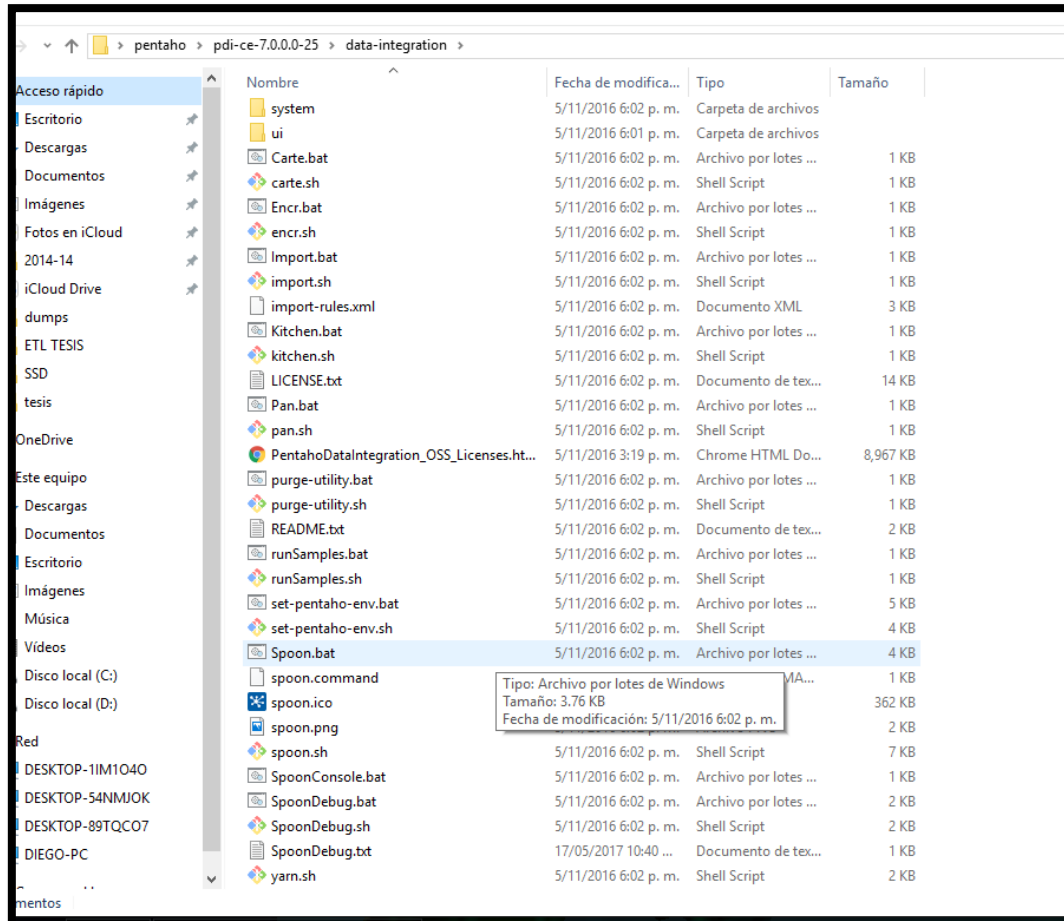


Ilustración 70. Archivos de Pentaho Data Integration

En la carpeta **C:\Users\USUARIO\Desktop\pentaho\pdi-ce-7.0.0-25\data-integration** se encuentra la herramienta spoon con la cual se realizara el proceso de ETL.

La siguiente herramienta que se instalara será el SCHEMA-WORKBENCH que será utilizada para la creación de cubos al igual que el data Integration este aplicativo de descarga como un archivo ZIP o RAR y se descomprime en la ubicación de la carpeta que el usuario desee.

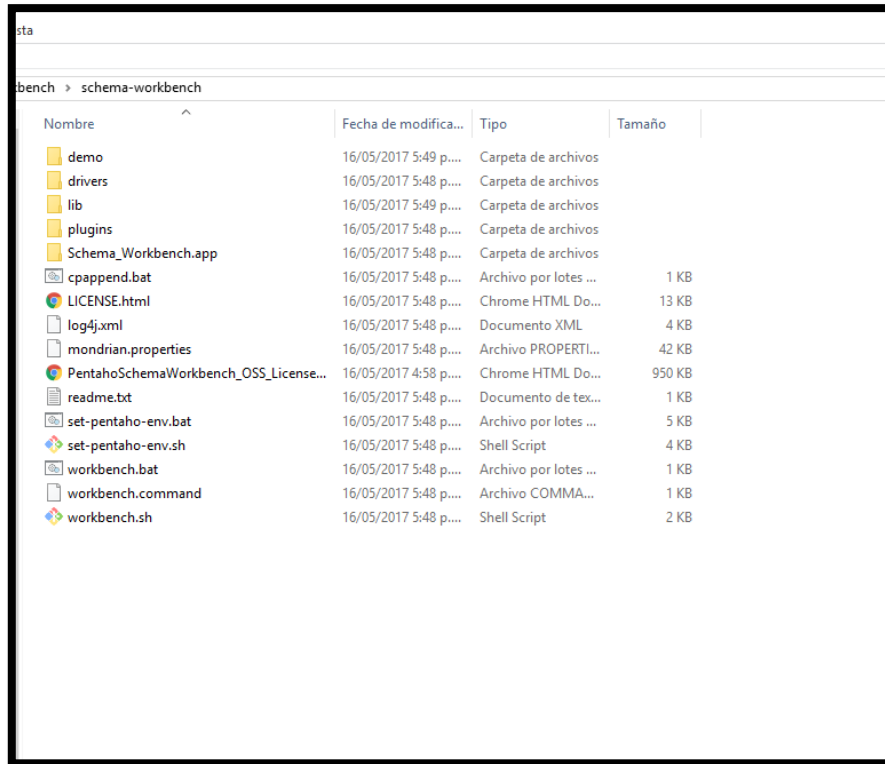


Ilustración 71. utilitarios del schema workbench

Para inicializar el Pentaho schema workbench primero se tendrá que inicializar el archivo llamada **set-pentaho-env.bat** para poder levantar los servicios del Pentaho luego para iniciar con la aplicación se debe activar el archivo llamado workbench.bat después de activar esto se abrirá la aplicación que nos ayudara para la construcción de los cubos .

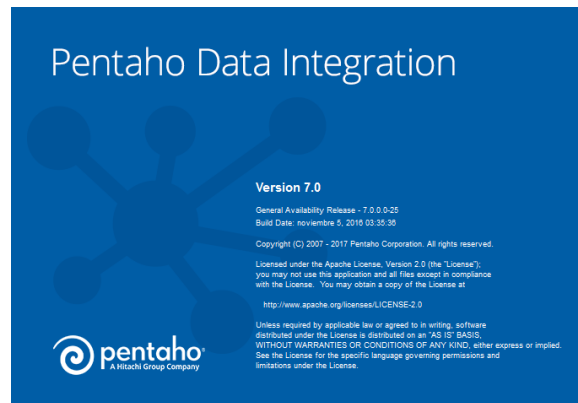


Ilustración 72.pentaho data Integration

El primer paso para empezar a trabajar con Pentaho es crear un repositorio de datos es una obligación para poder inicializar los controles que necesitaremos para realizar nuestro proceso ETL el proceso es muy sencillo es crear un repositorio con el nombre que desees y una contraseña, en este caso ya se tiene una cuenta creada en Pentaho llamada repositorio se pone el usuario y cuenta y ya se tiene acceso al data Integration.

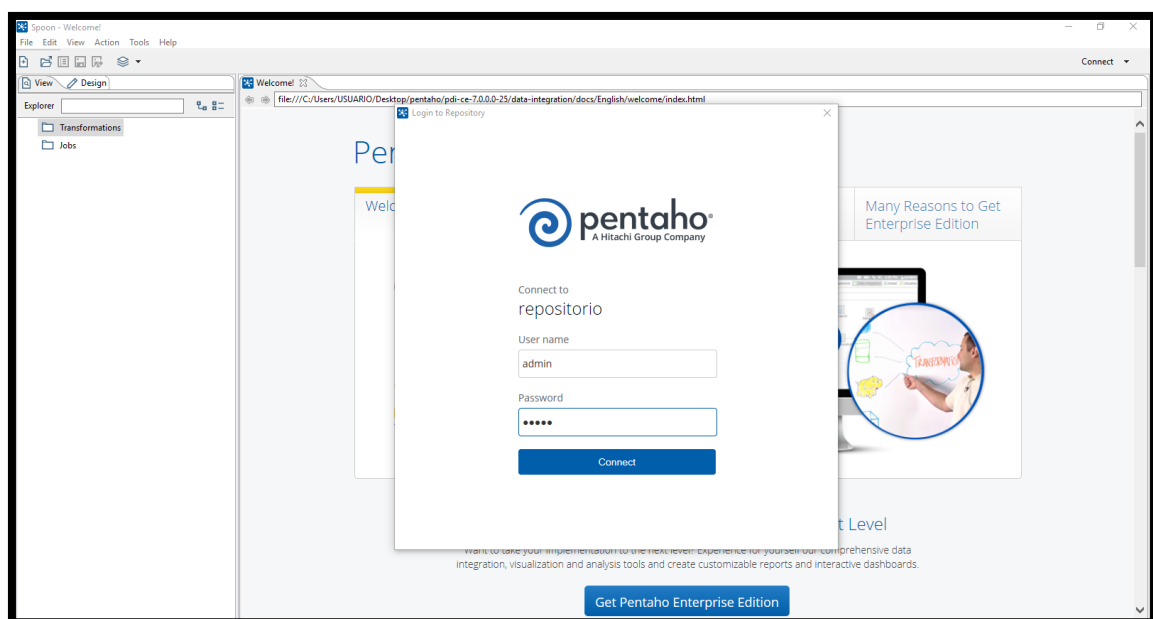


Ilustración 73.Inicio de sesión para el repositorio de Pentaho

4.17. Integración de datos

Después de haber realizado todos los pasos para la configuración del Pentaho se procederá a la realización del E

4.18. Proceso ETL para las dimensiones y hechos

Se procederá a realizar el proceso ETL de la aplicación BI iniciando con el poblamiento de las dimensiones Y terminando con el poblamiento de las tablas de hechos.

Dim_producto

Paso 1

El primer paso será poner un table input en el área de la transformación luego crear la conexión de la base de datos origen que se encuentra en el gesto SQL server

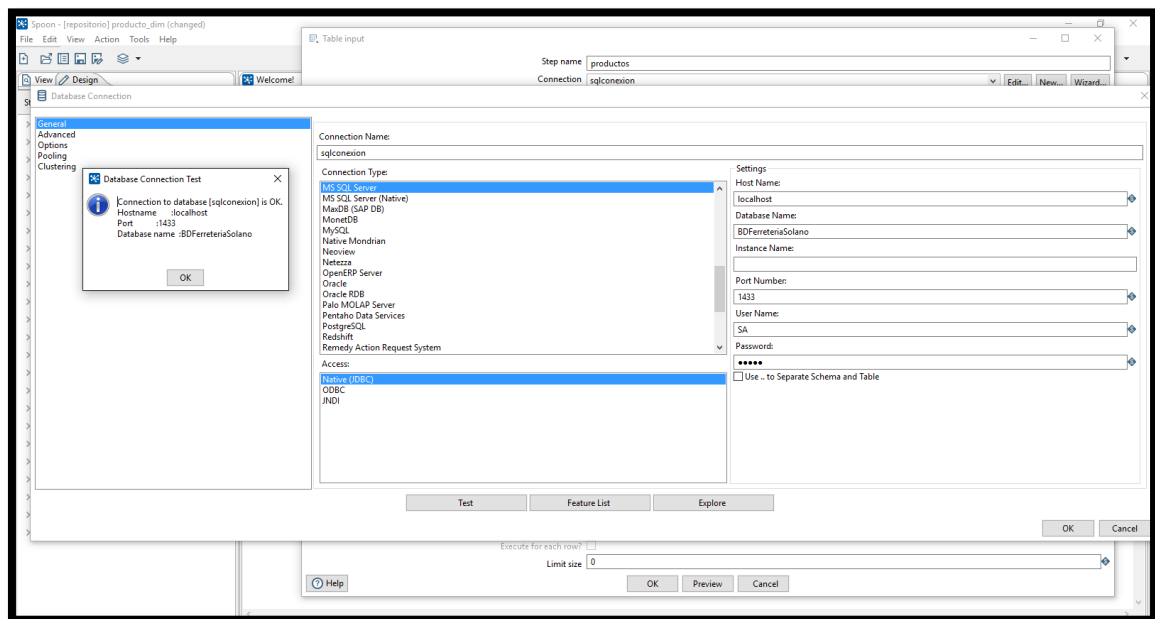
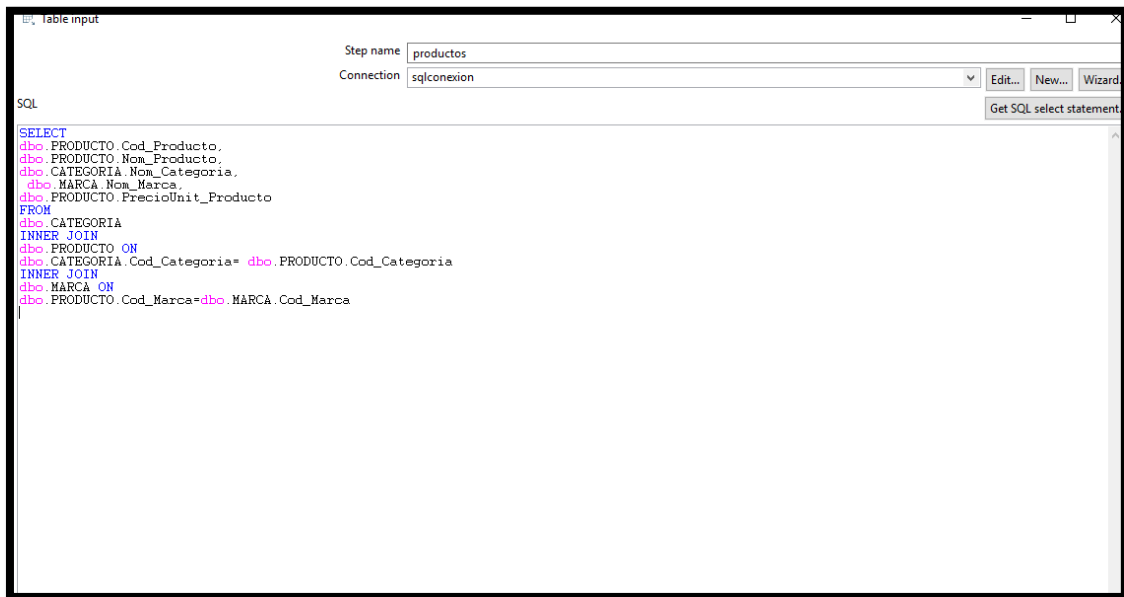


Ilustración 74.conexión de la base de datos origen

Pasó 2

Luego se eligen los campos necesarios mediante una consulta SQL en este caso para la dimensión producto elegiremos el Código de producto , nombre de producto , el precio del producto de la tabla productos , también elegiremos el nombre de la categoría de los productos de la tabla categoría y el nombre de la marca de la tabla Marca .



The screenshot shows a window titled "Table input" with a "Step name" of "productos" and a "Connection" of "sqlconexion". The SQL query is as follows:

```
SELECT
dbo.PRODUCTO.Cod_Producto,
dbo.PRODUCTO.Nom_Producto,
dbo.CATEGORIA.Nom_Categoria,
dbo.MARCA.Nom_Marca,
dbo.PRODUCTO.PrecioUnit_Producto
FROM
dbo.CATEGORIA
INNER JOIN
dbo.PRODUCTO ON
dbo.CATEGORIA.Cod_Categoria= dbo.PRODUCTO.Cod_Categoria
INNER JOIN
dbo.MARCA ON
dbo.PRODUCTO.Cod_Marca=dbo.MARCA.Cod_Marca
```

Ilustración 75.Consulta SQL para la elección de campos

Paso 3

se pone un table output en el área de la transformación y se hace la conexión correspondiente con la base de datos DW_Solano en el gesto MySQL se elige la tabla destino que será la tabla dim_producto de la base de datos DW_Solano que es el Data Warehouse

Paso 4

Se une el table input con el table output con la flecha de unión y se hace click derecho en la table output y damos click en la opción mapping unimos los campos elegidos de la table input de la base de datos transaccional y se unen con los campos de la tabla dimensional se hace click en la opción run en la parte superior y se procede con la carga de datos.

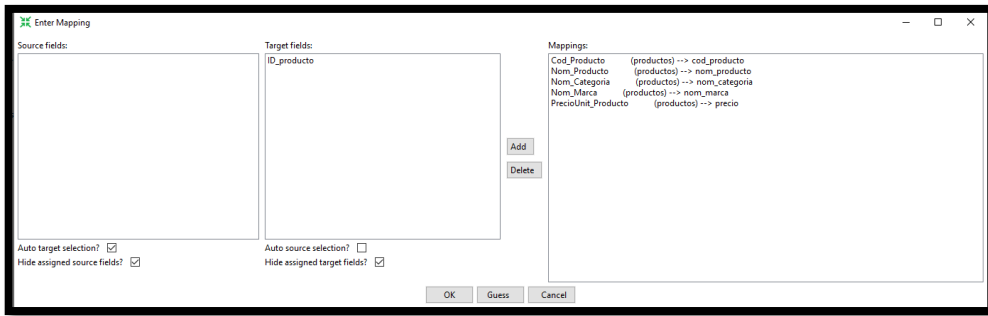


Ilustración 76. Mapeo de datos dim_producto

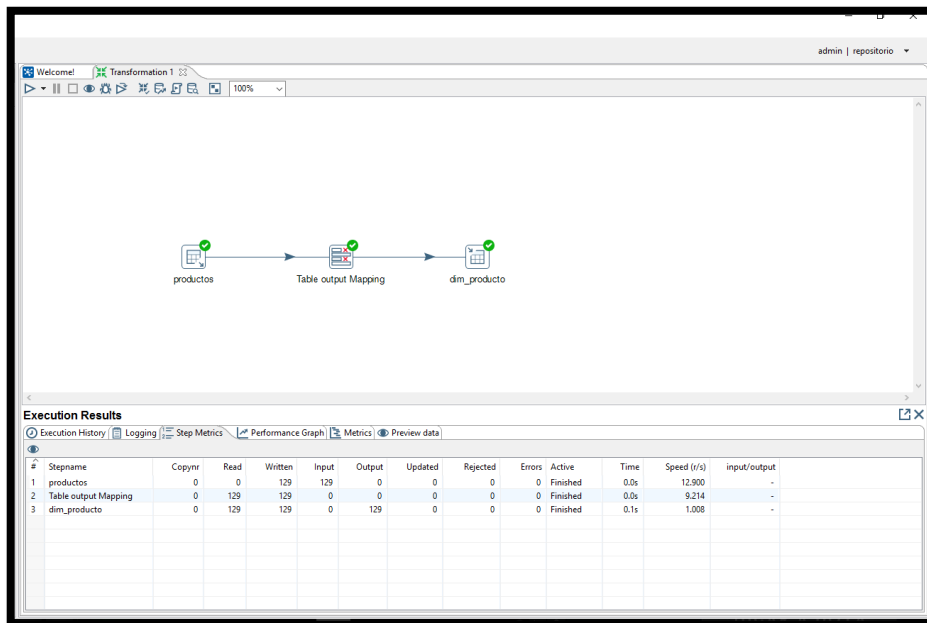


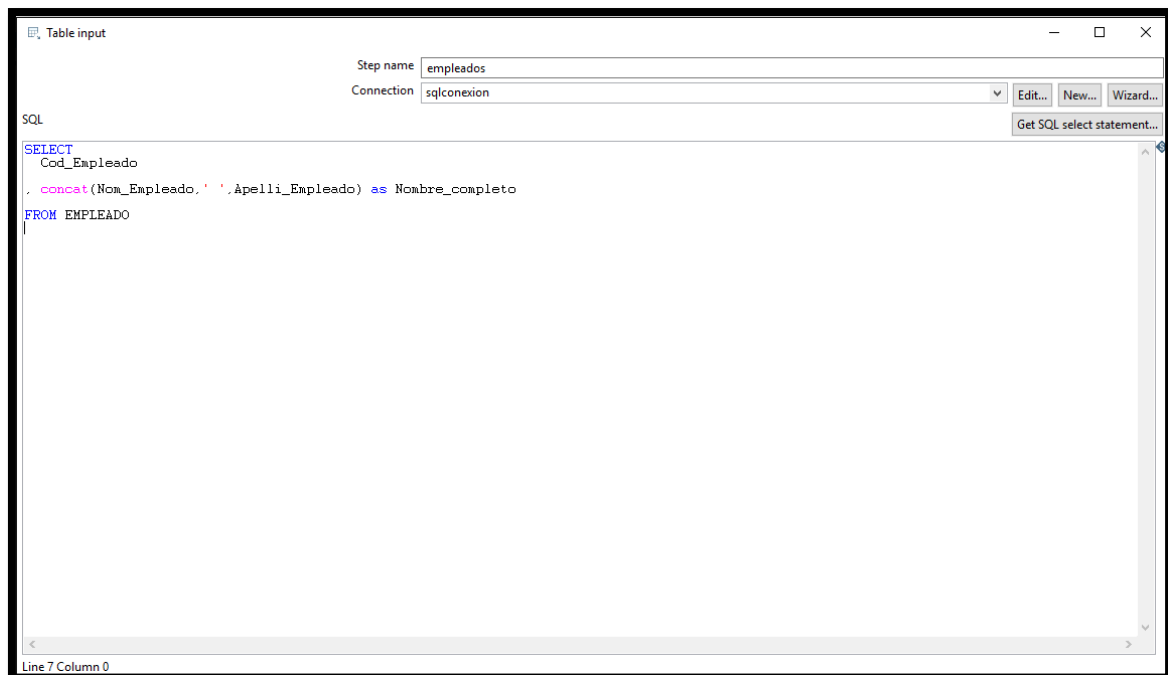
Ilustración 77. Proceso ETL culminado Dim_producto

Dim_empleado

Para las siguientes dimensiones se realizarán los mismos pasos de la dimensión antes explicadas para evitar mencionarlas de nuevo con diferencia que los campos elegidos para la conversión y carga serán distintos.

