



# ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA

**DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN**

**CARRERA DE INGENIERÍA EN SISTEMAS E INFORMÁTICA**

**TESIS PREVIO A LA OBTENCIÓN DEL TÍTULO DE  
INGENIERO EN SISTEMAS E INFORMÁTICA**

**AUTORES: ALEX PATRICIO DÁVILA MONTÚFAR,  
GONZALO WLADIMIR VILLARES DÁVILA**

**TEMA: ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UNA  
SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA EL  
CONTROL DE COMPRA DE MATERIA PRIMA Y VENTA DE  
SWEATERS PARA TEJIDOS ANAHÍ UTILIZANDO LA  
METODOLOGÍA HEFESTO V2.0.**

**DIRECTORA: ING. LORENA DUQUE  
CODIRECTOR: ING. PAUL DÍAZ**

**SANGOLQUÍ, JULIO 2015**

UNIVERSIDAD DE LAS FUERZAS ARMADAS – ESPE

CARRERA DE INGENIERÍA EN SISTEMAS E INFORMÁTICA

**CERTIFICADO**

Ing. Lorena Duque (DIRECTOR DE TESIS)

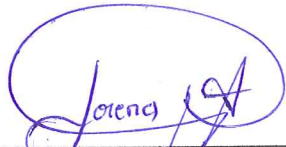
Ing. Paul Díaz (CODIRECTOR DE TESIS)

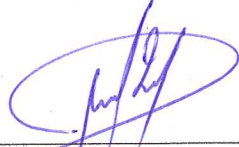
**CERTIFICAN**

Que el trabajo titulado “ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA EL CONTROL DE COMPRA DE MATERIA PRIMA Y VENTA DE SWEATERS PARA TEJIDOS ANAHÍ UTILIZANDO LA METODOLOGÍA HEFESTO V2.0.”, realizado por la Sr. Alex Patricio Dávila Montúfar y el Sr. Gonzalo Wladimir Villares Dávila, ha sido guiado y revisado periódicamente y cumple normas estatutarias establecidas de la Universidad de las Fuerzas Armadas - ESPE.

El mencionado trabajo consta de un documento empastado y un disco compacto el cual contiene los archivos en formato portátil de Acrobat (PDF). Se autoriza al Sr. Alex Patricio Dávila Montúfar y el Sr. Gonzalo Wladimir Villares Dávila, que el material se entregue al Ing. Mauricio Campaña, en su calidad de Director de la Carrera.

Sangolquí, 20 de Julio de 2015

  
\_\_\_\_\_  
Ing. Lorena Duque  
DIRECTOR

  
\_\_\_\_\_  
Ing. Paul Díaz  
CODIRECTOR

UNIVERSIDAD DE LAS FUERZAS ARMADAS – ESPE

CARRERA DE INGENIERÍA EN SISTEMAS E INFORMÁTICA

**DECLARACIÓN DE RESPONSABILIDAD**

Nosotros, Alex Patricio Dávila Montúfar y Gonzalo Wladimir Villares Dávila

DECLARAMOS QUE:

El proyecto de grado denominado “ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA EL CONTROL DE COMPRA DE MATERIA PRIMA Y VENTA DE SWEATERS PARA TEJIDOS ANAHÍ UTILIZANDO LA METODOLOGÍA HEFESTO V2.0.”, ha sido desarrollado con base a una investigación exhaustiva, respetando derechos intelectuales de terceros, conforme las citas que constan el pie de las páginas correspondiente, cuyas fuentes se incorporan en la bibliografía.

Consecuentemente este trabajo es de nuestra autoría.

En virtud de esta declaración, nos responsabilizamos del contenido, veracidad y alcance científico del proyecto de grado en mención.

Sangolquí, 20 de Julio de 2015

Alex Patricio Dávila Montúfar

Gonzalo Wladimir Villares Dávila

UNIVERSIDAD DE LAS FUERZAS ARMADAS – ESPE

CARRERA DE INGENIERÍA EN SISTEMAS E INFORMÁTICA

### AUTORIZACIÓN DE PUBLICACIÓN

Nosotros, Alex Patricio Dávila Montúfar y Gonzalo Wladimir Villares Dávila.

Autorizamos a la UNIVERSIDAD DE LAS FUERZAS ARMADAS – ESPE, la publicación, en la biblioteca virtual de la Institución del trabajo “ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA EL CONTROL DE COMPRA DE MATERIA PRIMA Y VENTA DE SWEATERS PARA TEJIDOS ANAHÍ UTILIZANDO LA METODOLOGÍA HEFESTO V2.0.”, cuyo contenido, ideas y criterios son de nuestra exclusiva responsabilidad y autoría.

Sangolquí, 20 de Julio de 2015

Alex Patricio Dávila Montúfar

Gonzalo Wladimir Villares Dávila

### **Dedicatoria**

Este trabajo va dedicado a mis padres Patricio y Amparito por el amor y cariño brindados a lo largo de mi vida y el sacrificio que han realizado para ayudarme a cumplir mis sueños, mis hermanos Paul y Anahí que han sido mi soporte y compañía incondicional en todo momento por más difícil que este sea, los amo infinitamente; y a mi hija Valentina fuente infinita de inspiración en cada uno de mis pasos, te agradezco por transformar en luz los momentos que han nublado mi existencia y enseñarme a vivir nuevamente con alegría, siempre serás mi motivación más grande.

**ALEX PATRICIO DÁVILA MONTÚFAR**

Este Proyecto de Investigación dedico a las personas más importantes MIS PADRES: Roque Villares y María Dávila.

Porque gracias a sus consejos, valores y amor han inculcado en mí la dedicación, esfuerzo y perseverancia, valores que se han convertido en herramientas para luchar por mis objetivos con ahínco y superarme cada día.

Dedico este trabajo a todas las personas que a lo largo de mi vida me han brindado su amistad, apoyo incondicional, comprensión y su experiencia a través de valiosos consejos

De la misma manera, a mis Maestros porque han sido las personas que junto a mis padres me han educado con cariño y respeto, formando a la persona que hoy puede ver cristalizado uno de mis proyectos.

**GONZALO WLADIMIR VILLARES DÁVILA**

## **Agradecimientos**

Quiero brindar un agradecimiento muy especial a todo el personal del Departamento de Ciencias de la Computación por la apertura y apoyo al proyecto planteado y la confianza depositada en su desarrollo, a la Universidad de las Fuerzas Armadas por ser una institución que ha estado al nivel requerido en esta etapa muy personal de mi formación académica, al Ingeniero Mauricio Campaña por su trabajo, a la Ingeniera Lorena Duque y el Ingeniero Paul Díaz por su apoyo y guía a lo largo del proyecto, a la Abogada Alexandra Duque por el apoyo y comprensión brindados y finalmente a todas las personas que me han acompañado con su amistad a lo largo de este camino.

**ALEX PATRICIO DÁVILA MONTÚFAR**

Mi agradecimiento sincero para mis asesores de Tesis: Ingeniera Lorena Duque e Ingeniero Paul Díaz, quienes con sus conocimientos, experiencia y motivación han sido el eje fundamental para la consecución del presente proyecto de investigación; profesionales de élite de quienes he tenido el honor de aprender magníficos aportes de conocimiento, responsabilidad, profesionalismo y seriedad.

Gracias a su dedicación y apoyo ahora tengo la oportunidad de abrir las puertas hacia nuevos retos profesionales con seguridad y confianza para iniciar la búsqueda permanente de mejoras constantes en varios niveles emprendimiento. Por esto y mucho más hacia ustedes mi espíritu de lealtad, admiración y respeto.

**GONZALO WLADIMIR VILLARES DÁVILA**



## Tabla de Contenidos

CAPÍTULO 1. INTRODUCCIÓN E INFORMACIÓN GENERAL.....	1
1.1 TEMA .....	1
1.2 INTRODUCCIÓN .....	1
1.3 FORMULACIÓN DEL PROBLEMA.....	2
1.4 JUSTIFICACIÓN .....	3
1.5 OBJETIVOS .....	3
1.5.1 Objetivo General .....	3
1.5.2 Objetivos Específicos.....	4
1.6 ALCANCE.....	4
CAPÍTULO 2. MARCO TEÓRICO.....	6
2.1 INTRODUCCIÓN AL BUSINESS INTELLIGENCE .....	6
2.1.1 Definiciones .....	6
2.1.2 Características Business Intelligence .....	7
2.1.3 Beneficios del Business Intelligence.....	7
2.1.4 Esquema de Business Intelligence .....	8
2.1.5 Fases de Business Intelligence .....	9
2.1.6 Modelización del Negocio .....	9
2.1.6.1 Indicadores de Negocio.....	9
2.1.6.2 Modelo Dimensional.....	10
2.1.6.3 Esquema Estrella.....	10
2.1.6.4 Esquema Copo de Nieve .....	11
2.1.6.5 Esquema Constelación .....	12
2.2 COMPONENTES DE BUSINESS INTELLIGENCE .....	13
2.2.1 Fuentes de Información.....	14
2.2.2 Proceso ETL.....	14
2.2.3 Datawarehouse .....	15
2.2.3.1 Estructura del Datawarehouse.....	16
2.2.3.2 Datamart.....	18
2.2.4 Herramientas para la explotación y visualización de la Información .....	19
2.2.4.1 OLAP .....	19
2.2.4.2 Query & Reporting.....	21
2.2.4.3 Cuadros de Mando .....	21
2.2.4.4 Datamining.....	22
2.2.4.5 Usuarios .....	23
2.3 METODOLOGÍA HEFESTO V2.0.....	23
2.3.1 Características de la Metodología.....	25
2.3.2 Análisis de requerimientos .....	26
2.3.2.1 Identificar Preguntas .....	26
2.3.2.2 Identificar indicadores y perspectivas .....	26
2.3.2.3 Modelo Conceptual .....	27
2.3.3 Análisis de los OLTP .....	27
2.3.3.1 Conformar Indicadores.....	28
2.3.3.2 Establecer Correspondencias .....	28
2.3.3.3 Nivel de Granularidad .....	28
2.3.3.4 Modelo Conceptual Ampliado .....	28
2.3.4 Modelo lógico del Datawarehouse.....	29

	vii
2.3.4.1	Tipo de Modelo Lógico del Datawarehouse ..... 29
2.3.4.2	Tablas de dimensiones ..... 29
2.3.4.3	Tablas de Hechos ..... 30
2.3.4.4	Uniones ..... 32
2.3.5	Integración de Datos ..... 32
2.3.5.1	Carga Inicial ..... 32
2.3.5.2	Actualización..... 33
2.4	HERRAMIENTAS DE DESARROLLO..... 34
2.4.1	Plataforma Pentaho Open Source Business Intelligence ..... 34
2.4.2	Pentaho Reporting..... 35
2.4.2.1	Pentaho Report Designer..... 35
2.4.2.2	Pentaho Report Designer Wizard..... 35
2.4.2.3	AD-HOC Reporting ..... 36
2.4.3	Pentaho Analysis..... 36
2.4.4	Community Dashboard Editor ..... 37
2.4.5	Pentaho Data Integration..... 38
2.4.6	Data Mining ..... 39
CAPÍTULO 3.	ANÁLISIS Y DISEÑO DEL PROYECTO ..... 41
3.1	DEFINICIÓN DEL PROYECTO ..... 41
3.1.1	Identificación de Roles..... 41
3.1.2	Metodología de Implementación..... 41
3.1.3	Situación Actual ..... 42
3.1.4	Fuentes de Datos ..... 42
3.2	ANÁLISIS DE REQUERIMIENTOS ..... 43
3.2.1	Identificar Preguntas ..... 43
3.2.2	Identificar Indicadores y Perspectivas ..... 44
3.2.3	Modelo Conceptual ..... 46
3.3	ANÁLISIS DE LOS OLTP..... 47
3.3.1	Construcción de los Indicadores ..... 47
3.3.2	Establecer Correspondencia ..... 48
3.3.3	Nivel de Granulidad ..... 51
3.3.4	Modelo Conceptual Ampliado ..... 53
3.4	MODELO LÓGICO DEL DATAWAREHOUSE ..... 54
3.4.1	Estándares de Diseño del Modelo Lógico..... 54
3.4.2	Tipo de Modelo Lógico del Datawarehouse ..... 55
3.4.3	Tablas de Dimensión..... 55
3.4.4	Tablas de Hechos ..... 58
3.4.5	Uniones ..... 60
CAPÍTULO 4.	DESARROLLO Y CONSTRUCCIÓN ..... 62
4.1	INSTALACIÓN Y CONFIGURACIÓN DE HERRAMIENTAS ..... 62
4.1.1	Instalación Power Designer 16 ..... 62
4.1.2	Instalación Java SE Development Kit 7..... 63
4.1.3	Instalación MySQL 5.5 ..... 65
4.1.4	Instalación Administrador de Base de Datos: SQLyog. .... 66
4.1.5	Instalación Servidor Pentaho Community Edition 5.4..... 68
4.1.6	Configuración Pentaho..... 69
4.2	INTEGRACIÓN DE DATOS..... 72
4.2.1	Estándares de Creación ..... 73



	viii
4.2.1.1 Almacenamiento físico del datawarehouse.....	73
4.2.1.2 Estándares de la Base de Datos del Datamart .....	75
4.2.1.3 Estándares de los Cubos Dimensionales .....	75
4.2.1.4 Estándares de Pentaho Data Integration.....	76
4.2.1.5 Estándares de Pentaho Community Dashboard Editor .....	76
4.2.1.6 Estándares de Pentaho Report Designer .....	77
4.2.2 Limpieza de Datos.....	77
4.2.3 Carga Inicial .....	80
4.2.4 Actualización.....	95
4.3 CREACIÓN DE CUBOS MULTIDIMENSIONALES.....	98
4.3.1 Diccionario de Datos.....	98
4.3.2 Creación de Cubos .....	99
4.3.3 Creación de Herramientas de Análisis .....	106
4.3.4 Creación de Tableros de Mando .....	107
4.3.5 Creación de Reportes .....	110
4.3.6 Creación de Usuarios .....	115
CAPÍTULO 5. CONCLUSIONES Y RECOMENDACIONES.....	117
5.1 CONCLUSIONES .....	117
5.2 RECOMENDACIONES .....	117
Bibliografía.....	119
Glosario de Términos .....	121
Curriculum Vitae.....	122

**Lista de tablas**

Tabla 1 Bases de Datos OLAP y OLTP.....	20
Tabla 2 Herramientas para el desarrollo del proyecto. ....	34
Tabla 3 Indicadores y Perspectivas del Proceso de Compras. ....	45
Tabla 4 Indicadores y Perspectivas del Proceso de Ventas.....	45
Tabla 5 Tablas de cada Proceso .....	50
Tabla 6 Estándares de Diseño del Modelo Lógico.....	55
Tabla 7 Estándares de Almacenamiento del Datawarehouse.....	73
Tabla 8 Estándares de la base de datos del datamart. ....	75
Tabla 9 Estándares de los Cubos Multidimensionales.....	75
Tabla 10 Estándares de Pentaho Data Integration.....	76
Tabla 11 Estándares de Pentaho Community Dashboard Editor. ....	76
Tabla 12 Estándares de Pentaho Report Designer. ....	77
Tabla 13 Lista de ETL para la carga inicial. ....	80
Tabla 14 Diccionario de datos del datawarehouse.....	98

## Lista de figuras

Figura 2.1: Esquema de BI.....	8
Figura 2.2: Fases de BI .....	9
Figura 2.3: Ejemplo Diagrama Entidad Relación .....	10
Figura 2.4: Ejemplo Esquema Estrella.....	11
Figura 2.5: Ejemplo Esquema Copo de Nieve.....	12
Figura 2.6: Ejemplo Esquema Constelación.....	12
Figura 2.7: Componentes de Business Intelligence.....	13
Figura 2.8: Proceso ETL .....	15
Figura 2.9: Datawarehouse.....	15
Figura 2.10: Características de un Datawarehouse .....	16
Figura 2.11: Estructura Datawarehouse .....	17
Figura 2.12: Cubo OLAP.....	19
Figura 2.13: Ejemplo de Cuadro de Mando.....	22
Figura 2.14: Usuarios BI.....	23
Figura 2.15: Metodología Hefesto, logotipo versión 2.0.....	24
Figura 2.16: Fases de la Metodología HEFESTO.....	24
Figura 2.17: Modelo Conceptual .....	27
Figura 2.18: Modelo Conceptual Ampliado .....	29
Figura 2.19: Tabla de Dimensión.....	30
Figura 2.20: Tabla de Hechos.....	31
Figura 2.21: Uniones.....	32
Figura 2.22: Ejemplo de flujo proceso ETL.....	33
Figura 2.23: Módulos de la Plataforma Pentaho.....	35
Figura 2.24: Herramienta de reportes.....	36
Figura 2.25: Vista de los diferentes paneles de análisis.....	37
Figura 2.26: Visualización de vistas de dashboard.....	38
Figura 2.27: Esquema Pentaho Data Integration.....	39
Figura 2.28: Módulo de Data Mining.....	40
Figura 3.1: Situación Actual .....	42
Figura 3.2: Modelo Conceptual Compra.....	46
Figura 3.3: Modelo Conceptual Venta .....	46
Figura 3.4: Diagrama Entidad-Relación .....	49
Figura 3.5: Correspondencia Proceso de Compras .....	50
Figura 3.6: Correspondencia Proceso de Ventas.....	51
Figura 3.7: Modelo Conceptual Ampliado Compras .....	53
Figura 3.8: Modelo Conceptual Ampliado Ventas .....	54
Figura 3.9: Tabla Dimensión Proveedor .....	56
Figura 3.10: Tabla Dimensión Tipo Compra .....	56
Figura 3.11: Tabla Dimensión Tiempo .....	56
Figura 3.12: Tabla Dimensión Producto .....	57
Figura 3.13: Tabla Dimensión Cliente.....	58
Figura 3.14: Tabla Dimensión Tiempo .....	58
Figura 3.15: Tabla Métrica Compra.....	59
Figura 3.16: Tabla Métrica Venta .....	59
Figura 3.17: Uniones Proceso de Compras .....	60
Figura 3.18: Uniones Proceso de Ventas .....	61

Figura 4.1: Formulario de suscripción SAP Sybase Power Designer.....	62
Figura 4.2: Instalador Power Designer 16.....	62
Figura 4.3: Selección de Productos Sybase Power Designer.....	63
Figura 4.4: Selección de productos Java.....	63
Figura 4.5: Selección carpeta de instalación JRE.....	64
Figura 4.6: Carpeta componentes Java instalados.....	64
Figura 4.7: Selección del tipo de instalación MySQL.....	65
Figura 4.8: Selección del tipo de configuración del Servidor MySQL.....	66
Figura 4.9: Instalación de SQLyog.....	67
Figura 4.10: Configuración SQLyog con la base de datos.....	67
Figura 4.11: Descarga de Business Analytics Platform.....	68
Figura 4.12: Descarga del Data Integration.....	69
Figura 4.13: Descarga del Report Designer.....	69
Figura 4.14: Carpeta de Instalación Pentaho.....	69
Figura 4.15: Creación de variables de entorno Java.....	70
Figura 4.16: Copia de drivers MySQL dentro de las herramientas Pentaho.....	71
Figura 4.17: Configuración JDNI.....	72
Figura 4.18: Almacenamiento Físico del Datawarehouse.....	74
Figura 4.19: Componentes de Limpieza de Datos.....	78
Figura 4.20: Fórmula fmLimpieza para evaluar los datos de entrada.....	78
Figura 4.21: Set Field Value svReemplazo.....	79
Figura 4.22: Switch/Case para seleccionar los resultados.....	79
Figura 4.23: Componente de Salida Excel seCedula.....	80
Figura 4.24: Componente Entrada Tabla.....	81
Figura 4.25: Carga Inicial Entrada Tabla.....	81
Figura 4.26: Pre visualización de los datos extraídos.....	82
Figura 4.27: Componente Calculadora.....	82
Figura 4.28: Operaciones Calculadora.....	83
Figura 4.29: Componente Java Script Modificado.....	83
Figura 4.30: Valor Java Script Modificado para extraer el número de Semana.....	84
Figura 4.31: Componente Almacén de Datos.....	84
Figura 4.32: Creación del almacén de datos.....	85
Figura 4.33: Carga inicial Compras.....	85
Figura 4.34: Componente de entrada et_Compra.....	86
Figura 4.35: Componente de transformación Calculadora.....	86
Figura 4.36: Componente Java Script jsSemana.....	87
Figura 4.37: Componente Java Script jsMesTexto.....	87
Figura 4.38: Creación de la Dimensión Comprobante dimComprobante.....	88
Figura 4.39: Creación de la Dimensión Tiempo dimTiempo.....	88
Figura 4.40: Creación de la Dimensión Proveedor dimProveedor.....	89
Figura 4.41: Creación de la Dimensión Tipo dimTipo.....	89
Figura 4.42: Creación de la Métrica MetricaCompra.....	90
Figura 4.43: Ejecución de carga inicial Compras.....	90
Figura 4.44: Carga Inicial Ventas.....	91
Figura 4.45: Componente de entrada et_Venta.....	91
Figura 4.46: Creación de la Dimensión Tiempo dimTiempo.....	92
Figura 4.47: Creación de la Dimensión Cliente dimCliente.....	92
Figura 4.48: Creación de la Dimensión Factura dimFactura.....	93

Figura 4.49: Creación de la Dimensión Producto dimProducto. ....	94
Figura 4.50: Creación de la Métrica MetricaVenta. ....	94
Figura 4.51: Ejecución de carga inicial Ventas. ....	95
Figura 4.52: Jobs para la Actualización del ETL. ....	95
Figura 4.53: Creación del Job trbDwhAnahi. ....	96
Figura 4.54: Herramienta START. ....	96
Figura 4.55: Detalles de transformación trfEtlCompra. ....	97
Figura 4.55: Ejecución del Job trbDwhAnahi. ....	97
Figura 4.56: Creación del Modelo de Cubo de Información. ....	99
Figura 4.57: Creación fuente de datos. ....	99
Figura 4.58: Parámetros de conexión al DW ....	100
Figura 4.59: Prueba de conexión. ....	100
Figura 4.60: Selección del tipo de creación de Cubo. ....	101
Figura 4.61: Selección de dimensiones y tabla de hecho. ....	101
Figura 4.62: Creación de las relaciones. ....	102
Figura 4.63: Relaciones entre la tabla de hecho y las dimensiones. ....	102
Figura 4.64: Personalizar el modelo. ....	103
Figura 4.65: Diseño del Modelo de análisis. ....	103
Figura 4.66 Creación de la Dimensión y Jerarquía. ....	104
Figura 4.67: Creación del nivel. ....	104
Figura 4.68: Resultado creación de dimensiones. ....	105
Figura 4.69: Creación de las métricas. ....	105
Figura 4.70: Creación de vista de análisis. ....	106
Figura 4.71: Selección del esquema y cubo. ....	106
Figura 4.72: Resultado Vista de Análisis. ....	107
Figura 4.73: Creación nuevo tablero de mando. ....	107
Figura 4.74: Estructura del Layout. ....	108
Figura 4.75: Consultas SQL para el dashboard. ....	108
Figura 4.76: Creación de Componentes. ....	109
Figura 4.77: Vista Previa del Dashboard. ....	109
Figura 4.78: Directorio para almacenar el Dashboard. ....	110
Figura 4.79: Dashboard Compra Venta Tejidos Anahí. ....	110
Figura 4.80: Directorio Report Designer. ....	111
Figura 4.81: Carga de la herramienta Report Designer. ....	111
Figura 4.82: Pantalla de inicio de Report Designer ....	111
Figura 4.83: Creación del DataSource ....	112
Figura 4.84: Creación nueva de fuente de datos. ....	112
Figura 4.85: Configuración fuente de datos. ....	113
Figura 4.86: Resultado de la configuración fuente de datos. ....	113
Figura 4.87: Sentencias SQL del reporte. ....	114
Figura 4.88: Diseño del reporte. ....	114
Figura 4.89: Publicación del reporte. ....	115
Figura 4.90: Nombre y ruta del reporte. ....	115
Figura 4.91: Creación de Usuarios. ....	116
Figura 4.92: Resultado de la creación de usuarios. ....	116

## **Resumen**

El presente proyecto propone el análisis, diseño e implementación de una solución de Business Intelligence (BI) para la empresa textil Tejidos Anahí. La compañía dispone de una base de datos donde almacena la información contable, pero carece de herramientas que permitan analizar esta información y obtenerla mediante los reportes respectivos, debido a su flujo de trabajo y a los requerimientos de usuario, el proyecto se enfoca en los procesos críticos de Compra y Venta. Para el desarrollo de la solución se utilizó la metodología Hefesto V2.0, por sus ventajas y escalabilidad a la hora del diseño y construcción. La implementación del proyecto fue realizada utilizando la herramienta de software libre Pentaho Community que incluye en su paquete las facilidades ETL, Analysis, Metadata y Reporting. El proyecto está dirigido a implementar un datawarehouse que permita analizar la información relevante, dando un soporte oportuno en el proceso de toma de decisiones que permitan resolver los problemas de la compañía y aumentar su competitividad.

### **PALABRAS CLAVES:**

**BUSINESS INTELLIGENCE**

**DATAWAREHOUSE**

**HEFESTO**

**PENTAHO COMMUNITY**

**SOFTWARE LIBRE**

## **Abstract**

This project proposes the analysis, design and implementation of a Business Intelligence (BI) solution for textile's company Tejidos Anahí. The company has a database which stores the accounting information but lacks of tools to analyze this information and retrieve it with the respective reports, because of its workflow and user requirements, the project focuses on critical processes Purchase and Sale. For the solution's development we use Hefesto V2.0 methodology, for its advantages and scalability in design and construction. The project implementation was performed using the open source tool Pentaho Community that includes in its package ETL, Analysis, Metadata and Reporting facilities. The project is focus in implementing a data warehouse for analyzing the relevant information, providing timely support in decision-making process to solve the problems of the company and increase its competitiveness.

### **KEY WORDS:**

**BUSINESS INTELLIGENCE**

**DATAWAREHOUSE**

**HEFESTO**

**OPEN SOURCE**

**PENTAHO COMMUNITY**



## **CAPÍTULO 1.**

### **INTRODUCCIÓN E INFORMACIÓN GENERAL**

#### **1.1 TEMA**

Análisis, diseño e implementación de una solución de inteligencia de negocios para el control de compra de materia prima y venta de sweaters para Tejidos Anahí utilizando la metodología Hefesto V2.0.

#### **1.2 INTRODUCCIÓN**

Las pequeñas y medianas empresas (pymes) representan uno de los sectores claves en el desarrollo de los países de América Latina, incluido el Ecuador. Una de las herramientas que promueven el desarrollo y fortalecimiento de estas empresas en la economía mundial globalizada son las Tecnologías de la Información y Comunicación (TIC), por lo cual es imprescindible que las pymes estén preparadas para aceptar nuevos paradigmas tecnológicos, herramientas administrativas y cambios en su estructura organizativa, el manejo de su información y sus comunicaciones.

La inteligencia de negocios, dentro de las TIC, surge como herramienta que permite explotar la información de una empresa para encontrar un mayor entendimiento del funcionamiento de ella misma. Esto permite al usuario manipular la información de una manera sencilla, entender su desempeño y plantear escenarios a futuro, lo cual ayudará a tomar mejores decisiones.

El Business Intelligence (BI) es una herramienta directa en la planeación y estrategia comercial de una organización, independientemente del tipo o tamaño, es así que cumple con la función empresarial de proporcionar información necesaria para la toma de decisiones. El proporcionar información a los usuarios mediante un sistema de inteligencia de negocios, apoya a la organización a detectar amenazas y oportunidades, ya que uno de las principales funcionalidades es la de recopilar datos de todas las fuentes relevantes, validarla, procesarla, entregarla en el formato requerido y de forma más interactiva para los usuarios.

Tejidos Anahí empresa textil de la ciudad de Atuntaqui, inicia sus labores en el año de 2001, su principal actividad es la producción y comercialización de sweaters. Dentro de la organización, es de gran importancia que la tecnología se encuentre alineada con los objetivos, razón por la cual, luego de realizar un análisis de la situación de la empresa, se opta por implementar una solución de inteligencia de negocios para proporcionar información necesaria para la toma de decisiones.

Los datos constituyen un elemento primordial para la empresa, sobre todo para el desarrollo de una solución BI, el objetivo central de la implementación es explotar la información relevante sobre el desempeño de la organización, misma que será de gran importancia en el área gerencial de Tejidos Anahí en los procesos de evaluación del rendimiento y toma de decisiones.

### **1.3 FORMULACIÓN DEL PROBLEMA**

Tejidos Anahí es una empresa textil, que se dedica a la elaboración y venta de sweaters de lana, uno de los inconvenientes a los que se enfrenta la organización, es no contar con informes empresariales precisos que le permitan conocer el estado de los movimientos que se realizan en un periodo de tiempo, el desempeño de la empresa, así como reportes de las compras de materia de prima y las ventas realizadas conforme han ido planificando, en la actualidad el proceso se realiza a partir de un software contable que no cumple con todos los requisitos generados por la organización y a su vez se apoyan en herramientas como Excel obteniendo resultados no siempre precisos, esto genera una pérdida de tiempo y recursos realizando estos procesos.

El contar con información precisa y oportuna, será de gran ayuda para que la organización pueda tomar decisiones importantes de forma inmediata, mismas que serán fundamentales para el cumplimiento de objetivos planteados en el plan estratégico de la organización y a su vez, tomar los respectivos correctivos necesarios sobre la marcha, finalmente permitirá visualizar en forma general el estado financiero de la organización.

## **1.4 JUSTIFICACIÓN**

Tejidos Anahí en la actualidad no cuenta con herramientas que le permitan obtener información financiera actualizada de forma rápida, lo que obliga al personal gerencial de Tejidos Anahí a realizar todo este tipo de cálculos e informes mediante tablas dinámicas en Excel, lo que implica costos de personal y demora en la obtención de resultados finales.

El tener el control sobre los movimientos de compras de materia prima y ventas de productos, que se realiza a diario permitirá visualizar el estado en la que se encuentra la organización, con la finalidad de definir estrategias de mejora basada en el análisis de información rápida y oportuna. Además esta información ayudará a detectar falencias en los procedimientos y mejorar los procesos de producción y ventas de la compañía.

La implementación de una solución de Inteligencia de Negocios para Tejidos Anahí, permitirá visualizar de una forma amplia la situación de la organización, de la misma forma, generar una reducción de costos operativos, puesto que la información será recibida de forma más rápida y todos los reportes generados serán hechos de manera automáticamente y en tiempo real, reduciendo tiempos de respuesta para la obtención de datos procesados y cálculos finales; finalmente ayudará a corregir errores en los procesos internos de la organización y mejorar la productividad de la compañía.

## **1.5 OBJETIVOS**

### **1.5.1 Objetivo General**

Analizar, diseñar e implementar una solución de inteligencia de negocios para Tejidos Anahí, basada en la metodología Hefesto V2.0, mediante la utilización de herramientas de software libre de inteligencia de negocios.

### 1.5.2 Objetivos Específicos

- Investigar y utilizar la metodología de Hefesto V2.0 para el desarrollo de una solución BI dentro de una empresa PYME en el ámbito textil.
- Analizar los procesos internos de compra y venta de Tejidos Anahí para definir los indicadores que necesita la compañía.
- Determinar las fuentes de datos, diseñar la estructura, diseñar el modelo, las dimensiones y los cubos de Compra y Venta.
- Diseñar la navegabilidad del contenido, reportes y el cuadro de mando para el control de procesos correspondiente a cada datamart.

### 1.6 ALCANCE

Se realizará el análisis, diseño e implementación de una solución de Inteligencia de Negocios, para el soporte y control en la toma de decisiones en Tejidos Anahí que estará orientado a los procesos de Compras y Ventas.

Debido a la dimensión de la compañía, los procesos clave dentro del negocio, su flujo de datos y la cantidad de transacciones y datos, la solución comprenderá las siguientes áreas:

- Compra.
- Venta.

El datawarehouse estará constituido por un cubo de información para cada una de las áreas mencionadas, ayudará a estructurar los datos y entregar la información solicitada por el usuario para la toma de decisiones, la gerencia será el responsable del análisis de los cubos.

La aplicación contempla la consolidación de información de la compañía y se desarrollarán los reportes para cada una de las áreas que se adapten a las necesidades

de los usuarios finales, estos se harán en base a los indicadores más importantes donde se detallan los siguientes:

- Monto total de compras a cada proveedor en un periodo de tiempo.
- Monto total de compras en un periodo de tiempo.
- Monto total de compras por tipo de adquisición en un periodo de tiempo.
- Monto total de IVA en Compras, Retención a la Fuente y Retención de IVA en un periodo de tiempo.
- Unidades vendidas de cada producto a cada cliente en un periodo de tiempo.
- Monto total de ventas de cada producto en un periodo de tiempo.
- Monto total de ventas a cada cliente en un periodo de tiempo.
- Monto total de IVA en ventas en un periodo de tiempo.
- Monto total de Compras y Monto Total de Ventas en un periodo de tiempo.

El proyecto se desarrollará en base a la metodología Hefesto V2.0 y se implementará con herramientas de software libre Pentaho Community V5.4 en todas sus etapas.

## CAPÍTULO 2. MARCO TEÓRICO

### 2.1 INTRODUCCIÓN AL BUSINESS INTELLIGENCE

#### 2.1.1 Definiciones

*“Business intelligence (BI) es un término paraguas que abarca los procesos, las herramientas y las tecnologías para convertir datos en información, información en conocimiento y planes para conducir de forma eficaz las actividades de los negocios. BI abarca las tecnologías de datawarehousing los procesos en el ‘back end’, consultas, informes, análisis y las herramientas para mostrar información (estas son las herramientas de BI) y los procesos en el ‘front end’.” (Eckerson, 2005).*

*“Business Intelligence es la habilidad para transformar los datos en información y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios.” (Sinnexus, 2012).*

*“Business Intelligence (BI) es un compendio de tecnologías y aplicaciones que permiten recopilar la información de los diferentes fuentes de su empresa, para almacenarla, analizarla y proveerla a todo tipo de usuarios de su empresa con el fin de que puedan tomar mejores decisiones de negocio.” (Intellego, 2015).*

Mediante las diferentes definiciones obtenidas, se llega a la siguiente conclusión:

Business Intelligence (BI) es un conjunto de tecnologías, estrategias y herramientas que transforman los datos en información, mismas que permiten optimizar procesos que intervienen en la toma de decisiones de negocios.

BI permite la recolección de información efectiva, de manera que combinando datos y análisis, se pueda obtener el conocimiento necesario para

formular estrategias de mejora continua, permitiendo optimizar los procesos de planificación para el cumplimiento de objetivos.

### **2.1.2 Características Business Intelligence**

El conjunto de herramientas, técnicas y metodologías que componen una solución de Business Intelligence se definen por las siguientes características:

- Soporte en la toma de decisiones mediante el uso de herramientas propias de BI, que proporcionan información eficiente y eficaz en el momento oportuno.
- Acceso de los usuarios a la información independientemente de la procedencia de la misma.
- Integración y análisis de la información desde fuentes de datos heterogéneas, de manera completa, coherente y correcta.
- Permite al usuario visualizar la información de una forma amigable, interactiva y fácil de utilizar; no es necesario un conocimiento técnico para su uso.
- Permite comprender la situación pasada, presente y futura de una compañía, a través del manejo histórico de la información.
- Apoya a la toma de decisiones estratégicas, tácticas y operativas de una empresa en todas las áreas involucradas en la toma de las mismas.
- A partir de los datos proporcionados por un sistema de Business Intelligence es posible descubrir conocimiento.
- Capacidad de retroalimentar el conocimiento adquirido.
- Las actividades de inteligencia de negocios pueden ser agrupadas en tres categorías: Reporteo, Procesamiento analítico en línea (OLAP) y Minería de datos.

### **2.1.3 Beneficios del Business Intelligence**

Uno de los objetivos básicos del Business Intelligence es que servirá como herramienta de ayuda en la toma de decisiones, además permite descubrir cosas que hasta el momento se desconocen, es decir, ayuda obtener conocimiento.



Los beneficios que se obtienen de un sistema de BI pueden ser de distintos tipos:

- Beneficios Tangibles: por ejemplo reducción de costos, disminución de gastos, generación de ingresos, reducción de tiempos de producción, etc.
- Beneficios Intangibles: tener la información disponible para la toma de decisiones en nuestra compañía hará que aumente la ventaja competitiva; por ejemplo optimizar la atención al cliente, aumentar la satisfacción de los clientes, mejorar el acceso a los datos, etc.
- Beneficios Estratégicos: facilitan la formulación de estrategias, a quién?, dónde? y con qué productos dirigirnos?; por ejemplo mayor habilidad para analizar estrategias de precios, diversificar la generación de productos, mayor visibilidad de la gestión, etc.

#### 2.1.4 Esquema de Business Intelligence

A continuación se muestra el esquema estructural de una solución BI:

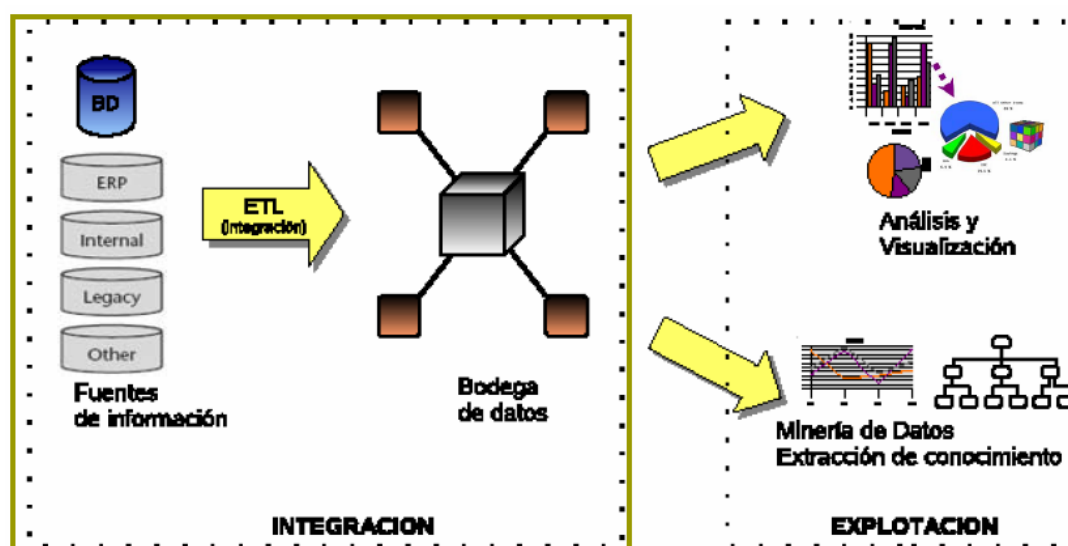


Figura 2.1: Esquema de BI.

Fuente: (Guzmán, 2010).

### 2.1.5 Fases de Business Intelligence

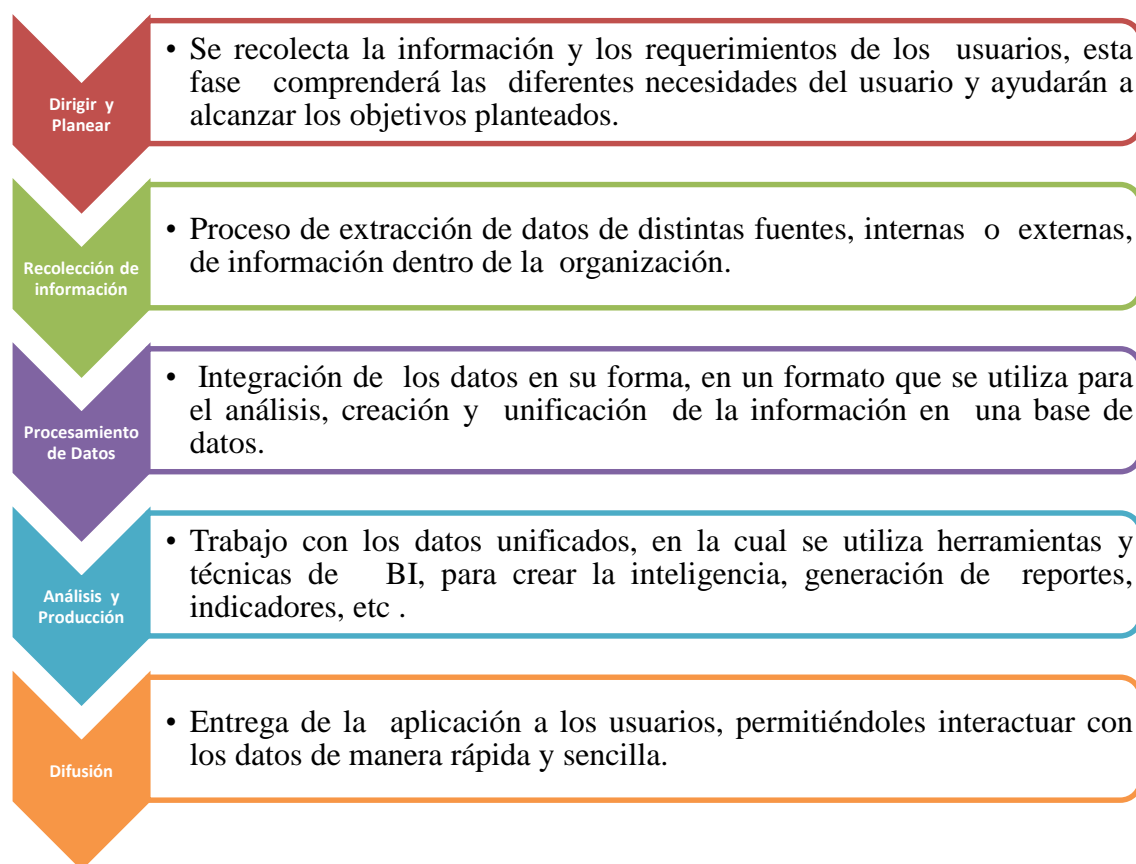


Figura 2.2: Fases de BI

### 2.1.6 Modelización del Negocio

Un modelo de negocio es una representación del funcionamiento del mismo a partir de la cual se puede analizar lo que sucede en la realidad. La construcción de un modelo de negocios ayuda a visualizar de qué manera afectarán los cambios que se introduzcan al resultado final buscado.

#### 2.1.6.1 Indicadores de Negocio

Los indicadores de negocios o KPI (Key Performance Indicator) son métricas cuantificables del nivel de desempeño de un proceso, relacionado con un objetivo específico. Estos factores se utilizan para evaluar si las organizaciones estas alcanzando sus objetivos, reflejar el estado actual de un negocio y definir una línea de acción futura.

Los KPI se caracterizan por ser medibles, alcanzables, relevantes y disponibles a tiempo en todas aquellas actividades que son críticas para conseguir el éxito dentro de una organización.

### 2.1.6.2 Modelo Dimensional

Una vez definido el modelo del negocio se debe transformar en un modelo dimensional, para ello es necesario determinar la información que se tiene disponible para comenzar a analizarlo.

Para el análisis se debe utilizar un modelo relacional, formado por tablas, compuestas a su vez de atributos o campos y las relaciones existentes entre las mismas. Cada tabla tiene una Clave primaria o PK (Primary Key) y se relaciona con las demás mediante las Claves externas o FK (Foreign Key) que a su vez son PK en sus propias tablas. Este modelo relacional está presente en la mayoría de sistemas de información actuales.

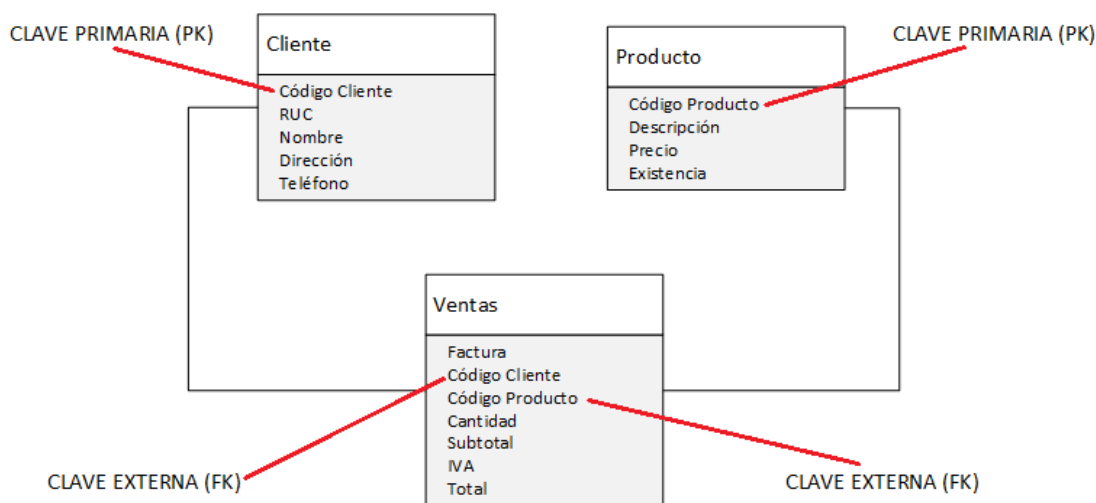


Figura 2.3: Ejemplo Diagrama Entidad Relación

### 2.1.6.3 Esquema Estrella

Este esquema es la arquitectura de almacén de datos más simple, se construye a partir del modelo relacional anteriormente revisado, donde el centro de una estrella

es una tabla conocida como Tabla de Hechos (aquello que se quiere medir o analizar) y los puntos de la estrella conocidos como Tablas de Dimensiones (como se quiere medir).

El esquema estrella implementa un diseño lógico relacional de base de datos, donde los Hechos representan la Tercera Forma Normal (3FN) y las dimensiones en Segunda Forma Normal (2FN).

La clave primaria de una tabla de hechos está formada por todas las columnas que forman las dimensiones, mientras que las columnas que contienen datos numéricos simplemente forman parte de los informes.

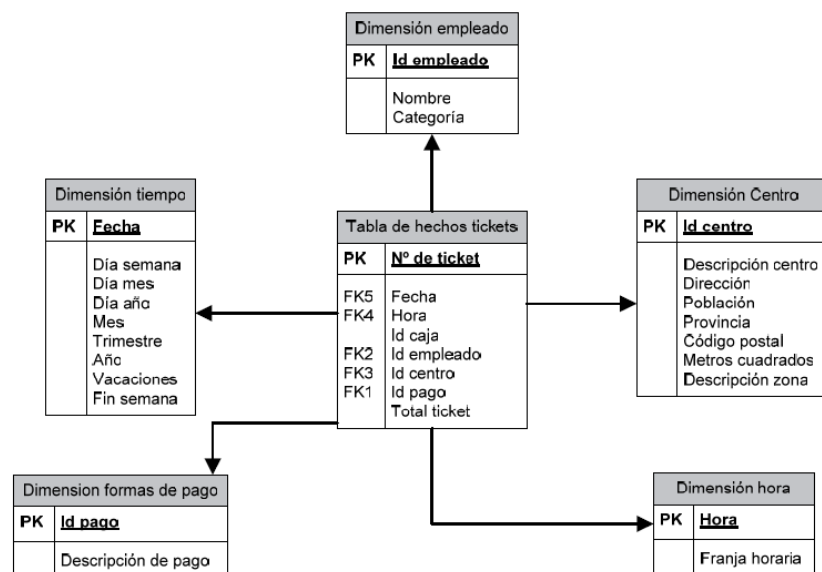


Figura 2.4: Ejemplo Esquema Estrella.

Fuente: (Cano, 2007)

#### 2.1.6.4 Esquema Copo de Nieve

Este esquema es una variedad más compleja del esquema estrella, las dimensiones se representan mediante relaciones normalizadas (3NF) y forman parte de un modelo relacional de base de datos.

Este esquema está enfocado a facilitar el mantenimiento de las dimensiones, además al estar las tablas totalmente normalizadas, reduce el espacio que ocupan.

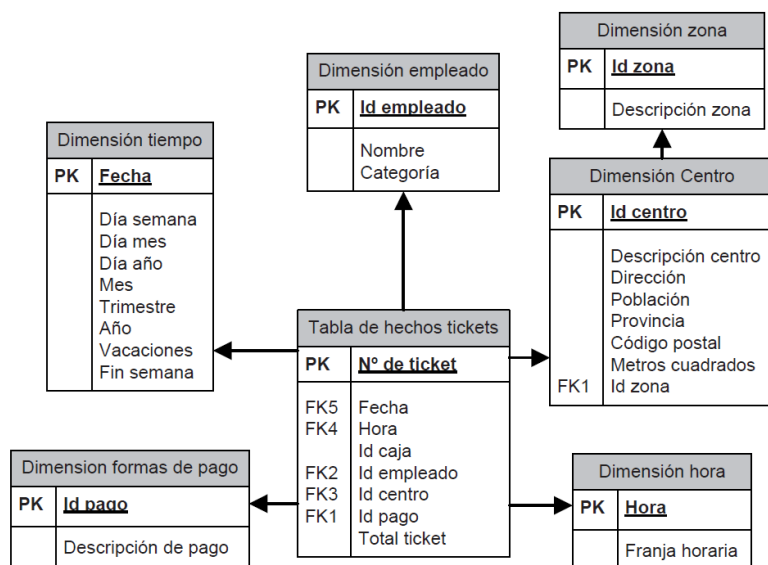


Figura 2.5: Ejemplo Esquema Copo de Nieve.

Fuente: (Cano, 2007)

### 2.1.6.5 Esquema Constelación

Este esquema está formado por una serie de esquemas estrella, consta de una tabla de hechos principal y una o más tablas de hechos auxiliares. Estas tablas se encuentran en el centro del modelo y están rodeadas con sus respectivas tablas de dimensiones.

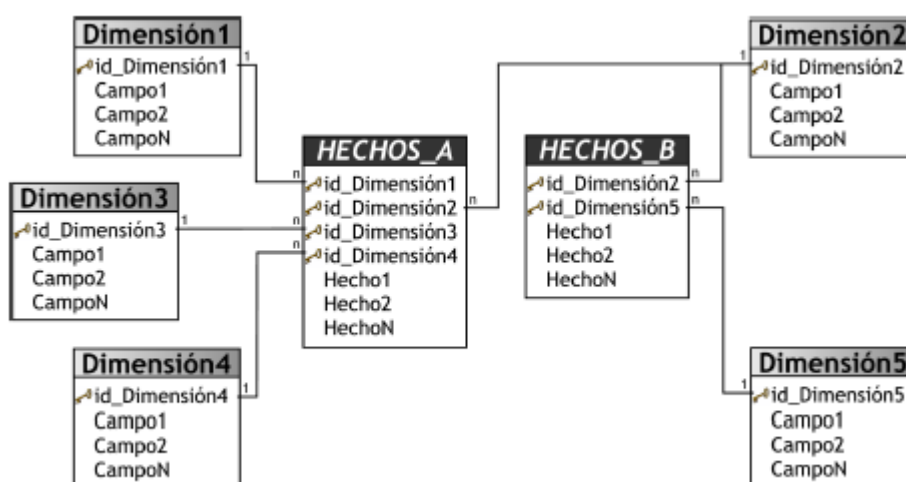


Figura 2.6: Ejemplo Esquema Constelación.

Fuente: (Cano, 2007)

## 2.2 COMPONENTES DE BUSINESS INTELLIGENCE

Es necesario conocer los componentes básicos que intervienen durante el desarrollo de BI, para lo cual se definen los siguientes términos:

- Fuentes de información que alimentan el datawarehouse.
- Procesos ETL (Extract Transform Load) de extracción, transacción y carga de datos al datawarehouse.
- Datawarehouse o almacén de datos.
- Motor OLAP, que provee capacidad de cálculo, consultas, funciones de planeamiento, pronóstico y análisis de datos.
- Herramientas de visualización, que permiten el análisis y navegación de la solución.

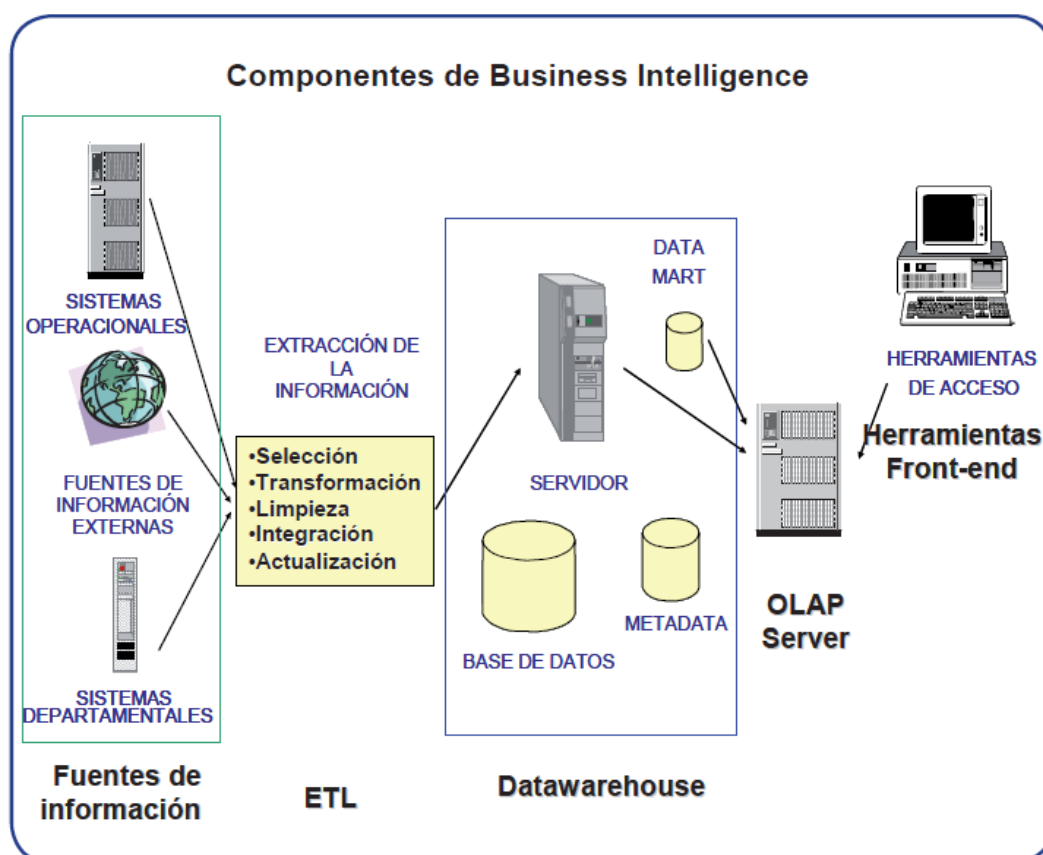


Figura 2.7: Componentes de Business Intelligence.

Fuente: (Cano, 2007)

### 2.2.1 Fuentes de Información

Las fuentes de información a las que se puede acceder para alimentar el almacén de datos son:

- Sistemas transaccionales y operacionales disponibles en una empresa, incluyen sistemas a la medida, ERP, CRM, SCM, etc.
- Sistemas de información departamentales, normalmente disponibles en hojas de cálculo.
- Fuentes externas de información, incorporadas por terceros a la compañía, por ejemplo estudios de mercado, estudios poblacionales, etc.

La información que se va a cargar en el datawarehouse es normalmente estructurada, es decir, aquella que se puede almacenar en tablas.

### 2.2.2 Proceso ETL

El proceso de extracción, transformación y carga (ETL) trata de recuperar los datos de las diferentes fuentes de información y alimentar al datawarehouse. El proceso ETL “consume entre el 60% y el 80% del tiempo de un proyecto de Business Intelligence, por lo que es un proceso clave en la vida de todo proyecto” (Eckerson, 2003).

Este proceso a su vez se divide en 5 subprocesos:

- Extracción de los datos en bruto de las diferentes fuentes de información.
- Limpieza de los datos eliminando valores duplicados y erróneos, para obtener datos de calidad.
- Transformación de datos limpios en datos consistentes para el análisis.
- Integración de los datos validados que se carga en el datawarehouse.
- Actualización periódica de los datos en el datawarehouse.



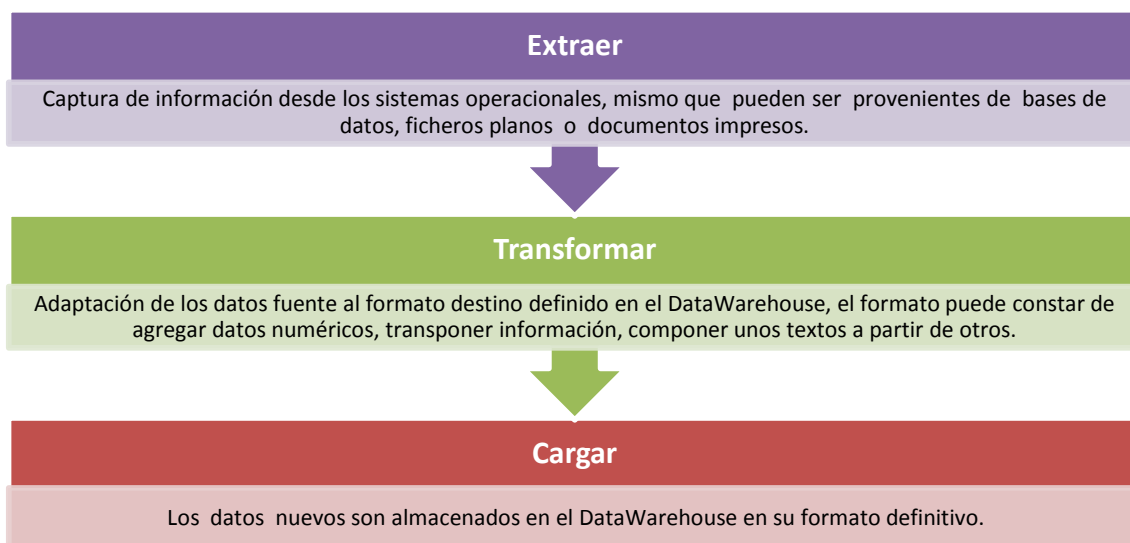


Figura 2.8: Proceso ETL

### 2.2.3 Datawarehouse

“Un datawarehouse es una colección de información creada para soportar las aplicaciones de toma de decisiones” (Inmon, 1992).

El datawarehouse es una base de datos que almacena información histórica del negocio la cual se integra y depura, esta información se obtiene de diferentes fuentes para luego procesarlas con la finalidad de convertirlas en información. La creación de un datawarehouse, permite implantar una solución Business Intelligence y se alimenta a partir de los datos operacionales mediante las herramientas ETL.

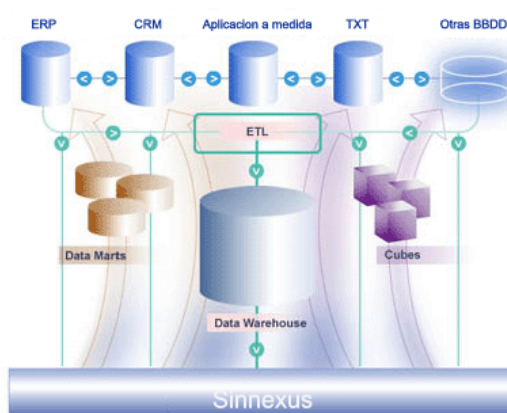


Figura 2.9: Datawarehouse.

Fuente: (Sinexus, 2015)

### Características de un Datawarehouse

El datawarehouse se caracteriza por ser: integrado, temático, de tiempo variable y no volátil, cada característica está definida en la Figura 2.10.



Figura 2.10: Características de un Datawarehouse

#### 2.2.3.1 Estructura del Datawarehouse

La estructura del datawarehouse está constituida en diferentes niveles de esquematización y detalle, los diferentes componentes que la conforman son los siguientes:

- Detalle actual de datos.
- Detalle histórico de datos.
- Datos ligeramente resumidos.
- Datos altamente resumidos.
- Meta data.

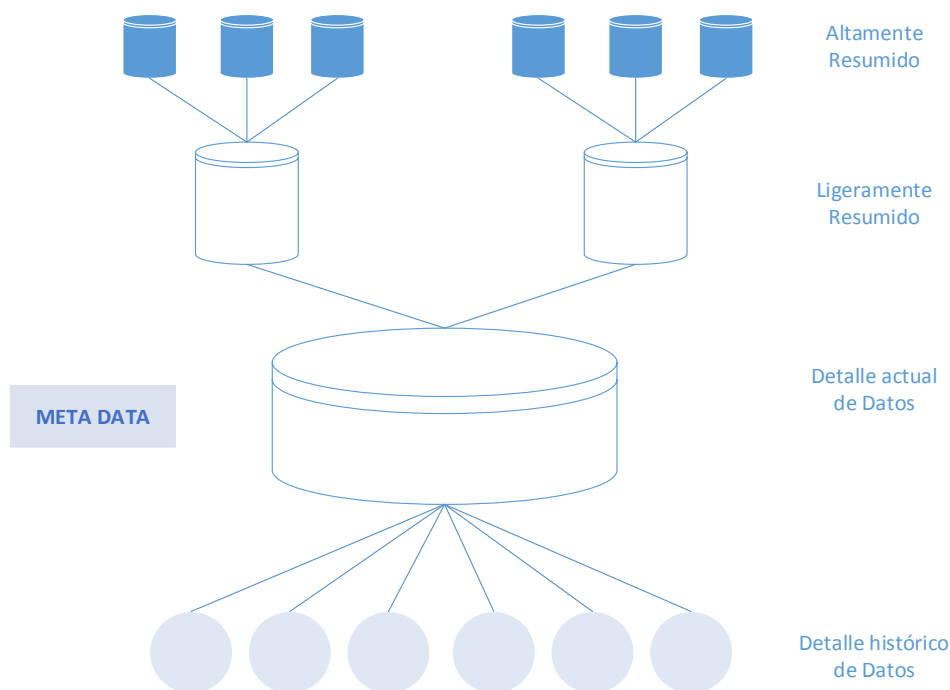


Figura 2.11: Estructura Datawarehouse

### **Detalle actual de datos**

Almacena a bajo nivel de granularidad, muestra datos y procesos recientes, es de fácil acceso pero la administración es costosa y compleja.

### **Detalle histórico de datos**

Es aquella almacenada masivamente, se almacena a nivel de detalle y su acceso no es frecuente por lo cual se encuentra ubicada en algún dispositivo de almacenamiento que no es el disco.

### **Datos ligeramente resumidos**

Por lo general estos datos provienen del nivel del detalle actual y se almacenan en el disco. Los puntos en los que se basa su construcción son:

- Unidad de tiempo esquematizada.
- Contenidos con data ligera resumida.

### **Datos altamente resumidos**

Datos completamente resumidos son accesibles y compactos, estos forman parte del datawarehouse independientemente de su ubicación física.

### **Metadata**

La metadata se ubica en una dimensión diferente al de otros datos, ya que el contenido no se lo toma directamente desde el ambiente operacional. De manera general *“Metadata es la información sobre los datos que se alimenta, transforma y existe en el datawarehouse”* (BI-Business Intelligence Blog, 2009).

La metadata es usada como directorio de contenidos del datawarehouse, guía de mapping y guía de algoritmos usados para esquematizar los datos.

#### **2.2.3.2 Datamart**

Un Datamart es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Funciona como un subconjunto del datawarehouse enfocado a un aspecto específico del negocio.

Un datamart 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. Puede ser alimentado desde los datos de un datawarehouse, o integrar por sí mismo un compendio de varias fuentes de información.

#### **Características del Datamart**

Un datamart tiene las siguientes características:

- Poco volumen de datos.
- Mayor rapidez de consulta.
- Consultas SQL y/o MDX sencillas.
- Validación directa de la información.
- Facilidad para la historización de los datos.

## 2.2.4 Herramientas para la explotación y visualización de la Información

Estas herramientas permiten manipular y visualizar la información que se almacena dentro del datawarehouse. Estas herramientas son:

- OLAP.
- Query & Reporting.
- Cuadros de Mando.
- Data Mining.
- Usuarios.

### 2.2.4.1 OLAP

OLAP (On-Line Analytical Processing) son herramientas que permiten realizar consultas complejas a grandes cantidades de datos. Proporciona y utiliza estructuras multidimensionales, cubos multidimensionales, que permiten al usuario explorar la información desde distintas perspectivas y periodos de tiempo; además provee indexación especializada y poderosas herramientas de cálculo.

La representación de grafica de los OLAP son los cubos.

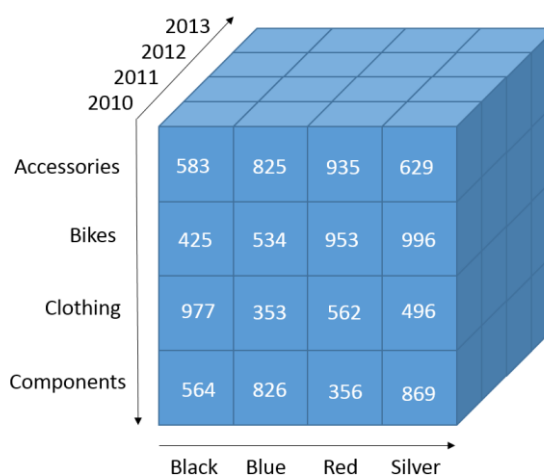


Figura 2.12: Cubo OLAP.

Fuente: (Component One Studio, 2013)

Las herramientas OLAP permiten rotar los cubos, es decir, cambiar el orden de las dimensiones dependiendo de nuestras necesidades.

## Tipos de Bases de Datos OLAP y OLTP

Tabla 1

*Bases de Datos OLAP y OLTP*

	OLAP	OLTP
DEFINICIÓN	Procesamiento Analítico en Línea.	Procesamiento de Transacciones En Línea.
OBJETIVOS	Asistir en el análisis del negocio. Identificando tendencias, comparando periodos, gestiones, mercados, índices mediante el almacenamiento de datos.	Asistir a aplicaciones específicas. Mantener integridad de los datos.
ALINEACIÓN DE DATOS	Están alineados por dimensión. Los datos son organizados definiendo dimensiones del negocio. Se focaliza en el cumplimiento de requerimientos del análisis del negocio.	Están alineados por aplicación. Se focaliza en el cumplimiento de requerimientos de una aplicación especial o una tarea específica.
INTEGRACIÓN DE DATOS	Los datos deben ser integrados. Son conocidos como datos derivados o DSS, dado que provienen de sistemas transaccionales y sistemas de archivos maestros.	Los datos no están integrados. Son calificados como datos primitivos, operacionales. Estructurados independientemente. Almacenados en diferentes formatos de archivos. Pueden residir en diferentes plataformas de hardware o RDBMS.
HISTORIA	Almacenan tanta historia como sea necesario para el análisis del negocio, son guardados por 2 a 5 años.	Retienen datos para 60 o 90 días después son resguardados por administradores de BDD en almacenamientos secundarios.
ACCESO Y MANIPULACIÓN DE DATOS	Tienen una carga y acceso masivo de datos, la carga y refresco es batch (bulk copy). La validación de datos se realiza antes o después de la carga, se realizan sentencias select sobre varios registros y tablas.	Realizan manipulación de datos registro por registro con inserts, updates y deletes. Necesitan rutinas de validación y transacciones a nivel de registro.
PATRONES DE USO	Patrón de uso liviano con picos de uso eventuales en el tiempo. Los picos de uso suceden diario o semanal.	Patrón de uso constante. Requiere grandes cantidades de recursos consumiendo solo el tiempo referido a la transacción.

Continúa ➡

PERFIL DE USUARIO	DE	El perfil de usuario corresponde a la comunidad gerencial para la toma de decisiones.	El perfil de usuario corresponde a los que interactúan con dichos sistemas, puesto que es la comunidad operativa.
-------------------	----	---	---

#### 2.2.4.2 Query & Reporting

Son herramientas que permite realizar consultas complejas e informes financieros que convierten los datos de la empresa en información esencial y pueden presentarse en distintos formatos adaptados a los destinatarios.

#### Características de las Herramientas Reporting

- Informes dinámicos.
- Información jerarquizada.
- Disponibilidad de formatos de salida de informes.
- Distintos canales de salida.

#### Beneficios de Query & Reporting

- El uso de parámetros predefinidos para que los usuarios no expertos puedan solicitar la ejecución de informes o consultas predefinidas.
- Los usuarios con experiencia, pueden generar consultas flexibles mediante aplicaciones que cuenten con interfaz gráfica.
- Los usuarios altamente experimentados, generan consultas en lenguaje de interrogación de datos.
- Hay una extensa gama de herramientas en el mercado para cumplir esta funcionalidad.