Paso 1

Para la tabla empleados solo necesitaremos el código y el nombre del empleado ya que en la base de datos transaccional el nombre está dividido en 2 campos utilizamos la función concat para unir el nombre y apellido en un solo campo



```
Table input
Step name empleados
Connection sqlconexion
Edit... New... Wizard...
Get SQL select statement...
SQL
SELECT
  Cod_Empleado
  , concat(Nom_Empleado, ' ', Apelli_Empleado) as Nombre_completo
FROM EMPLEADO
Line 7 Column 0
```

Ilustración 78.consulta para poblamiento Dim_empleado

Paso 2

Se mapean los campos origen con el destino

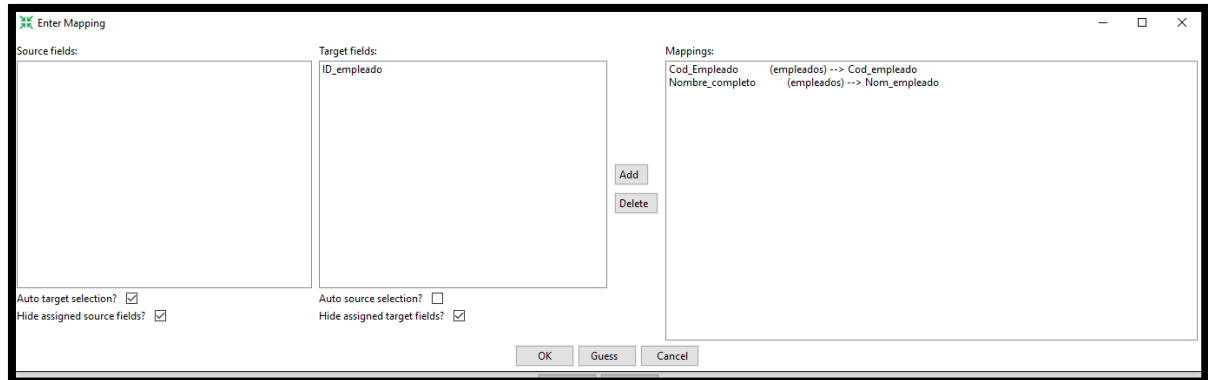


Ilustración 79. Mapeo Dim_Empleado

Paso 3

Se realiza la carga de datos en la tabla Dim_Empleado

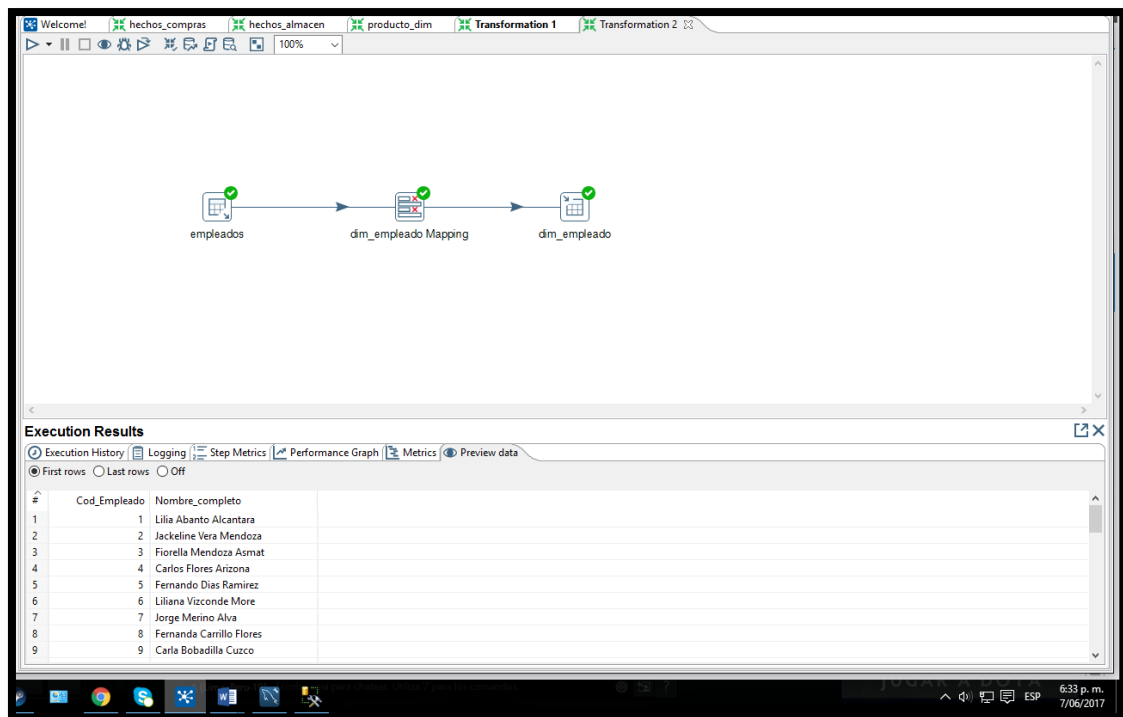


Ilustración 80. Carga de datos Dim_Empleado

Dim_sucursal

Se consultan los campos de la tabla empleado de la base de datos transaccional en el table input

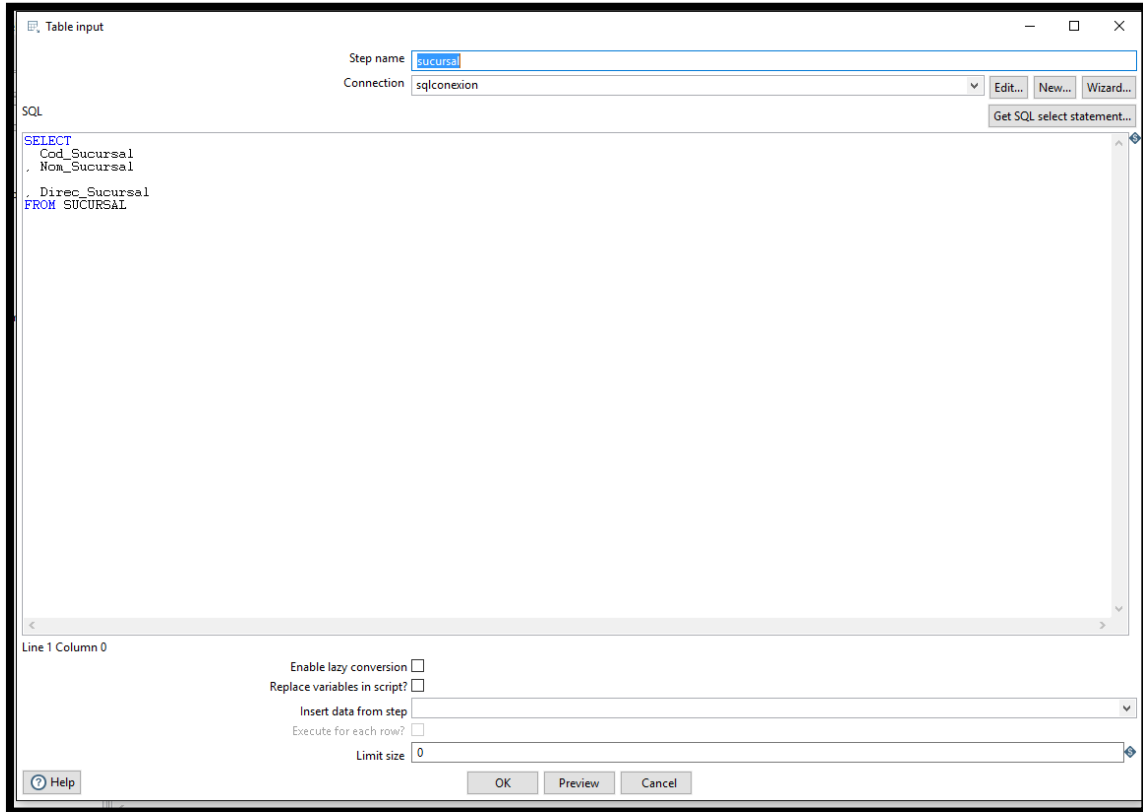


Ilustración 81, consulta para Dim_sucursal

Paso 2

Se mapean los campos del origen con el destino Dim_sucursal

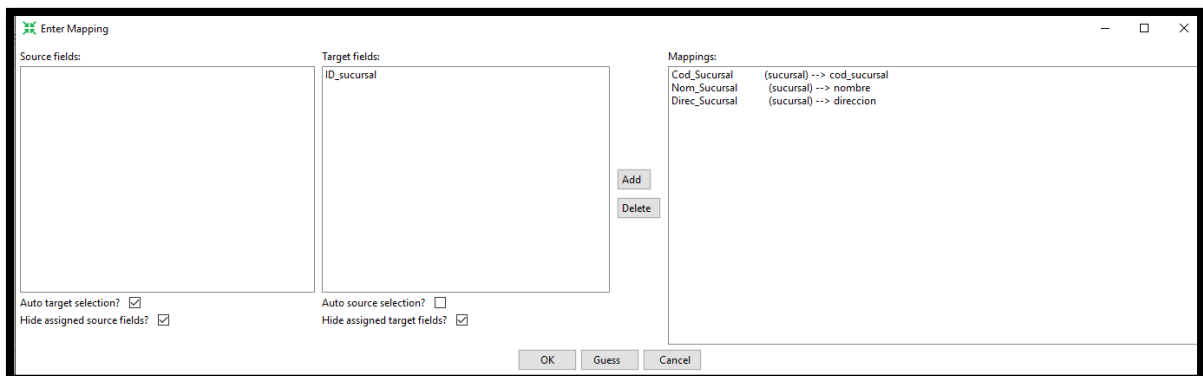


Ilustración 82.mapeo dim_producto

Paso 3

Se hace la carga de los datos

The screenshot displays the SQL Server Data Tools interface. At the top, a data flow diagram shows three steps: 'sucursal', 'dim_sucursal Mapping', and 'dim_sucursal', all marked with green checkmarks. Below the diagram, the 'Execution Results' window is open, showing a table with the following data:

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	sucursal	0	0	3	3	0	0	0	0	Finished	0.0s	375	-
2	dim_sucursal Mapping	0	3	3	0	0	0	0	0	Finished	0.0s	300	-
3	dim_sucursal	0	3	3	0	3	0	0	0	Finished	0.1s	25	-

Ilustración 83, carga de datos Dim_sucursal

Dim_cliente

Paso 1

Se eligen los campos de la base de datos transaccional mediante una consulta SQL

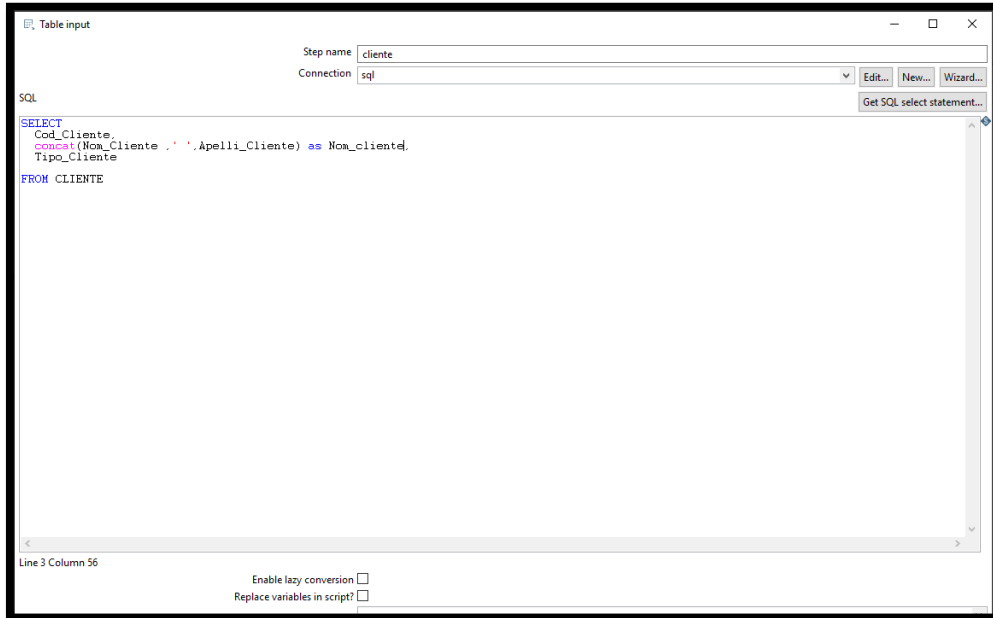


Ilustración 84.Consulta SQL para Dim_cliente

Paso 2

Se mapean los campos de origen con los de destino Dim_cliente

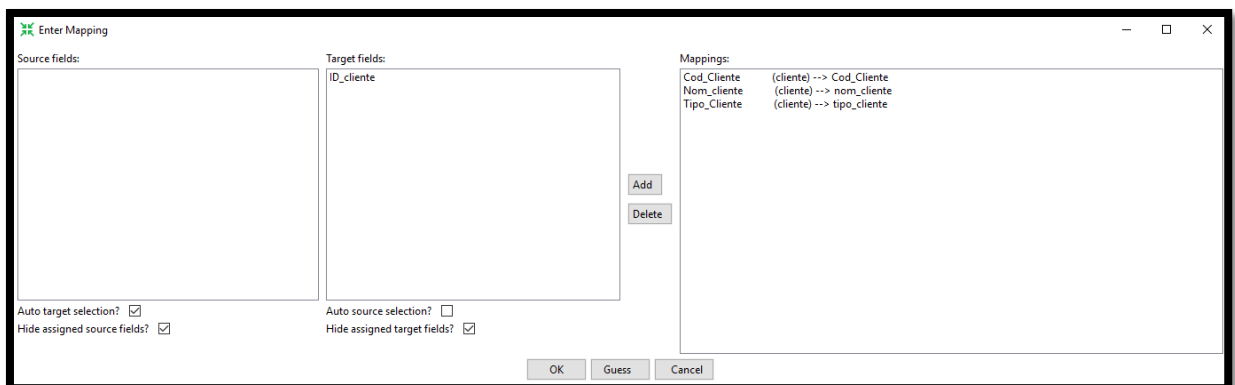


Ilustración 85.Mapeo Dim_cliente

Paso 3

Ejecutar la Carga de los datos hacia la dimensión Dim_cliente

The screenshot displays a data loading process in a BI tool. The top part shows a flow diagram with three components: 'cliente', 'dim_cliente Mapping', and 'dim_cliente', all with green checkmarks. The bottom part shows 'Execution Results' with a table of data.

#	Cod_Cliente	nom_cliente	tipo_cliente
1	1	GracielaAbanto Mendez	Ciente Natural
2	2	MariaAlgoo Carrion	Ciente Natural
3	3	MarthaAbraham Rondon	Ciente Natural
4	4	PabloAcosta Rivera	Ciente Natural
5	5	SergioVela Romero	Ciente Natural
6	6	HugoAlva Lecca	Ciente Natural
7	7	EmilioAforga Zegarra	Ciente Natural
8	8	MartinArgomedo Collantes	Ciente Natural
9	9	DiegoAyala Rodriguez	Ciente Natural

Ilustración 86.carga de datos Dim_cliente

Dim_tiempo_ventas, Dim_tiempo_almacen, Dim_tiempo_ventas

Paso 1

En este paso se extraerá la fecha de la tabla **Detalle_comprobante** de la base de datos transaccional y se hará una conversión de la fecha en varios niveles mediante SQL.

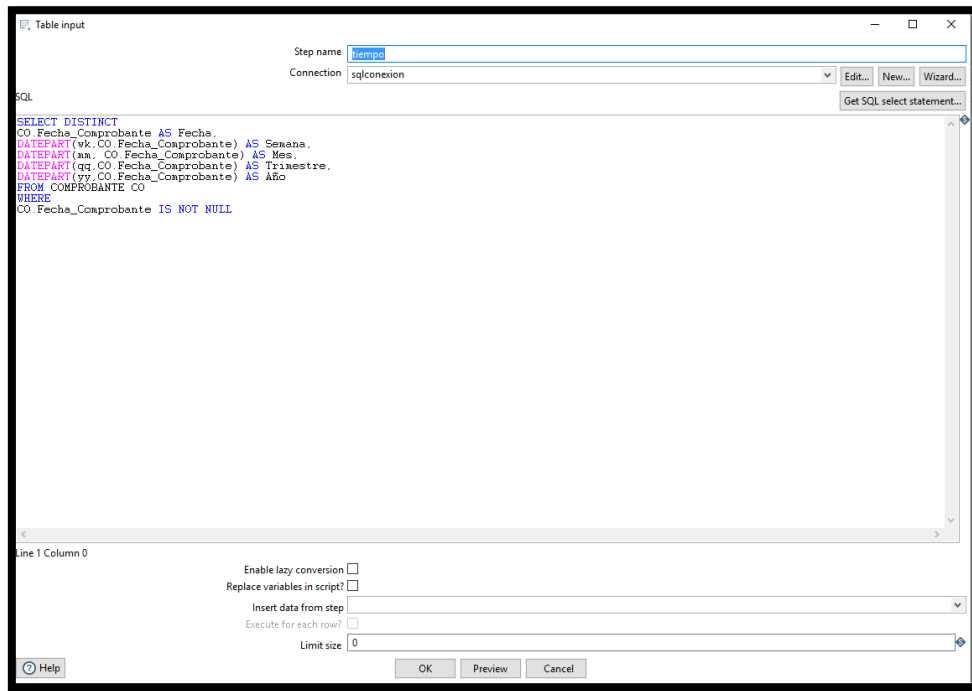


Ilustración 87. Consulta SQL para Dim_tiempo_ventas

Paso 2

Se hace una transformación de campos con el control value mapper cambiando algunos valores en mes y trimestre y se procede a mapear

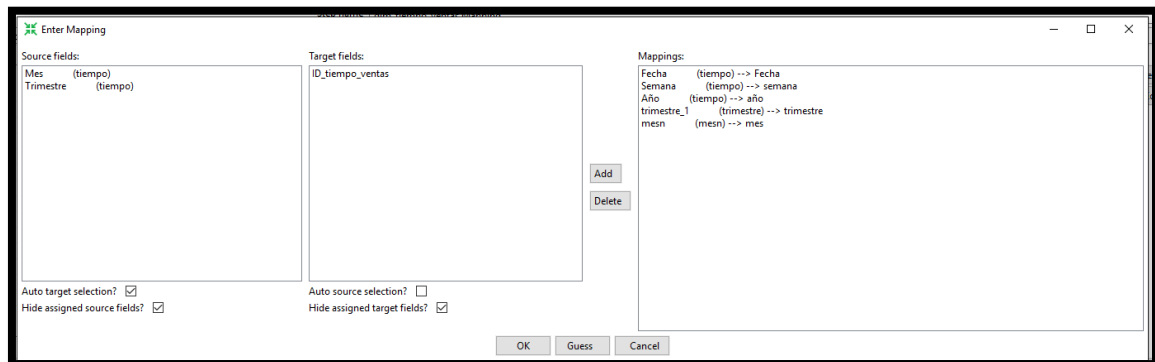


Ilustración 88. Mapeo tabla Dim_tiempo_ventas

Paso 3

Se ejecuta la carga de datos para la tabla Dim_tiempo_ventas

The screenshot displays a data load process in SQL Server Data Tools. The process flow consists of five steps: 'tiempo', 'mesn', 'trimestre', 'dim_tiempo_ventas Mapping', and 'dim_tiempo_ventas'. Each step is marked with a green checkmark, indicating successful execution. Below the flow, the 'Execution Results' window is open, showing a table with the following data:

#	Fecha	semana	año	trimestre	mes
1	2016-11-06	45	2016	T4	noviembre
2	2012-04-22	17	2012	T2	abril
3	2010-12-21	52	2010	T4	diciembre
4	2011-09-10	37	2011	T3	septiembre
5	2012-01-16	4	2012	T1	enero
6	2012-03-21	13	2012	T1	marzo
7	2011-08-12	33	2011	T3	agosto
8	2011-11-09	46	2011	T4	noviembre
9	2011-10-25	44	2011	T4	octubre

Ilustración 89. Carga de datos para Dim_tiempo_ventas

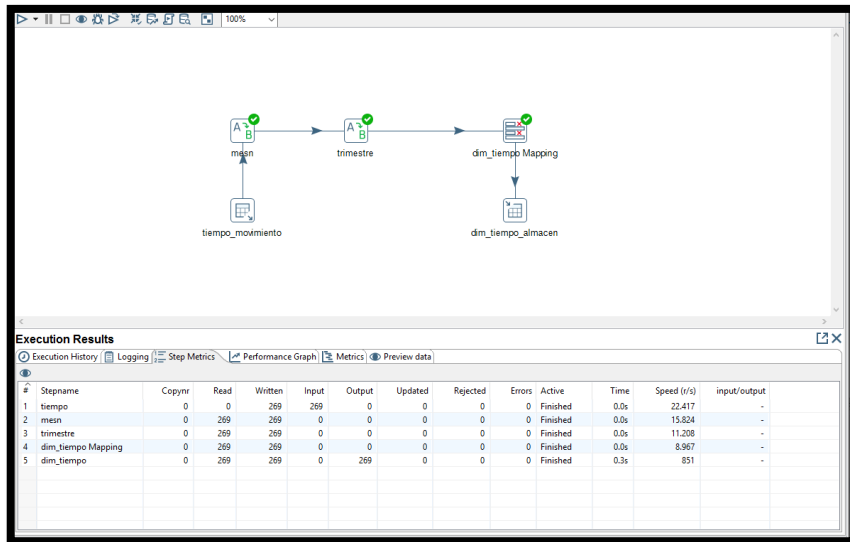


Ilustración 90. Carga de datos Dim_tiempo_almacen

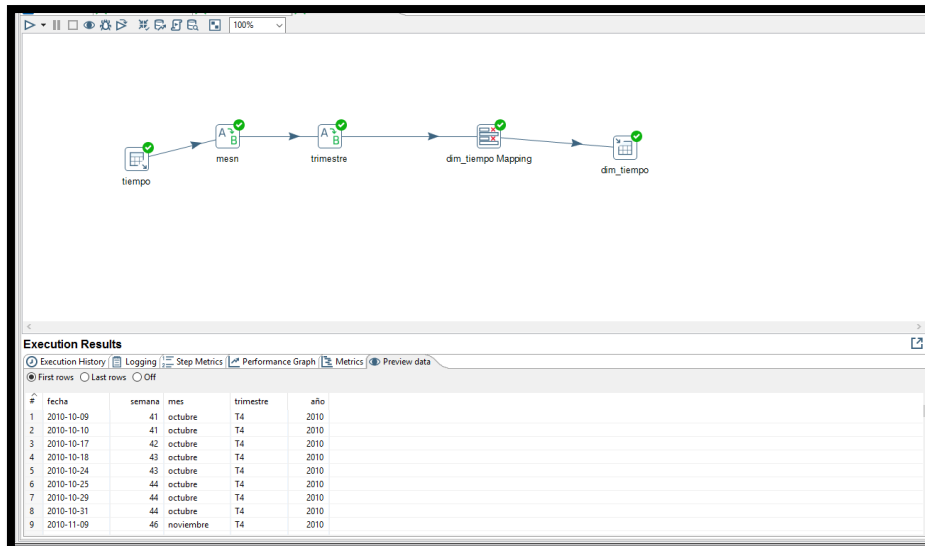


Ilustración 91. Carga de tabla Dim_tiempo_compras

Dim_proveedor

Paso 1

Se hace la conexión en el table input y se genera la consulta SQL que elegirá los campos que se van a extraer.