#### 2.2.4.3 Cuadros de Mando

Los cuadros de mando son una herramienta de gestión que traduce la estrategia de una empresa en un conjunto coherente de indicadores, es decir una metodología para diseñar, formular e implantar estrategias.

Los cuadros de mando ofrecen una visualización intuitiva de los indicadores empresariales para aquellos perfiles de usuarios responsables de monitorizar,

controlar y gestionar los procesos de negocio, además pueden ser personalizados para cada área de competencia.

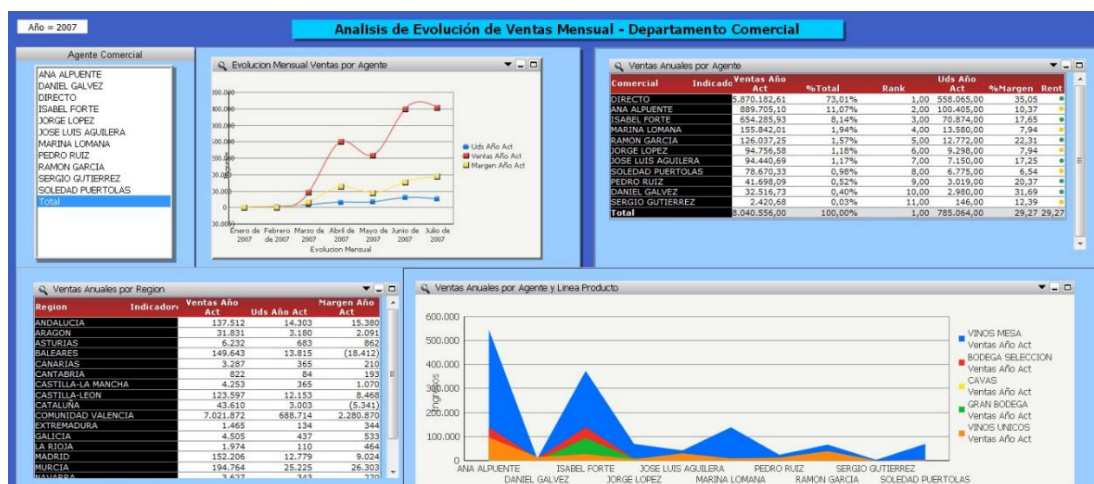


Figura 2.13: Ejemplo de Cuadro de Mando.

Fuente: (Espinoza, 2010)

Existen dos tipos de cuadros de mando:

- Cuadros de mando analíticos (Dashboard): a partir de los datamarts permiten obtener informes e indicadores claves, analizan áreas de negocio no relacionadas.
- Cuadros de mando integrales (Balanced Scorecard): permiten obtener una visión estratégico del conjunto de objetivos e indicadores de toda la organización, se desarrollan a nivel de todo el negocio.

#### 2.2.4.4 Datamining

Datamining o minería de datos es el proceso para encontrar e interpretar patrones y relaciones hasta ahora desconocidos de la información dentro del negocio.

Las herramientas datamining permiten analizar los datos de tal forma que ayuden a crear modelos estadísticos y probabilísticos de la organización, prediciendo futuras tendencias y comportamientos.



### 2.2.4.5 Usuarios

Algunos autores consideran a los usuarios como componentes del Business Intelligence, estos están encargados de tomar decisiones y planificar las actividades de negocio. Los usuarios están divididos en dos grupos: los productores de información y los consumidores de información.

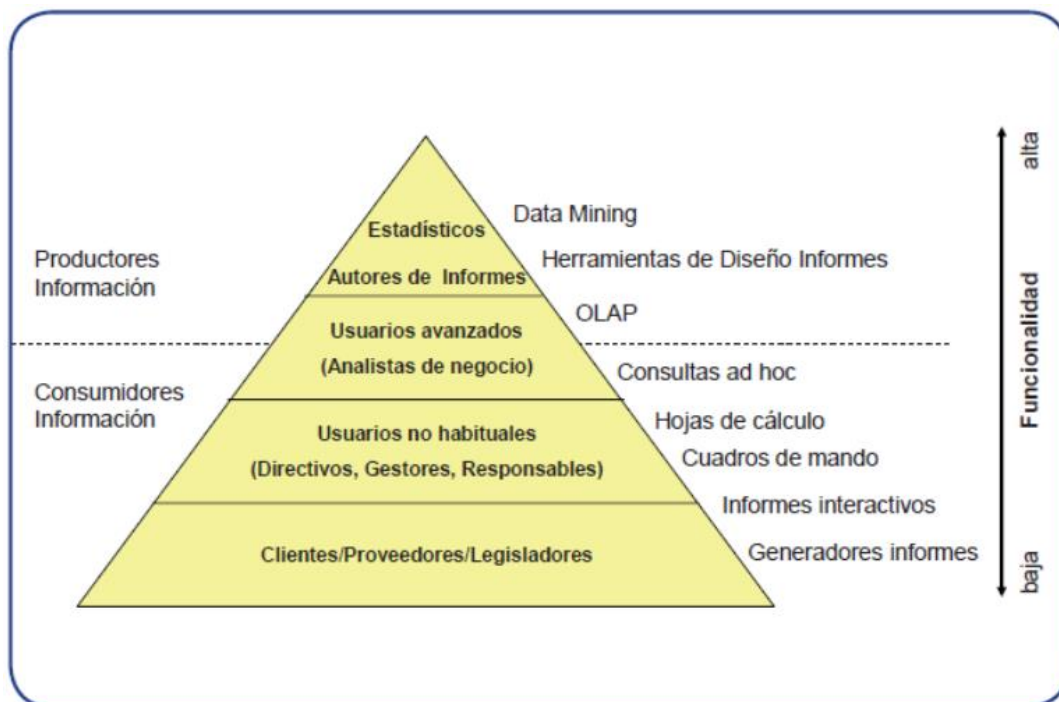


Figura 2.14: Usuarios BI.

Fuente: (Cano, 2007)

## 2.3 METODOLOGÍA HEFESTO V2.0

HEFESTO V2.0 es una metodología propia, de muy amplia investigación, que permitirá la construcción de un datawarehouse de forma sencilla, ordenada e intuitiva.

Esta metodología se encuentra en continua evolución y para su desarrollo se han tomado en cuenta todos los feedbacks que han aportado quienes la han utilizado en diversos países y con diversas finalidades. La construcción e implementación se puede aplicar a cualquier ciclo de vida del desarrollo de software.



Figura 2.15: Metodología Hefesto, logotipo versión 2.0.

Los pasos de la metodología HEFESTO puede resumirse a través del siguiente gráfico:



Figura 2.16: Fases de la Metodología HEFESTO.

Fuente: (Bernabeu, 2010)

El proceso inicia recolectando las necesidades de información de los usuarios y obteniendo las preguntas claves del negocio. Luego, se debe identificar los indicadores resultantes y sus respectivas perspectivas de análisis, mediante las cuales se construirá el modelo conceptual de datos del datawarehouse.

Inmediatamente, se analizarán los OLTP para determinar cómo se construirán los indicadores, señalar las correspondencias con los datos fuentes y seleccionar los campos de estudio de cada perspectiva.

Luego se realiza la construcción del modelo lógico del almacén de datos, en donde se definirá el tipo de esquema que se va a implementar, además, se establecerá las tablas de dimensiones y hechos con sus respectivas relaciones.

Finalmente, utilizando procesos y subprocesos ETL, se definirán políticas y estrategias para la carga inicial de datos en el datawarehouse y su actualización respectiva.

Los siguientes temas de este capítulo están desarrollados en base al libro del Ing. Ricardo Bernabeu: HEFESTO V2.0, específicamente en la sección HEFESTO: Metodología propia para la Construcción de un Data Warehouse, que es una recopilación actualizada de las mejores prácticas de la metodología.

### **2.3.1 Características de la Metodología.**

- Los objetivos y resultados esperados de cada fase son fácilmente distinguibles y de fácil comprensión.
- Está basada en los requerimientos del usuario, por lo cual su estructura es fácilmente adaptable.
- Disminuye los requerimientos de cambio, ya que involucra al usuario final en cada etapa del desarrollo.
- Utiliza modelos conceptuales y lógicos, sencillos de interpretar y analizar.
- Es independiente del tipo de ciclo de vida.
- Es independiente de las herramientas que se utilicen para su implementación.
- Es independiente de las estructuras físicas que contenga el datawarehouse y de su distribución.
- Al culminar una fase, los resultados obtenidos son el punto de partida de la fase siguiente.

- Se aplica tanto para datawarehouse como para datamart.

### **2.3.2 Análisis de requerimientos**

Se identifica los requerimientos del usuario a través de preguntas que expliquen los objetivos de la empresa, se procede a analizar estas inquietudes e identificar cuáles serán los indicadores que se tomaran en cuenta al momento de la construcción del datamart.

Se debe considerar para la construcción el número de datamarts, si se requiere dos por ejemplo, se deberá aplicar la metodología dos veces, es decir una para cada datamart.

#### **2.3.2.1 Identificar Preguntas**

La primera fase consiste en el acopio de la información, la cual puede obtenerse a través de diferentes formas y técnicas. El objetivo principal de esta fase es el de identificar y obtener las necesidades de información, que es esencial para llevar a cabo las metas y estrategias de la empresa ya que facilitara una eficaz y eficiente toma de decisiones.

Un punto importante a tenerse es que la información debe estar sustentada por algún OLTP, para poder realizar posteriormente la elaboración del datamart.

#### **2.3.2.2 Identificar indicadores y perspectivas**

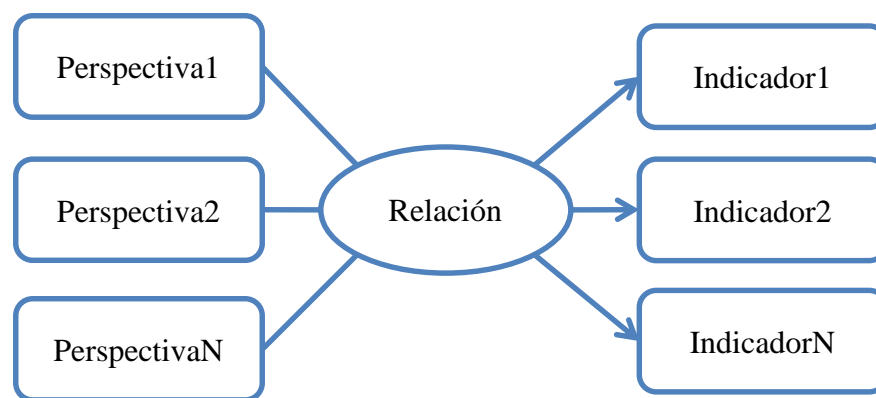
Una vez establecido las preguntas de negocio se procede a su descomposición para determinar los indicadores que se utilizaran y las perspectivas de análisis que intervendrán.

Se debe considerar que los indicadores, para que sean efectivos, son en general valores numéricos y representan lo que se quiere analizar, mientras que las perspectivas se refieren a los objetos con los cuales se va a examinar los indicadores.

### 2.3.2.3 Modelo Conceptual

*“Modelo Conceptual: descripción de alto nivel de la estructura de la base de datos, en la cual la información es representada a través de objetos, relaciones y atributos.” (Data Prix, 2009).*

En esta etapa se construye un modelo conceptual a partir de los indicadores y perspectivas obtenidas en el punto anterior, con el modelo se observara el alcance del datamart, para luego trabajar sobre el mismo.



*Figura 2.17: Modelo Conceptual*

Como se observa en la figura 2.17, el modelo conceptual permite observar y comprender cuáles serán los resultados que se obtendrán con el datamart, las variables que se utilizarán para analizarlos y la relación que existe entre ellos.

### 2.3.3 Análisis de los OLTP

En la segunda fase, se analizarán las fuentes OLTP para determinar cómo se calcularan los indicadores y establecer las respectivas correspondencias entre el modelo conceptual creado (Figura 2.17) y las fuentes de datos establecidas. También se definirá los campos que incluirán cada perspectiva y se ampliará el modelo conceptual.

### **2.3.3.1 Conformar Indicadores**

En este paso se determinará como serán calculados los indicadores, tomando en cuenta:

- Hecho y/o hechos que lo componen, con su respectiva fórmula de cálculo.
- Función de sumalización que se utilizará para su agregación. Por ejemplo: SUM, AVG, COUNT, etc.

### **2.3.3.2 Establecer Correspondencias**

Se examina los OLTP disponibles que contengan la información necesitada, como también sus características, para poder identificar las relaciones entre el modelo conceptual y las fuentes de datos disponibles.

### **2.3.3.3 Nivel de Granularidad**

En este paso se debe indicar el nivel de detalle en el análisis y seleccionar los campos que contendrá cada perspectiva.

Se expone al usuario los análisis disponibles para cada perspectiva, explicando su significado, valores posibles y características, estos decidirán cuál o cuáles son los más relevantes para examinar los indicadores y cuáles no.

Con respecto a la perspectiva Tiempo, es muy importante definir los periodos mediante el cual se agruparán o sumarán los datos. Estos pueden ser: día de la semana, quincena, mes, trimestres, semestre, año, etc.

### **2.3.3.4 Modelo Conceptual Ampliado**

En este paso, se procede a graficar los resultados obtenidos en los pasos anteriores, se amplía el modelo conceptual, colocando bajo cada perspectiva los campos definidos por el usuario y bajo cada indicador su respectiva fórmula de cálculo.

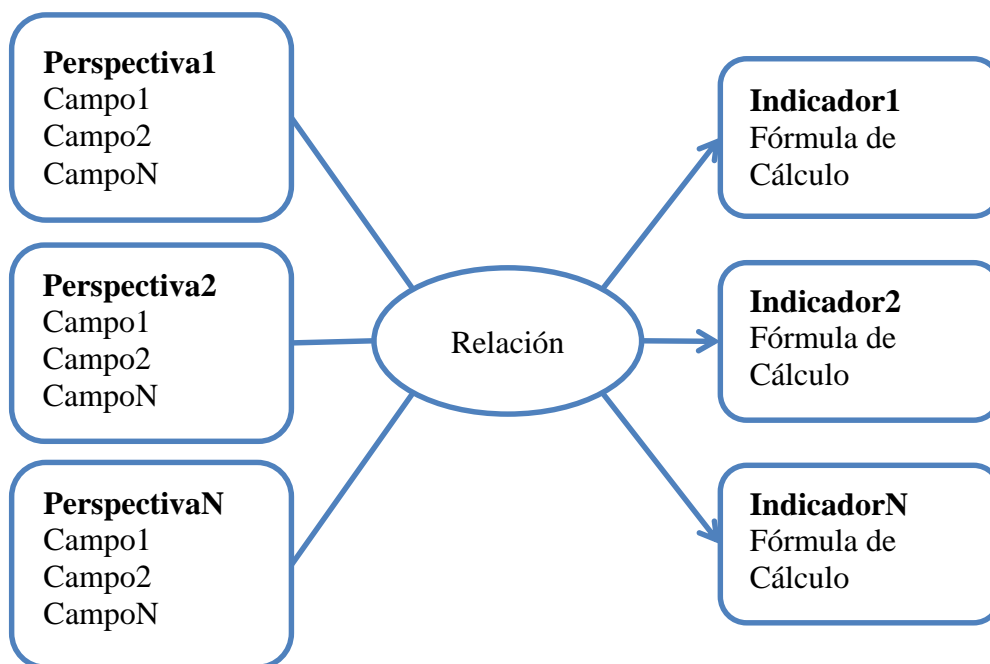


Figura 2.18: Modelo Conceptual Ampliado

### 2.3.4 Modelo lógico del Datawarehouse

En esta fase se confecciona el modelo lógico de la estructura del datawarehouse, a partir del modelo conceptual base, para ello, primero se define el tipo de modelo que se utilizará y luego se diseña las tablas de dimensiones y métricas; finalmente se realiza las asociaciones entre las tablas.

#### 2.3.4.1 Tipo de Modelo Lógico del Datawarehouse

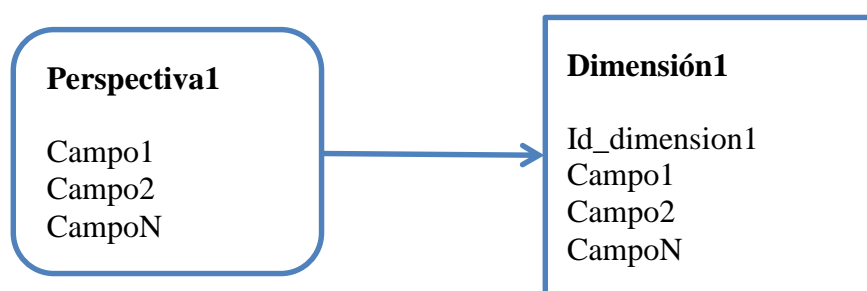
Se debe establecer el tipo de esquema que se utilizará dependiendo de los requerimientos y necesidades del usuario. Es muy importante el tipo de esquema que se empleará (Capítulo 2.1.6) ya que esta decisión afectará directamente a la elaboración del modelo lógico.

#### 2.3.4.2 Tablas de dimensiones

En este paso se debe diseñar las tablas de dimensiones que formaran parte del datawarehouse.

Para cualquiera de los tres tipos de esquemas definidos, cada perspectiva del modelo conceptual será una tabla de dimensión. Debe tomarse en cuenta cada perspectiva con sus campos relacionados y realizar el siguiente proceso:

- Elegir un nombre que identifique la tabla de dimensión.
- Añadir un campo que represente su clave principal.
- Definir los nombres de los campos con un modelo estandarizado.



*Figura 2.19: Tabla de Dimensión.*

### 2.3.4.3 Tablas de Hechos

En este paso se definen las tablas que contendrán los hechos con los que se construirán los indicadores del modelo.

Para los esquemas en estrella y copo de nieve, se debe considerar lo siguiente:

- Asignar un nombre a la tabla de hechos que represente la información analizada, área de investigación, negocio enfocado, etc.
- Definir su clave primaria, que se compone de la combinación de las claves primarias de cada tabla de las dimensiones relacionadas.
- Crear tantos campos de hechos como indicadores se hayan definido en el modelo conceptual.



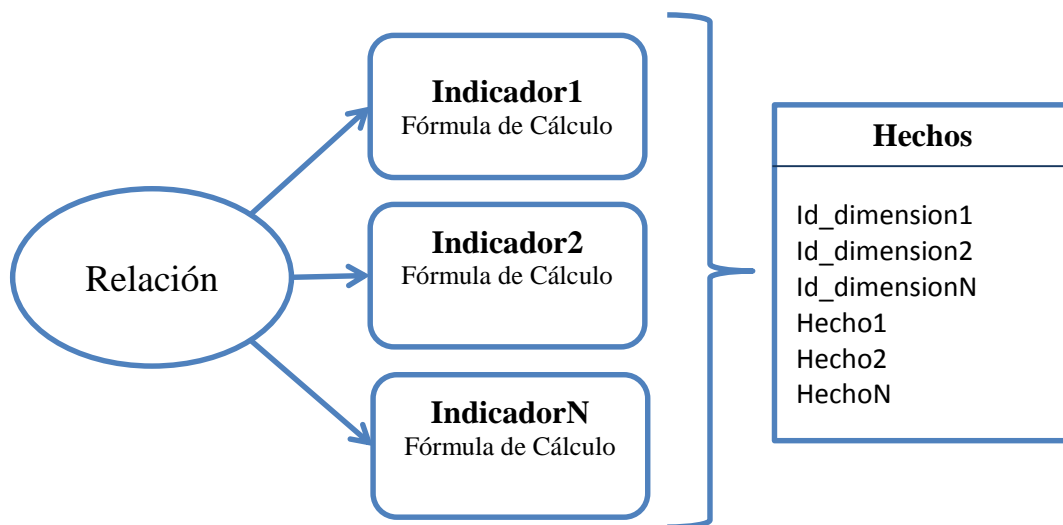


Figura 2.20: Tabla de Hechos.

Para los esquemas constelación se debe considerar lo siguiente:

- Las tablas de hechos se deben elaborar teniendo en cuenta el análisis de las preguntas realizadas por el usuario y sus respectivos indicadores y perspectivas.
- Cada tabla de hechos debe tener un nombre que la identifique, contener todos sus hechos y su clave debe estar formada por la combinación de las claves primarias de las tablas de dimensiones relacionadas.

Al elaborar las tablas de hechos, se debe tener en cuenta:

- *Caso 1: Si en dos o más preguntas de negocio existen los mismos indicadores pero con diferentes perspectivas de análisis, existirán tantas tablas de hechos como preguntas cumplan esta condición.*
- *Caso 2: Si en dos o más preguntas de negocio figuran diferentes indicadores con diferentes perspectivas de análisis, existirán tantas tablas de hechos como preguntas cumplan esta condición.*
- *Caso 3: Si el conjunto de preguntas de negocio cumplen con las condiciones de los dos puntos anteriores se deberán unificar aquellos interrogantes que posean diferentes indicadores pero iguales perspectivas de análisis, para luego reanudar el estudio de las preguntas. (Bernabeu, 2010).*

### 2.3.4.4 Uniones

Se realizan las uniones correspondientes entre las tablas de dimensiones y sus tablas de hechos.

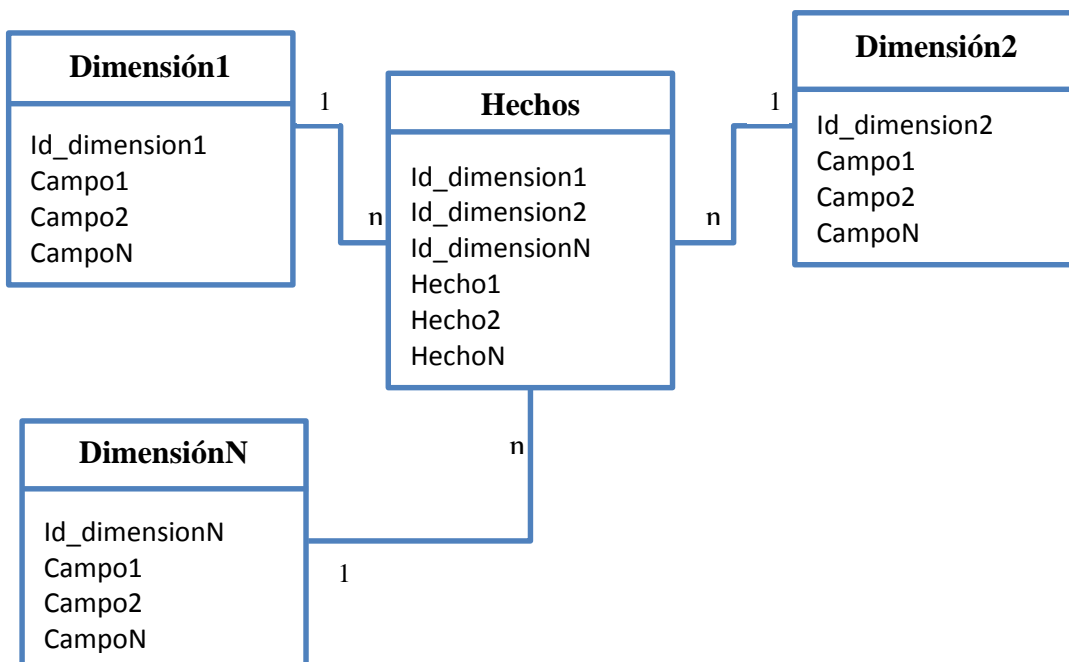


Figura 2.21: Uniones

### 2.3.5 Integración de Datos

Una vez construido el modelo lógico, se debe proceder a poblarlo con datos, utilizando procesos ETL, luego se definirán las reglas y políticas para su respectiva actualización, así como los procesos que la llevarán a cabo.

#### 2.3.5.1 Carga Inicial

En esta fase se realiza la carga inicial al datawarehouse o datamart, es decir, alimentar el modelo lógico de datos que ha sido construido, para lo cual se debe llevar a cabo una serie de tareas básicas, tales como limpieza de datos, calidad de datos y procesos ETL.

Estas tareas puede contener una lógica compleja, actualmente existen varios software que se pueden utilizar con tal fin. Se debe evitar que el datawarehouse o datamart sea cargado con valores faltantes o anómalos.

El primer paso es cargar los datos de las dimensiones y posteriormente los de las tablas de hechos, teniendo en cuenta la correcta correspondencia entre cada elemento. En el caso de un esquema tipo copo de nieve, cada vez que existan jerarquías de dimensiones se comenzarán cargando las tablas de dimensiones del nivel más general al más detallado.

Es importante tener presente, que al cargar los datos en las tablas de hechos pueden utilizarse pre agregaciones, ya sea al nivel de granularidad de la misma o a otros niveles diferentes.

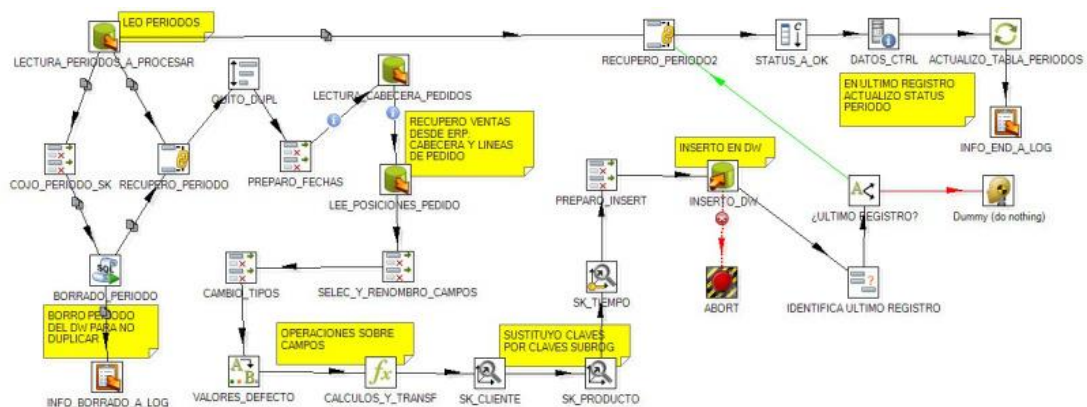


Figura 2.22: Ejemplo de flujo proceso ETL.

Fuente: (Espinosa, 2010)

### 2.3.5.2 Actualización

Cuando se haya cargado en su totalidad el datawarehouse o datamart, se deben establecer las políticas y estrategias de actualización, dependiendo del diseño de carga.

Se debe tener en cuenta las siguientes acciones:

- Especificar las tareas de limpieza de datos, calidad de datos y procesos ETL para actualizar los datos.
- Especificar de forma general y detallada las acciones que deberán realizar cada software o herramienta a utilizar.

## 2.4 HERRAMIENTAS DE DESARROLLO

Las herramientas a utilizar en el desarrollo del proyecto se muestran en la siguiente tabla.

*Tabla 2*

*Herramientas para el desarrollo del proyecto.*

<b>SOFTWARE</b>	<b>CARACTERÍSTICA</b>
PENTAHO BI SERVER 5.4	Servidor OLAP.
PENTAHO REPORT DESIGNER	Reporteador.
PENTAHO DATA INTEGRATION 5.4	Integración de Datos (ETL).
BASE DE DATOS MYSQL 5.5	Repositorio.
POWER DESIGNER 16.1	Diseño de esquemas.

### 2.4.1 Plataforma Pentaho Open Source Business Intelligence

La plataforma Open Source Pentaho envuelve amplias necesidades de análisis de datos e informes empresariales, la misma tiene un ambiente de desarrollo e implementación basada en Java, esto hace que Pentaho sea una solución flexible que cubre una amplia gama de necesidades empresariales tanto las básicas como las sofisticadas y a su vez específicas del negocio.

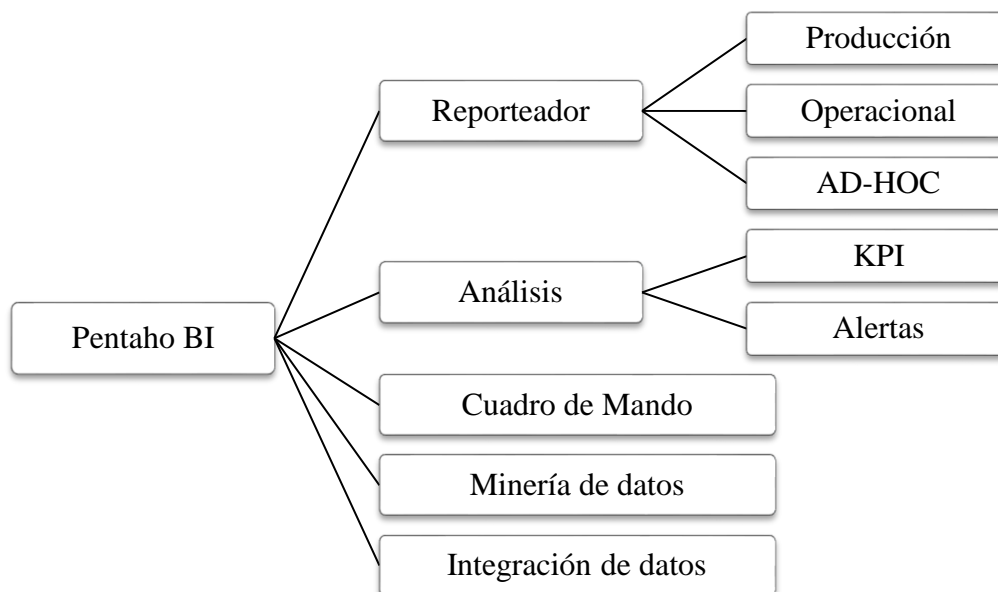


Figura 2.23: Módulos de la Plataforma Pentaho.

## 2.4.2 Pentaho Reporting

La solución proporcionada por la plataforma Pentaho CE se encuentra integrada en su suite con un desarrollo de informes llamado Pentaho Reporting el cual es un potente generador de informes que permite la distribución de los resultados del análisis en múltiples formatos.

### 2.4.2.1 Pentaho Report Designer

Pentaho Report Designer se encuentra estructurada de tal forma que los desarrolladores puedan acceder a sus prestaciones de forma inmediata, el mismo incluye un editor de consultas para facilitar la preparación de los datos que serán utilizados en un informe.

### 2.4.2.2 Pentaho Report Designer Wizard

El wizard es una herramienta de diseño de informes que permite y facilita el trabajo a los usuarios para obtener resultados de forma inmediata, la misma está diseñada para los usuarios con poco conocimiento técnico en inteligencia de negocios.

### 2.4.2.3 AD-HOC Reporting

Es una herramienta similar a Pentaho Report Designer Wizard con la diferencia que esta es vía web.



Figura 2.24: Herramienta de reportes.

### 2.4.3 Pentaho Analysis

Pentaho Analysis permite operar con efectividad el diseño del cubo de información, para entender y descubrir la información necesaria para tomar buenas decisiones y ganar perspicacia en el uso de la herramienta.

Las características generales de esta herramienta son:

- Navegar y explorar.
- Vista dimensional de los datos.
- Interactuar con alto rendimiento mediante tecnologías optimizadas para la rápida respuesta interactiva.

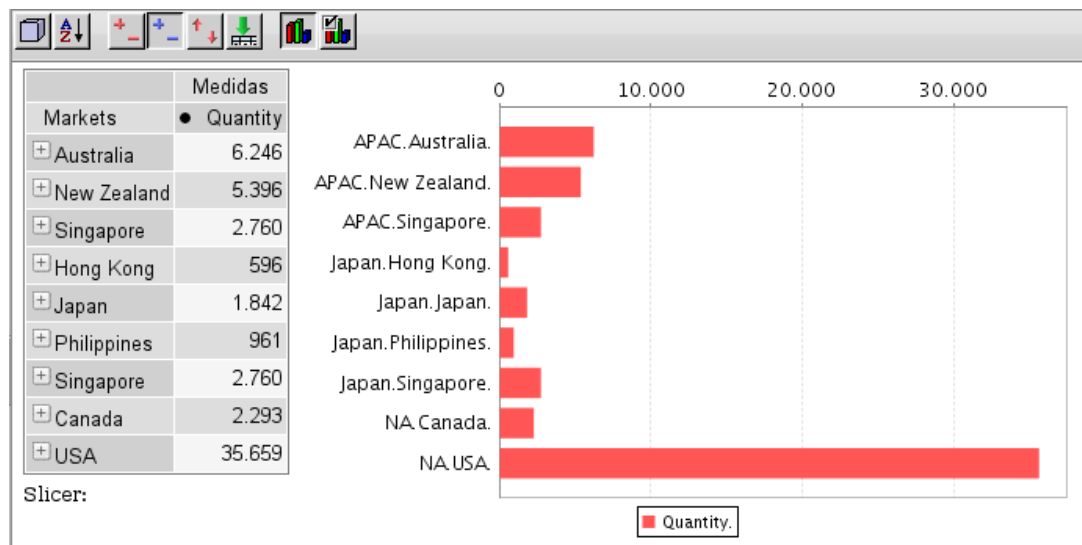


Figura 2.25: Vista de los diferentes paneles de análisis.

Fuente:(Open BI, 2015)

#### 2.4.4 Community Dashboard Editor

Los dashboard o cuadros de mando proveen una inmediata perspectiva individual, departamental o empresarial y así facilita a los usuarios obtener información crítica del negocio para entender y mejorar el rendimiento organizacional.

Community dashboard editor es una herramienta muy potente que cuenta con las siguientes características:

- Prestablecer alertas basadas en reglas de negocios.
- Seguimiento de excepciones.
- Detalles subyacentes para la realización de investigaciones, con reportes de soporte.
- Determinación e identificación de métricas clave para el negocio más conocidos como KPI.