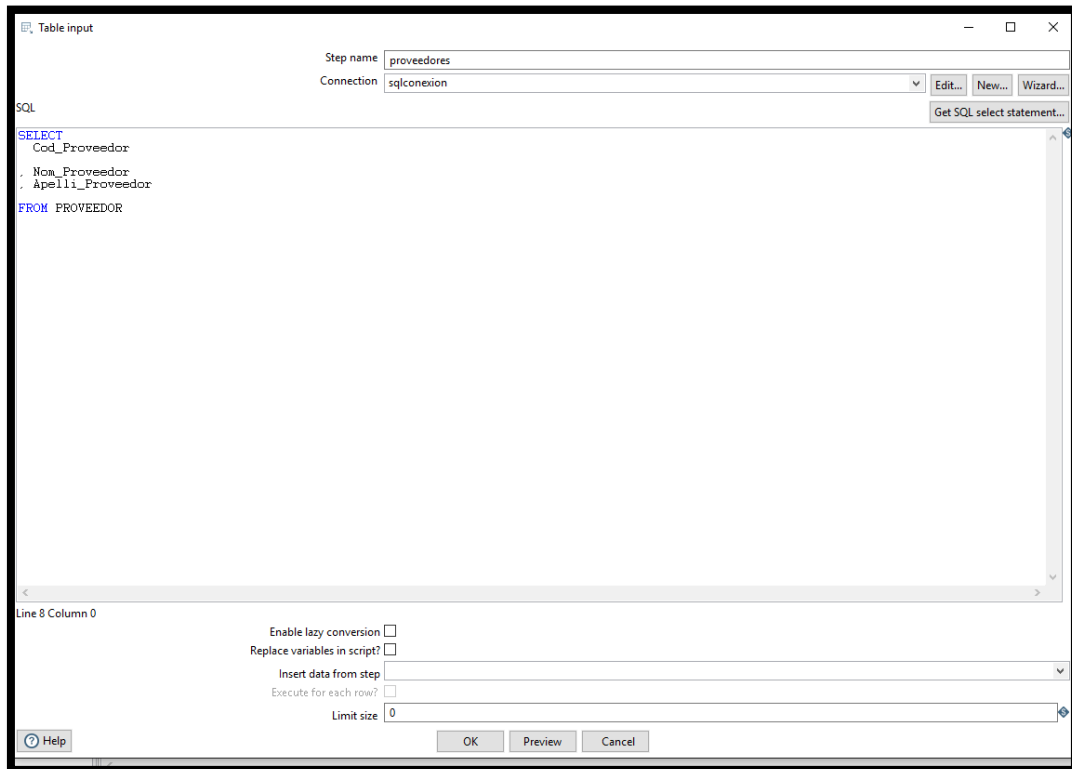


Ilustración 92.consulta SQL para la tabla Dim_proveedor

Paso 2

Se mapean los datos en el table output Dim_proveedor

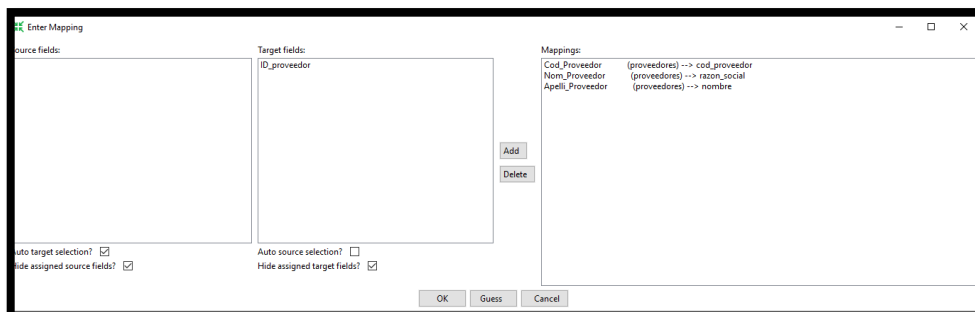


Ilustración 93.mapeo tabla Dim_proveedor

Paso 3

Carga de datos a la tabla Dim_proveedor

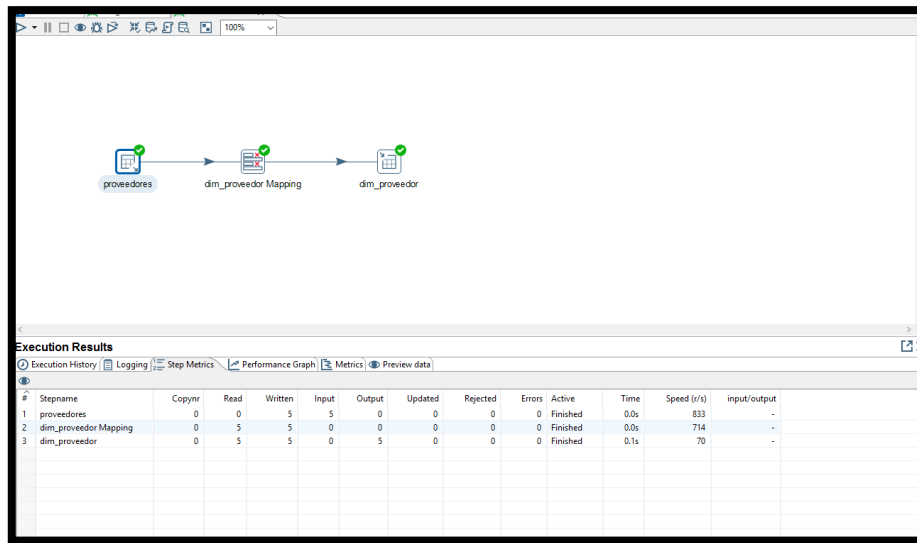


Ilustración 94.carga de datos Dim_proveedor

Dim_material

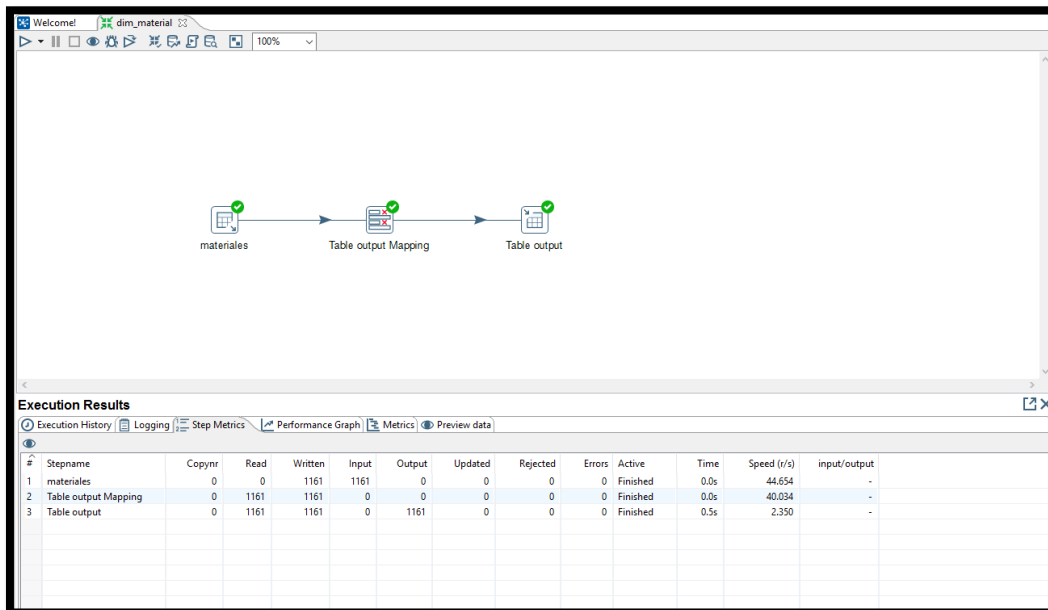


Ilustración 95.Carga de datos de tabla Dim_material

Dim_ubigeo

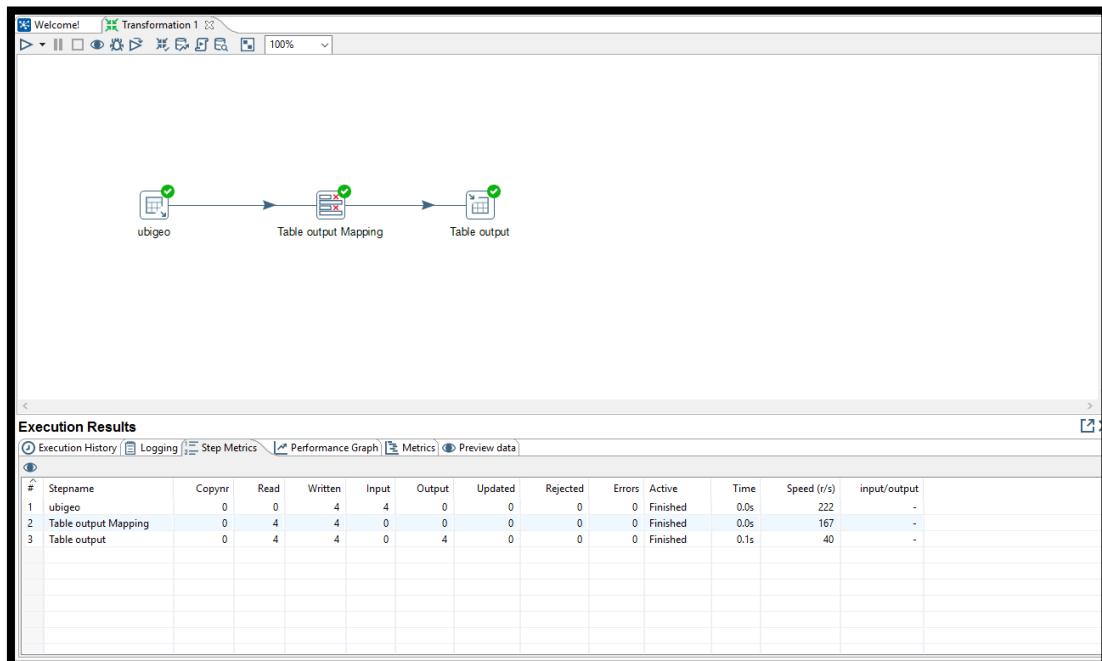


Ilustración 96. carga de datos de la tabla Dim_ubigeo

Compra_hechos

Para la población de la tabla de hechos se realizaron muchos procesos de los cuales mencionaremos los principales primero elegimos datos de las dimensiones involucradas con la tabla de hechos para realizar una conexión de datos dimensionales con datos transaccionales con tal de conectar las base de datos transaccional con la dimensional poder realizar los cálculos necesarios para poder extraer todas las filas que poblaran la tabla de hechos los atributos que irán en cada tabla ya fueron descritos anteriormente en el punto **Establecer correspondencias**.

Paso 1

Conectamos una dimensión con el control database join para hacer el join de las tablas de la base de datos transaccional y elegir los campos que se calcularán.

Database Join

Step name: Database join

Connection: sqlconexion

SQL

```
SELECT
  Nom_Proveedor,
  Nom_Sucursal,
  DATEPART(wk, CM.Fecha_Compra) AS Semana,
  DATEPART(mn, CM.Fecha_Compra) AS Mes,
  DATEPART(qt, CM.Fecha_Compra) AS Trimestre,
  DATEPART(yy, CM.Fecha_Compra) AS Año,
  Distrito,
  CT.Cod_Compra AS Cantidad,
  CM.monto_compra AS Total_Compras,
FROM
  PROVEEDOR INNER JOIN COMPRAS AS CM ON PROVEEDOR.Cod_Proveedor = CM.Cod_Proveedor
  INNER JOIN COTIZACION AS CT ON CM.Cod_Compra = CT.Cod_Compra
  INNER JOIN SUCURSAL ON PROVEEDOR.Cod_Ubigeo = SUCURSAL.Cod_Ubigeo
  ON CT.Cod_Sucursal = SUCURSAL.Cod_Sucursal
WHERE
  CT.Cod_Compra = ? AND CT.Descrip_ProdCom = ?
```

Line 12 Column 1

Number of rows to return: 0

Outer join?

Replace variables

The parameters to use:

#	Parameter fieldname	Parameter Type
1	nro_pedido	String
2	Descripcion	String

Help OK Cancel Get Fields

Ilustración 97.consulta para la tabla de compra_hechos

Paso 2

Igualar los campos elegidos de las dimensiones con los campos transaccionales generados por el database join con el control stream value lookup después de eso ordenamos las claves con el control sort rows y con el control Group By elegimos los campos que se van a calcular.

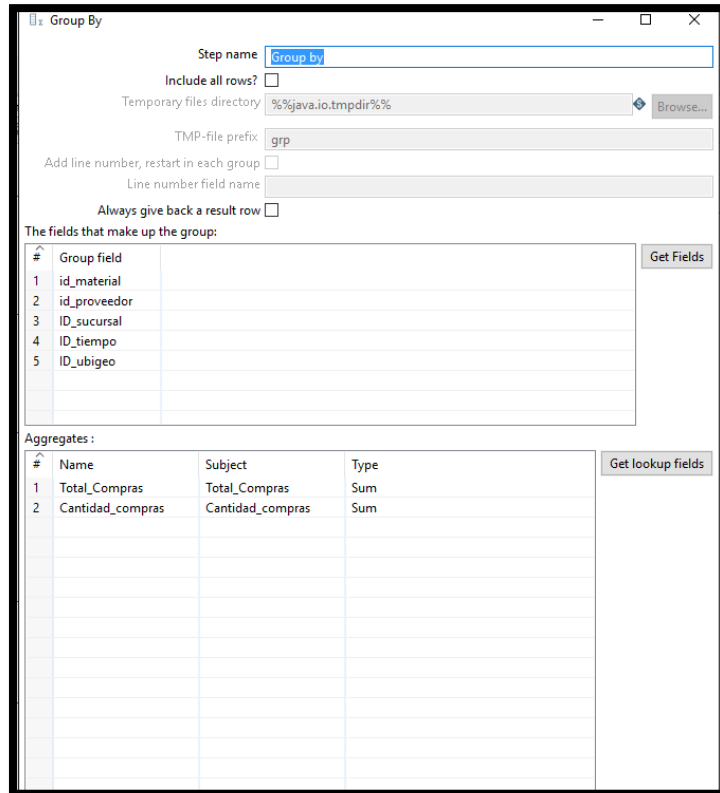


Ilustración 98.ordenar filas y cálculos compra_ hechos

Paso 3

Se mapean los campos con la tabla de hechos

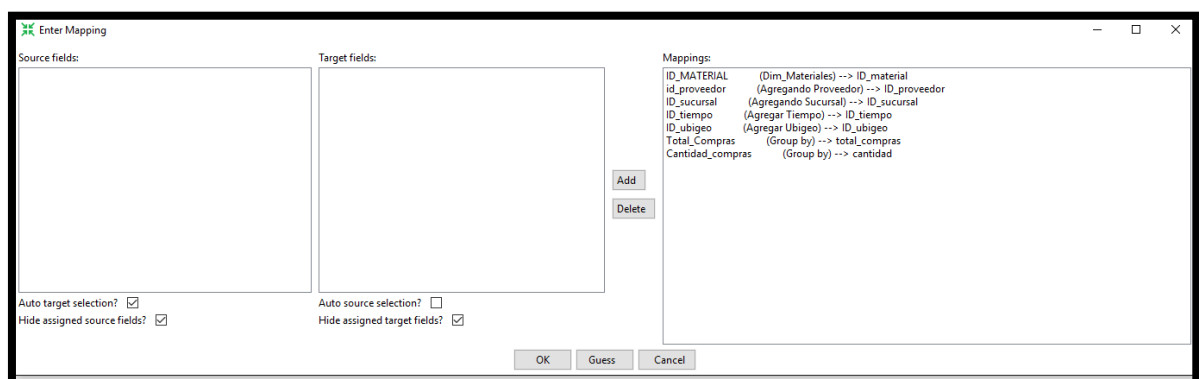


Ilustración 99.mapeo compra_ hechos

Se procede con la ejecución de la solución para la carga de datos en la tabla Hechos_compras.

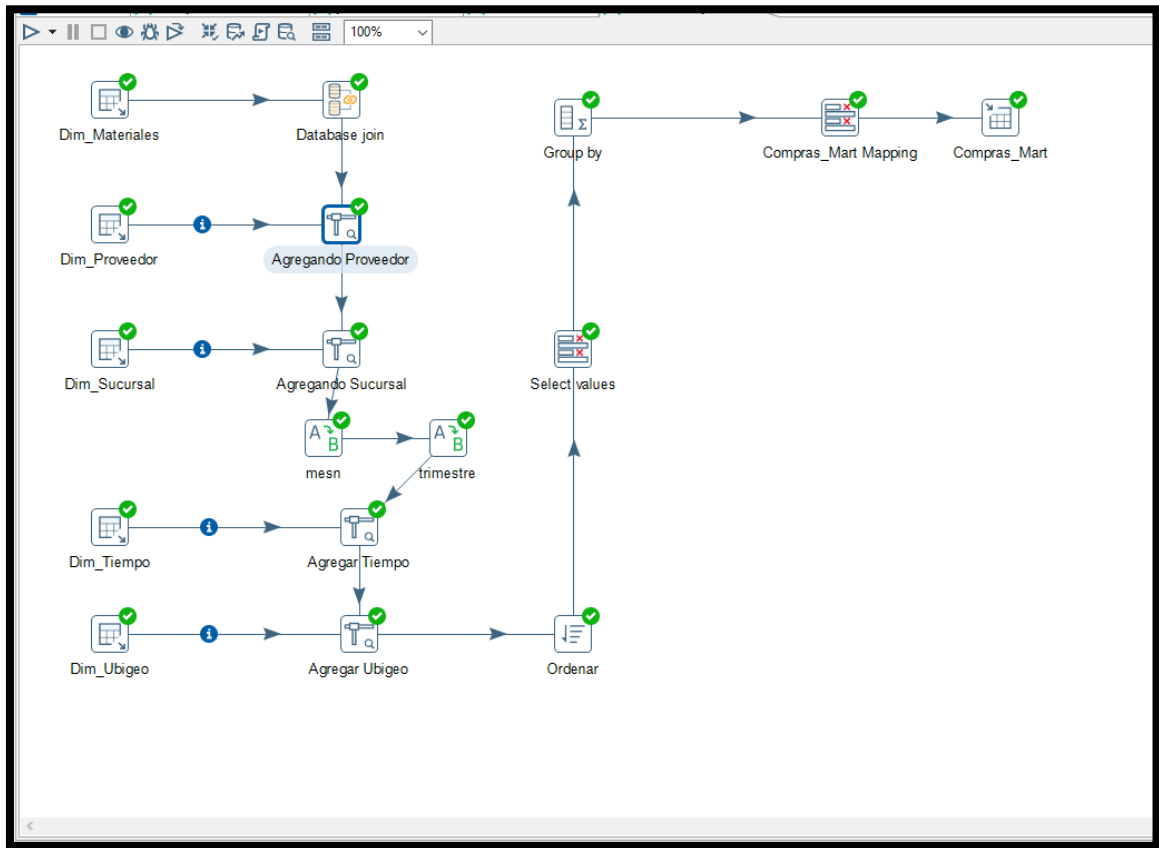
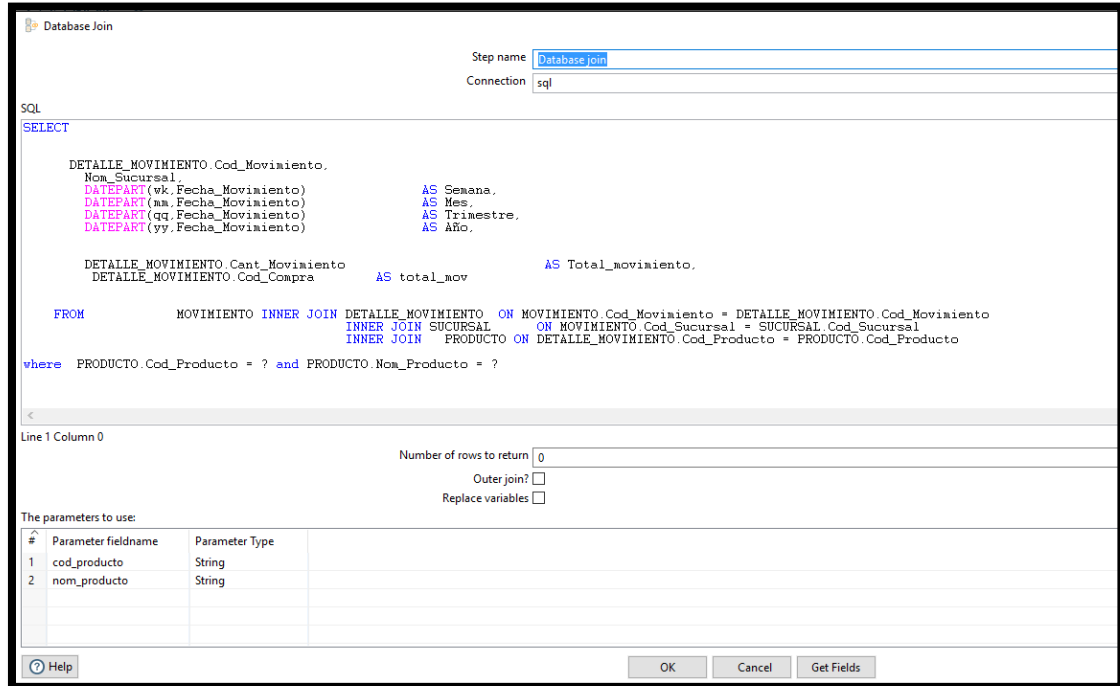


Ilustración 100.carga de datos compra_hechos

Almacen_hechos

Paso 1

Al igual que en la tabla de hechos anterior de eligen los campos que se igualaran a las dimensiones y los campos que se utilizaran para los cálculos con el control database join



The screenshot shows a 'Database Join' dialog box with the following content:

Step name: Database join
Connection: sql

SQL

```
SELECT
DETALLE_MOVIMIENTO.Cod_Movimiento,
Nom_Sucursal,
DATEPART(wk, Fecha_Movimiento) AS Semana,
DATEPART(ma, Fecha_Movimiento) AS Mes,
DATEPART(qq, Fecha_Movimiento) AS Trimestre,
DATEPART(yy, Fecha_Movimiento) AS Año,
DETALLE_MOVIMIENTO.Cant_Movimiento AS total_mov,
DETALLE_MOVIMIENTO.Cod_Compra AS Total_movimiento.
FROM
MOVIMIENTO INNER JOIN DETALLE_MOVIMIENTO ON MOVIMIENTO.Cod_Movimiento = DETALLE_MOVIMIENTO.Cod_Movimiento
INNER JOIN SUCURSAL ON MOVIMIENTO.Cod_Sucursal = SUCURSAL.Cod_Sucursal
INNER JOIN PRODUCTO ON DETALLE_MOVIMIENTO.Cod_Producto = PRODUCTO.Cod_Producto
where PRODUCTO.Cod_Producto = ? and PRODUCTO.Nom_Producto = ?
```

Line 1 Column 0

Number of rows to return: 0
Outer join?:
Replace variables:

The parameters to use:

#	Parameter fieldname	Parameter Type
1	cod_producto	String
2	nom_producto	String

Buttons: Help, OK, Cancel, Get Fields

Ilustración 101.consulta para almacen_hechos

Paso 2

Ordenamos los campos de la tabla de hechos y elegimos la función de con la que se va a calcular los hechos.

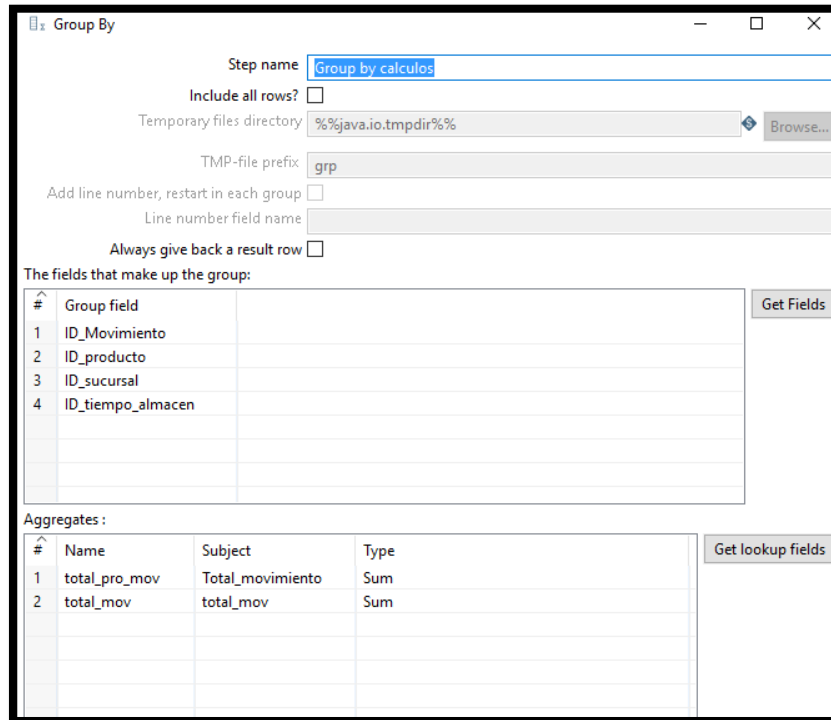


Ilustración 102. Ordenar campos Almacen_ hechos

Paso 3

Se mapean los campos de la bd transaccional con los campos de la tabla de hechos

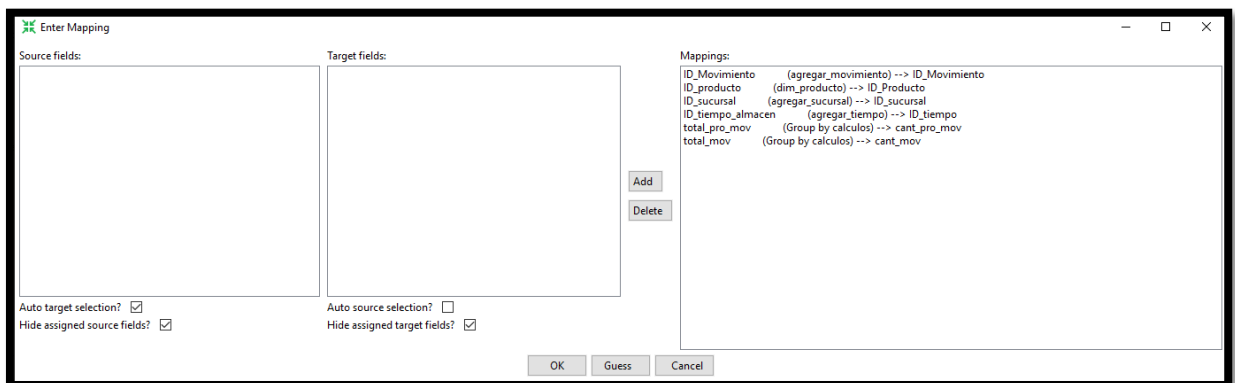


Ilustración 103.mapeo de la tabla Almacen_ hechos

Se ejecuta la carga de datos para la tabla de Hechos_almacen

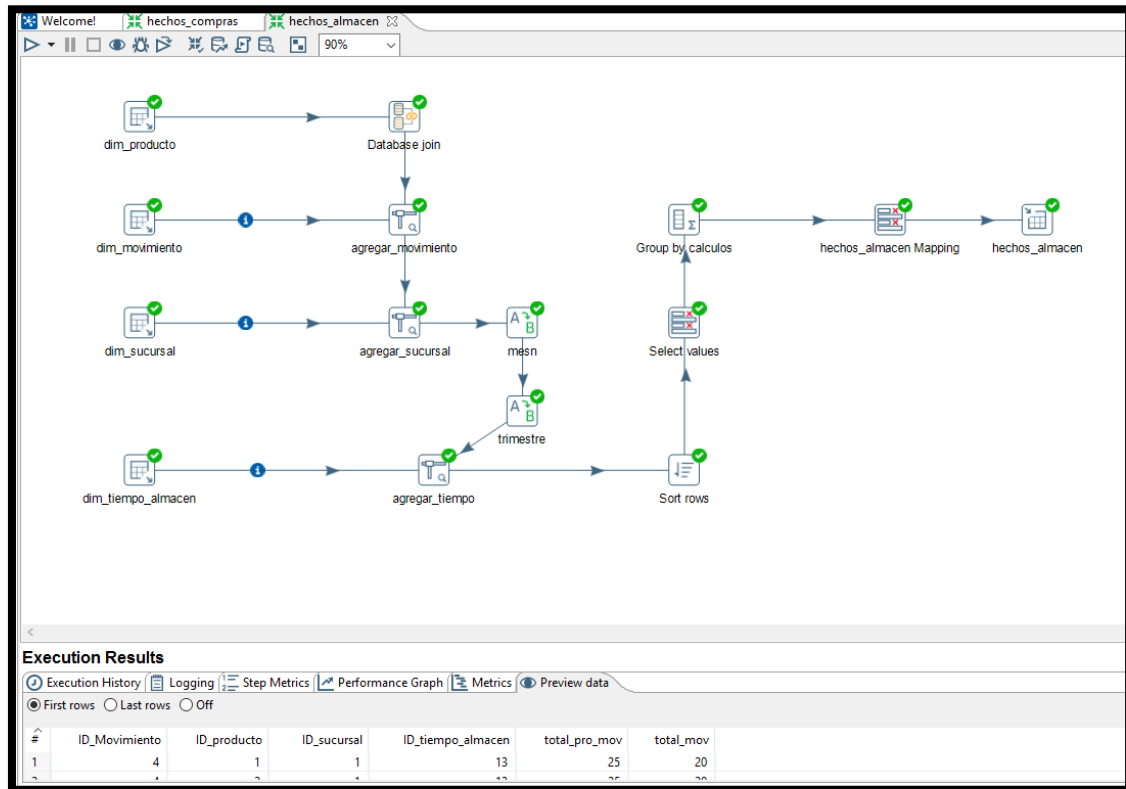


Ilustración 104.carga de datos de la tabla Almacen_hechos

Ventas_hechos

Paso 1

Se eligen las dimensiones involucradas elegimos la primera dimensión y la unimos con el database join ingresamos la consulta que generara los campos de la base de datos transaccional con las dimensiones.