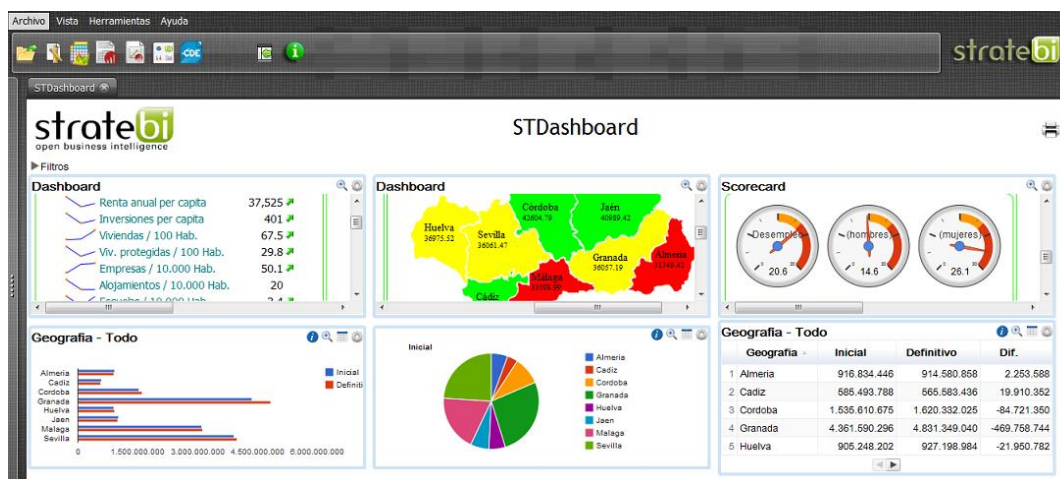


Figura 2.26: Visualización de vistas de dashboard.

Fuente:(StrateBI, 2011)

## 2.4.5 Pentaho Data Integration

Pentaho Data Integration o más conocido como Pentaho ETL se encarga de alimentar los datos en un datawarehouse o un datamart desde las fuentes que provienen de ubicaciones diversas, estas fuentes pueden ser todos los tipos de sistemas operacionales que la empresa posee, generalmente las fuentes no son homogéneas entre sí ya que son muy diversas, por lo que es necesario realizar adaptaciones pertinentes entre las distintas fuentes para obtener una buena calidad de datos.

Los ETL leen, procesan y limpian los datos fuentes para lograr llegar al usuario final de forma ordenada y optima, provee consistencia que es uno de los principales desafíos para las empresas.

El principal uso de las herramientas ETL permiten evitar grandes cargas de trabajo manual y automatizar este proceso.



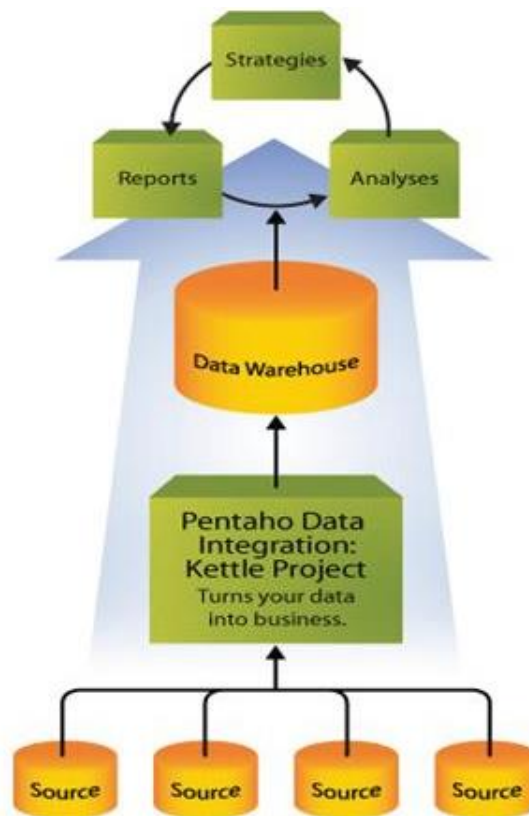


Figura 2.27: Esquema Pentaho Data Integration.

Fuente: (SIVSA, 2015)

#### 2.4.6 Data Mining

La plataforma de Pentaho ofrece diferentes tipos de soluciones para el desarrollo de un proyecto de inteligencia de negocios, en este caso se hace referencia al paquete para desarrollo de proyectos de data mining que consisten en detectar patrones en la información de los distintos conjuntos de datos.

El data mining se base en patrones y tendencias que se pueden aplicar a situaciones empresariales como por ejemplo predecir ventas, dirigir correos a clientes específicos, determinar productos que se pueden vender juntos, buscar secuencias en el orden en que los clientes agregan productos a un carro de compras, etc.

El generar un modelo de minería de datos forma parte de un proceso que envuelve desde la formulación de preguntas acerca de los datos hasta la creación de un modelo para resolver dichas inquietudes hasta la implementación del modelo.

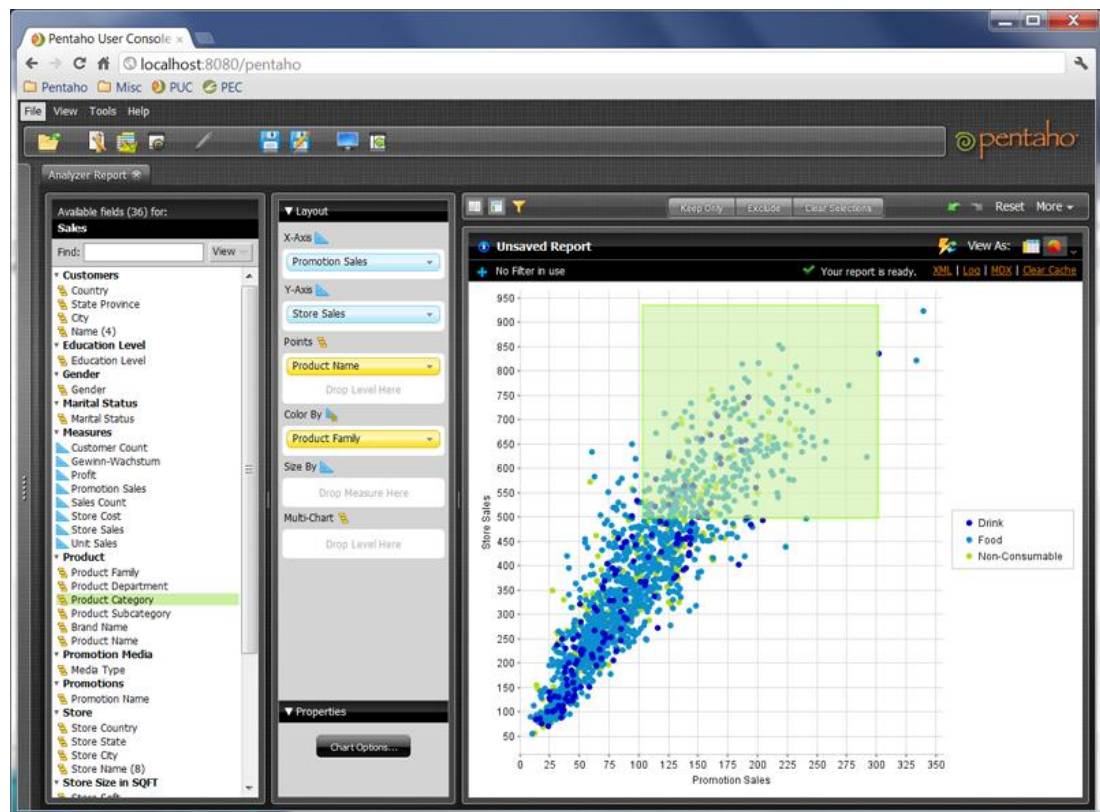


Figura 2.28: Módulo de Data Mining.

Fuente: (Data Runs the World, 2012)

## **CAPÍTULO 3.**

### **ANÁLISIS Y DISEÑO DEL PROYECTO**

#### **3.1 DEFINICIÓN DEL PROYECTO**

Debido a las características de Tejidos Anahí, una PYME dedicada a la producción en el área Textil, se ha determinado la necesidad, conjuntamente con la gerencia, de desarrollar una solución BI para los procesos estratégicos de Compra y Ventas dentro de la organización, con la finalidad de diseñar una herramienta de control y ayuda en la toma de decisiones de estos procesos, de una forma más eficaz, por parte de la compañía.

El desarrollo del proyecto permitirá dar mayor independencia a la manipulación de la información útil, permitiendo obtenerla de manera más fácil y rápida, ya que con las deficiencias actuales, se desperdician recursos en la obtención de información, hasta ahora extraída y procesada de forma manual, retrasando así los procesos de análisis de la gerencia.

##### **3.1.1 Identificación de Roles**

- Patrocinador del proyecto: Tejidos Anahí
- Administrador de Bases de Datos: Ing. Patricio Garzón
- Personal Involucrado en el Negocio: Patricio Dávila, Ing. Paul Dávila.
- Líder, Analista, Modelador, Diseñador, Desarrollador del Proyecto: Alex Dávila, Gonzalo Villares.

##### **3.1.2 Metodología de Implementación**

Se realizará la construcción de una solución Business Intelligence, para los procesos de Compra y Ventas en la compañía Tejidos Anahí.

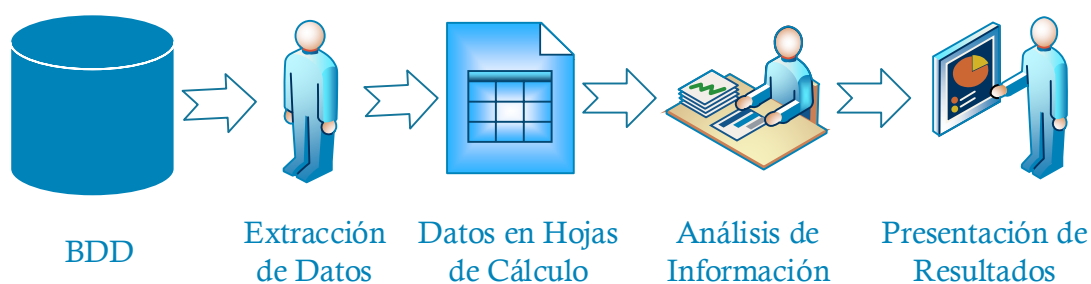
Para el análisis, diseño y desarrollo de la solución se utilizará la metodología Hefesto V2.0, metodología ágil propia para el desarrollo de soluciones BI y compatible con las herramientas Pentaho a utilizar.

La implementación de la solución se realizará con herramientas de software libre, siendo Pentaho Community V5.4 la herramienta base del desarrollo. Pentaho Report Designer permitirá la integración y el diseño de reportes en la solución, Pentaho Data Integration nuestra herramienta ETL y Community Dashboard Editor la herramienta de diseño y desarrollo de los cuadros de mando previstos.

Se implementaran dos cubos: Compra y Ventas, para obtener la solución requerida por el usuario.

### 3.1.3 Situación Actual

Actualmente el proceso de análisis de información de los procesos de Compra y Ventas se realiza de forma manual por parte de la gerencia de la compañía, de forma no técnica e implementada en plantillas y hojas de cálculo.



*Figura 3.1: Situación Actual*

### 3.1.4 Fuentes de Datos

Las fuentes de datos serán recopiladas desde la base de datos estructural del sistema contable Tiny con el que cuenta la compañía, que se encuentra alojado en el servidor. Las tablas se encuentran con datos estructurados y están alojados en el motor de base de datos MySQL.

## 3.2 ANÁLISIS DE REQUERIMIENTOS

### 3.2.1 Identificar Preguntas

Para identificar las preguntas relevantes del proyecto, se realizó una serie de entrevistas con el personal gerencial y los involucrados en el departamento financiero de Tejidos Anahí, en primer lugar, para conocer el modelo de negocio, los flujos de trabajo y los procesos involucrados en las actividades diarias de la organización.

Se analizó de forma general los procesos importantes en la vida diaria de la organización y se determinó con los involucrados, que los procesos más importantes en el flujo comercial son los de Compra y Ventas.

A continuación se procedió a indagar que es lo que se necesitaba conocer en estos procesos y las variables más importantes para tomar decisiones en torno a ellos. Entre las respuestas obtenidas se identifica que los valores más importantes que se desean conocer son de los flujos de compras en determinados lapsos de tiempo, así como las cantidades de ventas a través del tiempo.

Las preguntas de negocio obtenidas fueron las siguientes:

#### **Proceso de Compras**

- Se desea conocer el monto de compras realizadas a determinado proveedor en un periodo de tiempo. “Monto total de compras a cada proveedor en un periodo de tiempo”.
- Se desea conocer el monto total de compras realizadas en un periodo de tiempo. “Monto total de compras en un periodo de tiempo”.
- Se desea conocer el monto total de compras por tipo en un periodo de tiempo. “Monto total de compras por tipo en un periodo de tiempo”.
- Se desea conocer el monto total de IVA, Retención a la Fuente y Retención de IVA en un periodo de tiempo. “Monto total de IVA, Retención a la Fuente y Retención de IVA en un periodo de tiempo”.

**Proceso de Ventas:**

- Se necesita conocer cuántas unidades de cada producto fueron vendidas a los clientes en un periodo determinado. “Unidades vendidas de cada producto a cada cliente en un periodo de tiempo”.
- Se desea conocer el monto total de ventas de productos en un periodo determinado. “Total de ventas de cada producto en un periodo de tiempo”.
- Se desea conocer el monto total de ventas a cada cliente en un periodo determinado. “Monto total de ventas a cada cliente en un periodo de tiempo”.
- Se desea conocer el monto total de IVA en un periodo determinado. “Monto total de IVA en un periodo de tiempo”.
- Se desea conocer un monto total de Compras y el monto total de Ventas en un periodo de tiempo. “Monto total de Compras y Monto Total de Ventas en un periodo de tiempo”.

La dimensión Tiempo se ha convertido en un elemento fundamental y se le dará vital importancia en la construcción del datawarehouse.

Las necesidades de información recuperadas se encuentran alineadas con los objetivos y estrategias de la empresa, lo que permitirá hacer un análisis sobre estos elementos claves y serán herramienta fundamental en la toma de decisiones de estos procesos.

**3.2.2 Identificar Indicadores y Perspectivas****Proceso de Compras:**

Monto total de compras a cada proveedor en un periodo de tiempo.

INDICADOR

PERSPECTIVAS

Monto total de compras en un periodo de tiempo.

INDICADOR

PERSPECTIVAS

Monto total de compras por tipo de compra en un periodo de tiempo.

INDICADOR

PERSPECTIVAS

Monto total de IVA en un tiempo determinado.

INDICADOR                      PERSPECTIVAS

Monto total de Retención Fuente en un tiempo determinado.

INDICADOR                      PERSPECTIVAS

Monto total de Retención IVA en un tiempo determinado.

INDICADOR                      PERSPECTIVAS

*Tabla 3*

*Indicadores y Perspectivas del Proceso de Compras.*

INDICADORES	PERSPECTIVAS
<ul style="list-style-type: none"> <li>• MONTO TOTAL DE COMPRAS</li> <li>• MONTO TOTAL IVA</li> <li>• MONTO TOTAL RETENCIÓN FUENTE</li> <li>• MONTO TOTAL RETENCIÓN IVA</li> </ul>	<ul style="list-style-type: none"> <li>• Proveedor</li> <li>• Tiempo</li> <li>• Tipo</li> </ul>

**Proceso de Ventas:**

Unidades vendidas de cada producto a cada cliente en un periodo de tiempo.

INDICADOR                      PERSPECTIVAS

Total de ventas de cada producto en un periodo de tiempo.

INDICADOR                      PERSPECTIVAS

Monto total de ventas a cada cliente en un periodo de tiempo.

INDICADOR                      PERSPECTIVAS

Monto total de IVA en un periodo de tiempo.

INDICADOR                      PERSPECTIVAS

*Tabla 4*

*Indicadores y Perspectivas del Proceso de Ventas.*

INDICADORES	PERSPECTIVAS
<ul style="list-style-type: none"> <li>• UNIDADES VENDIDAS</li> <li>• TOTAL VENTA</li> <li>• MONTO TOTAL DE VENTAS</li> <li>• MONTO TOTAL DE IVA</li> </ul>	<ul style="list-style-type: none"> <li>• Producto</li> <li>• Cliente</li> <li>• Tiempo</li> </ul>

### 3.2.3 Modelo Conceptual

#### Proceso de Compras:

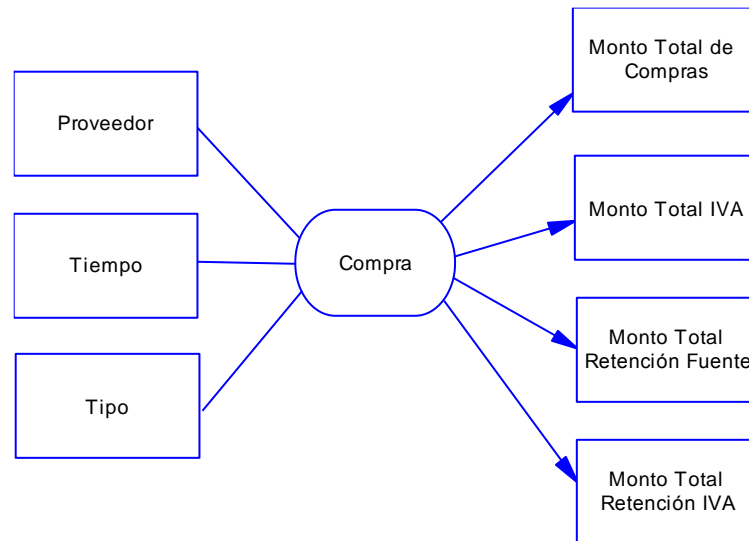


Figura 3.2: Modelo Conceptual Compra

#### Proceso de Ventas:

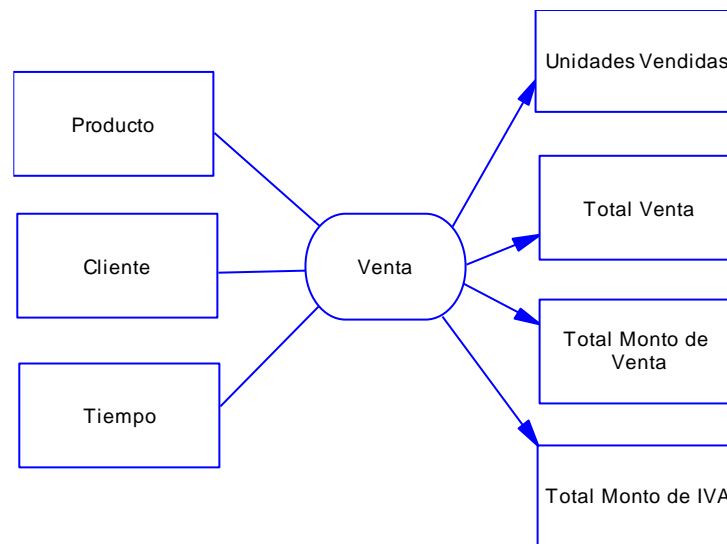


Figura 3.3: Modelo Conceptual Venta



### 3.3 ANÁLISIS DE LOS OLTP

#### 3.3.1 Construcción de los Indicadores

##### **Proceso de Compras**

##### Monto Total de Compras:

- Hechos: Total de Compras
- Función de sumarización: SUM.

Aclaración: el indicador Monto Total de Compras representa la sumatoria de las unidades de cada producto que han sido vendidas.

##### Monto Total IVA:

- Hechos: Total de IVA
- Función de sumarización: SUM.

Aclaración: el indicador Monto Total IVA representa la sumatoria del IVA en las facturas de compra.

##### Monto Total Retención Fuente:

- Hechos: Total de Retención Fuente
- Función de sumarización: SUM.

Aclaración: el indicador Monto Total Retención Fuente representa la sumatoria de la Retención a la Fuente en las facturas de compra.

##### Monto Total Retención IVA:

- Hechos: Total de Retención IVA
- Función de sumarización: SUM.

Aclaración: el indicador Monto Total Retención IVA representa la sumatoria de la Retención al IVA en las facturas de compra.

##### **Proceso de Ventas**

##### Unidades Vendidas:

- Hechos: Unidades Vendidas
- Función de sumarización: SUM.

Aclaración: el indicador Unidades Vendidas representa la sumatoria del Total de las facturas de compra a un determinado proveedor.

Total de Venta:

- Hechos: Total de Venta
- Función de sumarización: SUM.

Aclaración: el indicador Total de Venta representa la sumatoria del Total en el detalle de facturas de un determinado producto.

Monto Total de Ventas:

- Hechos: Total de Ventas
- Función de sumarización: SUM.

Aclaración: el indicador Monto Total de Ventas representa la sumatoria del Total de las facturas de venta a un determinado cliente.

Monto Total de IVA:

- Hechos: Total de IVA
- Función de sumarización: SUM.

Aclaración: el indicador Monto Total de IVA representa la sumatoria del IVA de las facturas de venta.

### **3.3.2 Establecer Correspondencia**

Los OLTP disponibles en la empresa analizada, se encuentran en una base de datos MySQL con datos estructurados, a continuación se presenta el diagrama Entidad-Relación de los procesos de compra y venta en la Figura 3.4.

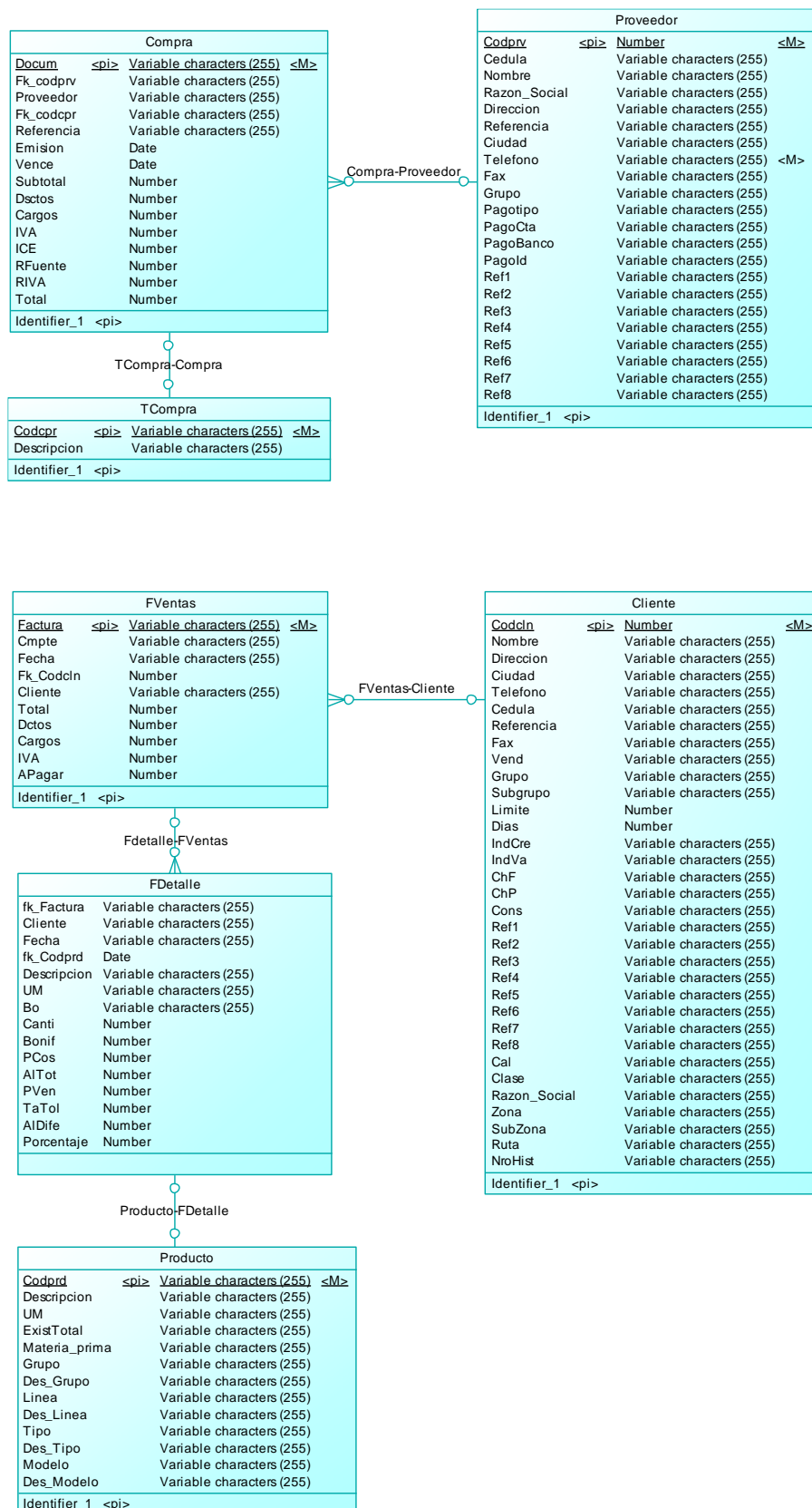


Figura 3.4: Diagrama Entidad-Relación

A continuación se determinará la correspondencia entre el modelo Entidad-Relación de las fuentes de Datos y el Modelo Lógico. Se procede a dividirlo en los dos procesos utilizando para el proceso de compra y ventas las tablas indicadas en el siguiente gráfico:

Tabla 5

Tablas de cada Proceso

PROCESO	TABLAS
1 Compras	<ul style="list-style-type: none"> <li>• Compra: Tabla cabecera con la información de compras.</li> <li>• Proveedor: Tabla con la información del Proveedor.</li> <li>• TCompra: Tabla con la descripción del tipo de compra.</li> </ul>
2 Ventas	<ul style="list-style-type: none"> <li>• Fventas: Tabla cabecera de la factura de ventas.</li> <li>• FDetalle: Tabla Detalle de la factura de ventas.</li> <li>• Cliente: Tabla con la información de los clientes.</li> <li>• Producto: Tabla con la información de Productos.</li> </ul>

**Proceso de Compras**

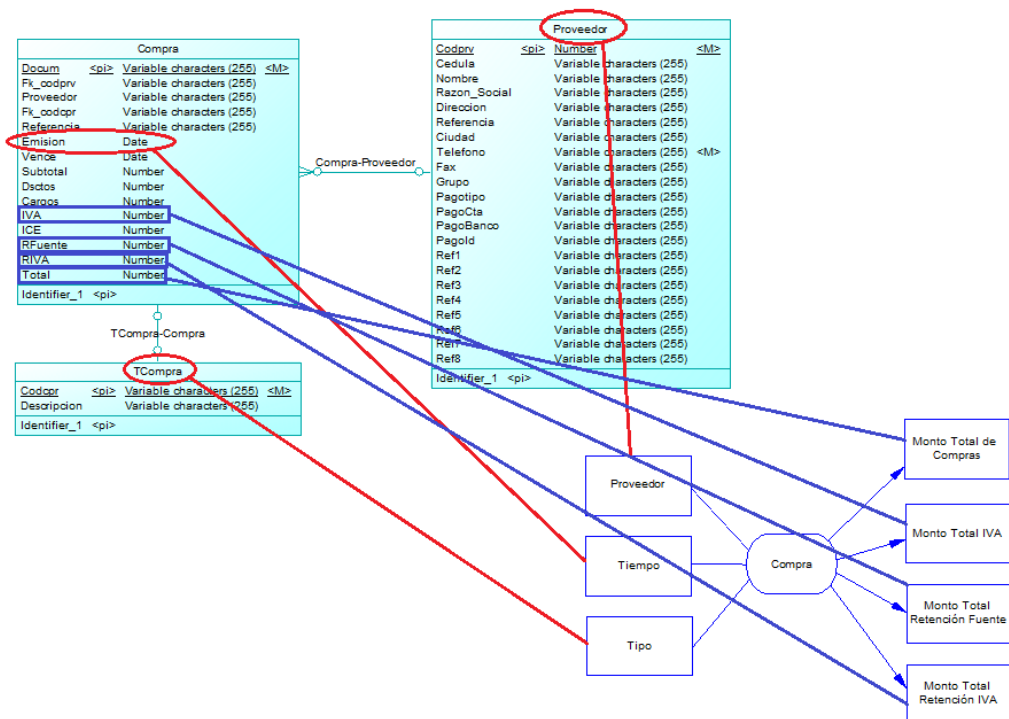


Figura 3.5: Correspondencia Proceso de Compras

## Proceso de Ventas

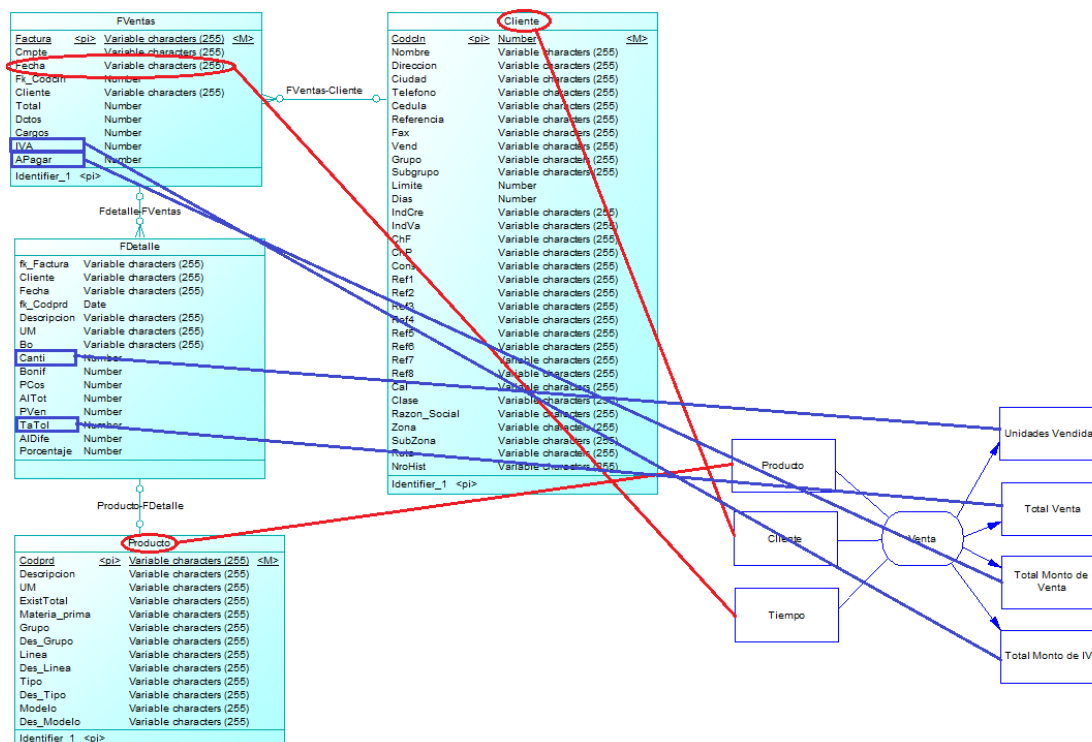


Figura 3.6: Correspondencia Proceso de Ventas

### 3.3.3 Nivel de Granulidad

Después de analizar las correspondencias entre la base de datos y el modelo y de establecer con el usuario cuales eran los campos requeridos para analizar los diferentes indicadores, se selecciona la siguiente información de cada una de las perspectivas:

#### Proceso de Compras:

##### Perspectiva Proveedor:

- “Cedula”: de la tabla “Proveedor” que hace referencia al número de cédula del proveedor.
- “Nombre”: de la tabla “Proveedor” que hace referencia al nombre del proveedor.
- “Direccion”: de la tabla “Proveedor” que hace referencia a la dirección del proveedor.

- “Ciudad”: de la tabla “Proveedor” que hace referencia al nombre de la ciudad del proveedor.
- “Telefono”: de la tabla “Proveedor” que hace referencia al número telefónico del proveedor.
- “Grupo”: de la tabla “Proveedor” que hace referencia al nombre del grupo geográfico al que pertenece el proveedor.

#### Perspectiva Tipo:

- “Descripcion”: de la tabla “TCompra” que hace referencia al tipo de Compra realizada.

#### Perspectiva Tiempo:

- “Dia”: valor numérico correspondiente al día.
- “Semana”: referido al número de semana del mes.
- “Mes”: referido al nombre del mes.
- “Anio”: valor numérico correspondiente al año.

### **Proceso de Ventas**

#### Perspectiva Cliente:

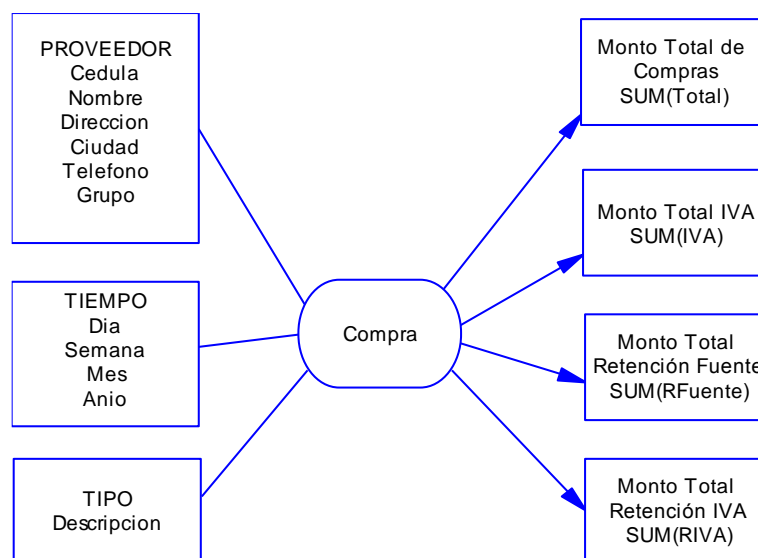
- “Nombre”: de la tabla “Cliente” que hace referencia al nombre del cliente.
- “Cedula”: de la tabla “Cliente” que hace referencia al número de cédula del cliente.
- “Ciudad”: de la tabla “Cliente” que hace referencia al nombre de la ciudad del cliente.
- “Grupo”: de la tabla “Cliente” que hace referencia al nombre del grupo al que pertenece el cliente.
- “SubZona”: de la tabla “Cliente” que hace referencia al nombre de la región a la que pertenece el cliente.

Perspectiva Producto:

- “Descripcion”: de la tabla “Producto” que hace referencia al nombre del producto.
- “Des\_Grupo”: de la tabla “Producto” que hace referencia al nombre del grupo al que pertenece el producto.
- “Des\_Linea”: de la tabla “Producto” que hace referencia a la línea a la que pertenece el producto.
- “Des\_Tipo”: de la tabla “Producto” que hace referencia al nombre al tipo de producto.
- “Des\_Modelo”: de la tabla “Producto” que hace referencia al nombre al modelo del producto.

Perspectiva Tiempo:

- “Dia”: valor numérico correspondiente al día.
- “Semana”: referido al número de semana del mes.
- “Mes”: referido al nombre del mes.
- “Anio”: valor numérico correspondiente al año.