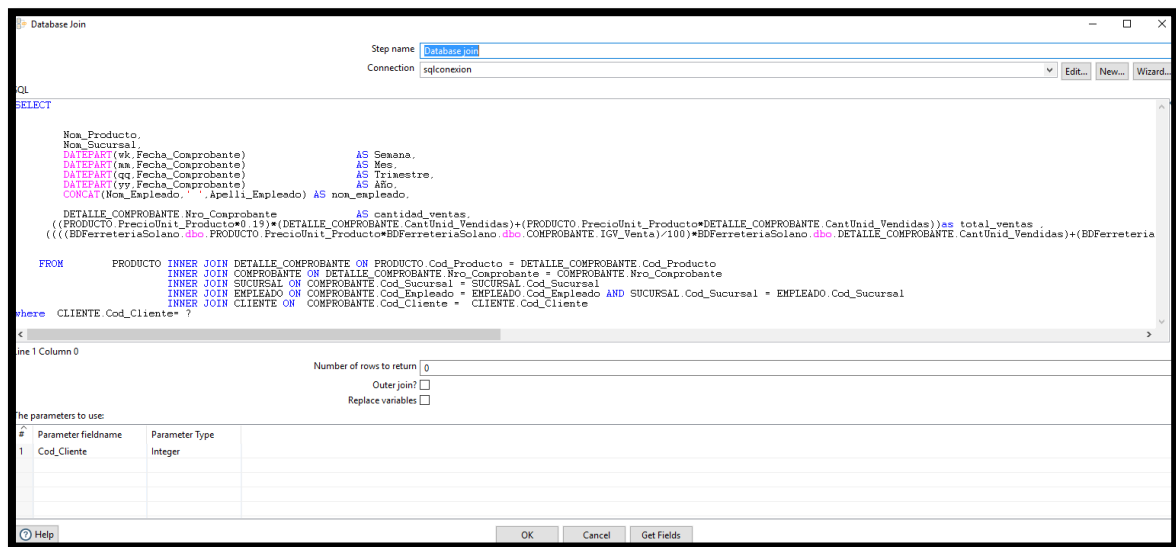


Ilustración 105.consulta SQL para ventas_hechos

Pasó 2

Se ordenan las filas y se eligen los campos que se calcularán con su respectiva función

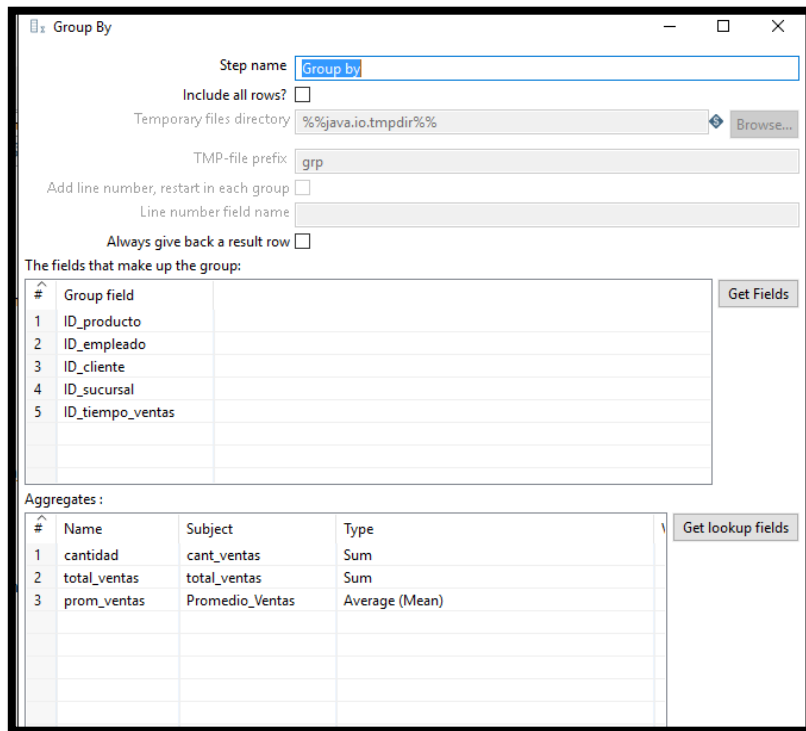


Ilustración 106.ordenar campos ventas_hechos

Paso 3

Se mapean los campos transaccionales con los campos de la tabla Hechos_Ventas

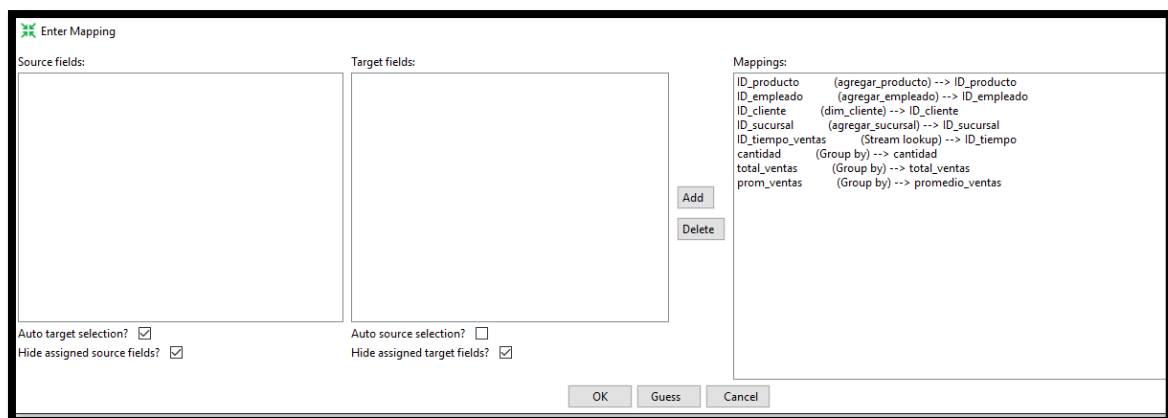


Ilustración 107.mapeo ventas_hechos

Y se procede a la carga de datos

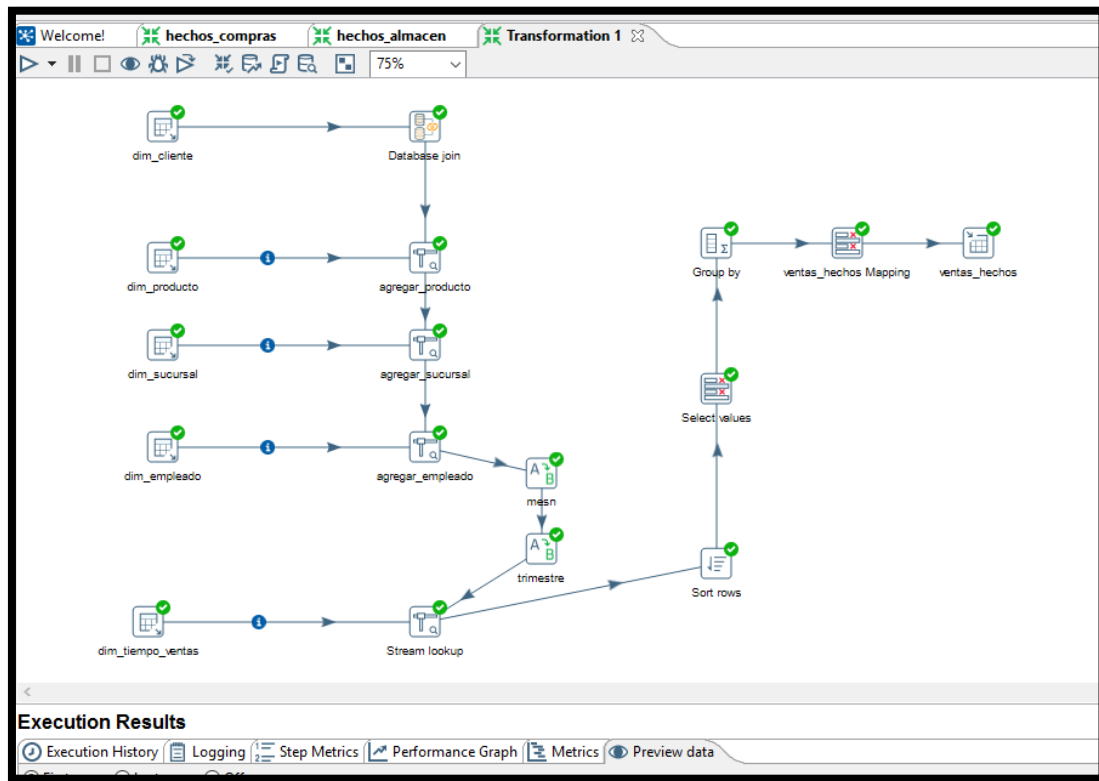


Ilustración 108.carga de datos ventas_hechos

4.19. Creación del cubo OLAP

Para el siguiente paso del proyecto se procederá a crear el cubo OLAP de los procesos de la empresa se crearan 3 cubos de acuerdo a las tablas de hechos que existen en la base de datos dimensional para la creación del proyecto usaremos la herramienta de Pentaho SCHEMA-WORKBENCH. El primer paso para la creación del cubo será la conexión a la base de datos dimensional.

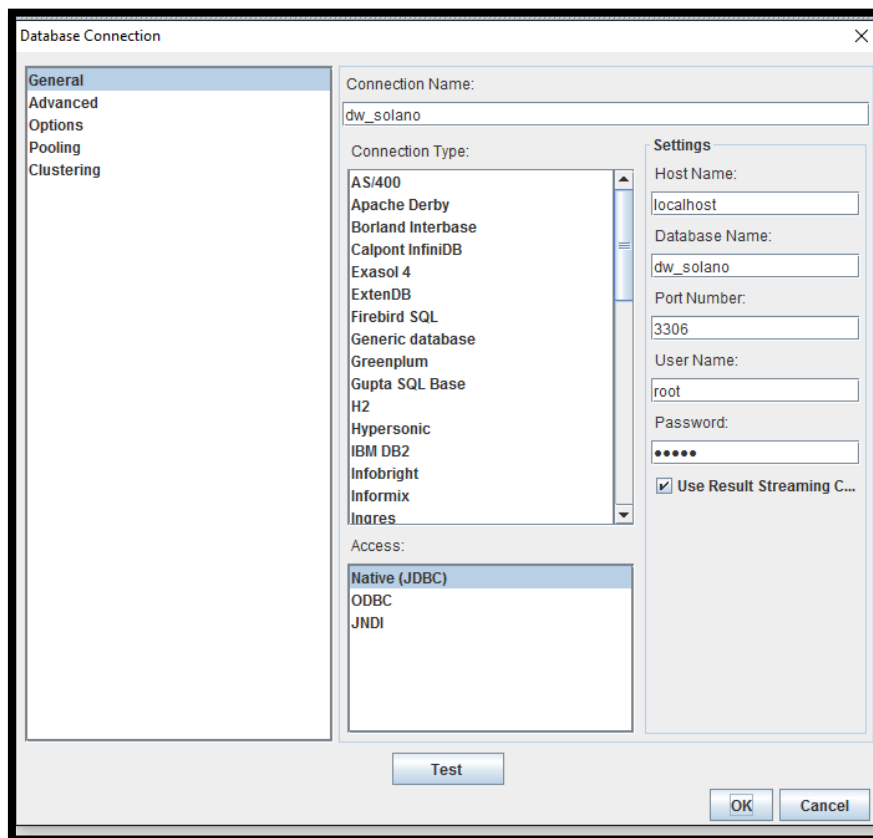


Ilustración 109. conexión a la base de datos para el cubo

El siguiente paso es crear el primer cubo que será llamado **ventas_hechos** perteneciente al proceso de ventas de la ferretería solano. El proceso para la creación del cubo es sencillo creamos un cubo agregamos la tabla de hechos con la que se trabajara y creamos las dimensiones correspondientes con sus jerarquías y las métricas que se utilizaran

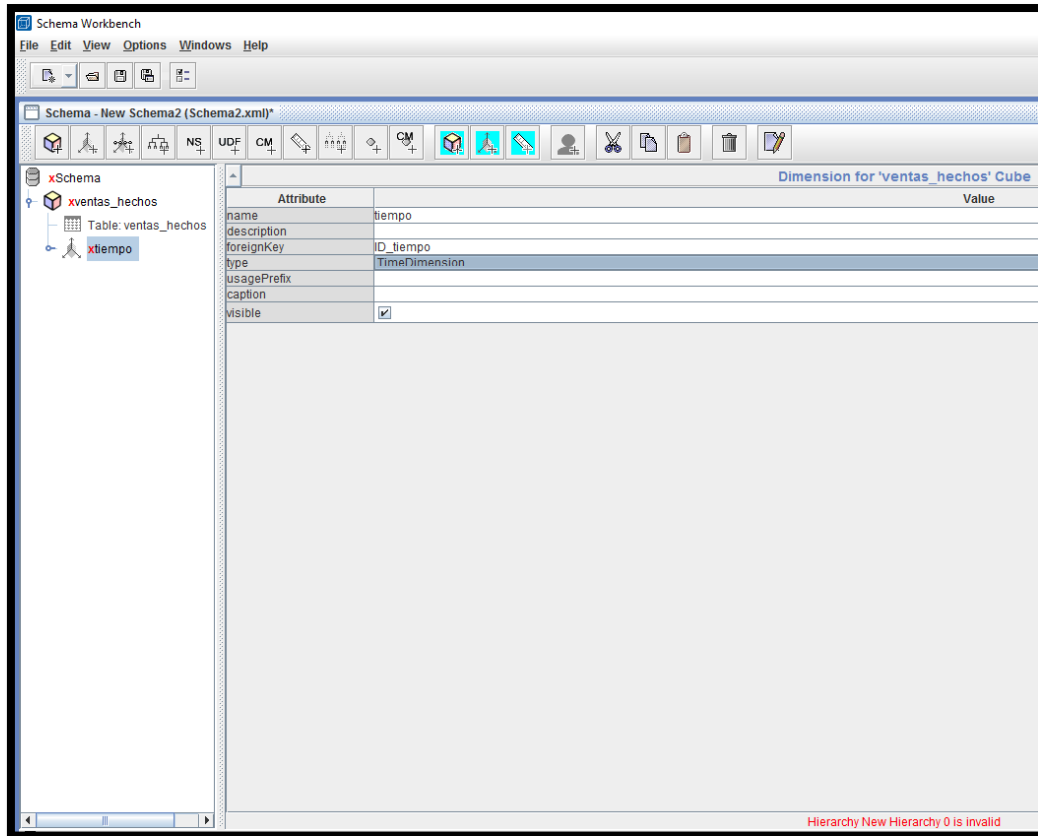


Ilustración 110.vista previa de creación del cubo ventas_hechos

Luego de agregar las dimensiones con sus jerarquías agregamos las métricas que necesitan una configuración sencilla de elección de campos y formatos de texto y al terminar la construcción del cubo se obtendrá una vista como la de la ilustración 113 .

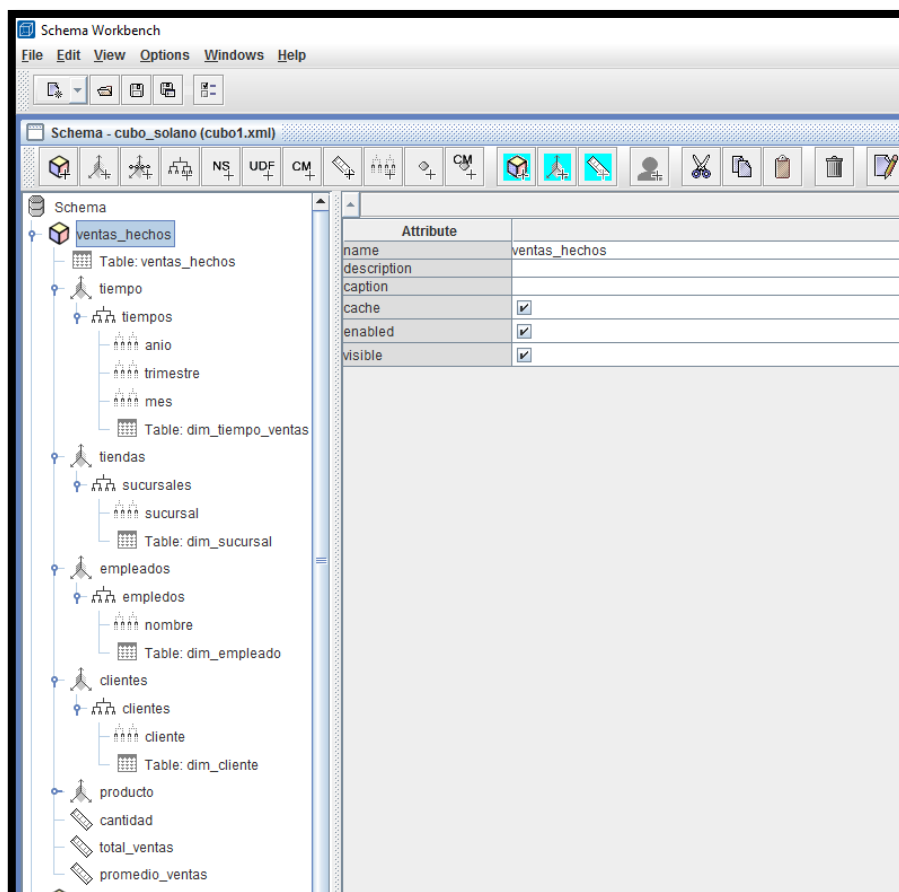


Ilustración 111.esquema del cubo ventas_ hechos

Para el proyecto fue necesario crear 3 cubos ya que son distintas áreas involucradas en el negocios a continuación se mostrara la estructura del cubo del proceso de compras y almacén.

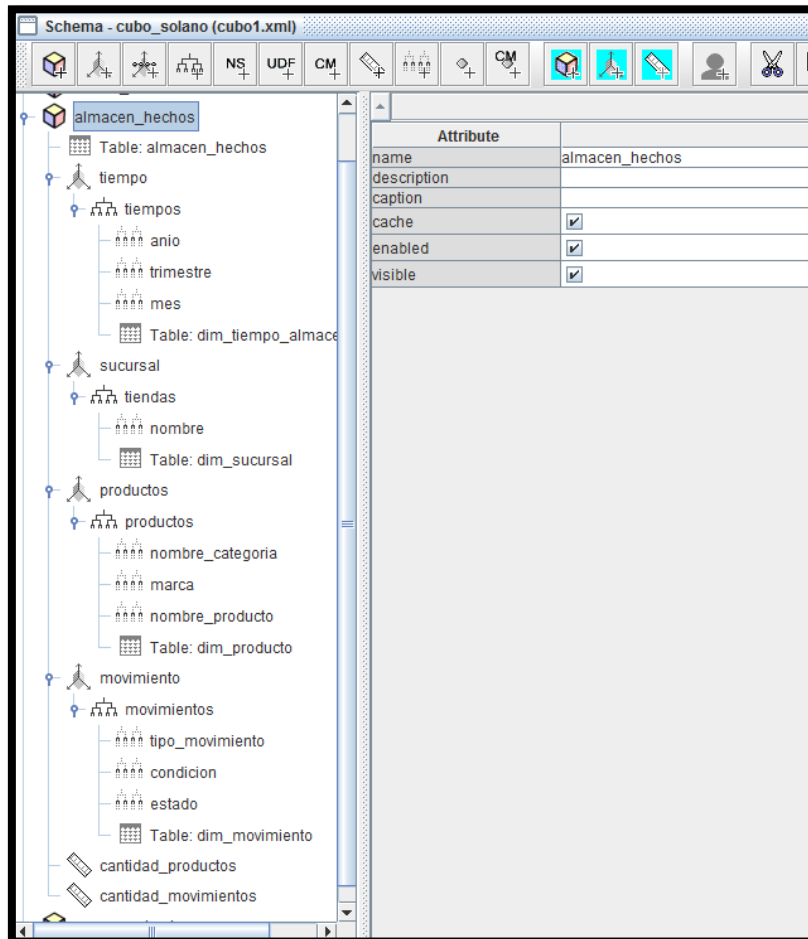


Ilustración 112.estructura del cubo almacen_hechos

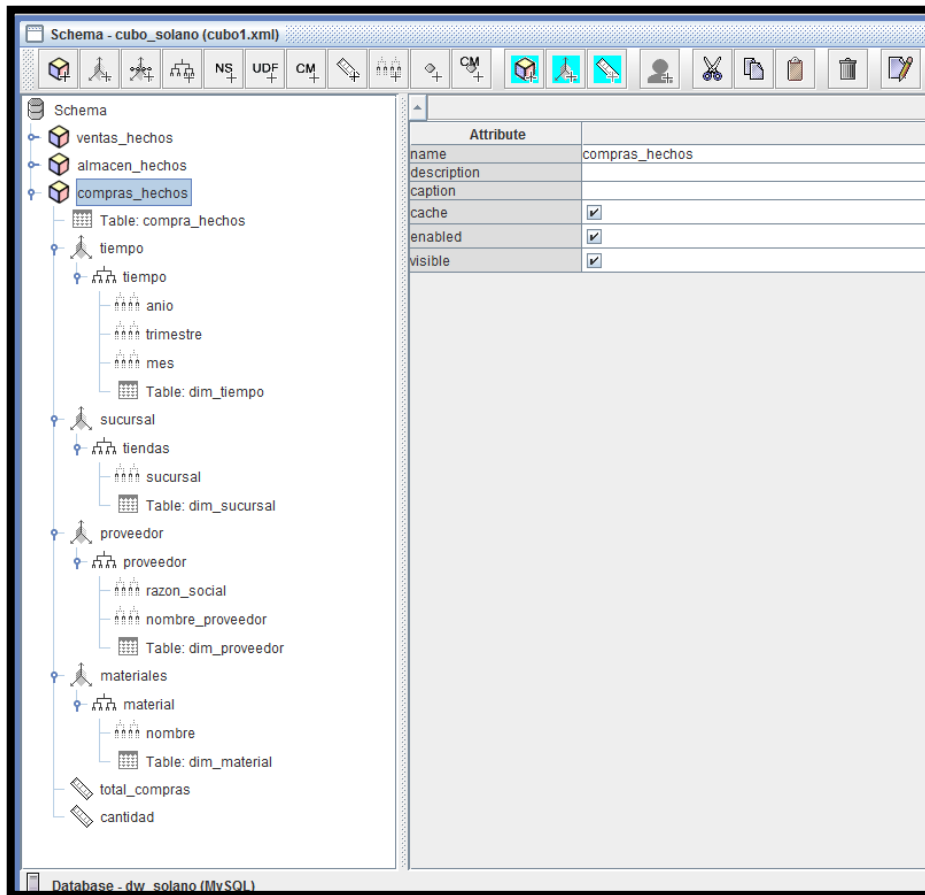


Ilustración 113.estructura del cubo Compras_hechos

Después de haber creado los cubos correctamente los publicaremos en el Pentaho BI server que es otra herramienta de Pentaho la cual se descargó previamente y se instaló

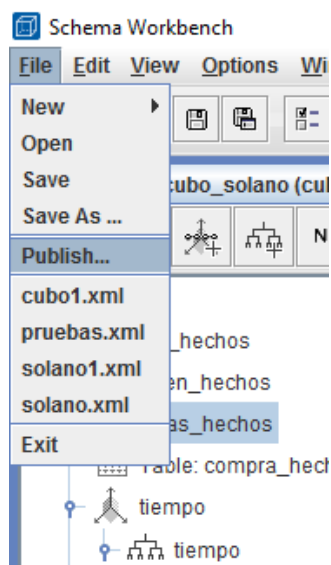


Ilustración 114.publicación del cubo

Después de haberlo publicado ejecutamos el archivo start-pentaho.bat de la carpeta Pentaho BI server y nos vamos a nuestro navegador y ponemos la siguiente dirección <http://localhost:8080/pentaho/Home> y nos direccionara a la página del BI server local de Pentaho donde se ha publicado nuestros cubos.