**3.3.4 Modelo Conceptual Ampliado****Proceso de Compras**

*Figura 3.7: Modelo Conceptual Ampliado Compras*

### Proceso de Ventas

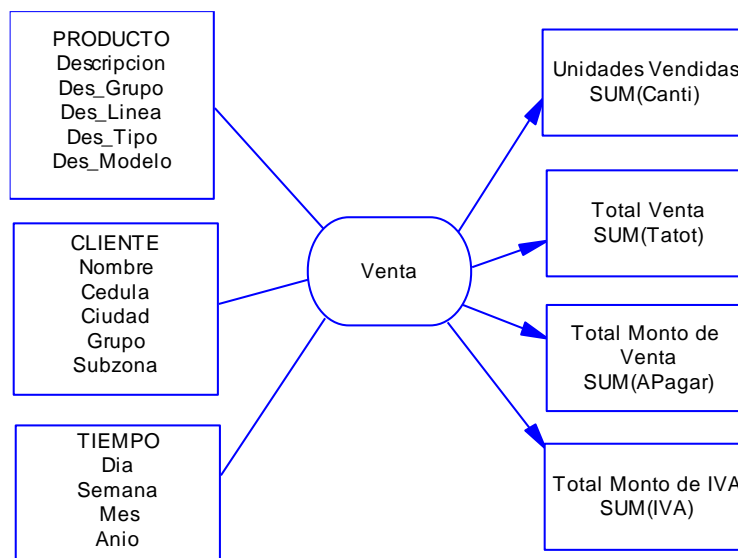


Figura 3.8: Modelo Conceptual Ampliado Ventas

## 3.4 MODELO LÓGICO DEL DATAWAREHOUSE

En base al modelo conceptual, el siguiente paso es desarrollar el modelo lógico del datawarehouse, donde se evalúa el modelo, se selecciona el más indicado, se desarrollan las tablas que lo componen y se define su correspondencia.

### 3.4.1 Estándares de Diseño del Modelo Lógico

Se define los estándares para el diseño y posterior creación de todos los componentes del datawarehouse. Los estándares son el formato que se van a utilizar en el diseño del modelo lógico del datawarehouse, estos se definen mediante un esquema que tendrá varias columnas que se especifican a continuación:

- **Componente:** se indica el nombre del componente que se va a estandarizar.
- **Prefijo:** se almacena el nombre del prefijo que va a tener el objeto.
- **Número de caracteres:** se indica el número máximo de caracteres que tendrá el componente.
- **Descripción:** se indica la descripción del nombre que se va a emplear para cada componente.
- **Ejemplo:** es un ejemplo del nombre estandarizado que tiene cada componente.



Es posible visualizar los estándares para cada:

*Tabla 6*

*Estándares de Diseño del Modelo Lógico*

<b>Componente</b>	<b>Prefijo</b>	<b>Número de Caracteres</b>	<b>Descripción</b>	<b>Ejemplo</b>
Dimensiones	dim	30	Nombre que identifique la dimensión.	dimTiempo
Métricas	met	30	Nombre que identifique la métrica.	metVenta
Secuencial dimensiones	id	30	Nombre que identifique la dimensión.	idDimTipo
Secuencial métricas	id	30	Nombre que identifique la métrica.	idMetVenta
Campos de la Tabla		30	Nombre del campo perteneciente a cada una de las tablas.	Ciudad

### 3.4.2 Tipo de Modelo Lógico del Datawarehouse

El esquema que se va a utilizar en el modelamiento y construcción del datawarehouse es un modelo estrella.

### 3.4.3 Tablas de Dimensión

#### Proceso de Compras

Las perspectivas analizadas en el proceso de compras son: Proveedor, Tipo de Compra y Tiempo.

- **Perspectiva Proveedor**

Nombre de la tabla de dimensión: dimProveedor

Nombre de la Clave Principal: idDimProveedor

Campos de la tabla:

- Cedula, proviene de la columna Cedula.
- Nombre, proviene de la columna Nombre.
- Direccion, proviene de la columna Direccion.
- Ciudad, proviene de la columna Ciudad.
- Telefono, proviene de la columna Telefono.
- Grupo, proviene de la columna Grupo.

Descripción de la tabla:

dimProveedor			
<u>idDimProveedor</u>	<pi>	<Undefined>	<M>
Cedula		Variable characters (255)	
Nombre		Variable characters (255)	
Direccion		Variable characters (255)	
Ciudad		Variable characters (255)	
Telefono		Variable characters (255)	
Grupo		Variable characters (255)	
idDimProveedor	<pi>		

Figura 3.9: Tabla Dimensión Proveedor

- **Perspectiva Tipo de Compra**

Nombre de la tabla de dimensión: dimTipo

Nombre de la Clave Principal: idDimTipo

Campos de la tabla:

- Descripción, proviene de la columna Descripción.

Descripción de la tabla:

dimTipo			
<u>idDimTipo</u>	<pi>	<Undefined>	<M>
Descripcion		Variable characters (255)	
idDimTipo	<pi>		

Figura 3.10: Tabla Dimensión Tipo Compra

- **Perspectiva Tiempo**

Nombre de la tabla de dimensión: dimTiempo

Nombre de la Clave Principal: idDimTiempo

Descripción de la tabla:

dimTiempo			
<u>idDimTiempo</u>	<pi>	<Undefined>	<M>
Anio		<Undefined>	
Mes		<Undefined>	
Dia		<Undefined>	
DiaSemana		<Undefined>	
Semana		<Undefined>	
MesTexto		<Undefined>	
idDimTiempo	<pi>		

Figura 3.11: Tabla Dimensión Tiempo

## Proceso de Ventas

Las perspectivas analizadas en el proceso de ventas son: Producto, Cliente y Tiempo.

- **Perspectiva Producto**

Nombre de la tabla de dimensión: dimProducto

Nombre de la Clave Principal: idDimProducto

Campos de la tabla:

- Producto, proviene de la columna Descripcion.
- Grupo, proviene de la columna Des\_Grupo.
- Linea, proviene de la columna Des\_Linea.
- Tipo, proviene de la columna Des\_Tipo.
- Modelo, proviene de la columna Des\_Modelo.

Descripción de la tabla:

dimProducto			
<u>idDimProducto</u>	<pi>	<Undefined>	<M>
Producto		Variable characters (255)	
Grupo		Variable characters (255)	
Linea		Variable characters (255)	
Tipo		Variable characters (255)	
Modelo		Variable characters (255)	
idDimProducto	<pi>		

Figura 3.12: Tabla Dimensión Producto

- **Perspectiva Cliente**

Nombre de la tabla de dimensión: dimCliente

Nombre de la Clave Principal: idDimCliente

Campos de la tabla:

- Nombre, proviene de la columna Nombre.
- Cedula, proviene de la columna Cedula.
- Ciudad, proviene de la columna Ciudad.
- Grupo, proviene de la columna Grupo.
- Zona, proviene de la columna SubZona.

Descripción de la tabla:

dimCliente			
<u>idDimCliente</u>	<pi>	<Undefined>	<M>
Nombre		Variable characters (255)	
Cedula		Variable characters (255)	
Ciudad		Variable characters (255)	
Grupo		Variable characters (255)	
Zona		Variable characters (255)	
idDimCliente	<pi>		

Figura 3.13: Tabla Dimensión Cliente

- **Perspectiva Tiempo**

Nombre de la tabla de dimensión: dimTiempo

Nombre de la Clave Principal: idDimTiempo

Descripción de la tabla:

dimTiempo			
<u>idDimTiempo</u>	<pi>	<Undefined>	<M>
Anio		<Undefined>	
Mes		<Undefined>	
Dia		<Undefined>	
DiaSemana		<Undefined>	
Semana		<Undefined>	
MesTexto		<Undefined>	
idDimTiempo	<pi>		

Figura 3.14: Tabla Dimensión Tiempo

### 3.4.4 Tablas de Hechos

#### Proceso de Compras

- **Tabla de Hechos Compra**

Nombre de la tabla de hechos: MetricaCompra

Nombre de la Clave Principal: idMetricaCompra

Campos de la tabla:

- IVA, proviene de la columna IVA.
- RFuente, proviene de la columna RFuente.
- RIVA, proviene de la columna RIVA.
- Total, proviene de la columna Total.

Descripción de la tabla:

metCompra			
<u>idMetCompra</u>	<pi>	<Undefined>	<M>
IVA		Number	
RetencionFuente		<Undefined>	
RetencionIVA		<Undefined>	
Total		Number	
idDimProveedor.		<Undefined>	
idDimTipo.		<Undefined>	
idDimTiempo.		<Undefined>	
idMetCompra	<pi>		

Figura 3.15: Tabla Métrica Compra

### Proceso de Ventas

- **Tabla de Hechos Venta**

Nombre de la tabla de hechos: MetricaVenta

Nombre de la Clave Principal: idMetricaVenta

Campos de la tabla:

- Cantidad, proviene de la columna Canti.
- TotalVenta, proviene de la columna Tatol.
- IVA, proviene de la columna IVA.
- Total, proviene de la columna Total.

Descripción de la tabla:

metVenta			
<u>idMetCompra</u>	<pi>	<Undefined>	<M>
Cantidad		Number	
TotalVenta		Number	
IVA		Number	
Total		Number	
idDimProducto.		<Undefined>	
idDimCliente.		<Undefined>	
idDimTiempo.		<Undefined>	
idMetVenta	<pi>		

Figura 3.16: Tabla Métrica Venta

### 3.4.5 Uniones

#### Proceso de Compras

En esta fase se define las uniones de las dimensiones de Proveedor, Tipo de Compra y Tiempo con la tabla de métricas. El modelo de las uniones se representa en el siguiente gráfico:

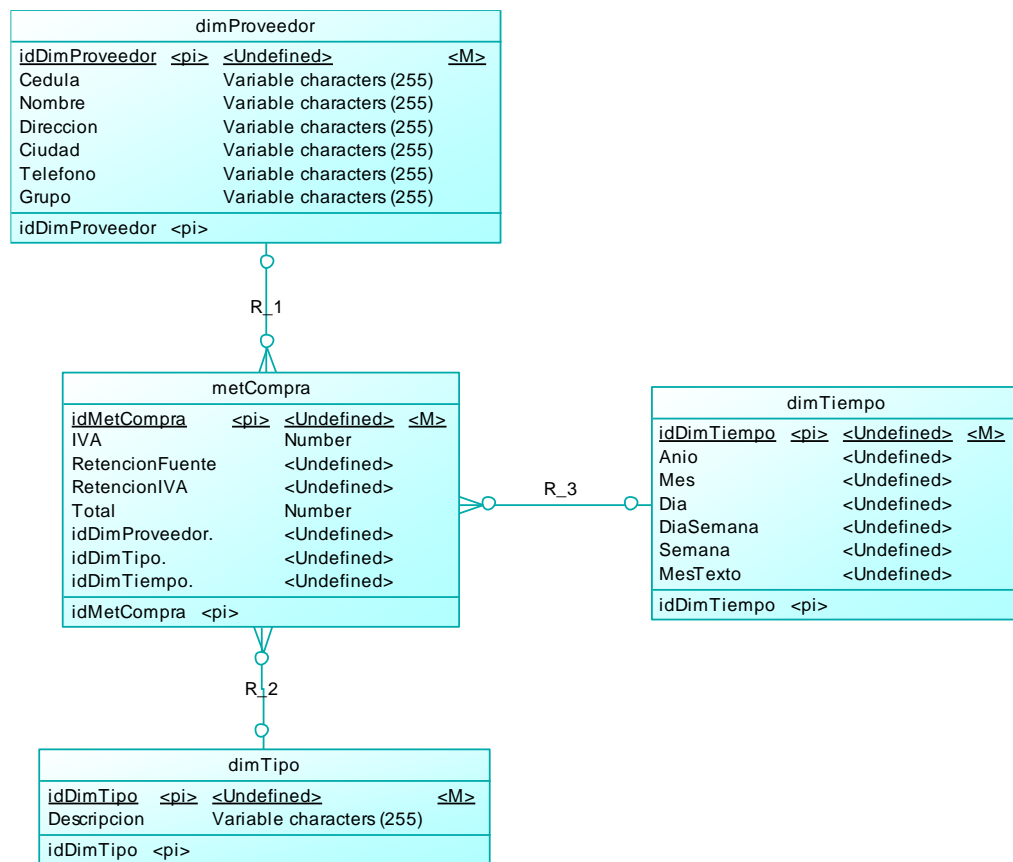


Figura 3.17: Uniones Proceso de Compras

#### Proceso de Ventas

En esta fase se define las uniones de las dimensiones de Cliente, Producto y Tiempo con la tabla de métricas. El modelo de las uniones se representa en el siguiente gráfico:

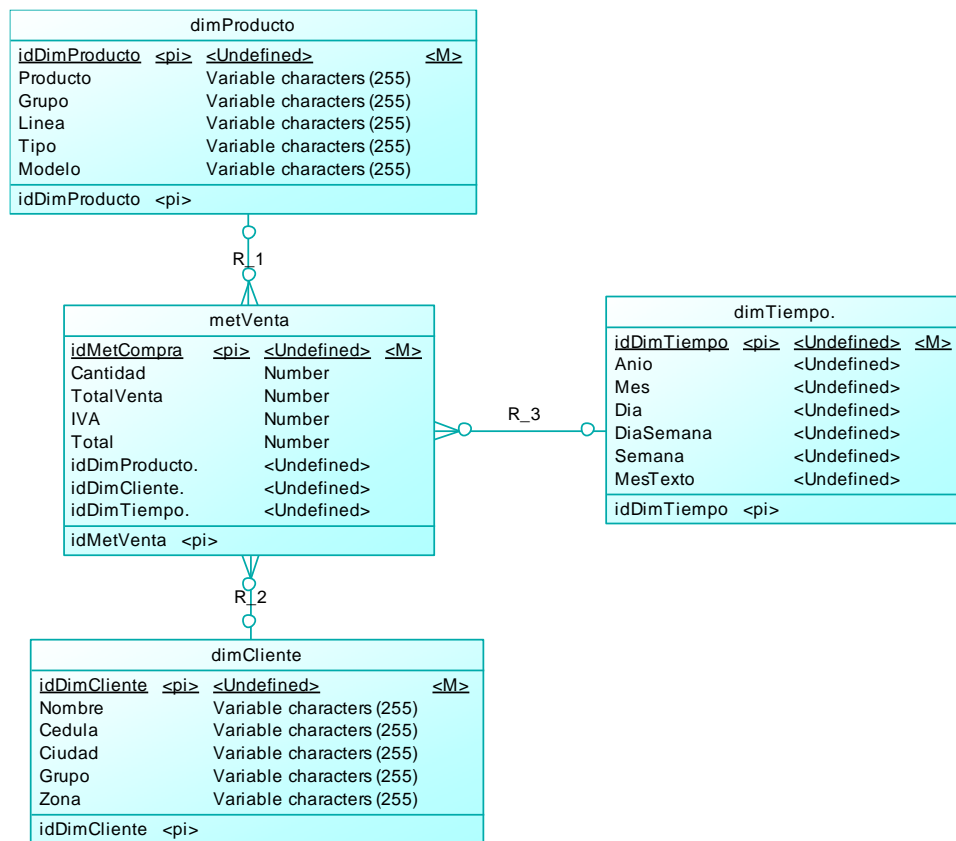


Figura 3.18: Uniones Proceso de Ventas

## CAPÍTULO 4. DESARROLLO Y CONSTRUCCIÓN

### 4.1 INSTALACIÓN Y CONFIGURACIÓN DE HERRAMIENTAS

#### 4.1.1 Instalación Power Designer 16

Para obtener una versión trial, valida por dos meses se debe ir al sitio web: <http://www.powerdesigner.de/en/trial-version-2/> y llenar el siguiente formulario que permite obtener el enlace de descarga en el correo electrónico.

The screenshot shows the 'Free SAP Sybase PowerDesigner download' form on the 'toolpark' website. The form includes fields for Title (Mr.), First Name, Last Name, E-Mail, Phone, and Company. A 'Your message' text area is also present. The right sidebar contains a 'KUNDEN' section listing companies like BLUEFORTE GmbH, BTB GmbH, BSIU, and BWinvest GmbH, a 'TESTIMONIAL' section with a quote, a 'LANGUAGES' section with options for English, Deutsch, and 简体中文, and a 'SEARCH' bar.

Figura 4.1: Formulario de suscripción SAP Sybase Power Designer.

Una vez descargado el software se procede a ejecutar el instalador.

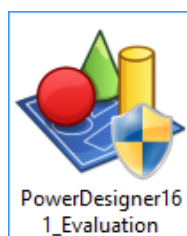


Figura 4.2: Instalador Power Designer 16.

Se inicia el instalador, se da clic en siguiente (Next), se selecciona los productos que van a instalarse y se vuelve a dar clic en siguiente.



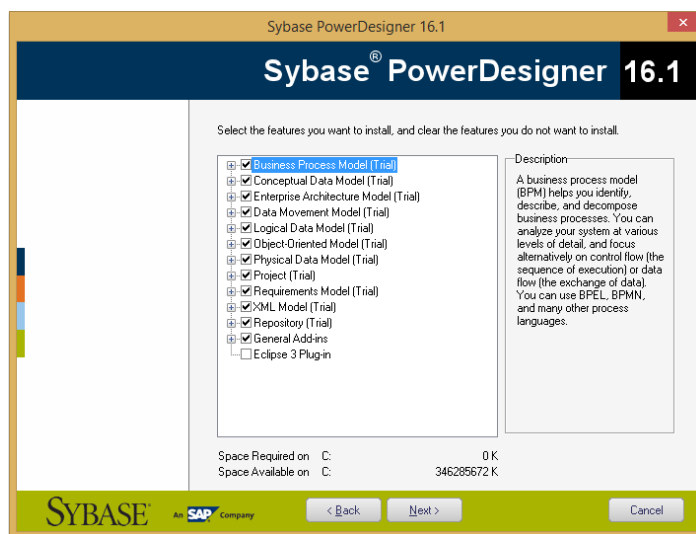


Figura 4.3: Selección de Productos Sybase Power Designer.

Al finalizar se tiene el producto instalado en el menú inicio y su correspondiente icono en el escritorio, la licencia Trial tiene una duración de 2 meses, si se ve la necesidad de insertar una licencia full, simplemente se adquiere y se inserta en la aplicación, esta automáticamente se actualiza.

#### 4.1.2 Instalación Java SE Development Kit 7

Para instalar la plataforma Java Development Kit (JDK) se debe ingresar al sitio web <http://www.oracle.com/technetwork/es/java/javase/downloads/index.html>, donde se selecciona el producto, el sistema operativo que se necesita y se descarga; a continuación se procede con la instalación.

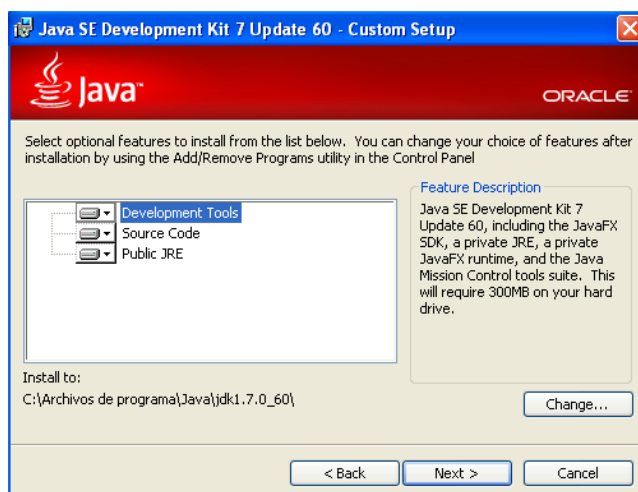


Figura 4.4: Selección de productos Java.

En el utilitario se selecciona los elementos a instalar y se da clic en siguiente. Se debe seleccionar la opción de Public JRE (Java Runtime Environment), plugin que es importante para que el software pueda ejecutarse, este puede instalarse por separado, pero por facilidad se recomienda instalarlo ahora.

Se debe escoger la carpeta para la instalación del JRE, por facilidad se selecciona la opción que viene predeterminada.

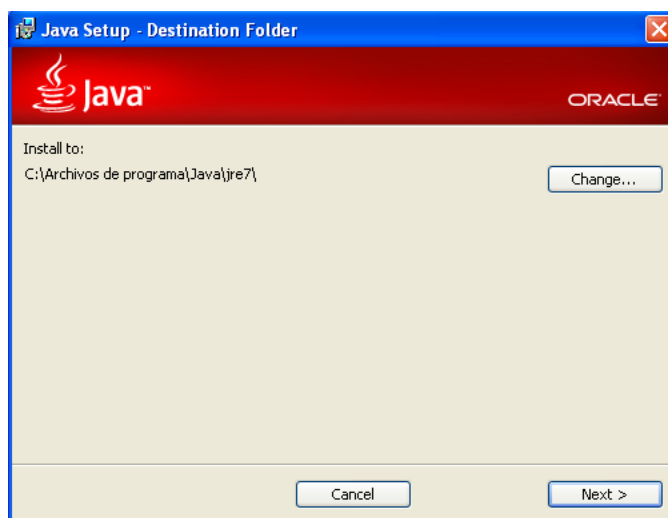


Figura 4.5: Selección carpeta de instalación JRE.

Finalmente es conveniente verificar la carpeta de instalación donde están los componentes Java.

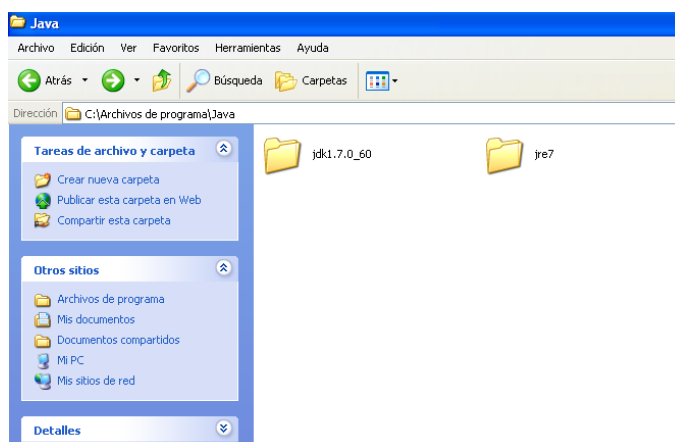


Figura 4.6: Carpeta componentes Java instalados.

### 4.1.3 Instalación MySQL 5.5

Para instalar MySQL se debe acceder al siguiente enlace: <https://dev.mysql.com/downloads/mysql/5.5.html> donde se escoge la versión del software y el sistema operativo. En este caso se selecciona la versión 5.5 y la plataforma Windows XP 32 bits.

Una vez descargado el archivo, se ejecuta el instalador. Se da clic en siguiente, se selecciona el tipo de instalación como se observa en la figura 4.7. Se escoge la opción más adecuada, en este caso la completa para obtener todas las funcionalidades del software.

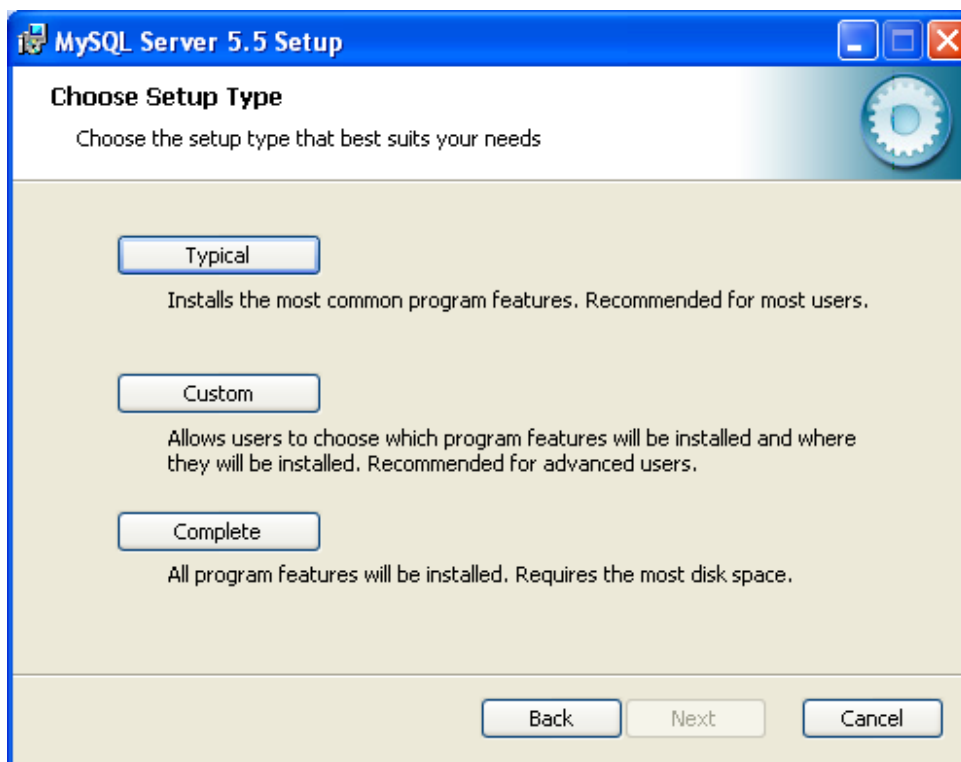


Figura 4.7: Selección del tipo de instalación MySQL.

Se da clic en siguiente hasta que la instalación finalice. Una vez terminado se procede a la configuración del producto, en este caso se selecciona la configuración estándar.

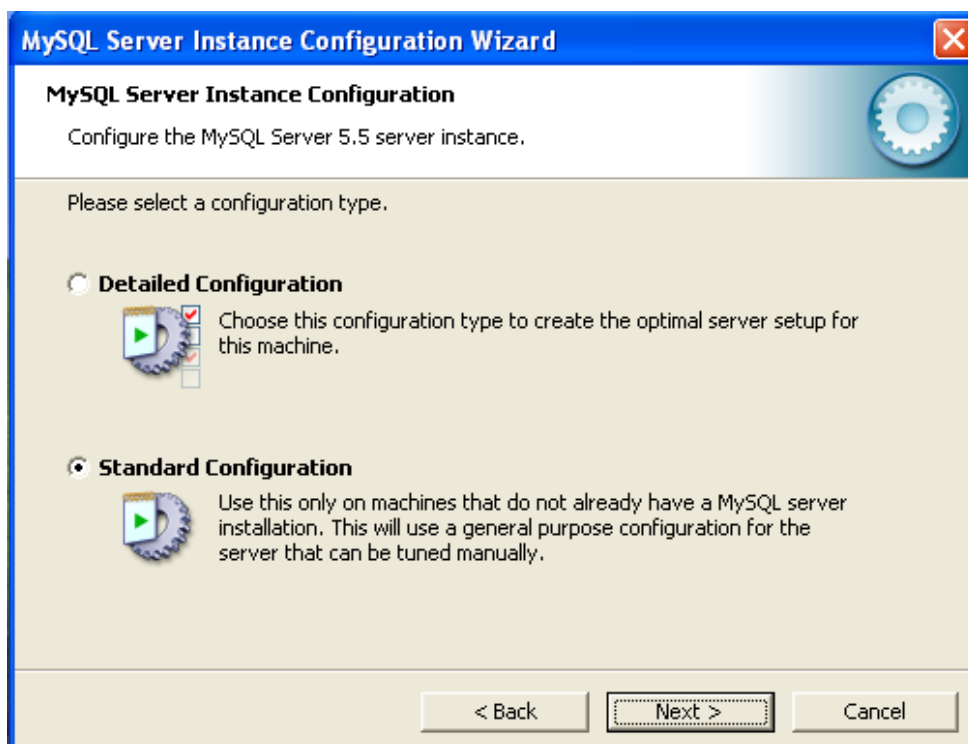


Figura 4.8: Selección del tipo de configuración del Servidor MySQL.

Seguido se selecciona la instalación como un servicio de Windows y finalmente se asigna un Usuario y Contraseña de Root para su administración.

#### 4.1.4 Instalación Administrador de Base de Datos: SQLyog.

Este software se encuentra en el siguiente sitio web, en su versión gratuita SQLyog Community Edition: <https://github.com/webyog/sqlyog-community/wiki/Downloads>, donde se descarga la versión para la plataforma necesaria.

Una vez descargado, se procede a ejecutar el instalador donde se selecciona el idioma y se da clic en siguiente. Se acepta la licencia de instalación y se escoge los productos a instalar.



Figura 4.9: Instalación de SQLyog.

Al finalizar, se procede a la configuración de la conexión con la base de datos. Primero, se selecciona una nueva conexión y se le da un nombre. Luego se procede a ingresar los datos de la conexión, usuario y contraseña de root antes ingresados en la base de datos MySQL y el puerto 3306 como predeterminado.

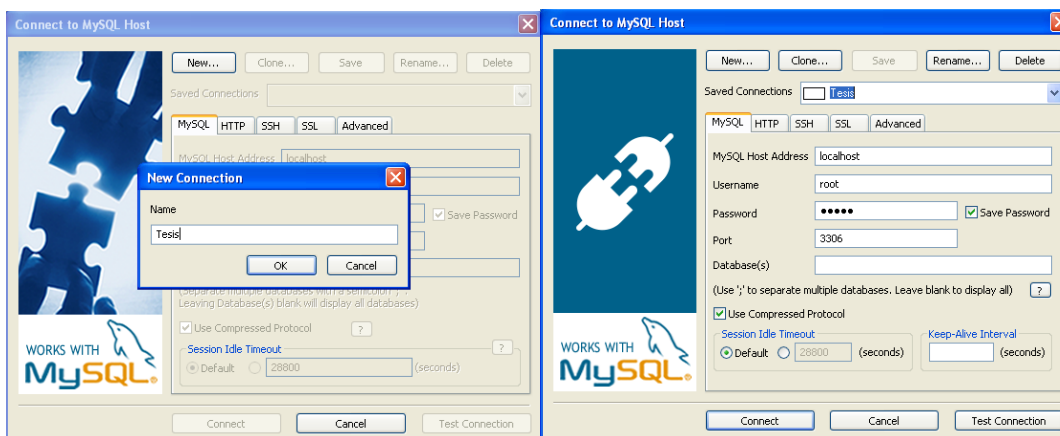


Figura 4.10: Configuración SQLyog con la base de datos.

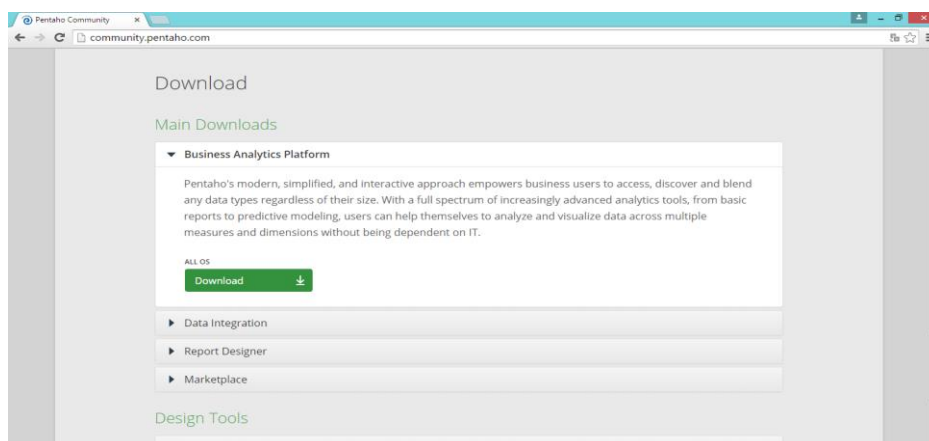
Finalmente se da un clic en Test Connection para verificar la conexión y Connect para enganchar a la base de datos.

#### 4.1.5 Instalación Servidor Pentaho Community Edition 5.4

Para instalar Pentaho Community Edition se necesita ingresar al sitio web de desarrollo, <http://community.pentaho.com/> y descargar el software en su versión libre.

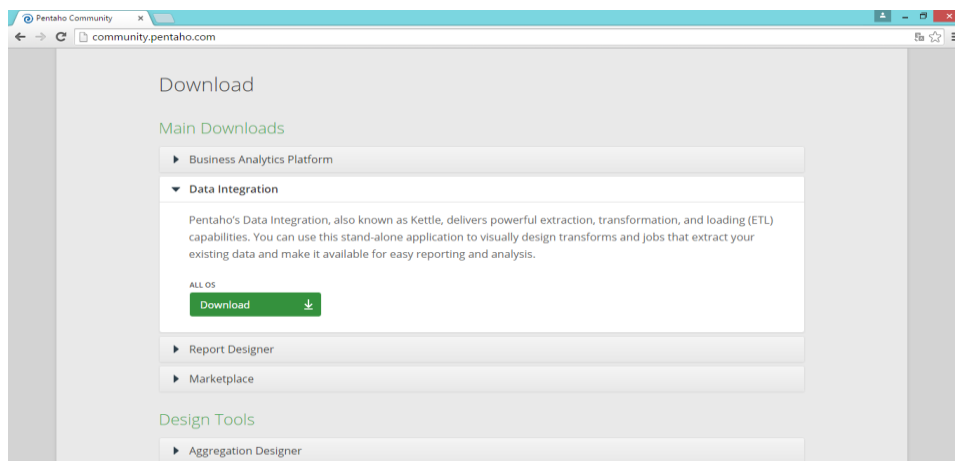
En esta versión de Pentaho ya no es necesario descargar e instalar la herramienta CTools, herramienta utilizada para realizar los cuadros de mando, ya que la misma viene ya incluida en la plataforma del servidor.

Primero hay que seleccionar y descargar Business Analytics Platform que incluye la plataforma de análisis y el servidor Pentaho.