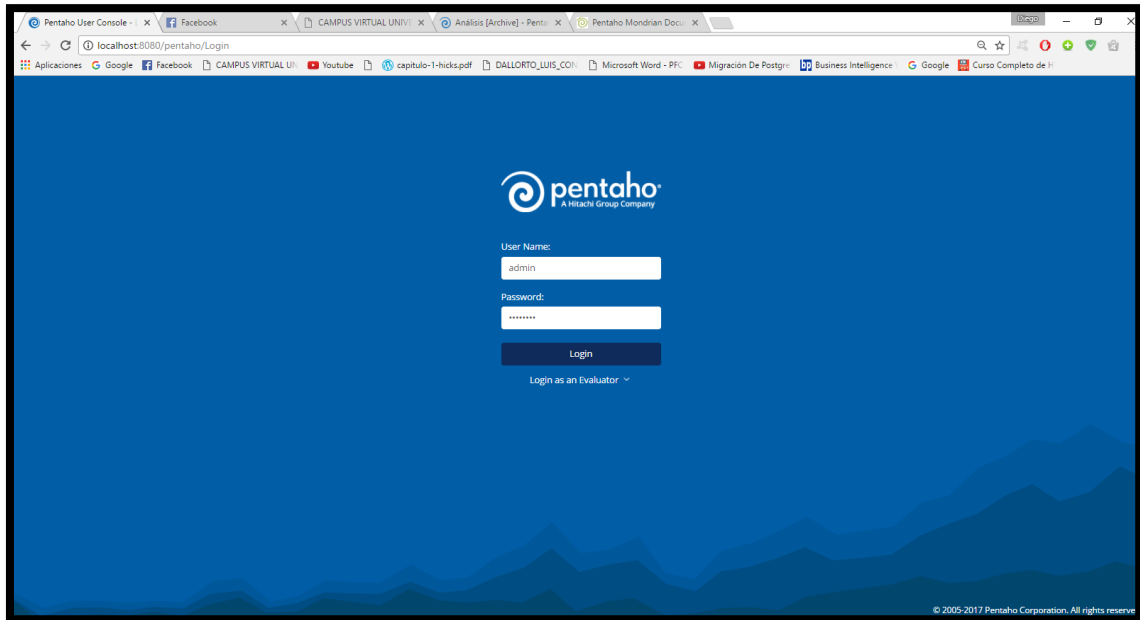


Ilustración 115.página principal de Pentaho BI server

Al igual que en los demás aplicativos se necesita crear la conexión en la pestaña data source.

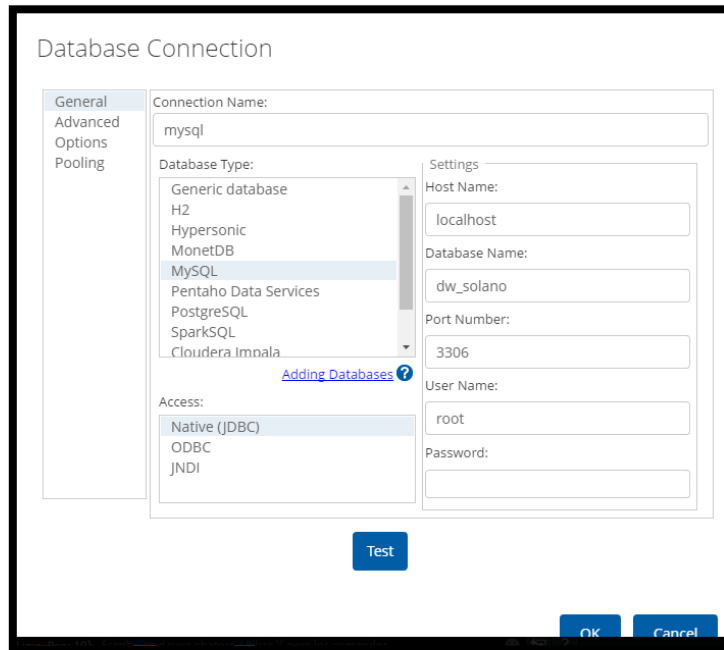
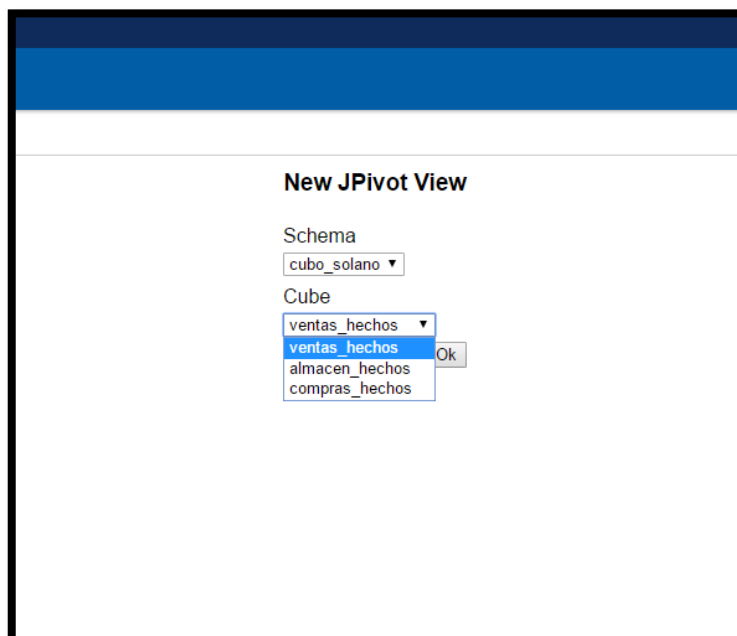


Ilustración 116. conexión de la base de datos en BI SERVER

Luego se elige la opción crear nueva jpivot view y seleccionas el cubo que se desea visualizar.



Después se proceden a generar la vista del cubo terminado

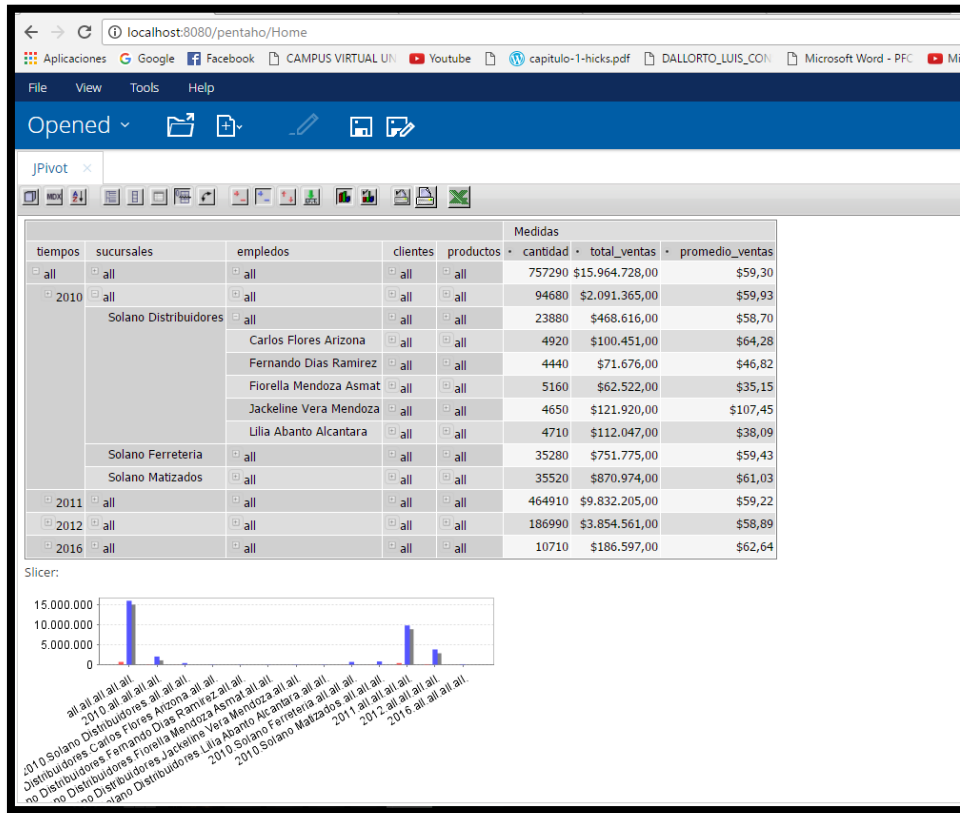


Ilustración 117, cubo ventas_ hechos

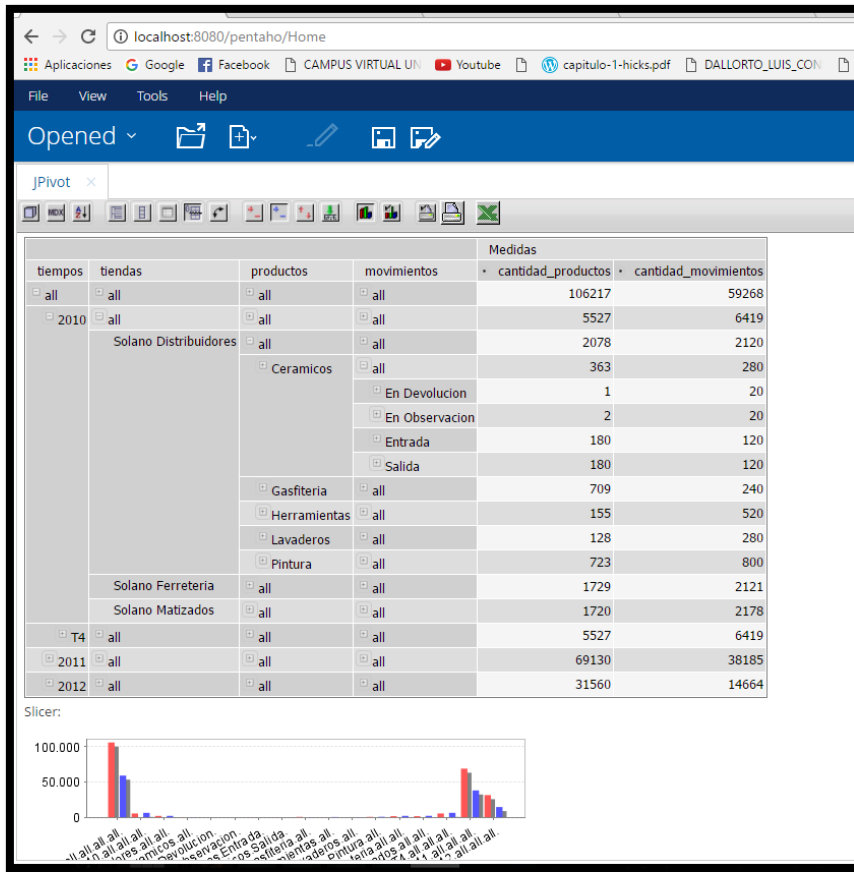


Ilustración 118.cubo Almacen_ hechos

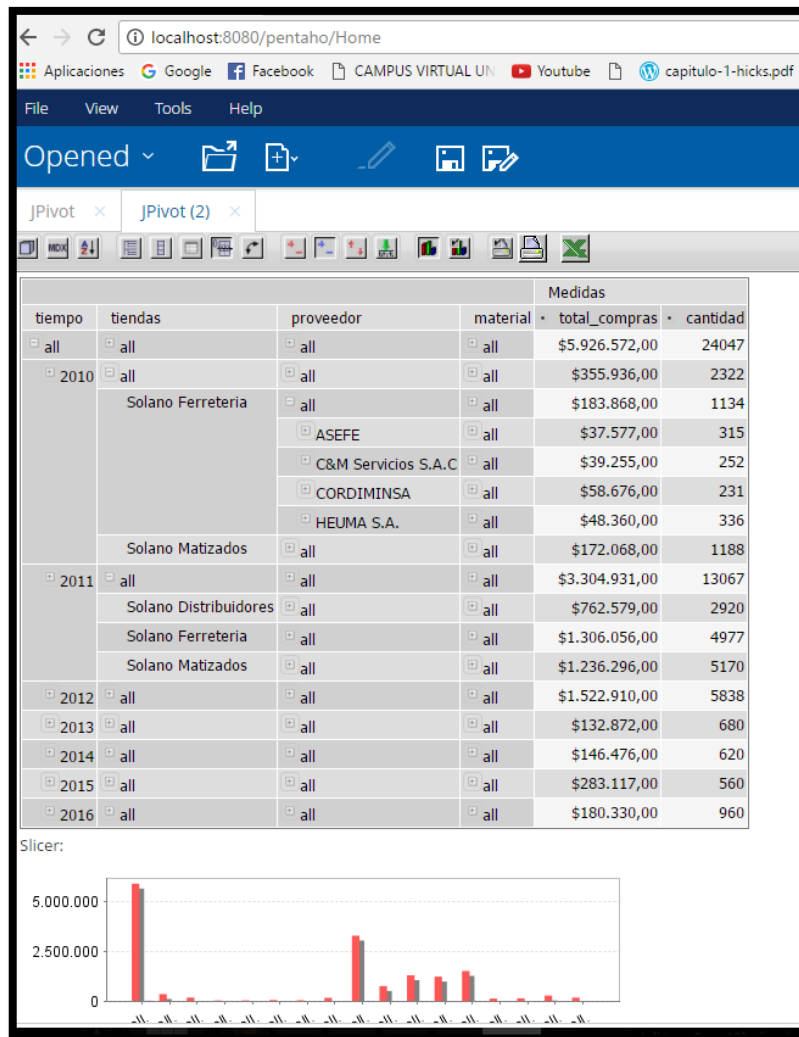


Ilustración 119.cubo Compras_ hechos

4.20. Reportes

Los requerimientos son el producto obtenido de este largo trabajo a continuación se mostraran los reportes por requerimiento especificado, se utilizó la herramienta Qlikview conectando el cubo y la base de datos dimensional para la creación del origen de datos que es la que permitirá poder acceder a toda la información para la creación de los siguientes gráficos.

Reporte de requerimiento N°1

Para este requerimiento se utilizó las dimensiones tiempo, proveedor y cantidad de productos de la tabla Compras_hechos este reporte permite visualizar un record de unidades de productos comprados a cada proveedor de la empresa.

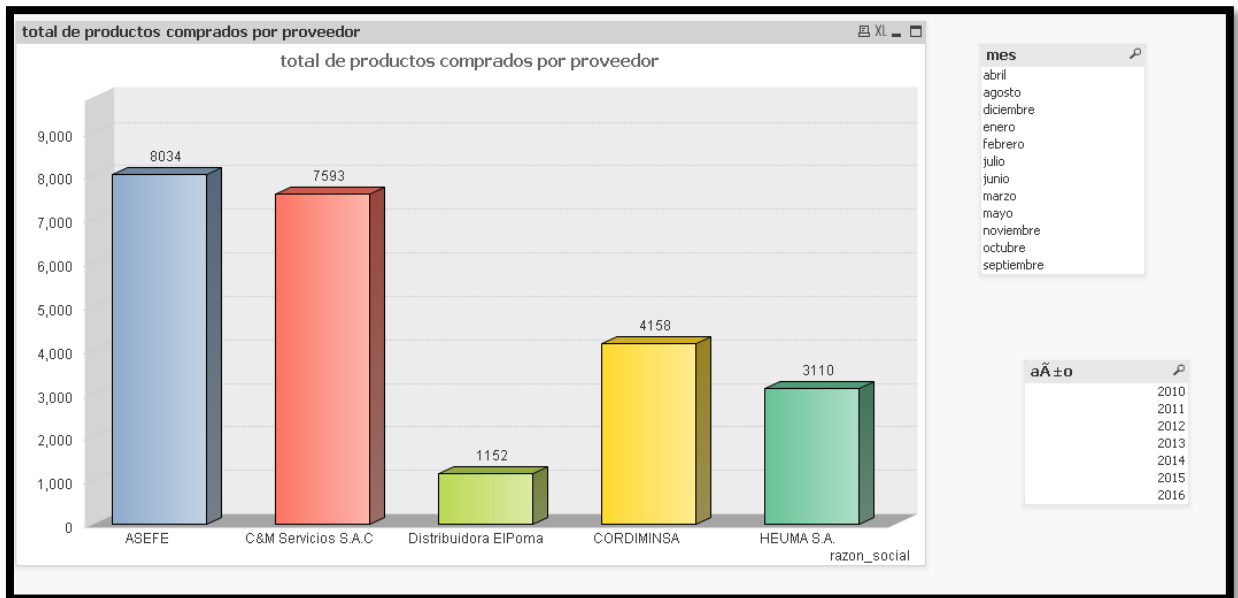


Ilustración 120. reporte requerimiento N°1

Reporte de requerimiento N°2

Este requerimiento permite tener un control del dinero que se ha invertido por cada proveedor por distintos años.

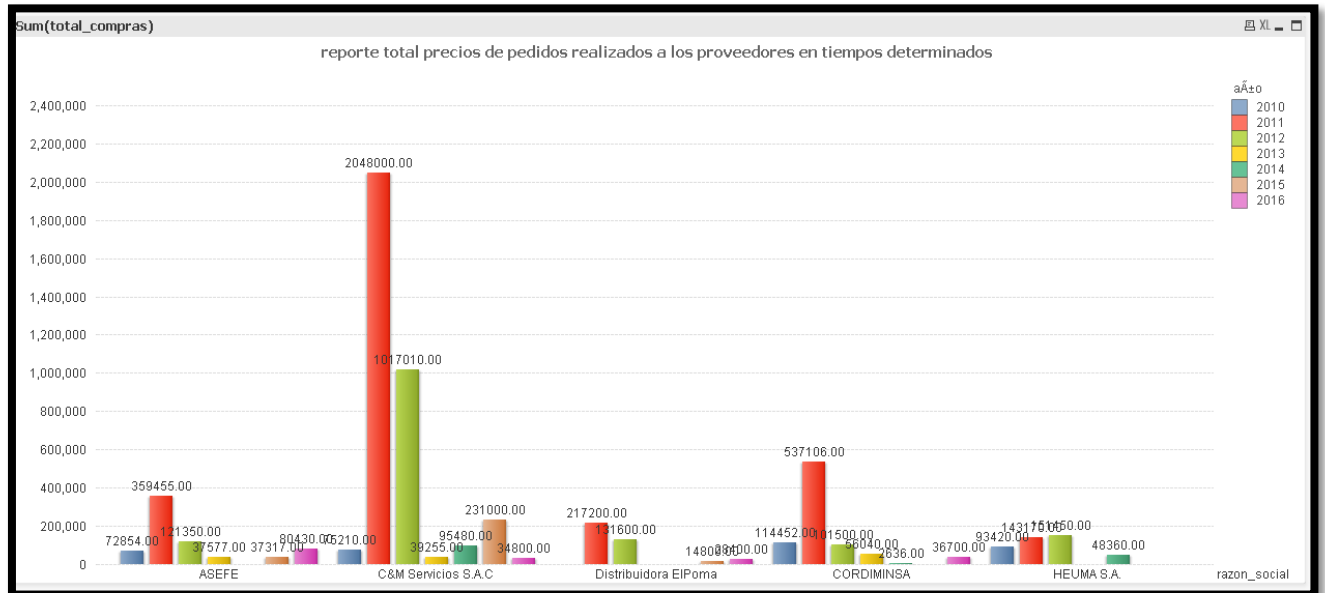


Ilustración 121. reporte de requerimiento N°2

Reporte de requerimiento N°3

Este reporte nos permite obtener la información sobre la cantidad de productos se han comprado por sucursal en los distintos años.

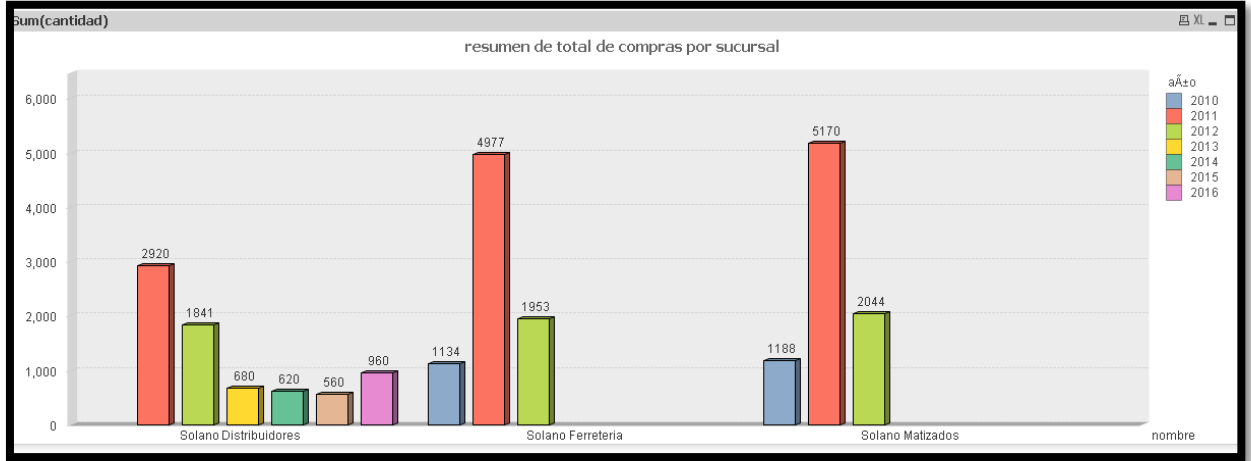


Ilustración 122. reporte de requerimiento N°3

Reporte de requerimiento N°4

Este reporte nos permite conocer las veces que se ha hecho compras a los proveedores por los distintos meses del año

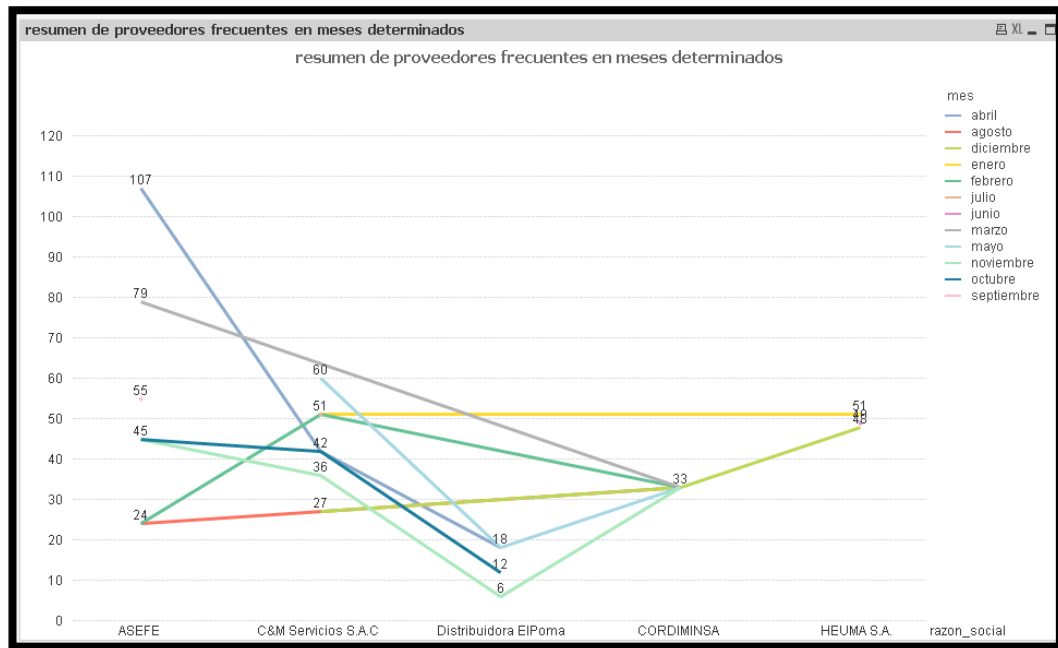


Ilustración 123. reporte de Requerimiento N°4

Reporte de requerimiento N°5

Este requerimiento nos permite obtener una vista de la cantidad por producto que ha sido pedido en un determinado mes.

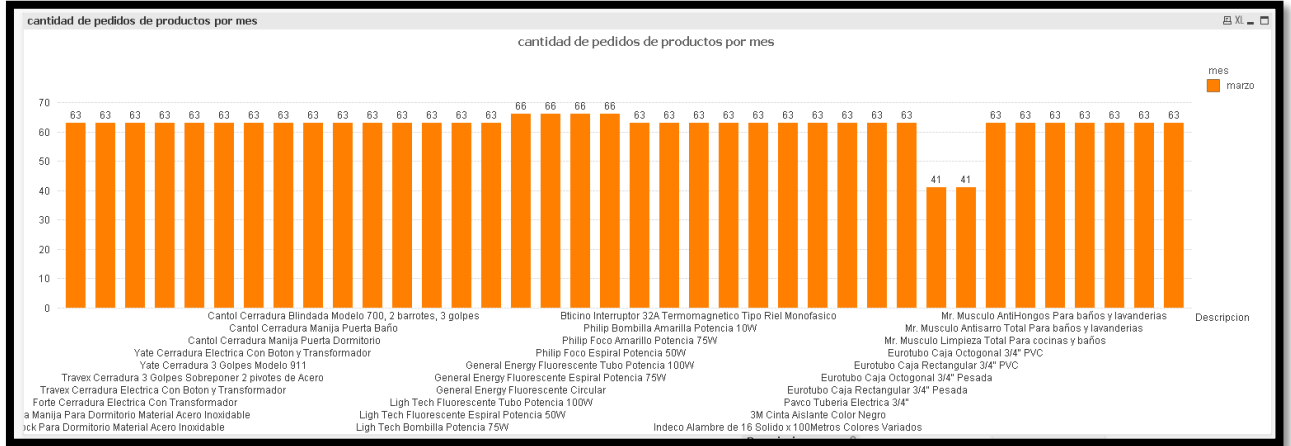


Ilustración 124. reporte de requerimiento N°5

Reporte de requerimiento N°6

El requerimiento 6 nos trae un total de egresos que ha obtenido las distintas sucursales por semestres en distintos años.



Ilustración 125. reporte de requerimiento N°6

Reporte de requerimiento N°7

En este reporte se aprecia los productos que son más vendidos en la tienda por porcentaje.

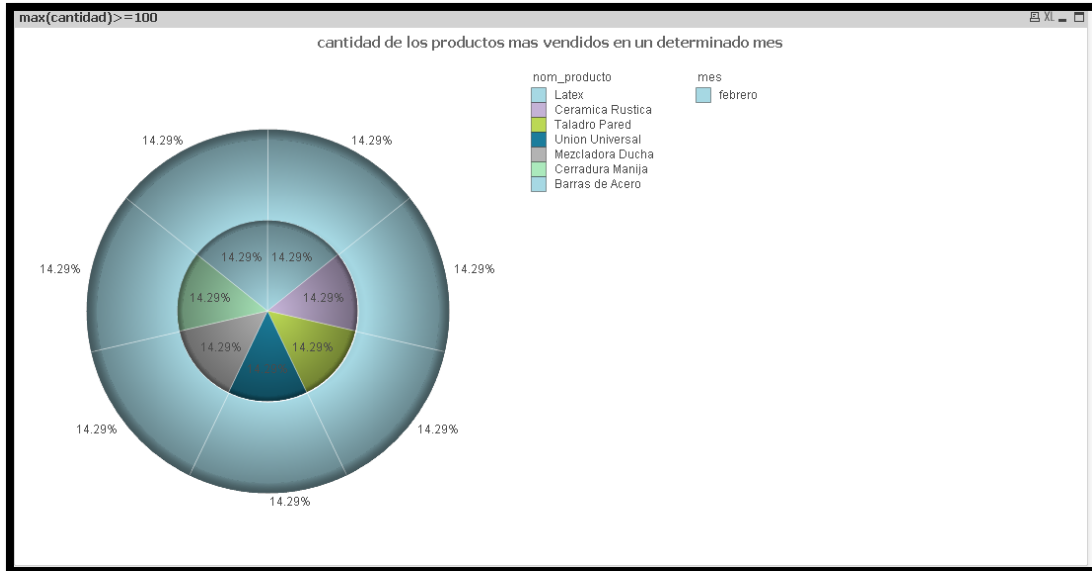


Ilustración 126, reporte de Requerimiento N°7

Reporte de requerimiento N°8

En el siguiente reporte se muestran el record de cantidad por producto recibidos en meses determinados y sucursales.

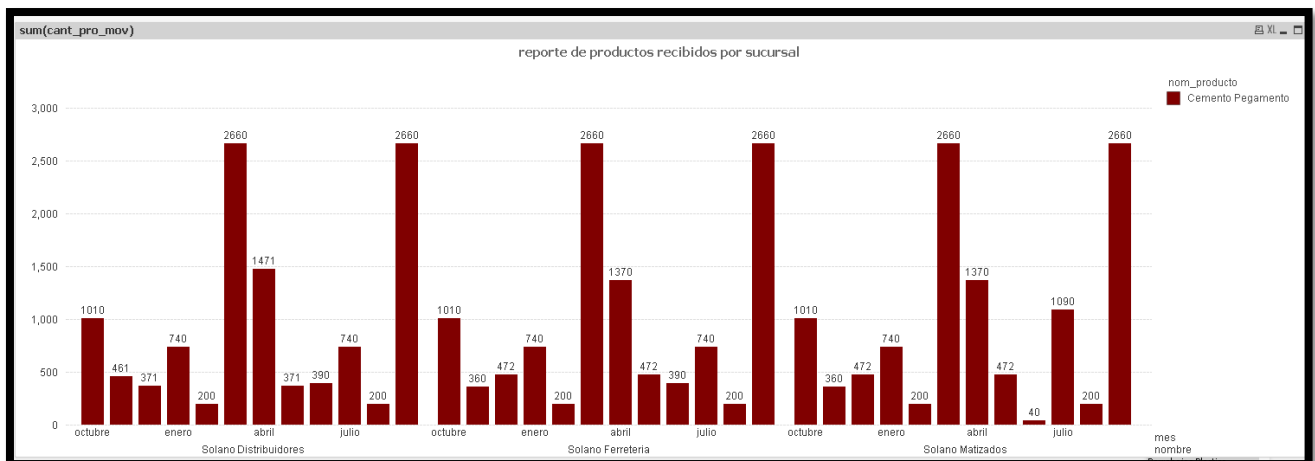


Ilustración 127. reporte de Requerimiento N°8

Reporte de requerimiento N°9

En el siguiente reporte se visualiza un gráfico de líneas de las demandas de las categorías de los productos que se dan por meses determinados en este caso por el mes de diciembre.

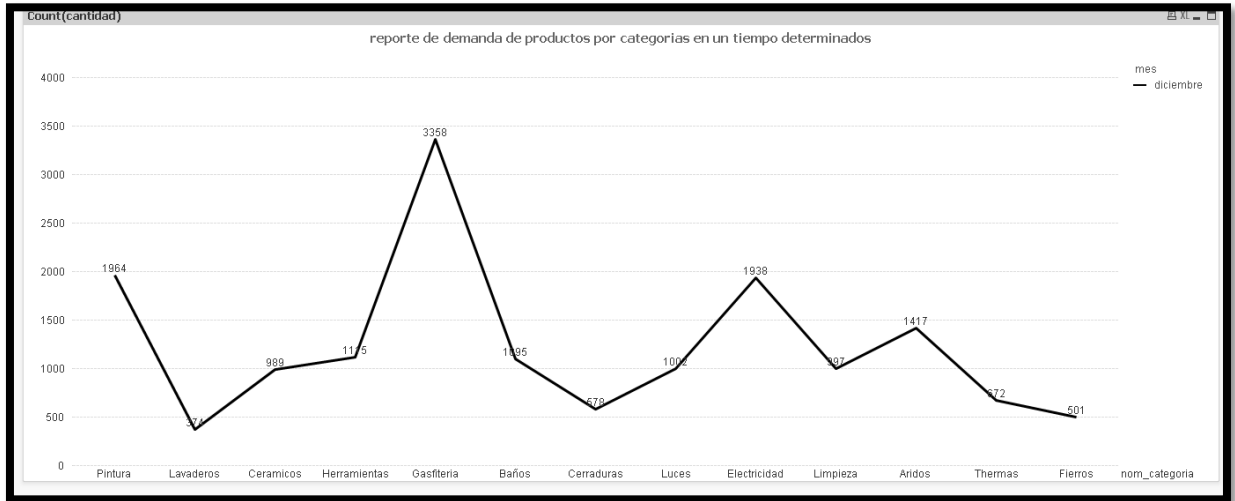


Ilustración 128. reporte de Requerimiento N°9

Reporte de requerimiento N°10

En este reporte nos crea una vista de la cantidad de los productos y condición en las que se encontraban en los distintos meses pueden ser productos en venta en procesos de devolución o en almacén esta vista también se puede generar con datos actuales.

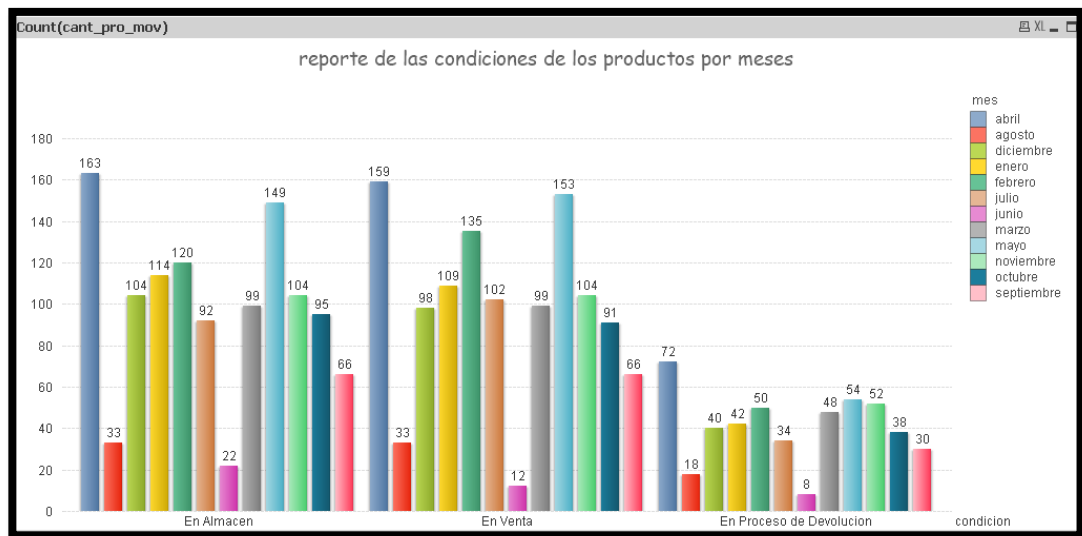


Ilustración 129. reporte de requerimiento N°10

Reporte de requerimiento N°11

En este reporte se visualizan la cantidad de productos que se han recibido por mes y por sucursal se tomaron solo 3 meses para hacer la vista más amigable.

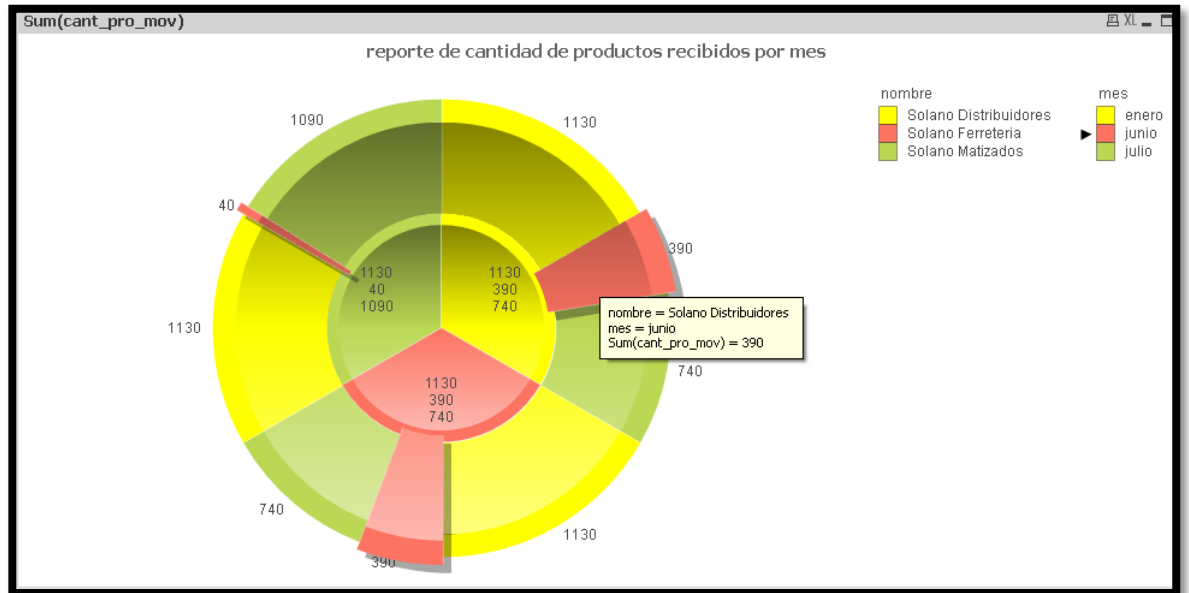


Ilustración 130. reporte de requerimiento n°11

Reporte de requerimiento N°12

En este reporte se puede apreciar la cantidad de productos que están en buen estado en el área de almacén de la empresa por mes.

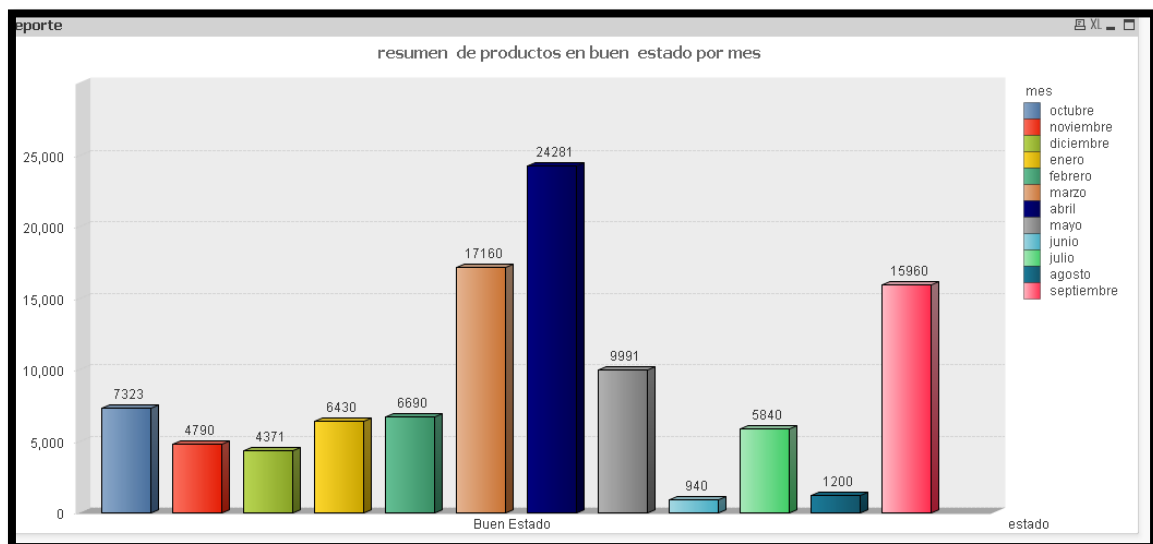


Ilustración 131. Reporte de requerimiento N°12

Reporte de requerimiento N°13

En este reporte se aprecia el record total de ventas que ha tenido cada empleado por distintas sucursales.

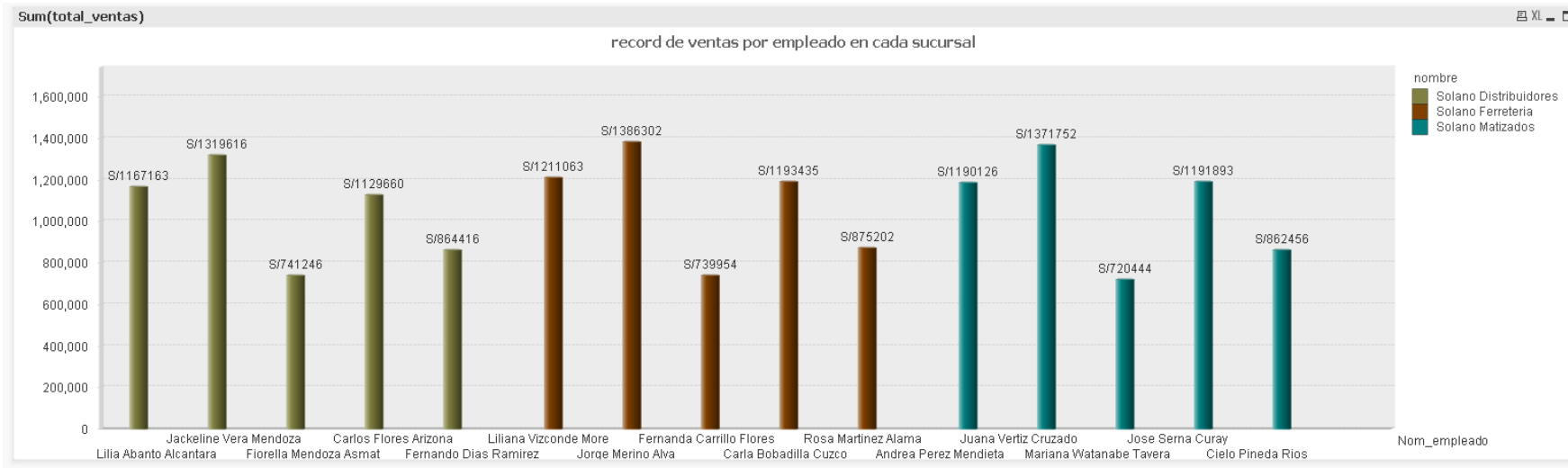


Ilustración 132. reporte de requerimiento N°13

Reporte de requerimiento N°14

En este indicador se genera una vista de la cantidad de productos que se han vendido por trimestres y años.

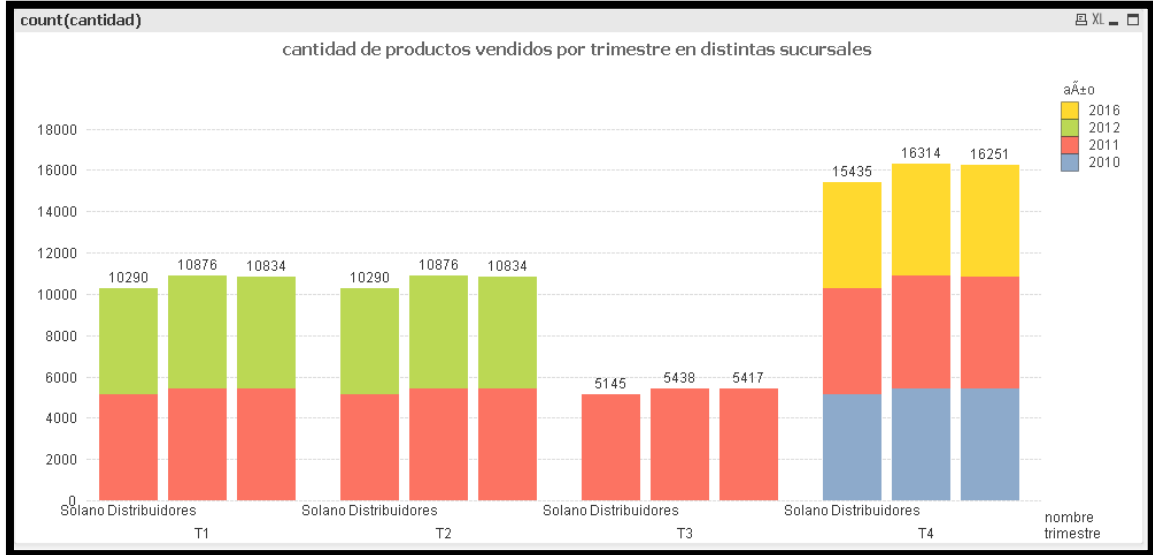


Ilustración 133. reporte de requerimiento N°14

Reporte de requerimiento N°15

En este indicador se puede apreciar la cantidad de productos que han sido comprados por los distintos clientes de la empresa

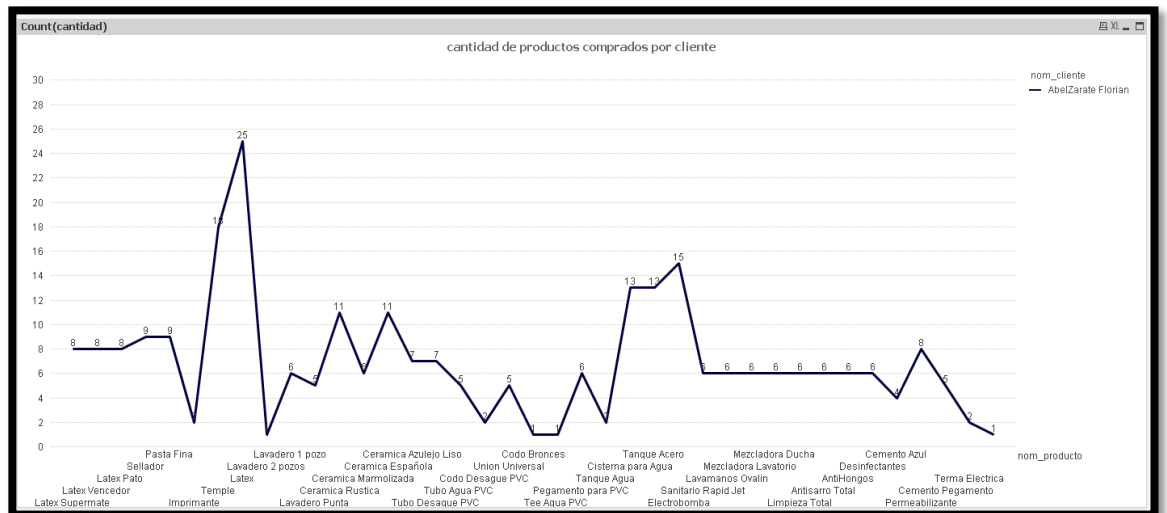


Ilustración 134. reporte de requerimiento N°15

Reporte de requerimiento N°16

En el siguiente grafico se enfatiza en la cantidad por distintos productos comprados por un cliente.

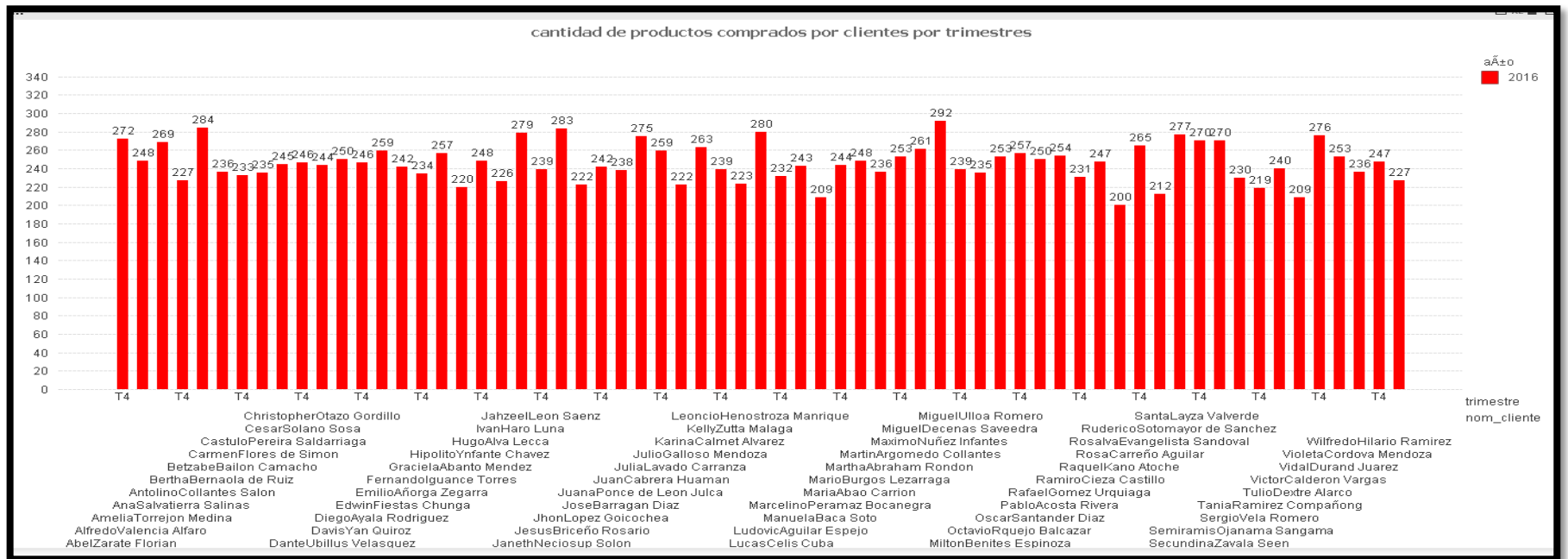


Ilustración 135. reporte de requerimiento N°16

Reporte de requerimiento N°17

En el siguiente grafico se obtiene la vista de las ventas realizadas por mes con el filtro del mes de setiembre

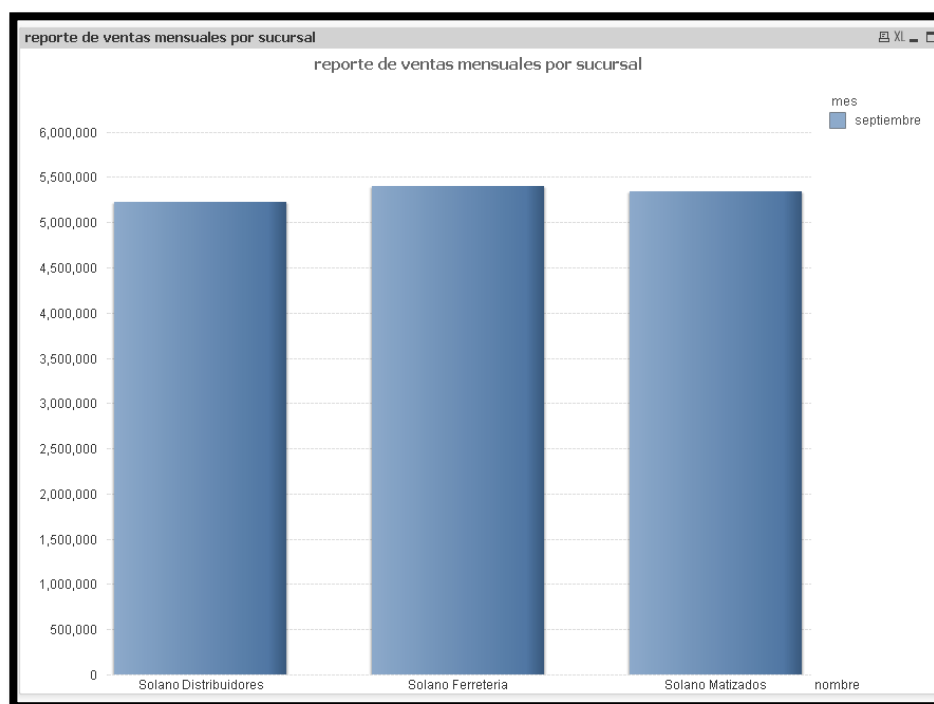


Ilustración 136. Reporte de requerimiento N°17

Reporte de requerimiento N°18

En el siguiente grafico se obtiene una vista del promedio de las ventas por sucursal y por marca de producto en el grafico se aprecian 3 marcas para generar una vista más entendible.