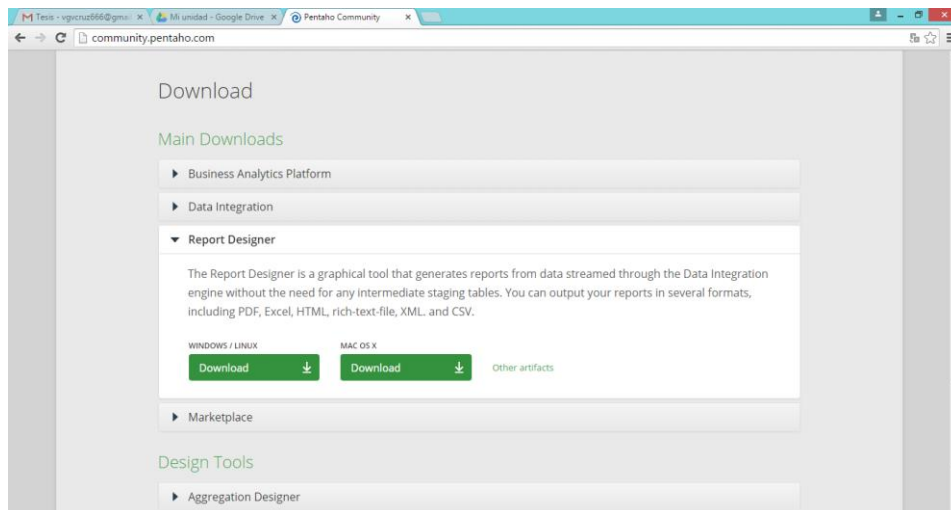
*Figura 4.11: Descarga de Business Analytics Platform.*

El siguiente componente a descargar es el Data Integration, componente que contienen las herramientas ETL de la plataforma.



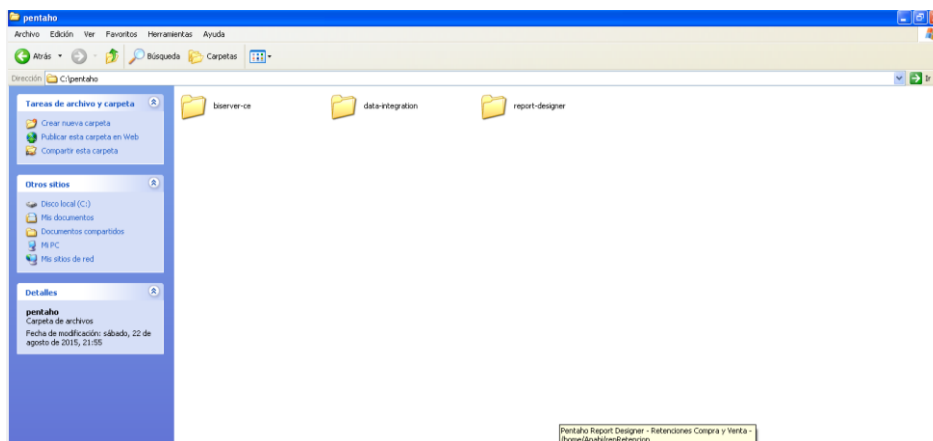
*Figura 4.12: Descarga del Data Integration.*

Finalmente seleccionar y descargar el Report Designer, herramienta que permite realizar los reportes de la solución.



*Figura 4.13: Descarga del Report Designer.*

La instalación es sencilla, primero se debe crear en la unidad C: del servidor una carpeta llamada Pentaho, en su interior se debe copiar los archivos descargados y descomprimirlos; una vez finalizado se procede a la configuración del servidor y los servicios detallado en el siguiente tema.



*Figura 4.14: Carpeta de Instalación Pentaho.*

#### 4.1.6 Configuración Pentaho

La configuración de las herramientas Pentaho Community Edition, no requieren un proceso de instalación al ser archivos ejecutables dentro de cada uno de los paquetes de desarrollo.

Para su configuración se deben añadir los drivers para conexión a las diferentes fuentes de datos.

Lo primero a realizar es la configuración de las variables de entorno, que permiten crear la variable de entorno java. Para esto, hay que entrar a la creación de variables del sistema, seleccionar nueva variable de entorno, establecer el nombre de la misma y la respectiva ruta a la carpeta que contiene el JDK.

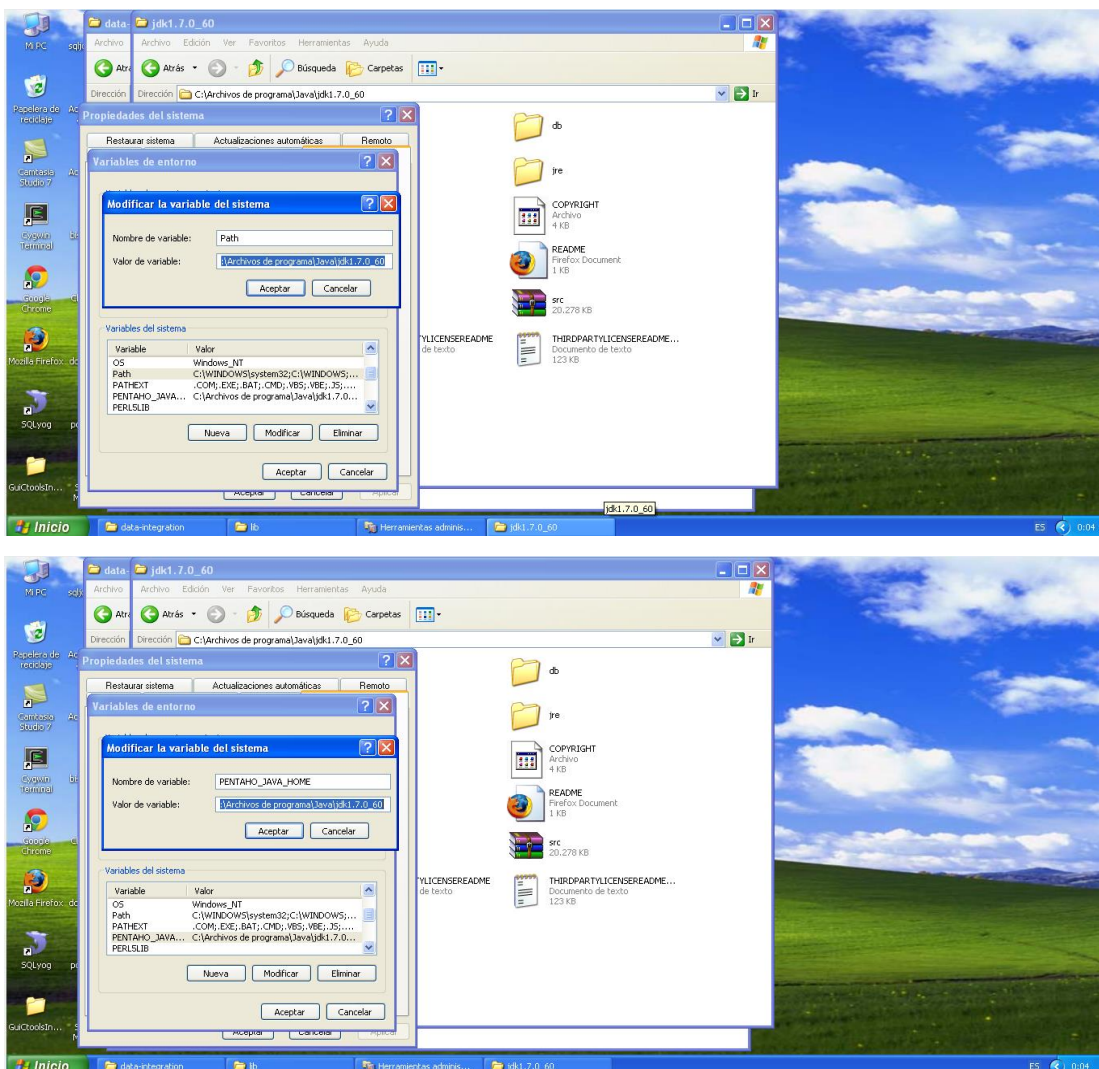


Figura 4.15: Creación de variables de entorno Java.

La segunda etapa de la configuración requiere copiar los drivers de la base de datos MySQL en las librerías de Pentaho ubicada dentro de cada una de las



herramientas con el nombre de carpeta **lib**. Los drivers de la base de datos se deben descargar del sitio web de cada uno de los desarrolladores, en este caso de la web de MySQL <https://dev.mysql.com/downloads/connector/j/5.0.html> en su versión libre.

Se debe pegar los drivers en cada una de las librerías de las diferentes herramientas Pentaho que han sido descargadas.

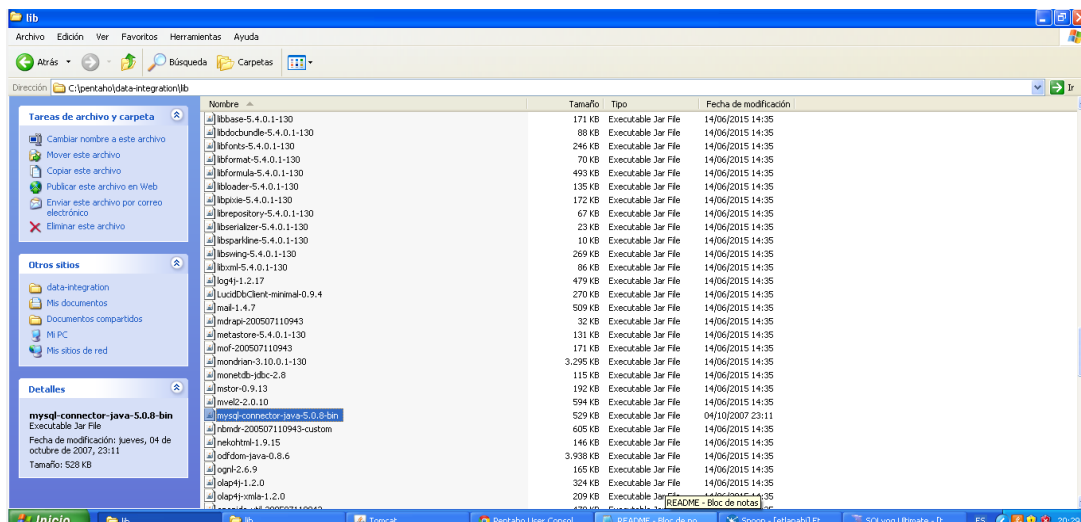


Figura 4.16: Copia de drivers MySQL dentro de las herramientas Pentaho.

La siguiente etapa es la configuración del Pool de Conexiones de las fuentes y destinos de datos, llamada configuración JDNI. Para esto, se ingresa dentro de las carpetas `C:\Pentaho\data-integration\simple-jdni` y `C:\Pentaho\biserver-ce\pentaho-solutions\system\simple-jdni`, donde se encuentra un archivo JDBC; dar clic derecho y editar.

Al final del archivo se agrega la información de las conexiones de datos, donde se ingresa los siguientes datos:

```
etlanahi/type=javax.sql.DataSource
etlanahi/driver=com.mysql.jdbc.Driver
etlanahi/url=jdbc:mysql://localhost:3306/etlanahi
etlanahi/user=root
etlanahi/password=tesis
anahi/type=javax.sql.DataSource
```

```

anahi/driver=com.mysql.jdbc.Driver
anahi/url=jdbc:mysql://localhost:3306/anahi
anahi/user=root
anahi/password=tesis
dwanahi/type=javax.sql.DataSource
dwanahi/driver=com.mysql.jdbc.Driver
dwanahi/url=jdbc:mysql://localhost:3306/dwanahi
dwanahi/user=root
dwanahi/password=tesis

```

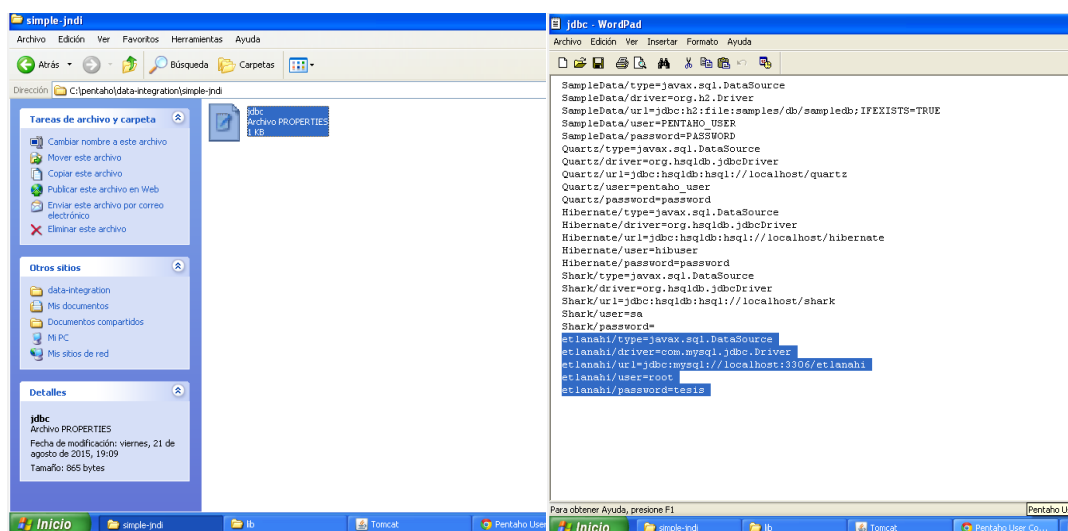


Figura 4.17: Configuración JNDI.

## Seguridad

Los niveles de seguridad de la aplicación se establecen por medio de la administración de Pentaho con el uso de usuarios y roles. Este procedimiento es descrito en el capítulo 4.2.5 Creación de Usuarios.

## 4.2 INTEGRACIÓN DE DATOS

En este capítulo se hace un repaso de los procesos de estandarización, carga inicial de datos, actualizaciones y creación de los componentes lógicos del datawarehouse utilizando las herramientas Pentaho Community.

### 4.2.1 Estándares de Creación

Los estándares son el formato que se van a utilizar en el diseño de cada componente de la solución BI. Se definen los estándares para la creación de todos los componentes físicos del datawarehouse, estos se estructuran en base a cada herramienta mediante un esquema que tendrá varias columnas que se especifican a continuación:

- **Componente:** se indica el nombre del componente que se va a estandarizar.
- **Prefijo:** se almacena el nombre del prefijo que va a tener el objeto.
- **Número de caracteres:** se indica el número máximo de caracteres que tendrá el componente.
- **Descripción:** se indica la descripción del nombre que se va a emplear para cada componente.
- **Ejemplo:** es un ejemplo del nombre estandarizado que tiene cada componente.

#### 4.2.1.1 Almacenamiento físico del datawarehouse.

El almacenamiento físico es el sistema de carpetas que se va a manejar para la creación de nuestra solución. Este sistema físico de carpetas permitirá alojar de manera ordenada los archivos de cada uno de los entregables de las diferentes herramientas que se utilizan.

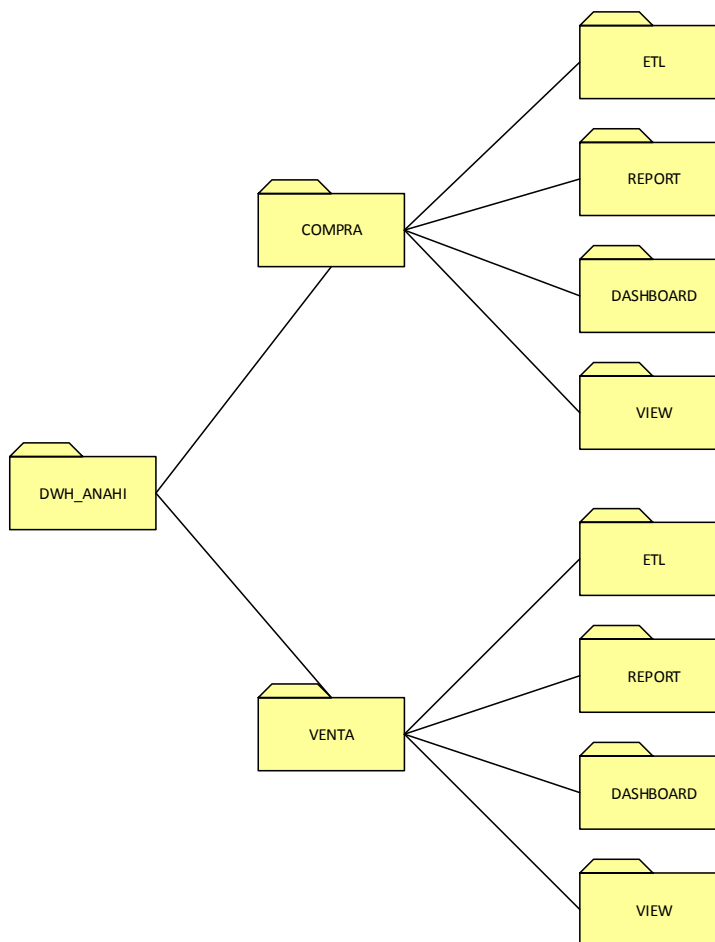
El estándar que se utiliza esta descrito en la siguiente tabla:

*Tabla 7*

*Estándares de Almacenamiento del Datawarehouse.*

Componente	Prefijo	Número de Caracteres	Estándar	Ejemplo
Carpeta General	DWH_	30	Nombre que identifique al datawarehouse.	DWH_ANAHI
Subcarpetas		30	Nombre que identifique el módulo o los componentes que almacena.	COMPRA DASHBOARD

El sistema de carpetas para la solución se muestra en la Figura 4.18.



*Figura 4.18: Almacenamiento Físico del Datawarehouse.*

- **COMPRAS:** Carpeta que almacena toda la información del Datamart de Compras.
- **VENTAS:** Carpeta que almacena toda la información del Datamart de Ventas.
- **ETL:** Carpeta que almacena los ETL y los Jobs.
- **REPORT:** Carpeta que almacena los reportes de cada proceso, archivos de extensión prpt que provee la herramienta Pentaho Report Designer.
- **DASHBOARD:** Carpeta que almacena los dashboard que son implementados por la herramienta Community Dashboard Editor.
- **VIEW:** Carpeta que almacena las vistas de análisis realizadas desde las herramientas Jpivot o el plugin Sayku Analytics, sirven para manipular los cubos de información según las necesidades del usuario.

#### 4.2.1.2 Estándares de la Base de Datos del Datamart

Los estándares para la Base de Datos del Datamart deben seguir los siguientes parámetros:

*Tabla 8*

*Estándares de la base de datos del datamart.*

Componente	Prefijo	Número de Caracteres	Estándar	Ejemplo
Datamart	dwh	30	Nombre que identifique al Datamart.	dwhCompras
Tabla de Dimensiones	Dim	30	Nombre que identifique la dimensión.	dimProveedor
Tabla de Métricas	Metrica	30	Nombre que identifique la métrica.	MetricaVenta
Campos de las Tablas		30	Nombre que identifique a cada campo dentro de las tablas.	Cedula

#### 4.2.1.3 Estándares de los Cubos Dimensionales

Los estándares para la creación de los cubos multidimensionales en este caso, están basadas en el diseño lógico del datawarehouse, estas deben seguir los siguientes parámetros:

*Tabla 9*

*Estándares de los Cubos Multidimensionales.*

Componente	Prefijo	Número de Caracteres	Estándar	Ejemplo
Cubo Multidimensional	dwh	30	Nombre que identifique al cubo multidimensional.	dwhCompra
Dimensiones		30	Nombre que identifique la dimensión.	Tiempo
Métricas		30	Nombre que identifique la métrica.	Total
Niveles		30	Nombre que identifique los campos de la dimensión.	Producto
Jerarquías		30	Nombre que identifique la jerarquía.	Ubicacion

#### 4.2.1.4 Estándares de Pentaho Data Integration

Los estándares de la herramienta Pentaho Data Integration deben seguir los siguientes parámetros:

*Tabla 10*

*Estándares de Pentaho Data Integration.*

Componente	Herramienta PDI	Prefijo	Número de Caracteres	Estándar	Ejemplo
ETL	Spoon	etl	30	Nombre que identifique al ETL.	EtlAnahiVenta
Fuentes de entrada	Entrada Tabla	et_	30	Nombre que identifique a la fuente de información.	et_Compra
Calculadora	Calculadora		30	Nombre que identifique la herramienta de cálculo.	Calculadora
Componente Java Script	Valor Java Script Modificado	js	30	Nombre que identifique al script.	jsSemana
Almacén de datos	Búsqueda/Actualización en Combinación		30	Nombre que identifique la dimensión.	dimTiempo

#### 4.2.1.5 Estándares de Pentaho Community Dashboard Editor

Los estándares de la herramienta Pentaho Dashboard Editor deben seguir los siguientes parámetros:

*Tabla 11*

*Estándares de Pentaho Community Dashboard Editor.*

Componente	Herramienta CDE	Prefijo	Número de Caracteres	Estándar	Ejemplo
CDE	CDE Dashboard	cmi	30	Nombre que identifique a los cuadros de mando.	cmiVentas
Layouts	Layouts	pnl	30	Nombre que identifique al elemento ubicado en el cuadro de mando.	pnlTitulo
Fuentes SQL	SQL Queries	Sql	30	Nombre de la consulta SQL para el dashboard.	sqlVentaXproducto
Componentes de selección.	Components	sel	30	Nombre de los parámetros de selección del dashboard.	selMes, selProveedor
Diagramas y Gráficos	Charts	Bc, lc, pc	30	Nombre de los diagramas y componentes gráficos.	bcVentaXProducto lcComVenMes pcComVenAnu

#### 4.2.1.6 Estándares de Pentaho Report Designer

Los estándares para la creación de reportes con la herramienta Pentaho Data Integration deben seguir los siguientes parámetros:

Tabla 12

*Estándares de Pentaho Report Designer.*

Componente	Prefijo	Número de Caracteres	Estándar	Ejemplo
Reportes	rep	30	Nombre de los archivos de reportes.	repVentasProClie

#### 4.2.2 Limpieza de Datos

Para la limpieza de datos, al establecer y usar los componentes de combinación de dimensión estos crean relaciones únicas para los campos de cada dimensión y en las métricas los valores son establecidos a 0, este proceso es generado automáticamente por el motor de mapeo llamado Hibernate que realiza el mapeo interno de datos.

Además, se debe establecer un sistema de limpieza utilizando las herramientas propias de Pentaho, para este caso se utiliza los siguientes componentes del Pentaho Data Integration:

- **Fórmula:** este componente permite calcular expresiones o fórmulas que pueden hacer cálculos matemáticos simples o expresiones lógicas complejas utilizando condicionales como el If/Then.
- **Set Field Value:** este componente permite cambiar una variable por otra.
- **Swith/Case:** este componente permite dividir el flujo de datos en dos direcciones diferentes. Se debe configurar la condición para evaluar el flujo de datos, luego evalúa cada fila y determina si cumplió o no con la condición; si la cumple será enviada hacia el siguiente paso establecido (True), caso contrario será enviada hacia otro paso (False).
- **Salida Excel:** este componente permite enviar datos a archivos Excel.

Los pasos para limpiar la información para los procesos de compra y venta son los mismos, en este ejemplo se analiza los pasos de limpieza del proceso de ventas, donde se describe el desarrollo de cada uno.

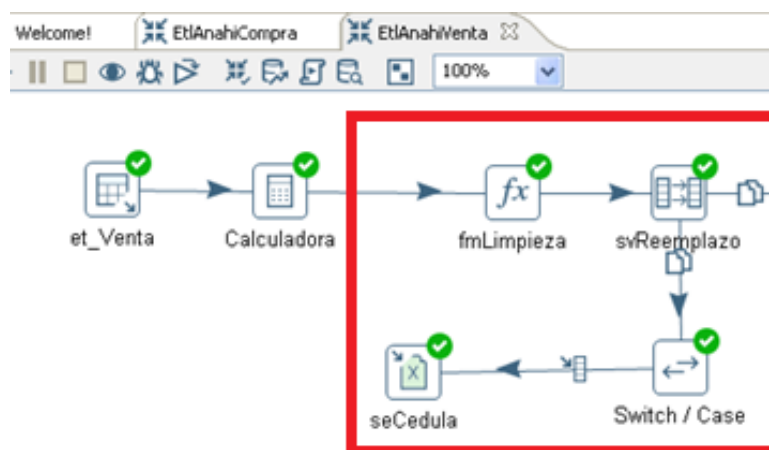


Figura 4.19: Componentes de Limpieza de Datos.

- **fmLimpieza**

Este paso permite ingresar las fórmulas que van a evaluar los valores de las variables de entrada que se van a recibir desde el almacén de datos. En este caso se evalúa el correcto número de caracteres para el número telefónico, número de factura, número de comprobante, cédula, etc.; finalmente se almacena dentro de una nueva variable.

#	New Field	Formula	Value type	Length	Precision	Replace value
1	TelefonoV	$f(\text{len}(\text{trim}([\text{telefono}])) = 0; '9999999999'; [\text{telefono}])$	String	11		
2	FacturaV	$f(\text{len}(\text{trim}([\text{factura}])) = 0; 'N0'; [\text{factura}])$	String	255		
3	compteV	$f(\text{len}(\text{trim}([\text{compte}])) = 0; 'N0'; [\text{compte}])$	String	255		
4	cedulaOk	$f(\text{OR}(\text{len}(\text{trim}([\text{cedula}])) = 10; \text{len}(\text{trim}([\text{cedula}])) = 13); 1; 0)$	Integer			

Figura 4.20: Fórmula fmLimpieza para evaluar los datos de entrada.



- **svReemplazo**

Este paso permite actualizar el valor de los campos originales, con los campos actualizados obtenidos desde la fórmula.

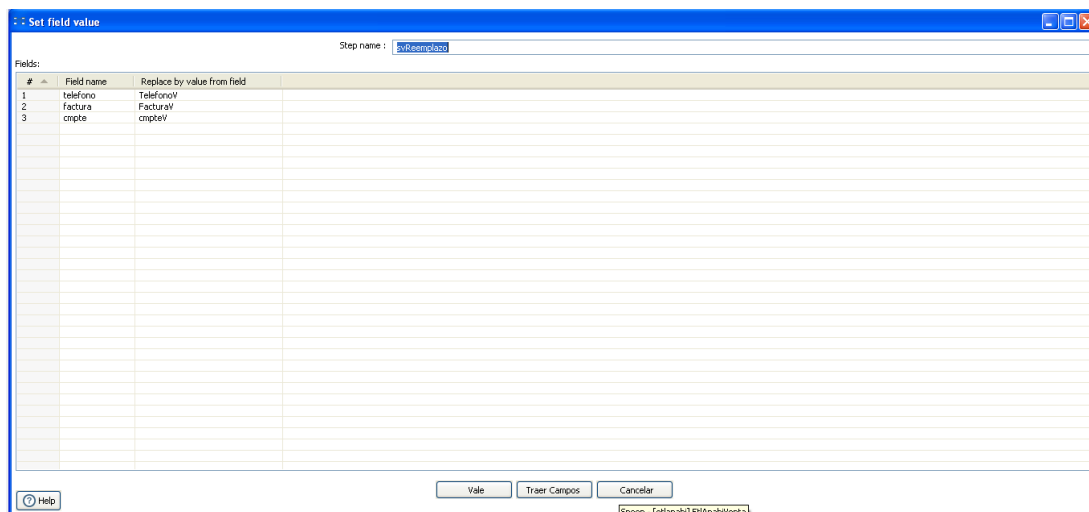


Figura 4.21: Set Field Value svReemplazo.

- **Swith / Case**

Permite identificar si el número de cedula tiene el formato indicado, si lo hace sigue el proceso, caso contrario se llenara con valor 0.

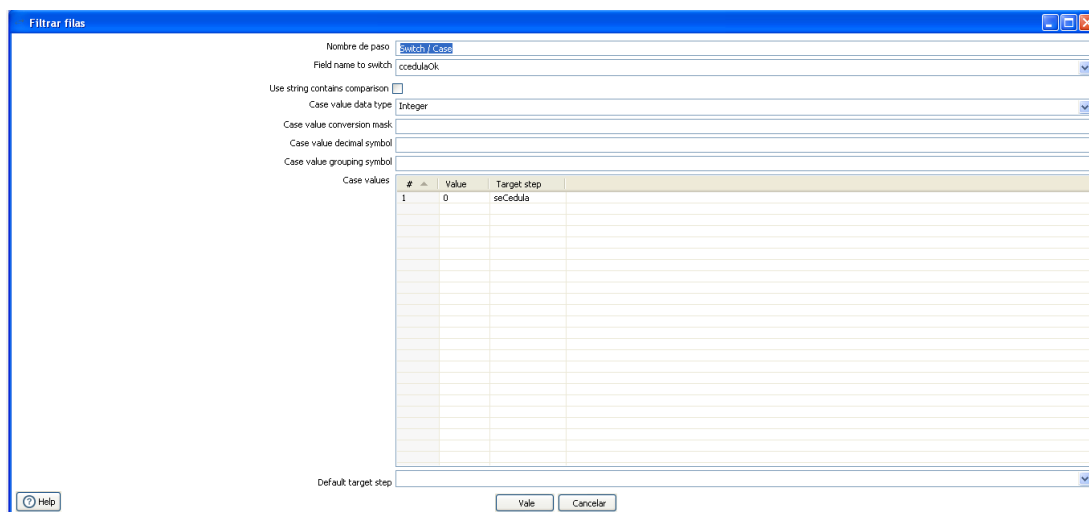


Figura 4.22: Switch/Case para seleccionar los resultados.

- **seCedula**

Permite identificar los números de cedula con valor 0 y mostrarlos en una tabla de Excel para que el usuario pueda visualizar, verificar y llenar la información de forma correcta.

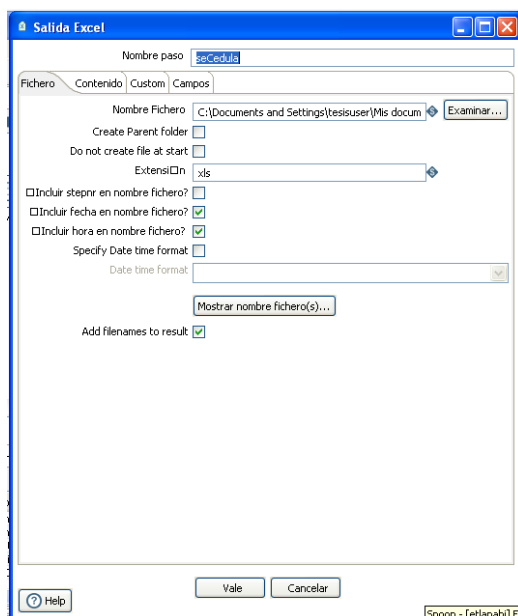


Figura 4.23: Componente de Salida Excel seCedula.

### 4.2.3 Carga Inicial

En este proceso se realiza la carga inicial de datos en el datawarehouse, para lo cual se define los componentes del ETL a utilizar, los procesos que realiza y la construcción del mismo.

Los ETL que se crean están descrito en la siguiente tabla:

Tabla 13

Lista de ETL para la carga inicial.

Nombre	Proceso	Descripción
ETLAnahiCompra	Compra	ETL para el proceso de Compras que contiene la carga, limpieza de información, tablas de dimensiones y métricas.
ETLAnahiVenta	Venta	ETL para el proceso de Ventas que contiene la carga, limpieza de información, tablas de dimensiones y métricas.

## Fuentes de Entrada

Las fuentes de entrada, permiten definir la fuente de información a utilizar para alimentar el datawarehouse.

Para este estudio se utiliza el componente **Entrada Tabla** el cual permite conectar a la base de datos y extrae la información necesaria para la carga.

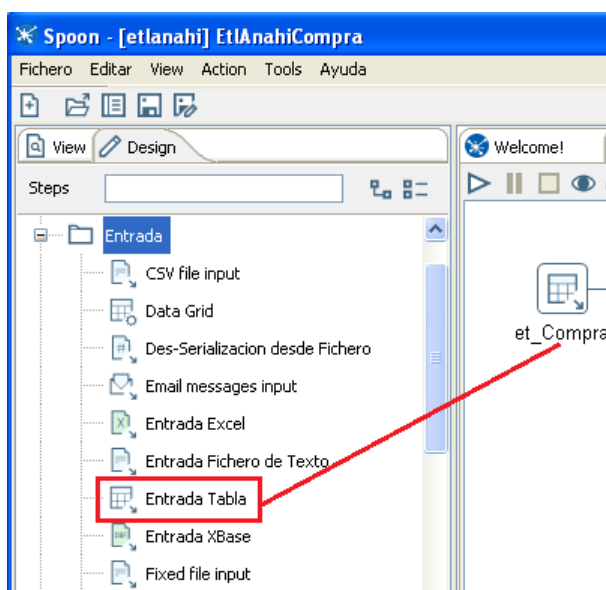


Figura 4.24: Componente Entrada Tabla.

En este componente se selecciona la conexión a la base de datos de entrada, para este caso **Anahí**, y se establece el procedimiento SQL necesario para la extracción de la información necesaria para el estudio. En la figura 4.19 se analiza la estructura de este componente.

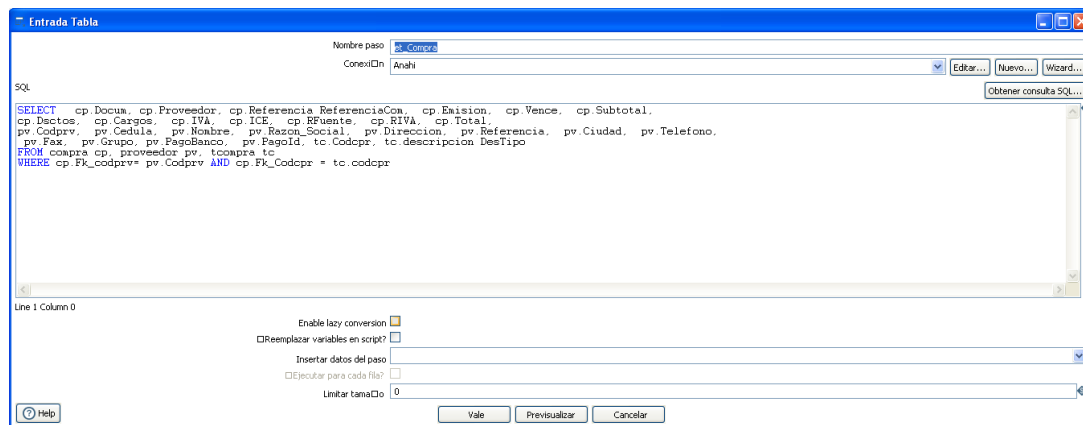


Figura 4.25: Carga Inicial Entrada Tabla.

A continuación mediante el botón **Previsualizar** se puede observar y verificar la información que se extrae de la base de datos para este estudio.