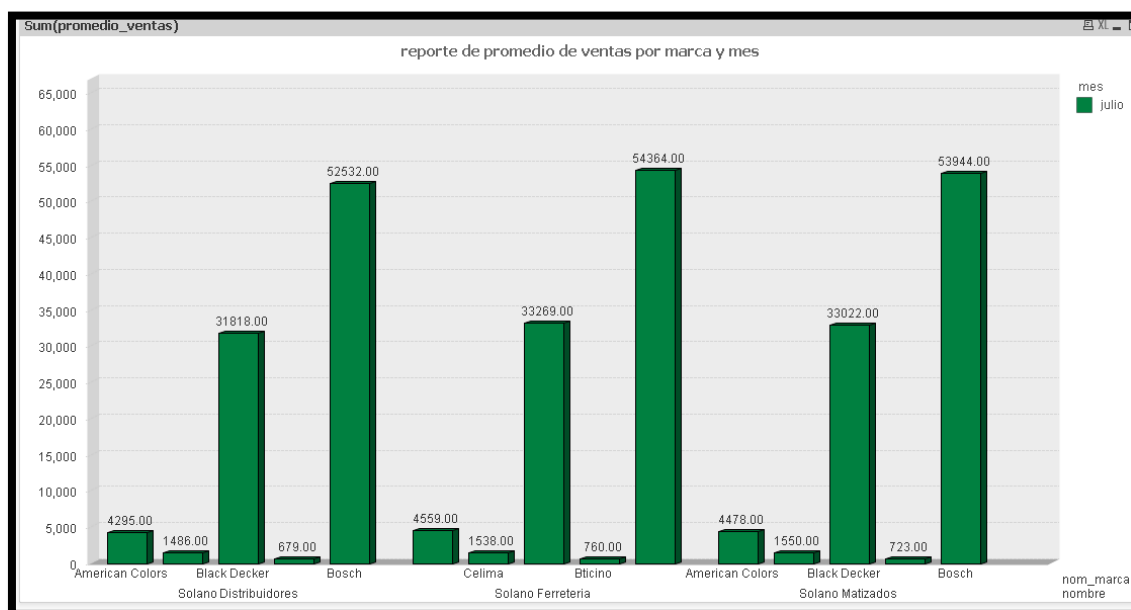


Ilustración 137. reporte de requerimiento 18

5. DISCUSION DE RESULTADOS

Para la constatación de la hipótesis se ha considerado lo siguiente:

Formulación del Problema

¿Cómo mejorar el análisis de la información para generar indicadores que den soporte a la toma de decisiones al área de ventas de la ferretería “solano”?

Formulación de la hipótesis

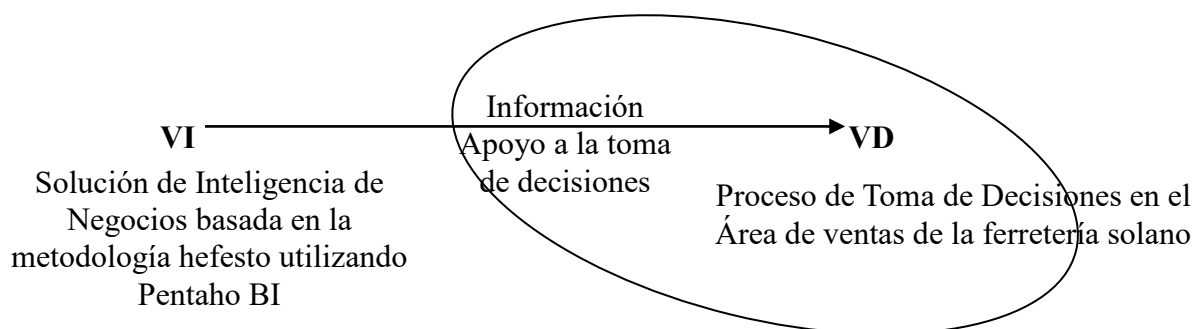
El desarrollo de una aplicación de Inteligencia de Negocios usando la metodología Hefesto y Pentaho BI en el Área de Ventas de la Ferretería Solano permitirá mejorar el análisis de la información generando indicadores que den soporte a la toma de decisiones.

Luego se definen las variables que intervienen en la veracidad o falsedad de la hipótesis:

Variable Independiente (VI): Solución de Inteligencia de Negocios Basada en la metodología hefesto utilizando Pentaho BI.

Variable Dependiente (VD): Indicadores que dan soporte a la toma de decisiones en el Área de ventas de la Ferretería Solano

5.1. MANERA PRESENCIAL



5.1.1. DISEÑO PREEXPERIMENTAL PRE-PRUEBA Y POST-PRUEBA

PRE-PRUEBA (O₁): Es la medición previa de X a G

POST-PRUEBA (O₂): Corresponde a la nueva medición de X a G

Se determinó usar el Diseño Pre Experimental Pre-Prueba y Post-Prueba, porque nuestra hipótesis se adecua a este diseño. Este diseño experimenta con un solo grupo de sujetos el cual es medido a través de un cuestionario antes y después de presentar el estímulo (Data Warehouse). Este diseño se presenta de la siguiente manera:

G O₁ X O₂

Donde:

X: Tratamiento, estímulo (Data Warehouse)

O: Medición a sujetos (Cuestionario)

G: Grupo de sujetos (Empleados)

El espacio Muestral que se tomó para la medición de los indicadores de la hipótesis, correspondió al total de personas que operarán el Data Warehouse, siendo estos 2: El jefe de área de multas y el administrador de la base de datos; a estas dos personas se le aplicó un cuestionario, antes de interactuar con el Data Warehouse (**O₁**) y después de interactuar con el mismo (**O₂**).

Al concluir la investigación se establecen las diferencias entre O₁ y O₂ para determinar si hay o no incremento en los resultados obtenidos.

5.1.2. CÁLCULO DE LOS INDICADORES DE LA HIPÓTESIS

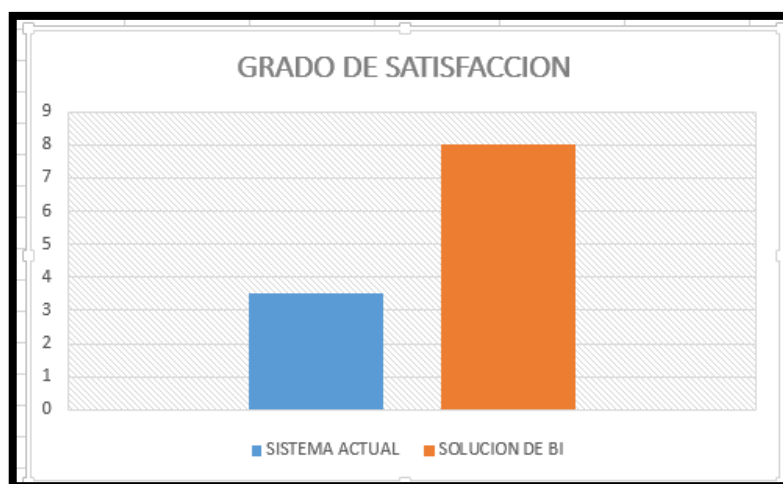
Para el cálculo de los indicadores de la hipótesis en la solución de inteligencia de negocios para la toma de Decisiones - Data Warehouse Propuesto (D.W.P) y el Sistema Actual – Sistemas Transaccional (S.T), se realizó un cuestionario donde se evaluó a los usuarios luego de haber interactuado con el Data Mart.

Los valores que los usuarios dieron a las respuestas del cuestionario fueron aplicados según el rango de satisfacción que muestran en la siguiente tabla:

RANGO	GRADO DE SATISFACCION
0 – 2.5	Insatisfecho
2.5 – 5.0	Medianamente Satisfecho
5.0 – 7.5	Satisfecho
7.5 – 10.0	Muy Satisfecho

Tabla 26.tablas de rangos de satisfacción

Pregunta n°1 ¿se puede saber cuántos productos han sido comprados por proveedor en un tiempo determinado?



Resultado: Los valores obtenidos son: 3.5 (S.T) y 8 (D.W.P) el Data Warehouse mejora los reportes acerca de la cantidad de productos comprados por proveedores con comparativas mediante gráficos dinámicos.

Pregunta n°2 ¿se puede conocer el precio total de los pedidos por proveedores que se realizaron en un tiempo determinado?



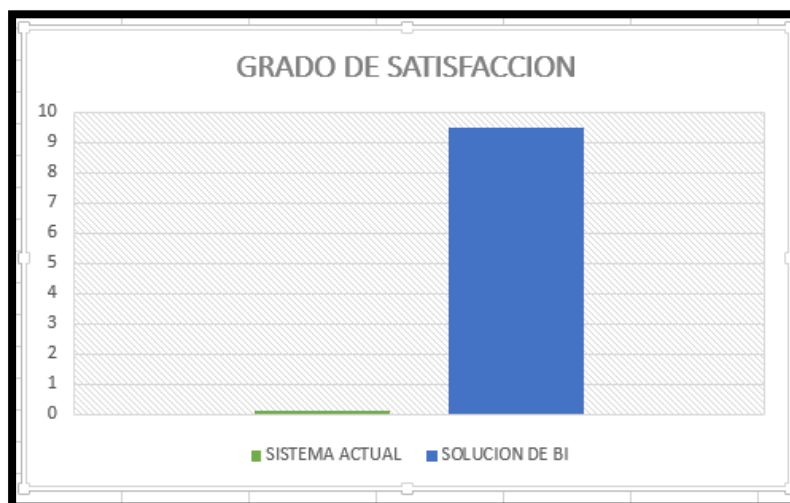
Resultados: Los valores que se obtuvieron son: 2 (S.T) y 8(D.W.P) EL Data Warehouse permite obtener un record total de todos los pedidos que se han realizado a lo largo de los años y meses haciendo comparaciones entre ellos.

Pregunta n°3 ¿se puede obtener un resumen de cantidad de compras en las diferentes sucursales en un mes determinado?



Resultado: Los valores obtenidos son: 2.5 (S.T) y 8.5 (D.W.P) el data Warehouse permite tener una vista completa de la cantidad de compras que se han realizado en las sucursales especificados por gráficos.

Pregunta n°4 ¿se puede obtener un resumen de los proveedores más frecuentes en un mes determinado?



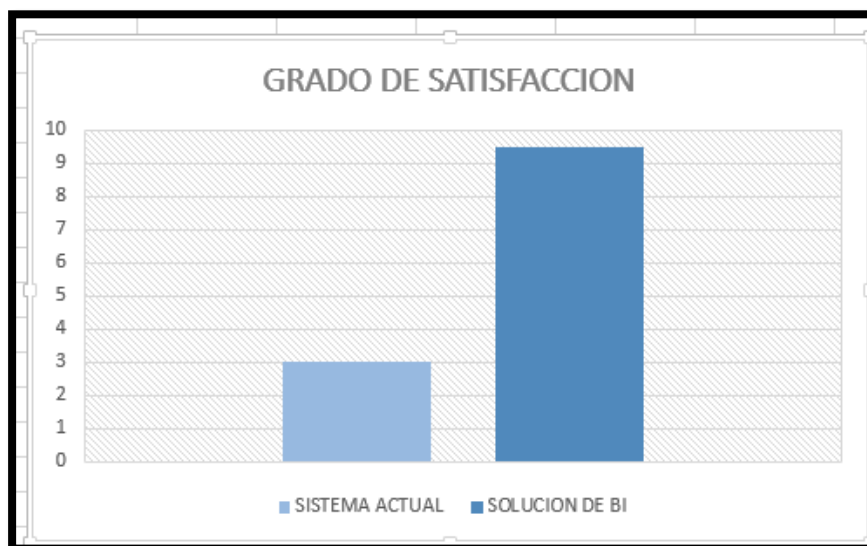
Resultados: los valores obtenidos fueron 0 (S.T) Y 9.5 (D.W.P) el data Warehouse permite obtener un resumen de los proveedores que han frecuentado las compras de la ferretería por mes, mediante gráficos comparativos se puede visualizar cuantas veces han sido solicitados por la ferretería el sistema actual no presenta esta consulta.

Pregunta n°5 ¿se puede saber cuántos pedidos de productos se hacen en un mes determinado?



Resultados: Los valores obtenidos fueron 2 (S.T) y 8 (D.W.P) el DW presenta reportes de la cantidad exacta de pedidos por productos realizados en distintos tiempos

Pregunta n°6 ¿Se puede saber los egresos que hay en las diferentes sucursales en un determinado semestre?



Resultados: los valores obtenidos fueron 3 (S.T) y 9.5 (D.W.P) el DW permite obtener un record total de egresos por sucursal de una manera más detallada con valores más exactos y comparativas mediante gráficos.

Pregunta n°7 ¿se puede saber cuáles son los productos más vendidos en un tiempo determinado?



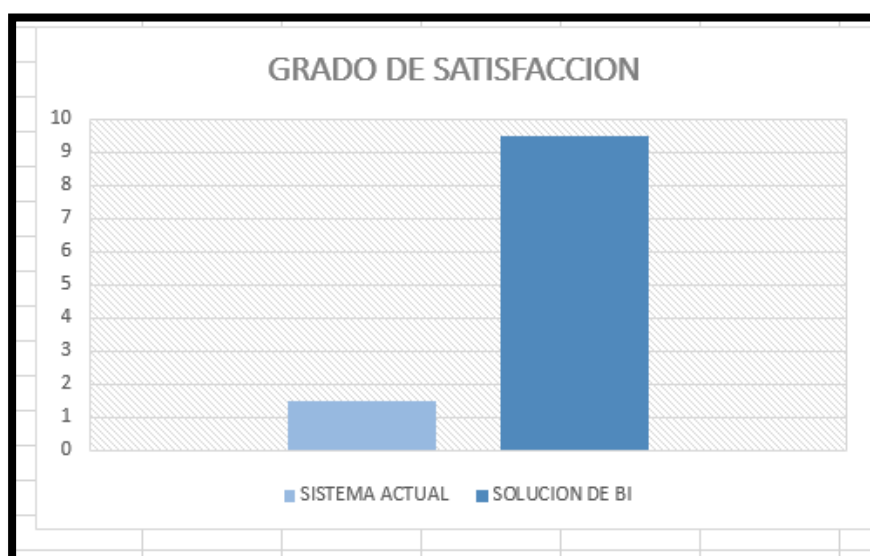
Resultados: los valores obtenidos son : 1.5 (S.T) 10 (D.W.P) el Dw en este cuestionario obtuvo la calificación más alta de satisfacción ya que la solución permite conocer cuáles son los productos más comprados por los clientes y pueden crear conocimiento para ofertas y demanda

Pregunta n°8 ¿se puede saber cuáles son los productos recibidos en distintas sucursales en un tiempo determinado?



Resultados: Los valores obtenidos fueron 1 (S.T) y 8(D.W.P) el Dw permite conocer los productos que se han recibido en cada sucursal el sistema transaccional solo genera reportes de entradas diarias el DW genera reportes dinámicos y detallados por distintos años y meses.

Pregunta n°9 ¿se puede conocer la categoría de producto y su demanda en un trimestre determinado?



Resultado: Se obtuvieron los siguientes valores 1.5 (S.T) y 9.5 (D.W.P) el Dw permite conocer las categorías de productos con más demandas en trimestres detalladamente.

Pregunta n°10 ¿se puede conocer las condiciones de los productos en un tiempo determinado?



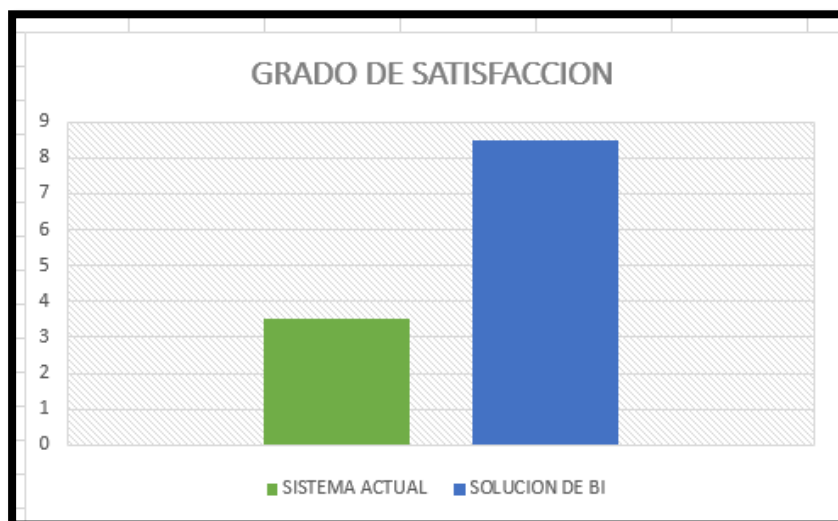
Resultados: Los valores obtenidos son 2 (S.T) y 7.5 (D.W.P) el DW permite tener reportes de la condiciones de todos los productos de las 3 sucursales mediante un reporte completo.

Pregunta n°11 ¿se puede obtener un resumen de la cantidad de productos que entraron en las diferentes sucursales en tiempos determinados?



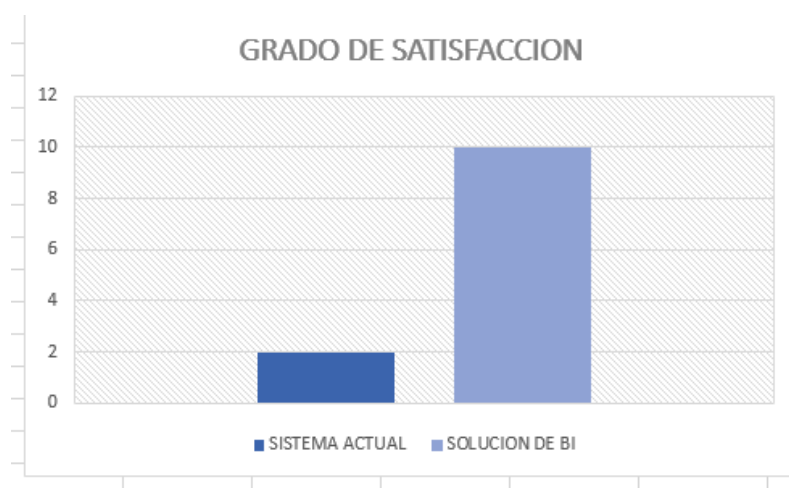
Resultados: Los valores obtenidos fueron 3 (S.T) y 6.5 (D.W.P) los 2 sistemas pueden obtener este requerimiento pero la solución BI lo genera a una mayor escala.

Pregunta N°12 ¿se puede obtener un resumen de los productos en buen estado en tiempos determinados?



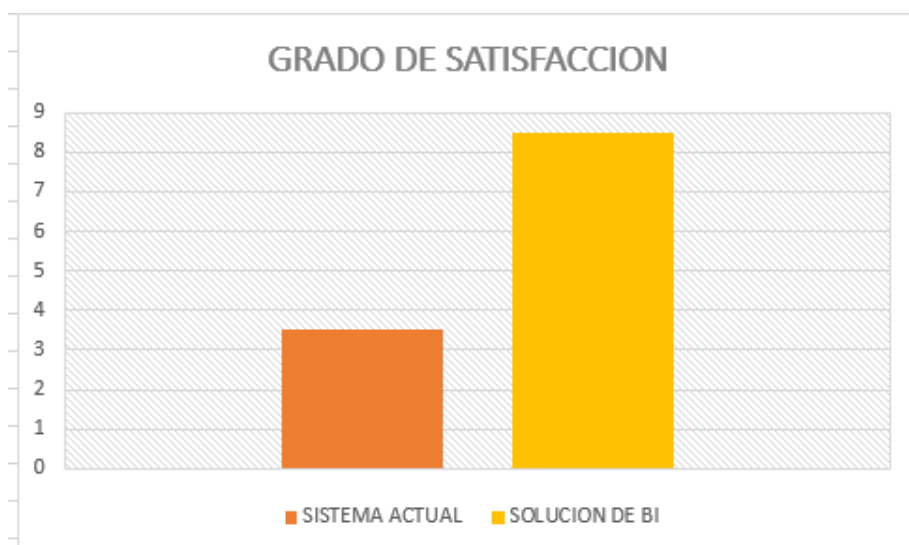
Resultado: Los valores obtenidos son: 3.5 (S.T) y 8.5 (D.W.P) La diferencia entre el sistema transaccional y el BI es demasiado grande el DW tiene un detalle y un nivel de conocimiento mucho mayor que el sistema transaccional se puede hacer comparaciones de los distintos meses en el BI y en el S.T no es posible.

Pregunta N°13 ¿se puede conocer el record de ventas totales por cada empleado en las diferentes sucursales?



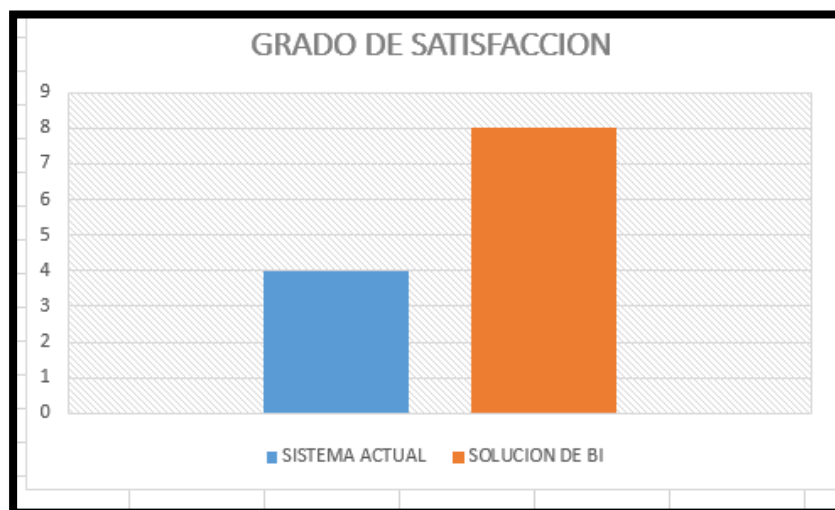
Resultado: Los valores obtenidos son 2 (S.T) 10 (D.W.P) es mucha la diferencia de satisfacción ya que en el sistema BI podemos obtener records de ventas por empleado de casa sucursal en el sistemas transaccional no tiene esa opción tan completa.

Pregunta n°14 ¿se puede conocer la cantidad de productos vendidos en cada sucursal por trimestre y año?



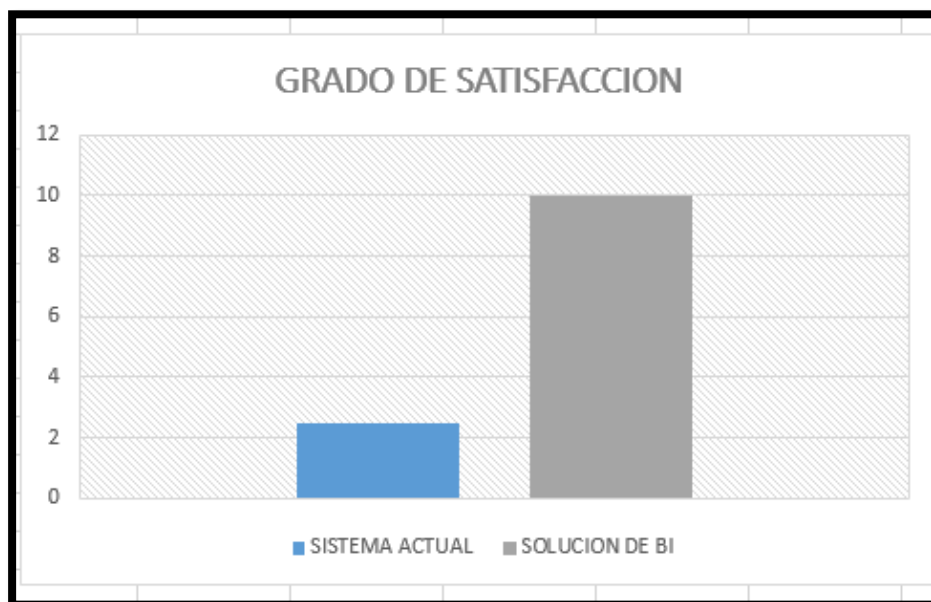
Resultado: Los valores obtenidos son: 3.5 (S.T) 9.5 (D.W.P) el Dw genera los reportes dinámicos por trimestre año y mes puede ser por cualquier fecha en cambio el sistema transaccional solo puede generar reportes semanales sin tanto detalle.

Pregunta n°15 ¿se puede conocer la cantidad total de productos comprados por un cliente en un tiempo determinado?