#	Docum	Proveedor	ReferenciaCom	Emision	Vence	Subtotal	Dsctos	Cargos	IVA	ICE	RFuente	RIVA
1	FP000000073	ESPINOSA VALLEJOS GLADYS ESTHE	131583	2015/03/03 00:00:00.000	15/03/03	12,2	0	0	1,5	0	0	0
2	FP010200021	ESPINOSA VALLEJOS GLADYS ESTHE	134756	2015/05/08 00:00:00.000	15/05/08	8,5	0	0	0,9	0	0,1	0
3	FP140000440	ECUACOMEX	36753	2014/09/22 00:00:00.000	14/09/22	357	0	0	42,8	0	3,6	0
4	FP000000114	MAQUINARIAS HENRIQUES C.A.	2276	2015/03/24 00:00:00.000	15/03/24	88,4	0	0	10,6	0	0,9	0
5	FP140000007	QUITO MERCANTIL AUTOMOTRIZ MER	3699	2014/01/17 00:00:00.000	14/01/17	166,8	0	0	20	0	1,8	0
6	FP140000008	QUITO MERCANTIL AUTOMOTRIZ MER	3698	2014/01/17 00:00:00.000	14/01/17	187,3	0	0	22,5	0	2,5	0
7	FP140000266	QUITO MERCANTIL AUTOMOTRIZ MER	4649	2014/06/24 00:00:00.000	14/06/24	103,6	0	0	12,4	0	1	0
8	FP140000467	QUITO MERCANTIL AUTOMOTRIZ MER	5108	2014/09/17 00:00:00.000	14/09/17	389,2	0	0	46,7	0	0	0
9	FP140000468	QUITO MERCANTIL AUTOMOTRIZ MER	5101	2014/09/17 00:00:00.000	14/09/17	4,5	0	0	0,5	0	0	0
10	FP140000017	ENTREGAS ESPECIALES ESPENTREGA	8006	2014/01/03 00:00:00.000	14/01/03	74,5	0	0	8,9	0	1,5	0
11	FP140000021	BORDADOS FABIOLITA	201	2014/01/10 00:00:00.000	14/02/09	207	0	0	0	0	0	0
12	FP140000031	BORDADOS FABIOLITA	202	2014/01/29 00:00:00.000	14/02/28	365,4	0	0	0	0	0	0
13	FP140000120	BORDADOS FABIOLITA	207	2014/03/10 00:00:00.000	14/03/10	284,4	0	0	0	0	0	0
14	FP140000180	BORDADOS FABIOLITA	209	2014/04/16 00:00:00.000	14/04/16	102	0	0	0	0	0	0
15	FP140000461	BORDADOS FABIOLITA	242	2014/10/03 00:00:00.000	14/11/02	244,4	0	0	0	0	0	0
16	FP140000493	BORDADOS FABIOLITA	245	2014/10/23 00:00:00.000	14/11/22	97,8	0	0	0	0	0	0
17	FP140000544	BORDADOS FABIOLITA	249	2014/11/27 00:00:00.000	14/11/27	96,6	0	0	0	0	0	0
18	FP000000147	AGUIRRE CALDERON SEGUNDO CESAR	28744	2015/04/08 00:00:00.000	15/06/07	194,4	0	0	23,3	0	0	0
19	FP000000148	AGUIRRE CALDERON SEGUNDO CESAR	45212	2015/04/10 00:00:00.000	15/06/09	62,5	0	0	7,5	0	0	0
20	FP000000153	AGUIRRE CALDERON SEGUNDO CESAR	129165	2015/04/12 00:00:00.000	15/04/12	21,4	0	0	2,6	0	0	0
21	FP000000165	AGUIRRE CALDERON SEGUNDO CESAR	29011	2015/04/15 00:00:00.000	15/06/14	194,4	0	0	23,3	0	0	0
22	FP000000166	AGUIRRE CALDERON SEGUNDO CESAR	29012	2015/04/15 00:00:00.000	15/04/15	71	0	0	8,5	0	0	0
23	FP000000188	AGUIRRE CALDERON SEGUNDO CESAR	29316	2015/04/22 00:00:00.000	15/06/21	194,4	0	0	23,3	0	0	0
24	FP000000213	AGUIRRE CALDERON SEGUNDO CESAR	29603	2015/04/30 00:00:00.000	15/06/29	194,4	0	0	23,3	0	0	0
25	FP010200006	AGUIRRE CALDERON SEGUNDO CESAR	29781	2015/05/05 00:00:00.000	15/05/05	65,6	0	0	7,9	0	0	0
26	FP010200008	AGUIRRE CALDERON SEGUNDO CESAR	29857	2015/05/06 00:00:00.000	15/05/06	194,4	0	0	23,3	0	0	0
27	FP010200025	AGUIRRE CALDERON SEGUNDO CESAR	30160	2015/05/14 00:00:00.000	15/05/14	194,4	0	0	23,3	0	0	0
28	FP010200036	AGUIRRE CALDERON SEGUNDO CESAR	30402	2015/05/21 00:00:00.000	15/07/20	194,4	0	0	23,3	0	0	0
29	FP010200037	AGUIRRE CALDERON SEGUNDO CESAR	30403	2015/05/21 00:00:00.000	15/06/21	17,9	0	0	2,1	0	0	0
30	FP010200048	AGUIRRE CALDERON SEGUNDO CESAR	30651	2015/05/28 00:00:00.000	15/07/27	194,4	0	0	23,3	0	0	0
31	FP010200058	AGUIRRE CALDERON SEGUNDO CESAR	31215	2015/06/04 00:00:00.000	15/06/04	194,4	0	0	23,3	0	0	0
32	FP010200060	AGUIRRE CALDERON SEGUNDO CESAR	31475	2015/06/12 00:00:00.000	15/07/12	194,4	0	0	23,3	0	0	0
33	FP010200092	AGUIRRE CALDERON SEGUNDO CESAR	31778	2015/06/20 00:00:00.000	15/07/20	194,4	0	0	23,3	0	0	0
34	FP010200103	AGUIRRE CALDERON SEGUNDO CESAR	32540	2015/06/26 00:00:00.000	15/07/26	194,4	0	0	23,3	0	0	0

Figura 4.26: Pre visualización de los datos extraídos.

## Calculadora

Esta herramienta es la encargada de realizar las operaciones matemáticas y de segmentación de datos que previamente se ha recibido de las diversas fuentes, brinda la posibilidad de hacer uso de funciones propias dependiendo del tipo de dato de entrada con el que se esté trabajando.

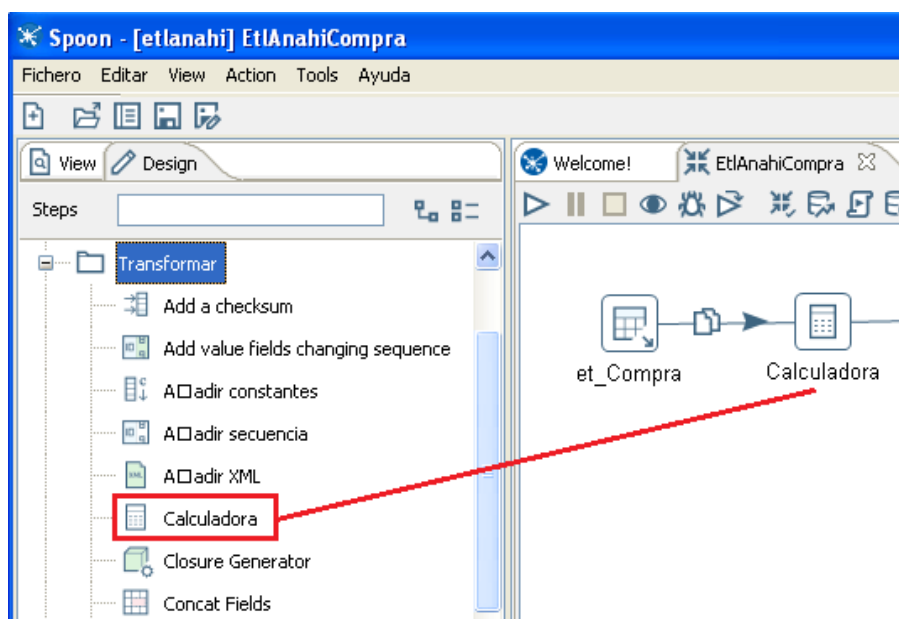


Figura 4.27: Componente Calculadora.

Este componente realiza los cálculos de conversión o extracción de datos de los campos de entrada; dependiendo del tipo de entrada se puede aplicar distintas funciones. En este caso para el campo de entrada **Fecha**, permite obtener las características segmentadas de la misma, pudiendo ser día, mes, año, semana, etc.

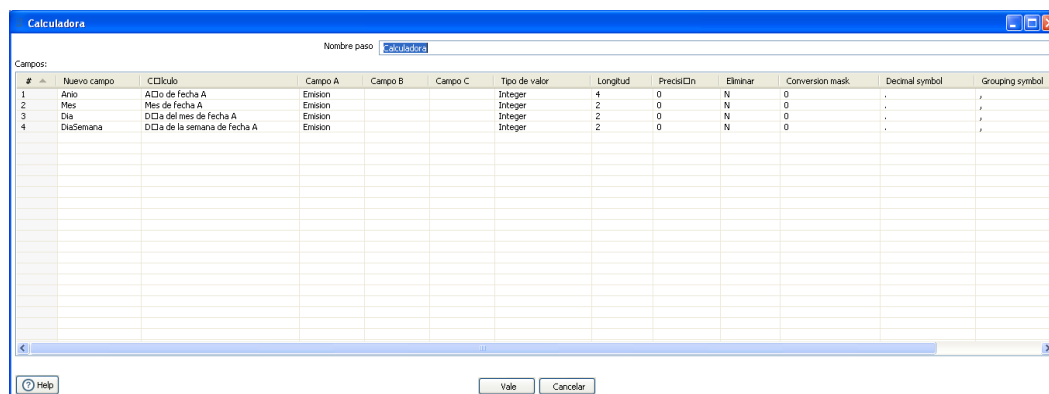


Figura 4.28: Operaciones Calculadora.

## Componente Java Script

Los componentes Java Script que provee la herramienta, permiten programar las validaciones necesarias para la implementación además la transformación de los datos necesarios para el análisis.

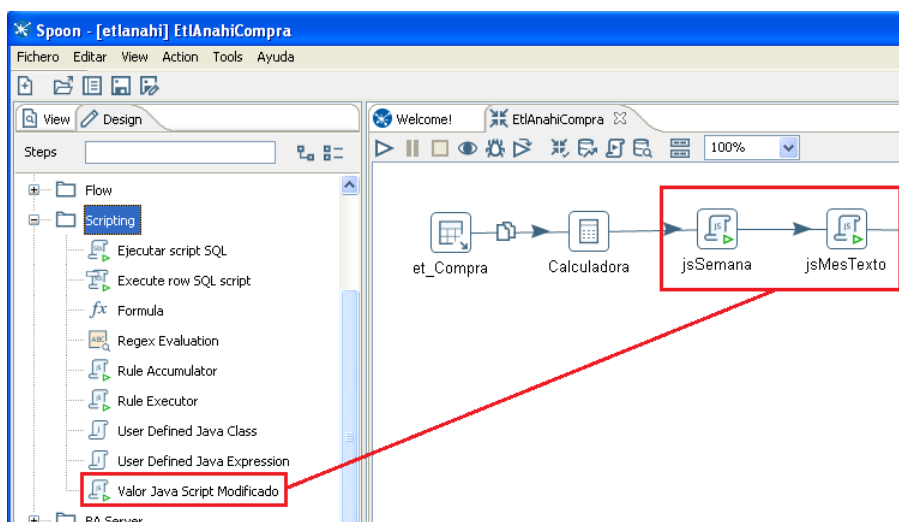


Figura 4.29: Componente Java Script Modificado.

En este caso el valor script **jsSemana** permiten extraer el número de semana del mes al que pertenece una fecha, para ello se utiliza un código que facilite el proceso basado en serie de condicionales. La programación de cada script es independiente de la operación que se va a realizar y del procedimiento que se necesite.

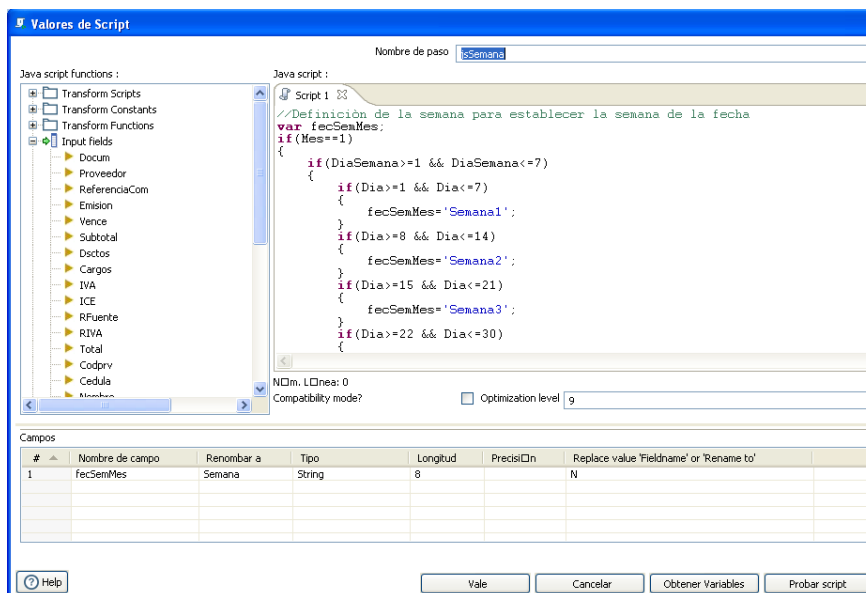


Figura 4.30: Valor Java Script Modificado para extraer el número de Semana.

### Almacén de Datos.

Este componente permite estructurar las dimensiones y cargar los datos dentro de cada una. Permite la creación de la base de datos, con los datos antes transformados y en las dimensiones definidas, para el datawarehouse, finalizando el proceso ETL.

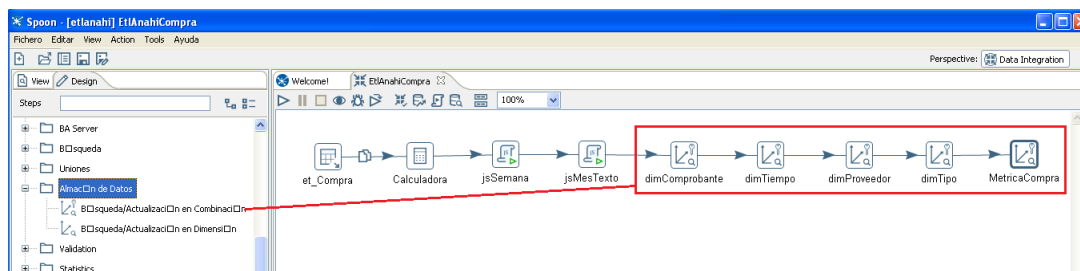


Figura 4.31: Componente Almacén de Datos.

Búsqueda/Actualización en Combinación, permite estructurar y crear las métricas y dimensiones que alimentan el datamart, para lo cual se define el nombre de la dimensión, se selecciona la conexión de destino, la nombre de la tabla de destino, los campos que la componen y crea el campo de clave técnica, conocido como identificador de dimensión, que será un secuencial único.

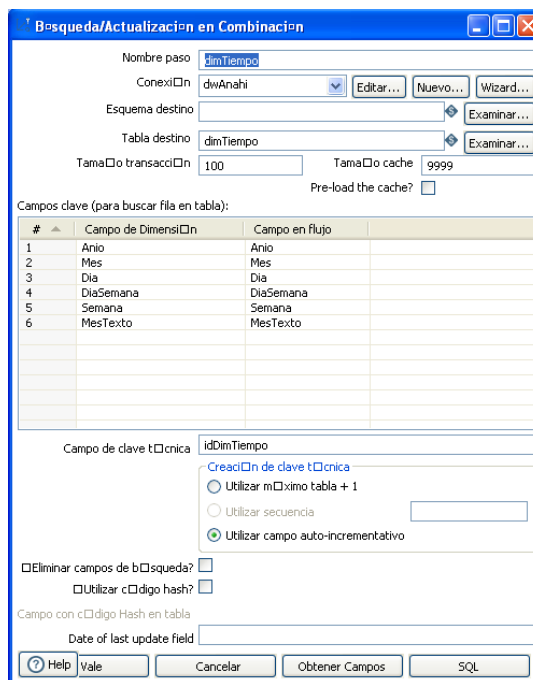


Figura 4.32: Creación del almacén de datos.

## Carga Inicial Compras

El proceso establecido para la carga inicial del ETL Compras es el siguiente:



Figura 4.33: Carga inicial Compras.

- **et\_Compra:** este componente de entrada permite la extracción de la información desde la base de datos correspondiente al proceso de Compras.

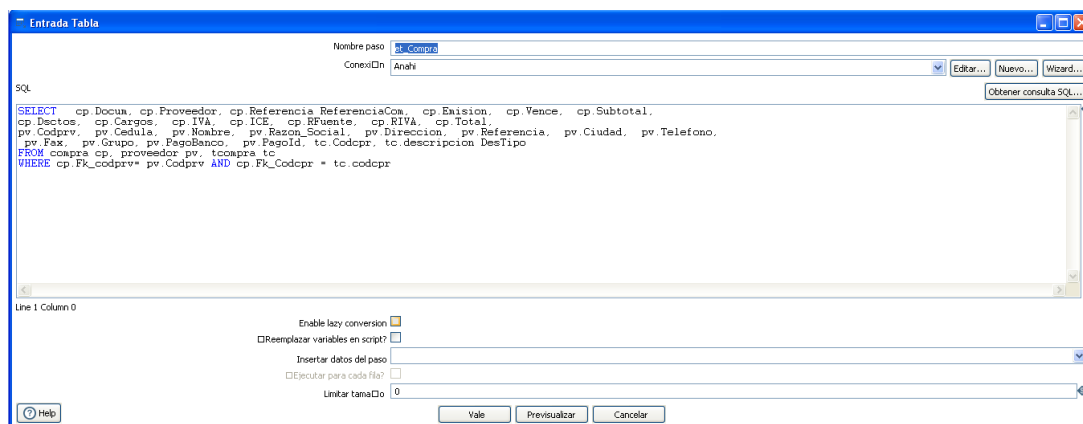


Figura 4.34: Componente de entrada et\_Compra.

- **Calculadora:** este componente permite extraer el año, mes, día del mes y la semana de la fecha de compra correspondiente al campo emisión de la fuente de datos.

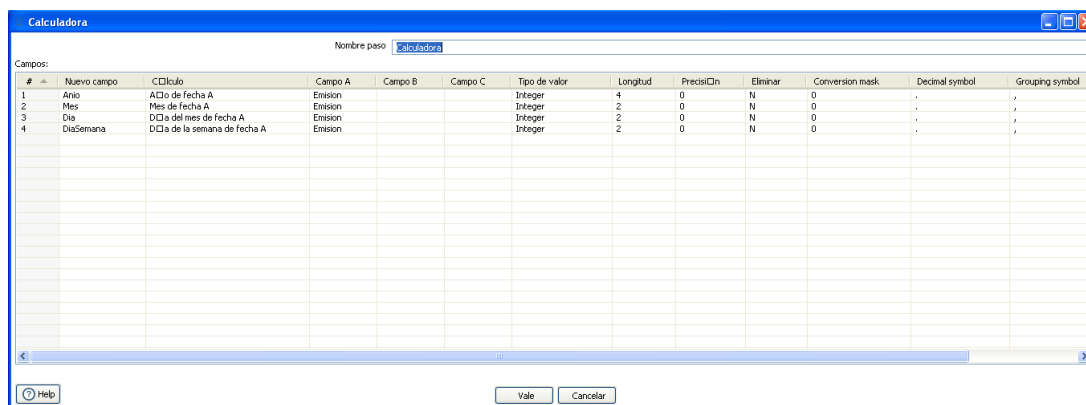


Figura 4.35: Componente de transformación Calculadora.

- **jsSemana:** componente Java Script que establece el cálculo de la semana del mes a la que pertenece una fecha.



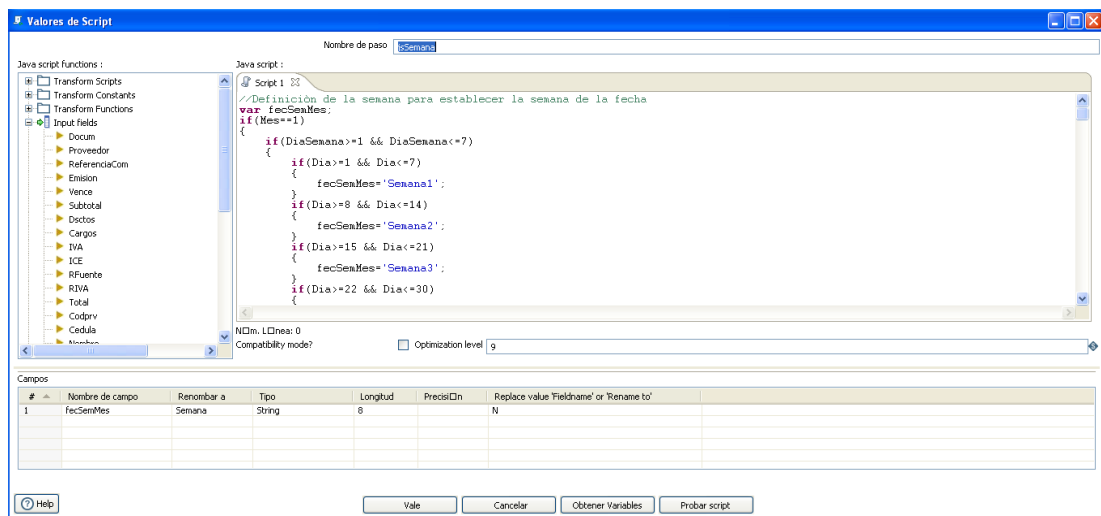


Figura 4.36: Componente Java Script jsSemana.

- **jsMesTexto:** componente Java Script que establece la transformación del mes numérico a uno valor de tipo texto.

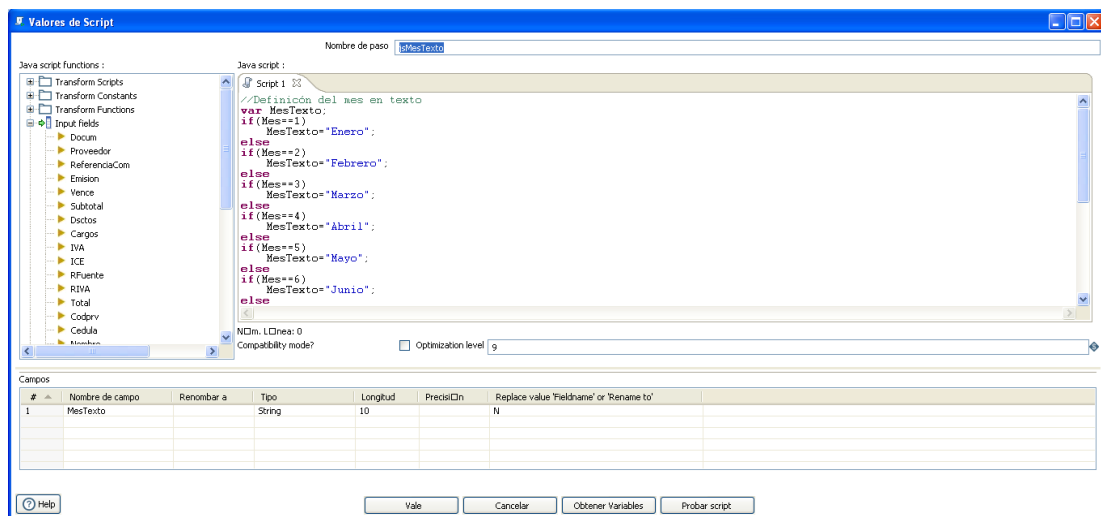


Figura 4.37: Componente Java Script jsMesTexto.

- **dimComprobante:** componente que ejecuta los pasos para la creación de la dimensión Comprobante. Se establece la conexión, la tabla de destino y los campos que la componen. Además se establece el nombre del identificador dimensional idDimComprobante.

The dialog box 'Búsqueda/Actualización en Combinación' is used for creating a dimension. The 'Nombre paso' field is set to 'dimComprobante'. The 'Conexión' is 'dwAnahi'. The 'Esquema destino' and 'Tabla destino' are both 'dimComprobante'. The 'Tamaño transacción' is 100 and 'Tamaño cache' is 9999. The 'Pre-load the cache?' checkbox is unchecked.

The 'Campos clave (para buscar fila en tabla):' table is as follows:

#	Campo de Dimensión	Campo en flujo
1	Docum	Docum
2	ReferenciaCom	ReferenciaCom
3	Vence	Vence

The 'Campo de clave técnica' is 'idDimComprobante'. The 'Creación de clave técnica' options are: 'Utilizar máximo tabla + 1' (unchecked), 'Utilizar secuencia' (unchecked), and 'Utilizar campo auto-incrementativo' (checked). There are also checkboxes for 'Eliminar campos de búsqueda?' and 'Utilizar código hash?'. The 'Date of last update field' is empty.

Figura 4.38: Creación de la Dimensión Comprobante dimComprobante.

- **dimTiempo:** componente que ejecuta los pasos para la creación de la dimensión Tiempo. Se establece la conexión, la tabla de destino y los campos que la componen. Además se establece el nombre del identificador dimensional idDimTiempo.

The dialog box 'Búsqueda/Actualización en Combinación' is used for creating a dimension. The 'Nombre paso' field is set to 'dimTiempo'. The 'Conexión' is 'dwAnahi'. The 'Esquema destino' and 'Tabla destino' are both 'dimTiempo'. The 'Tamaño transacción' is 100 and 'Tamaño cache' is 9999. The 'Pre-load the cache?' checkbox is unchecked.

The 'Campos clave (para buscar fila en tabla):' table is as follows:

#	Campo de Dimensión	Campo en flujo
1	Año	Año
2	Mes	Mes
3	Día	Día
4	DíaSemana	DíaSemana
5	Semana	Semana
6	MesTexto	MesTexto

The 'Campo de clave técnica' is 'idDimTiempo'. The 'Creación de clave técnica' options are: 'Utilizar máximo tabla + 1' (unchecked), 'Utilizar secuencia' (unchecked), and 'Utilizar campo auto-incrementativo' (checked). There are also checkboxes for 'Eliminar campos de búsqueda?' and 'Utilizar código hash?'. The 'Date of last update field' is empty.

Figura 4.39: Creación de la Dimensión Tiempo dimTiempo.

- **dimProveedor:** componente que ejecuta los pasos para la creación de la dimensión Proveedor. Se establece la conexión, la tabla de destino y los campos que la componen. El nombre del identificador dimensional es idDimProveedor.

Nombre paso: dimProveedor

Conexión: dwAnahi

Esquema destino:

Tabla destino: dimProveedor

Tamaño transacción: 100

Tamaño cache: 9999

Pre-load the cache?

Campos clave (para buscar fila en tabla):

#	Campo de Dimensión	Campo en flujo
1	Codigo	Codprv
2	Cedula	Cedula
3	Nombre	Nombre
4	Razon_Social	Razon_Social
5	Direccion	Direccion
6	Referencia	Referencia
7	Ciudad	Ciudad
8	Telefono	Telefono
9	Fax	Fax
10	Grupo	Grupo
11	PagoBanco	PagoBanco
12	PagoId	PagoId

Campo de clave primaria: idDimProveedor

Creación de clave primaria:

Utilizar el último tabla + 1

Utilizar secuencia

Utilizar campo auto-incrementativo

Eliminar campos de búsqueda?

Utilizar código hash?

Campo con código hash en tabla:

Date of last update field:

Buttons: Help, Vale, Cancelar, Obtener Campos, SQL

Figura 4.40: Creación de la Dimensión Proveedor dimProveedor.

- **dimTipo:** componente que ejecuta los pasos para la creación de la dimensión Tipo. Se establece la conexión, la tabla de destino y los campos que la componen. El nombre del identificador dimensional es idDimTipo.

Nombre paso: dimTipo

Conexión: dwAnahi

Esquema destino:

Tabla destino: dimTipo

Tamaño transacción: 100

Tamaño cache: 9999

Pre-load the cache?

Campos clave (para buscar fila en tabla):

#	Campo de Dimensión	Campo en flujo
1	CodTipo	Codprv
2	Descripcion	DesTipo

Campo de clave primaria: idDimTipo

Creación de clave primaria:

Utilizar el último tabla + 1

Utilizar secuencia

Utilizar campo auto-incrementativo

Eliminar campos de búsqueda?

Utilizar código hash?

Campo con código hash en tabla:

Date of last update field:

Buttons: Help, Vale, Cancelar, Obtener Campos, SQL

Figura 4.41: Creación de la Dimensión Tipo dimTipo.

- **MetricaCompra:** componente que ejecuta los pasos para la creación de la métrica Compra. Se establece la conexión, la tabla de destino y los campos que la componen, se debe añadir los identificadores dimensionales de todas las dimensiones antes creadas para establecer la relación correspondiente. El nombre del identificador dimensional es idMetricaCompra.

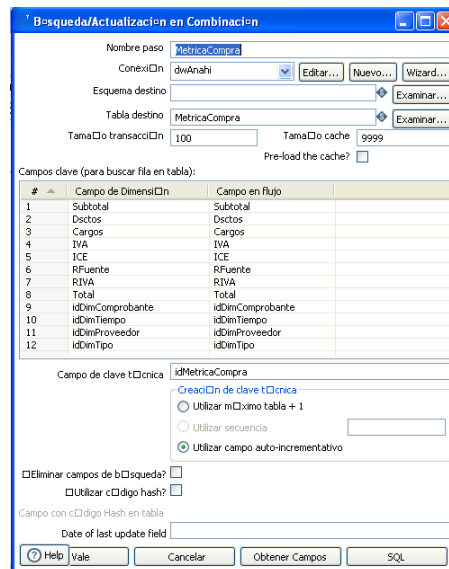


Figura 4.42: Creación de la Métrica MetricaCompra.

Para finalizar la carga inicial del proceso de Compras se da clic en el botón **Ejecutar esta transformación o trabajo** descrito en la Figura 4.43, donde si la ejecución es correcta se visualiza un check de color verde sobre cada uno de los componentes creados con éxito.

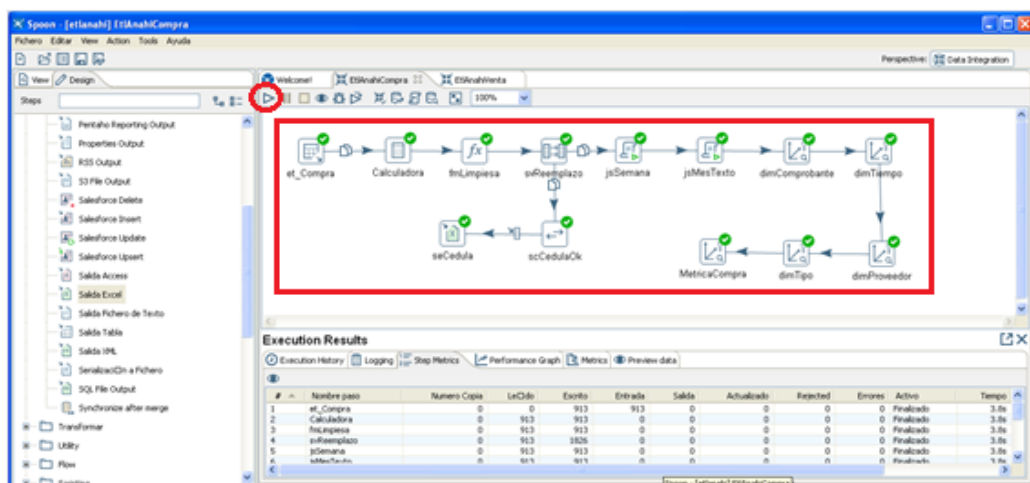


Figura 4.43: Ejecución de carga inicial Compras.

## Carga Inicial Ventas

El proceso establecido para la carga inicial del ETL Ventas es el siguiente:



Figura 4.44: Carga Inicial Ventas.

- **et\_Venta:** este componente de entrada permite la extracción de la información desde la base de datos correspondiente al proceso de Ventas.

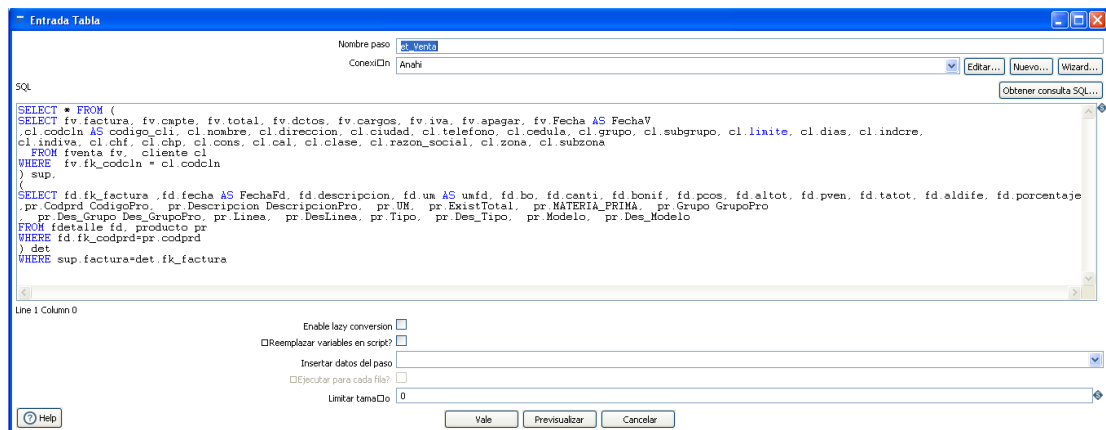


Figura 4.45: Componente de entrada et\_Venta.

- **Calculadora, jsSemana y jsMesTexto:** Estos componentes cumplen la misma funcionalidad y tienen igual estructura que en el proceso de Compras, el proceso de creación e identificadores es el mismo.
- **dimTiempo:** componente que ejecuta los pasos para la creación de la dimensión Tiempo. Se establece la conexión, la tabla de destino y los campos que la componen. Además se establece el nombre del identificador dimensional idDimTiempo.

Nombre paso: dimTiempo

Conexión: dwAnahi

Esquema destino:

Tabla destino: dimTiempo

Tamaño transacción: 100      Tamaño cache: 9999

Pre-load the cache?