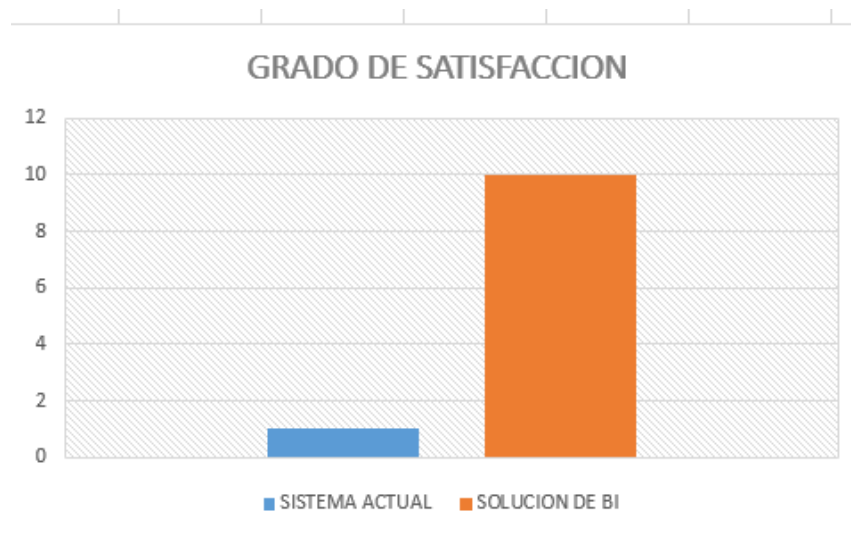
Resultados: los valores obtenidos son: 4(S.T) y 8 (D.W.P) el Dw permite tener la cantidad total de productos comprados por clientes mediante informes estadísticos y gráficos dinámicos.

Pregunta nº16 conocer los productos comprados por cliente en un mes determinado?



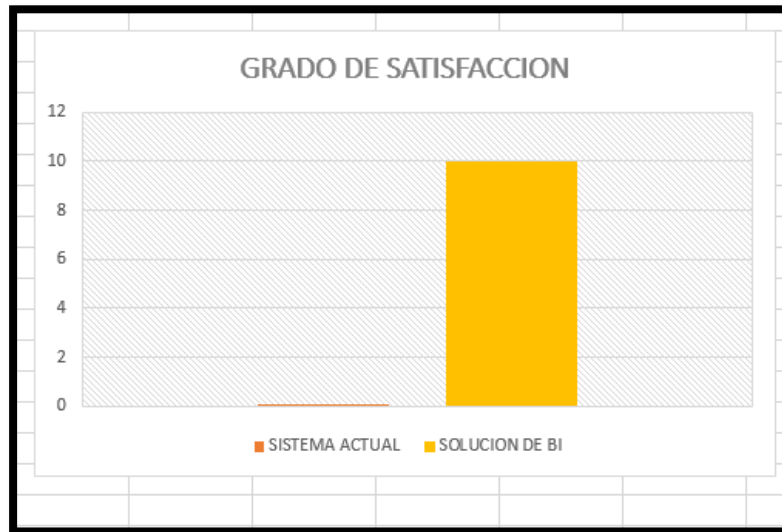
Resultados: los valores obtenidos son: 2 (S.T) y 10 (D.W.P) el sistema BI tuvo la máxima calificación de satisfacción debido al detalle que tiene el record de la relación de productos que un cliente ha comprado.

Pregunta nº17 ¿se puede obtener un resumen de las ventas totales en un determinado mes en las diferentes sucursales?



Resultado : los valores obtenidos fueron : 1 (S.T) y 10 (D.W.P) el S.T solo puede sacar records diarios de ventas de caja mientras que el BI puede obtener record anuales mensuales trimestrales .

Pregunta n°18 ¿se puede obtener el promedio de las ventas por sucursales en tiempos determinados?



Resultados: Los valores obtenidos son: 0 (S.T) y 9.5 (D.W.P) el sistema transaccional no puede obtener promedios de ventas mientras que el sistema BI puede obtener reportes de promedios por distintos tipos y gráficos comparativos.

5.1.3. APLICACIÓN DEL RANGO DE SATISFACCIÓN A LOS INDICADORES DE LA HIPÓTESIS

Los valores aplicados a los indicadores de la hipótesis tanto para el sistema Actual como para el Data Warehouse propuestos se muestran en la siguiente tabla:

Evaluación de los indicadores de la hipótesis:

INDICADORES	S.T	D.W.P
¿Se puede saber cuántos productos han sido comprados por proveedor en un tiempo determinado?	3	8
¿Se puede conocer el precio total de los pedidos por proveedores que se realizaron en un tiempo determinado?	2	8
¿Se puede obtener un resumen de cantidad de compras en las diferentes sucursales en un mes determinado?	2.5	8.5
¿Se puede obtener un resumen de los proveedores más frecuentes en un mes determinado?	0	9.5
¿Se puede saber cuántos pedidos de productos se hacen en un mes determinado?	2	8
¿Se puede saber los egresos que hay en las diferentes sucursales en un determinado semestre?	3	9.5
¿Se puede saber cuáles son los productos más vendidos en un tiempo determinado?	1.5	10
¿Se puede saber cuáles son los productos recibidos en distintas sucursales en un tiempo determinado?	1	8
¿Se puede conocer la categoría de producto y su demanda en un trimestre determinado?	1.5	9.5
¿Se puede conocer las condiciones de los productos en un tiempo determinado?	2	7.5
¿Se puede obtener un resumen de la cantidad de productos que entraron en las diferentes sucursales en tiempos determinados?	3	6.5

¿Se puede obtener un resumen de los productos en buen estado e tiempos determinados?	3.5	8.5
¿Se puede conocer el record de ventas totales por cada empleado en las diferentes sucursales?	2	10
¿Se puede conocer la cantidad de productos vendidos en cada sucursal por trimestre y año?	3.5	9.5
¿Se puede conocer la cantidad total de productos comprados por un cliente en un tiempo determinado?	4	8
¿Se puede conocer los productos comprados por cliente en un mes determinado?	2	10
¿Se puede obtener un resumen de las ventas totales en un determinado mes en las diferentes sucursales?	1	10
¿Se puede obtener el promedio de las ventas por sucursales en tiempos determinados?	0	9.5
PROMEDIO	2.08	8.80

Tabla 27. Tabla de indicadores promedios y valores

5.1.4. ANÁLISIS ESTADÍSTICO PARA LA PRUEBA PRESENCIAL DE LA HIPÓTESIS

Cálculo de la diferencia de dos medias:

DESCRIPCIÓN	MEDIAS	VARIANZAS
Fórmula	$\mu_i = \frac{\sum X_i}{N}$	$\sigma^2 = \frac{\sum (X_i - \mu_i)^2}{N}$
Cálculo	$\mu_1 = 2.08$	$\sigma^2_1 = 1.25$
N=18	$\mu_2 = 8.80$	$\sigma^2_2 = 1.03$

Tabla 28. Tabla diferencia de las medias

Cálculo de la Prueba de Hipótesis:

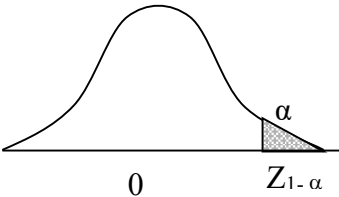
TIPO DE HIPÓTESIS	ESTADÍSTICA DE PRUEBA	REGIONES DE ACEPTACIÓN Y RECHAZO DE H ₀	VALOR CRÍTICO
Hipótesis Nula H ₀ : $\mu_1 - \mu_2 = 0$ Nivel de signif α	$z_0 = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma^2_1/n + \sigma^2_2/n}}$		$\alpha = 0.05$ $Z_{1-\alpha} = 0.97$
Hipótesis Alternativa H ₁ : $\mu_1 > \mu_2$	Z₀ = 17.60	Rechazar H₀ si, Z₀ > Z_{1-α}	17.60 > 0.97

Tabla 29. calculo prueba de la hipótesis

CONCLUSION: Ya que **Z₀ es mayor que Z_{1- α}** se rechaza la hipótesis nula por lo consiguiente que se acepta H₁ se llega a concluir que La implementación de una solución de inteligencia de negocios mejora el proceso de toma de decisiones en el área de ventas y logística de la ferretería solano E.I.R.L.

5.2.MANERA NO PRESENCIAL

La muestra que se tomó fue la misma a los cuales se les aplico un cuestionario el cual comprobará si el Data Warehouse brinda información valiosa para los procesos de Ventas y logística respondida por el gerente general de la corporación solano.

CUESTIONARIO DIRIGIDO AL GERENTE GENERAL y AL JEFE DE LOGISTICA DE LA FERRETERIA SOLANO.

Pregunta N°1 ¿Fueron entendibles los gráficos y los valores que se mostraron en los reportes de la solución BI fueron exactos?



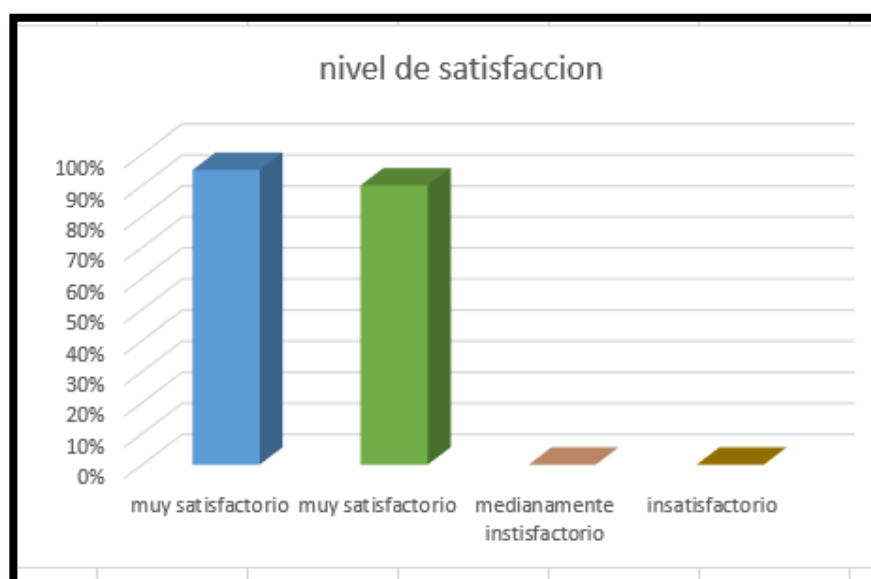
Resultado: el jefe de logística respondió que los gráficos y los valores mostrados que se mostraron en la solución BI fueron exactos muy satisfactoriamente en un 80% mientras que el gerente general respondió que los gráficos y los valores mostrados fueron satisfactoriamente en un 70%.

Pregunta N°2 ¿La solución BI permitió realizar las consultas requeridas por los diferentes jefes?



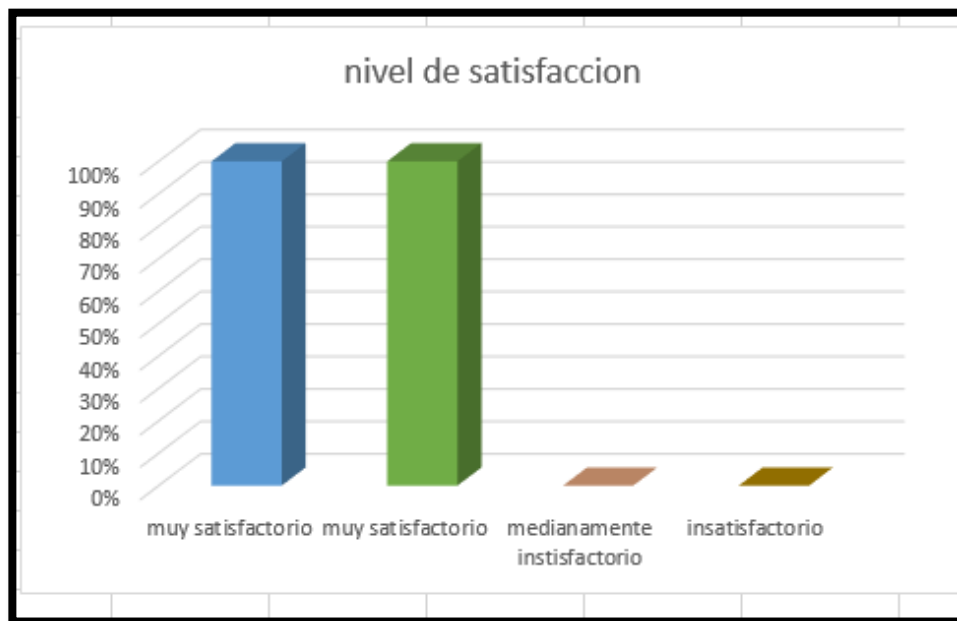
Resultado: El jefe de logística respondió que la solución BI si permitió realizar las consultas requeridas por los jefes muy satisfactoriamente (90%) de satisfacción mientras que el gerente general respondió que la solución BI permitió hacer toda las consultas requeridas satisfactoriamente (80%).

Pregunta N°3¿Se obtuvo un conocimiento nuevo acerca de las diferentes áreas jamás obtenido en la empresa?



Resultados: El jefe de logística respondió que si se obtuvo un conocimiento nuevo en las áreas muy satisfactoriamente (95%) mientras que el gerente general respondió que si se obtuvo un nuevo conocimiento de las áreas de la empresa jamás obtenido muy satisfactorio (90%).

Pregunta N°4 ¿La solución BI permitió un análisis profundo de la información mediante sus gráficos dinámicos?



Resultados: El jefe de logística respondió que la solución si permitió un análisis dinámico de los datos muy satisfactoriamente (100%) y el gerente general respondió que la solución si permitió el análisis profundo y dinámico mediante sus gráficos muy satisfactoriamente (100%).

CONCLUSION: Luego de evaluar a los usuarios que utilizaron la solución BI concluimos que la solución de inteligencia de negocios presenta información valiosa de una forma dinámica para un mejor análisis de los datos que pueda dar soporte en la toma de decisiones de la empresa.

6. CONCLUSIONES

- Se realizó una comparación de todas las posibles metodologías que se podían utilizar para la elaboración del presente proyecto para información más detallada ubicar la pagina 48, después de realizar las comparaciones adecuadas, la metodología escogida fue metodología Hefesto por ser una metodología ágil y completa para el desarrollo de proyectos a corto plazo.
- Se obtuvieron 18 requerimientos funcionales para el proyecto después de analizar el negocio y los procesos involucrados que fueron los de ventas, compras y almacén de acuerdo a las necesidades del personal involucrado y sus áreas esto esta referenciado en las páginas 73 y 74.
- Se desarrolló el proceso ETL del proyecto utilizando la herramienta Pentaho BI data Integration el cual permitió extraer los datos de la base de datos del sistema transaccional de la ferretería Solano para poblar la base de datos dimensional para el desarrollo de la aplicación BI propuesta en el trabajo de tesis referenciada en la página 113.
- Se diseñó la estructura del cubo OLAP de la aplicación BI utilizando la aplicación Pentaho BI server utilizando 14 dimensiones y 7 métricas necesarias para obtener un nivel de conocimiento amplio.
- Se obtuvieron 18 reportes fueron diseñados con la herramienta Qlikview ya que tiene mayor flexibilidad en la elaboración de reportes estadísticos utilizando la data de la aplicación BI y tiene una funcionalidad sencilla para la visualización de gráficos referenciados en la página 151.
- Se evaluaron los 18 reportes obtenidos con los administrativos de la empresa y se realizaron comparaciones mediante encuestas y nivel de satisfacción entre los reportes que se obtienen en su sistema transaccional y los reportes de la aplicación BI obteniendo resultados a favor de los reportes generados por la aplicación BI esto se puede apreciar en la discusión de resultados en la página 175.

- Se procedió a la implementación de la aplicación BI, siguiendo las 4 fases con sus sub-fases comprendidas en la metodología Hefesto, acoplando la solución a la infraestructura que tiene la ferretería Solano en sus sistemas utilizando la base de datos central de la empresa.

7. RECOMENDACIONES

- Para elegir una metodología correcta se debe realizar una investigación exhaustiva de diversas metodologías enfocadas al tema de estudio, se recomienda elegir estas metodologías y realizar cuadros comparativos evaluando su enfoque, su arquitectura y el la cantidad de fases que estas contienen para así poder evaluar cual nos conviene más.
- Para la obtención y comprensión de la lógica del negocio se debe mantener una relación estrecha con el usuario involucrado mantener una comunicación constante con los administrativos se recomienda realizar un cronograma de reuniones con el personal para comprometer a estos y sea posible recolectar la información necesaria de las áreas de donde se realizara el proyecto en el tiempo indicado.
- Para que el software Pentaho 7.0.0.0-25 pueda trabajar correctamente se necesita tener la última versión del java 8.0 y crear una variable de entorno llamada path y java home.
- Al momento de realizar el ETL se debe tener en cuenta los conectores apropiados de los gestores de las bases de datos para que el Pentaho funcione sin errores en el caso de este proyecto se trabajó con MYSQL se usó el conector mysql-connector-java-5.1.36.jar pero esto depende de los gestores de bases de datos hay un conector para cada tipo de ellos.

- Al momento de hacer la extracción de datos se debe usar una sintaxis correcta del lenguaje SQL para evitar redundancias en las tablas dimensionales.
- Definir un estándar de consulta para que al momento de realizar actualizaciones solo se cambie el archivo origen y destino por ejemplo al crear la consulta SQL para las tablas de hechos primero se deben igualar los campos por dimensión a los campos de la base de datos transaccional luego realizar los cálculos en consulta y utilizar la sintaxis ORDER BY y GROUP BY para las métricas con el fin de que no haya duplicidad de datos.
- Para la creación el cubo se recomienda usar una sola tabla de hechos por cubo, si se intenta usar 2 o más tablas de hechos en un solo cubo no se podrá realizar la implementación de este, se pueden usar las mismas dimensiones involucradas en la base de datos dimensional para diferentes cubos ya que estas pueden relacionarse con distintas tablas de hechos.
- Para evaluar la eficacia de los resultados se recomienda ver el nivel de granularidad o nivel de detalle y la exactitud de los reportes comparado con los reportes que se obtienen del sistema transaccional dicho esto se pueden evaluar distintos factores tales como el tiempo y especificaciones de acuerdo a los requerimientos.
- Para la implementación del proyecto es necesario seguir detalladamente los pasos de la metodología con la que se trabajara, puesto que se conoce cuáles serán los pasos a seguir para identificar gradualmente la información adecuada obteniendo un resultado efectivo sin cambios críticos que puedan arruinar el proyecto.

8. REFERENCIAS BIBLIOGRAFICAS

- alcami, R. I., & Carlos Devece carañana, J. G. (2009). Introduccion a la gestion de sistemas de informacion en la empresa . Universitat jaume .
- Arturo L, C. C. (2001). GUÍA PARA OBTENER EL RETORNO A LA INVERSION EN PROYECTOS DE DATAWAREHOUSE .
- austin, B. (10 de 02 de 2010). Benny austin. Obtenido de Benny austin:
<https://bennyaustin.wordpress.com/2010/05/02/kimball-and-inmon-dw-models/>
- bernabeu, R. (2010). HEFESTO , DATA WAREHOUSING: Investigación y Sistematización de Conceptos. argentina.
- Cano, J. L. (2007). BUSINESS INTELLIGENCE: competir con informacion. madrid: fundacion banesto.
- Curto, J. (2007). Introduccion al business intelligence . españa: UOC.
- Dataprix. (06 de 05 de 2009). Dataprix. Obtenido de Dataprix:
<http://www.dataprix.com/datawarehouse-manager>
- Ecured. (29 de 04 de 2017). Encured conocimiento de todos para todos. Obtenido de https://www.ecured.cu/Metodolog%C3%ADa_Hefesto
- facil, B. i. (25 de 6 de 2009). Business intelligence facil. Obtenido de <https://www.businessintelligence.info/productos/qlikview.html>
- Flores, R. (2004). Aplicación de minería de datos en ambiente universitario.
- gravitar. (8 de 5 de 2017). <http://gravitar.biz/pentaho/>. Obtenido de <http://gravitar.biz/pentaho/>
- Ian H. witten, E. F. (2000). Data mining practical Machine learning tools and techniques. san francisco: academic press.
- IBM. (29 de 04 de 2017). IBM Knowledge Center. Obtenido de https://www.ibm.com/support/knowledgecenter/es/SS9UM9_9.1.0/com.ibm.data.tools dimensional.ui.doc/topics/c_dm_star_schemas.html
- inmon, w. (2002). Building the data warehouse. En w. inmon, Building the data warehouse. canada: john wiley & sons.
- Kendall, K., & Kendall, J. (2005). Análisis y diseño de sistemas. Mexico: Pearson.
- Kimball, R. (2013). The Datawarehouse ToolKit, The Definitive Guide to Dimensional Modeling . Wiley.
- Laudon, K., & Laudon, J. (2012). Sistemas de Informacion Gerencial. New York:

Pearson.

Martínez, J. M. (21 de 4 de 2012). clavei. Obtenido de <https://www.clavei.es/blog/2012/que-es-qlikview-hablando-de-business-intelligence/>

microsoft. (12 de 07 de 2016). microsoft. Obtenido de microsoft: [https://msdn.microsoft.com/es-es/library/hh916536\(v=sc.12\).aspx](https://msdn.microsoft.com/es-es/library/hh916536(v=sc.12).aspx)

Nima Ramos , J. (2014). Soluciones OLAP con Microsoft SQL Server Analysis Services. Mexico: Person Eduaction.

pentaho. (8 de 5 de 2017). pentaho . Obtenido de <http://www.pentaho.com/>

Power Data. (08 de 08 de 2016). blog.powerdata.es. Obtenido de blog.powerdata.es: <http://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/405060/Qu-significa-la-integraci-n-de-datos>

Ramos, S. (25 de 07 de 2012). solidDq. Obtenido de solidDq: <http://blogs.solidq.com/es/business-analytics/curso-ms-business-intelligence-ssas-cubos-olap-conceptos-basicos-parte-i-27/>

rio, L. M. (2006). Mas alla del business intelligence . barcelona: gestion 2000.

robert stackowiak, j. r., & greenwald, r. (2007). Oracle Data Warehousing and Business Intelligence. wiley.

roussel, G. (2006). Decision support systems serving the company : the secrets to a succesfull project.

sanchez, L. z. (2008). Metodología para el Diseño. valencia.

sinnexus. (20 de 04 de 2014). sinnexus. Obtenido de https://www.sinnexus.com/business_intelligence/datawarehouse.aspx

Sinnexus. (01 de 07 de 2016). Sinnexus. Obtenido de Sinnexus: http://www.sinnexus.com/business_intelligence/sistemas_soporte_decisiones.aspx

Symantec. (02 de Noviembre de 2016). OLTP. Obtenido de https://www.symantec.com/es/mx/security_response/glossary/define.jsp?letter=o&word=oltp-online-transaction-processing

Wang, j. (2009). Encyclopedia of Data Warehousing and Mining, Second Edition.

9. ANEXOS

Carta de aceptación de la empresa



Ilustración 138. Carta de aceptación por la empresa

Cuestionarios para saber nivel de satisfacción de los usuarios sistema de inteligencia de negocios en la ferretería Solano E.I.R.L

SOLANO

Cuestionarios para saber nivel de satisfacción de los usuarios sistema de inteligencia de negocios en la ferretería Solano E.I.R.L.

Encuestado: Roger Córdoba (Jefe de logística) y Edgar Solano Ramirez (Gerente general)

Las siguientes preguntas contienen 4 alternativas de niveles de satisfacción pueden ser: **Muy satisfecho, satisfecho, medianamente satisfecho, insatisfecho** marca con un x su respuesta

Pregunta n°1 ¿se puede saber cuántos productos han sido comprados por proveedor en un tiempo determinado?

Muy satisfecho satisfecho medianamente satisfecho

insatisfecho

Pregunta n°2 ¿se puede conocer el precio total de los pedidos por proveedores que se realizaron en un tiempo determinado?

Muy satisfecho satisfecho medianamente satisfecho

insatisfecho

Pregunta n°3 ¿se puede obtener un resumen de cantidad de compras en las diferentes sucursales en un mes determinado?

Muy satisfecho satisfecho medianamente satisfecho

insatisfecho

Ilustración 139. Modelo de encuesta para hipótesis

SOLANO

Pregunta n°8 ¿se puede saber cuáles son los productos recibidos en distintas sucursales en un tiempo determinado?

Muy satisfecho satisfecho medianamente satisfecho

Insatisfecho

Pregunta n°9 ¿se puede conocer la categoría de producto y su demanda en un trimestre determinado?

Muy satisfecho satisfecho medianamente satisfecho

Insatisfecho

Pregunta n°10 ¿se puede conocer las condiciones de los productos en un tiempo determinado?

Muy satisfecho satisfecho medianamente satisfecho

Insatisfecho

Pregunta n°11 ¿se puede obtener un resumen de la cantidad de productos que entraron en las diferentes sucursales en tiempos determinados?

Muy satisfecho satisfecho medianamente satisfecho

Insatisfecho

Ilustración 140. Modelo de encuesta para hipótesis 2