Campos clave (para buscar fila en tabla):

#	Campos de Dimensión	Campo en flujo
1	Anio	Anio
2	Mes	Mes
3	Dia	Dia
4	DiaSemana	DiaSemana
5	Semana	Semana
6	MesTexto	MesTexto

Campo de clave técnica: idDimTiempo

Creación de clave técnica:

Utilizar máximo tabla + 1

Utilizar secuencia

Utilizar campo auto-incrementativo

Eliminar campos de búsqueda?

Utilizar código hash?

Campo con código Hash en tabla:

Date of last update field:

Figura 4.46: Creación de la Dimensión Tiempo dimTiempo.

- **dimCliente:** componente que ejecuta los pasos para la creación de la dimensión Cliente. Se establece la conexión, la tabla de destino y los campos que la componen. El nombre del identificador dimensional es idDimCliente.

Nombre paso: dimCliente

Conexión: dwAnahi

Esquema destino:

Tabla destino: dimCliente

Tamaño transacción: 100      Tamaño cache: 9999

Pre-load the cache?

Campos clave (para buscar fila en tabla):

#	Campos de Dimensión	Campo en flujo
1	codigo_cli	codigo_cli
2	nombre	nombre
3	direccion	direccion
4	ciudad	ciudad
5	telefono	telefono
6	cedula	cedula
7	grupo	grupo
8	subgrupo	subgrupo
9	limite	limite
10	dias	dias
11	indcre	indcre
12	indiva	indiva

Campo de clave técnica: idDimCliente

Creación de clave técnica:

Utilizar máximo tabla + 1

Utilizar secuencia

Utilizar campo auto-incrementativo

Eliminar campos de búsqueda?

Utilizar código hash?

Campo con código Hash en tabla:

Date of last update field:

Figura 4.47: Creación de la Dimensión Cliente dimCliente.

- **dimFactura:** componente que ejecuta los pasos para la creación de la dimensión Factura. Se establece la conexión, la tabla de destino y los campos que la componen. El nombre del identificador dimensional es idDimFactura.

Nombre paso: dimFactura

Conexión: dwAnahi

Esquema destino: dimFactura

Tabla destino: dimFactura

Tamaño transacción: 100

Tamaño cache: 9999

Pre-load the cache?

Campos clave (para buscar fila en tabla):

#	Campos de Dimensión	Campos en flujo
1	Factura	Factura
2	cmpte	cmpte
3	unfd	unfd

Campos clave técnica: idDimFactura

Creación de clave técnica:

- Creación de clave técnica
- Utilizar máximo tabla + 1
- Utilizar secuencia
- Utilizar campo auto-incrementativo

Eliminar campos de búsqueda?

Utilizar código hash?

Campos con código Hash en tabla

Date of last update field

Buttons: Help, Vale, Cancelar, Obtener Campos, SQL

Figura 4.48: Creación de la Dimensión Factura dimFactura.

- **dimProducto:** componente que ejecuta los pasos para la creación de la dimensión Producto. Se establece la conexión, la tabla de destino y los campos que la componen. El nombre del identificador dimensional es idDimProducto.

Nombre paso: dimProducto

Conexión: dwAnahi

Esquema destino: dimProducto

Tabla destino: dimProducto

Tamaño transacción: 100

Tamaño cache: 9999

Pre-load the cache?

Campos clave (para buscar fila en tabla):

#	Campos de Dimensión	Campos en flujo
1	CodigoPro	CodigoPro
2	Producto	DescripcionPro
3	UM	UM
4	Cod_grupo	GrupoPro
5	Grupo	Des_grupoPro
6	Cod_linea	Linea
7	Liema	DestLinea
8	Tipo	Tipo
9	Des_Tipo	Des_Tipo
10	Cod_Modelo	Modelo
11	Modelo	Des_Modelo

Campos clave técnica: idDimProducto

Creación de clave técnica:

- Creación de clave técnica
- Utilizar máximo tabla + 1
- Utilizar secuencia
- Utilizar campo auto-incrementativo

Eliminar campos de búsqueda?

Utilizar código hash?

Campos con código Hash en tabla

Date of last update field

Buttons: Help, Vale, Cancelar, Obtener Campos, SQL

Figura 4.49: Creación de la Dimensión Producto dimProducto.

- **MetricaVenta:** componente que ejecuta los pasos para la creación de la métrica Venta. Se establece la conexión, la tabla de destino y los campos que la componen, se debe añadir los identificadores dimensionales de todas las dimensiones antes creadas para establecer la relación correspondiente. El nombre del identificador dimensional es idMetricaVenta.

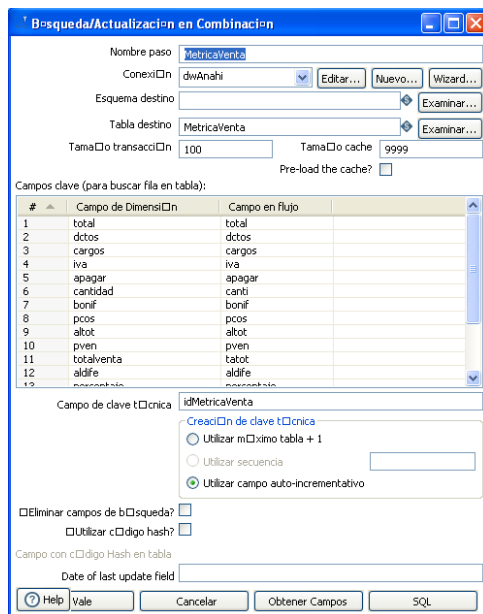


Figura 4.50: Creación de la Métrica MetricaVenta.

Para finalizar la carga inicial del proceso de Ventas se da clic en el botón **Ejecutar esta transformación o trabajo** descrito en la Figura 4.44, donde si la ejecución es correcta se visualiza un check de color verde sobre cada uno de los componentes creados con éxito.

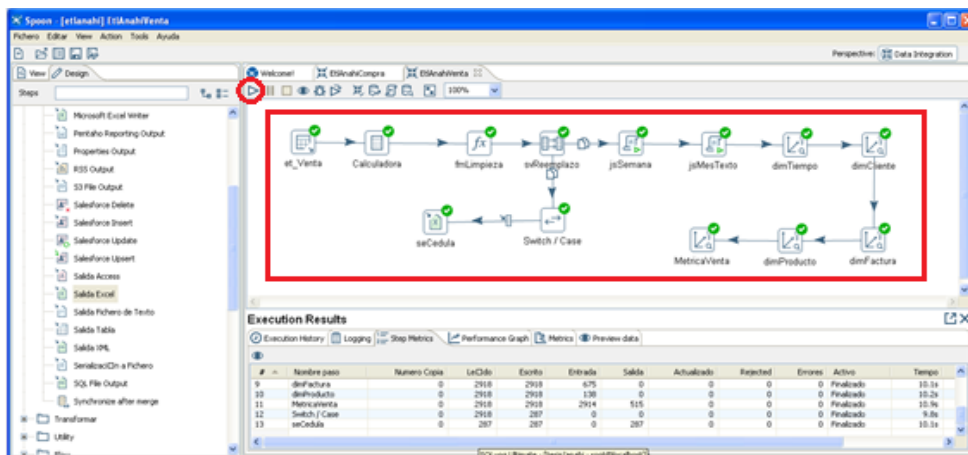




Figura 4.51: Ejecución de carga inicial Ventas.

#### 4.2.4 Actualización

Para la actualización de información de los ETL, es necesario trabajar con los llamados *Jobs*, herramientas que permiten realizar el trabajo de actualización automáticamente en un horario establecido y un intervalo de tiempo dependiendo los requerimientos de los usuarios.

Los Jobs permiten volver a ejecutar los procesos ETL, donde la herramienta Pentaho Data Integration permite reconocer automáticamente si la información de la fuente está actualizada o si existen nuevos registros los va añadiendo a la base de datos del datawarehouse.

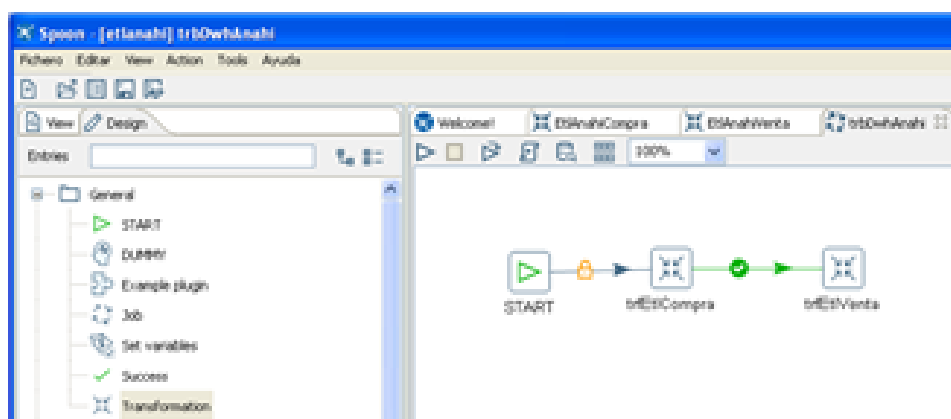


Figura 4.52: Jobs para la Actualización del ETL.

Para la creación del job de actualización se debe determinar la recurrencia de ejecución del trabajo y determinar el horario de actualización, donde se escoge uno que este en un rango que no produzca inconvenientes en el trabajo de la empresa. Para ello se debe seguir los siguientes pasos:

- **Creación del Job trbDwhAnahi**

Se selecciona en las herramientas General un nuevo Job, en las propiedades de creación se debe poner el nombre del Job y comenzar a configurar su contenido con los procesos que se necesite.

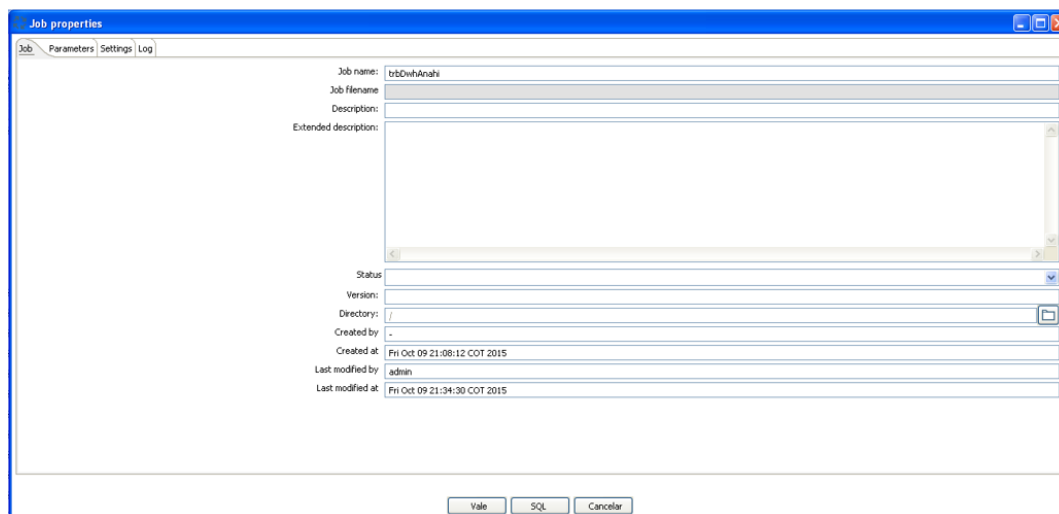


Figura 4.53: Creación del Job trbDwhAnahi.

- **START**

Este componente permite administrar el horario y la recurrencia de la actualización, se selecciona la opción Repeat para que la actualización sea constante, el tipo donde se selecciona Diario y la hora del día en la cual se va a ejecutar el trabajo establecido.

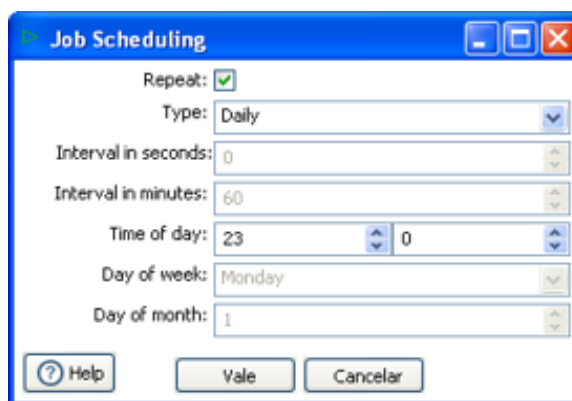


Figura 4.54: Herramienta START.

- **Trabajos trfEtlCompra y trfETLVenta**

Con esta herramienta se especifica la referencia para actualizar los ETL de Compra y Venta respectivamente. Aquí se selecciona cada ETL para que, automáticamente, se ejecuta todo el proceso de carga, limpieza de datos, transformación como se realizó en la carga inicial de datos.

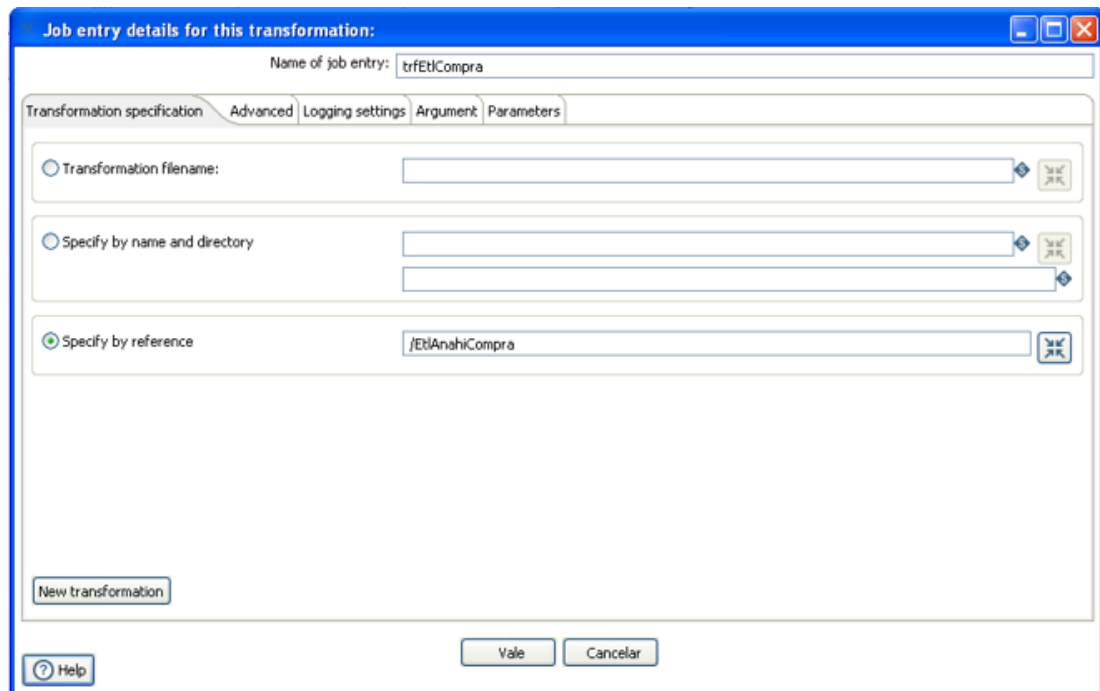


Figura 4.55: Detalles de transformación trfEtlCompra.

Finalmente se ejecuta el job para verificar que la actualización sea correcta, una vez hecho y verificado se obtiene un mensaje de ejecución exitoso. La siguiente actualización se hará de forma automática en el horario establecido.

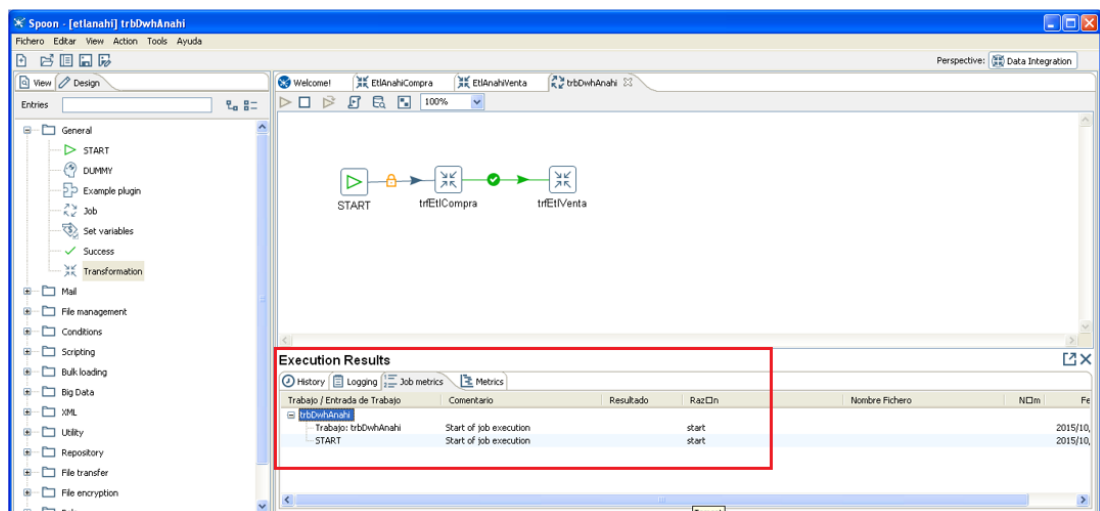


Figura 4.55: Ejecución del Job trbDwhAnahi.

### 4.3 CREACIÓN DE CUBOS MULTIDIMENSIONALES

#### 4.3.1 Diccionario de Datos

En este capítulo se describen los cubos multidimensionales que van a ser creados y las herramientas de explotación para cada uno.

*Tabla 14*

*Diccionario de datos del datawarehouse.*

TIPO	NOMBRE	PROCESO	DESCRIPCION
Cubo multidimensional	dwhCompra	Compra	Cubo multidimensional del proceso de Compras.
Cubo multidimensional	dwhVenta	Venta	Cubo multidimensional del proceso de Ventas.
Reporte	repTotalCompraPro	Compra	Reporte del Monto total de compras a cada proveedor en un periodo de tiempo.
Reporte	repTotalCompra	Compra	Reporte del Monto total de compras en un periodo de tiempo.
Reporte	repTotalCompraAdq	Compra	Reporte del Monto total de compras por tipo de adquisición en un periodo de tiempo.
Reporte	repTotalIva repTotalRetFuente repTotalRetIva	Compra	Reporte del Monto total de IVA en Compras, Retención a la Fuente y Retención de IVA en un periodo de tiempo.
Reporte	repUniVendidas	Venta	Reporte de Unidades vendidas de cada producto a cada cliente en un periodo de tiempo.
Reporte	repTotalClienteProducto	Venta	Reporte del Monto total de ventas de cada producto en un periodo de tiempo.
Reporte	repTotalVenta	Venta	Reporte del Monto total de ventas a cada cliente en un periodo de tiempo.
Reporte	repTotIvaVenta	Venta	Reporte del Monto total de IVA en ventas en un periodo de tiempo.
Reporte	repCompraVenta	Compra Venta	Reporte del Monto total de Compras y Monto Total de Ventas en un periodo de tiempo.
Vista de Análisis	AnalisisCompra	Compra	Vista de Análisis del cubo de Compras.
Vista de Análisis	AnalisisVenta	Venta	Vista de Análisis del cubo de Ventas.
Tablero de mando		Compra Venta	Indicadores claves de los procesos de compra y venta.
Usuarios	Admin	Compra Venta	Usuario Administrador General del sistema que puede ver y editar todos los elementos del DWH.

**Continúa** ➔

Usuarios	Gerencia	Compra Venta	Usuario de Gerencia que puede visualizar todos los elementos del datawarehouse.
Usuarios	Contabilidad	Compra Venta	Usuario de Contabilidad que tendrá acceso a los reportes.

### 4.3.2 Creación de Cubos

Para la creación de cubos de información, como primer paso se debe ingresar a la plataforma Pentaho, en el menú principal seleccionar la opción <New> y la opción <Manage Data Sources>.

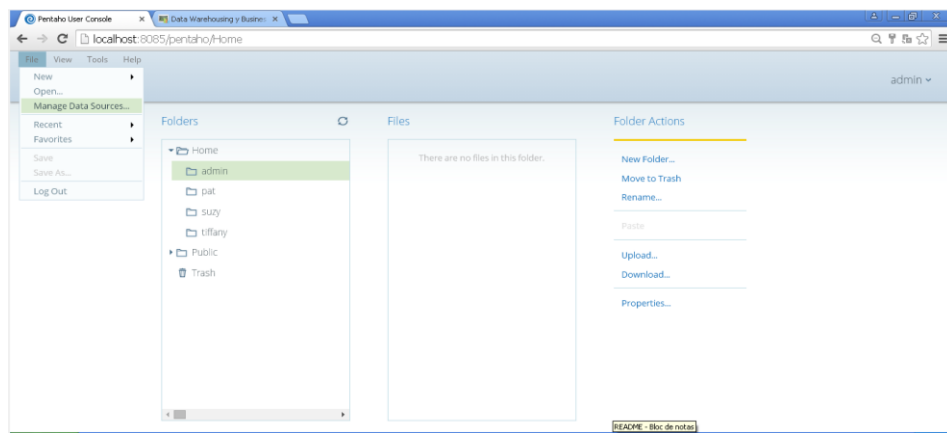


Figura 4.56: Creación del Modelo de Cubo de Información.

Seleccionar la opción del icono <+> para crear una fuente de datos, para la conexión a la base de datos que almacena el DW.

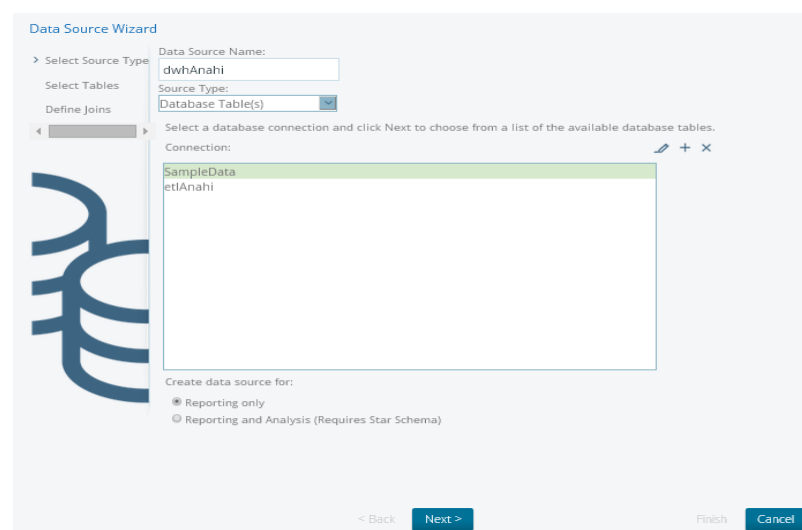


Figura 4.57: Creación fuente de datos

Establecer el nombre y los parámetros de conexión a la base de datos fuente del DW.

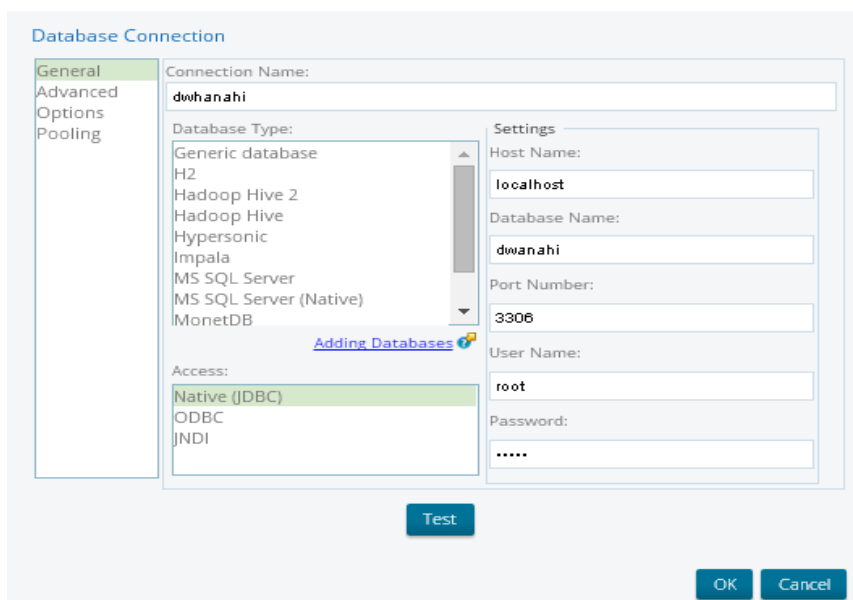


Figura 4.58: Parámetros de conexión al DW

Verificar si la conexión a la fuente de datos es correcta.

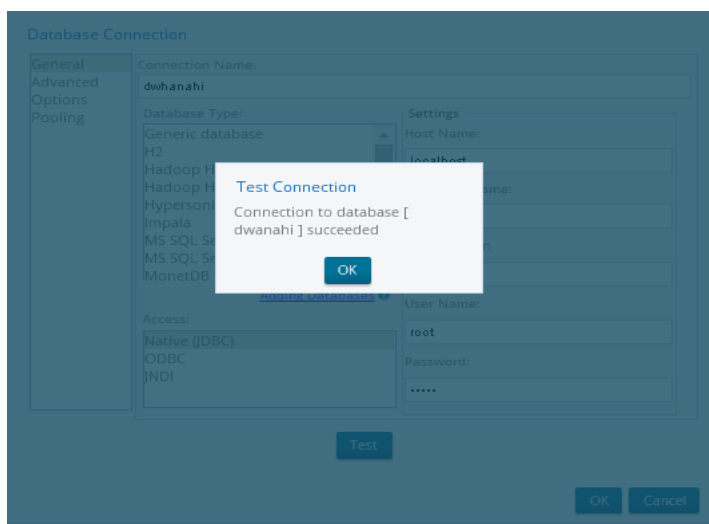


Figura 4.59: Prueba de conexión.

Una vez establecida la fuente de datos, establecer el nombre del Cubo, la fuente de datos que va a utilizar; de las opciones de la parte inferior seleccionar la opción <Reporting and Analysis>, presionar <Next>.

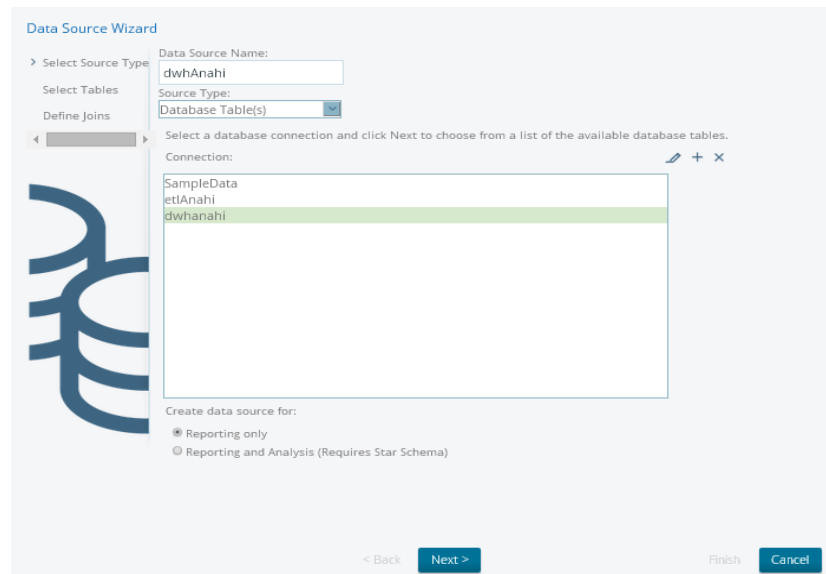


Figura 4.60: Selección del tipo de creación de Cubo.

Seleccionar las tablas que corresponden a las dimensiones y la tabla de métrica presionando la opción central < >, en la parte inferior seleccionar <Fact Table> del menú desplegable la tabla que corresponde a la FAC o tabla de hechos.

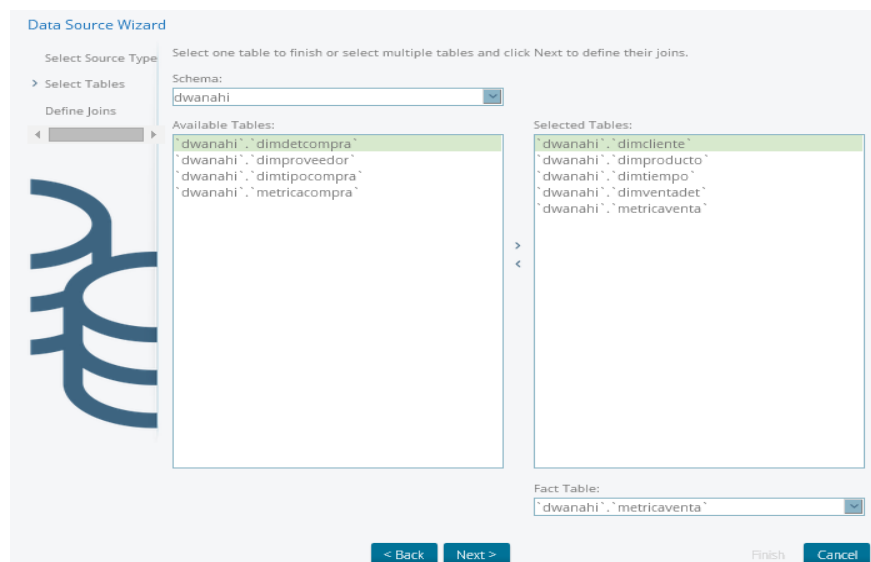


Figura 4.61: Selección de dimensiones y tabla de hecho.

Establecer las relaciones existentes entre la tabla de hecho y las dimensiones, por medio de los identificadores. De la opción <Left Table> seleccionar el campo identificador correspondiente a la dimensión, de la opción <Right Table> seleccionar

la dimensión y su campo identificador, presionar la opción <Create Join> para generar la relación entre la tabla de hecho y su dimensión, al finalizar las relaciones presionar la opción < Finish>.

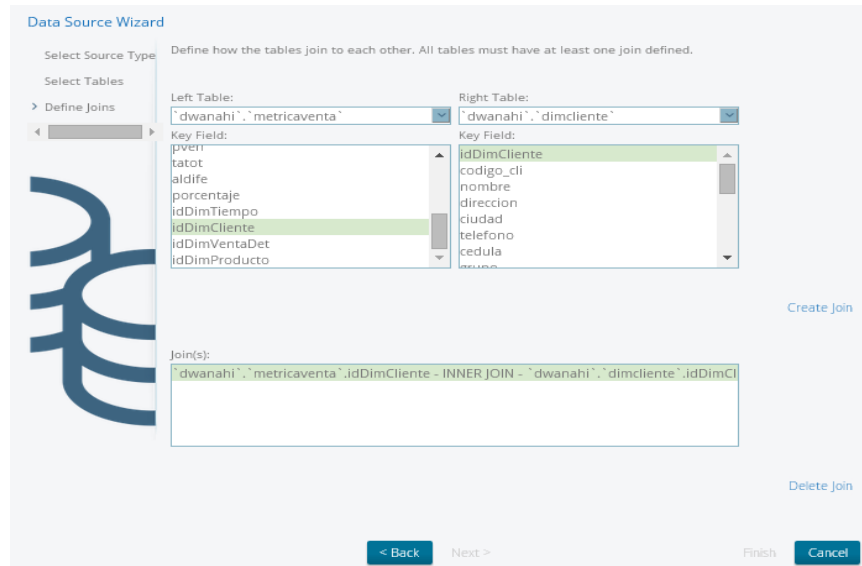


Figura 4.62: Creación de las relaciones.

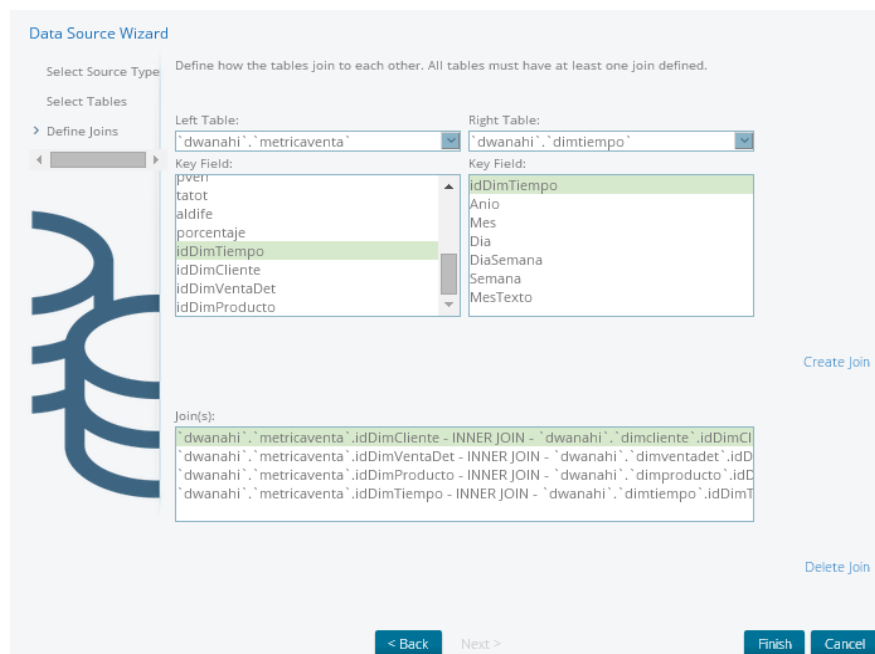


Figura 4.63: Relaciones entre la tabla de hecho y las dimensiones.

Seleccionar la opción <Customise Model Now> para personalizar el modelo.



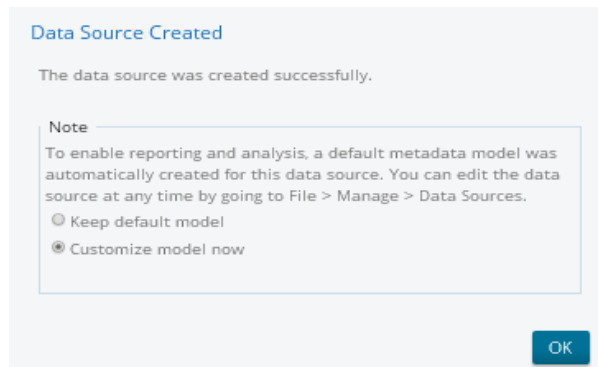


Figura 4.64: Personalizar el modelo.

Seleccionar la pestaña <Analysis> para proceder a personalizar las dimensiones, jerarquías y niveles.

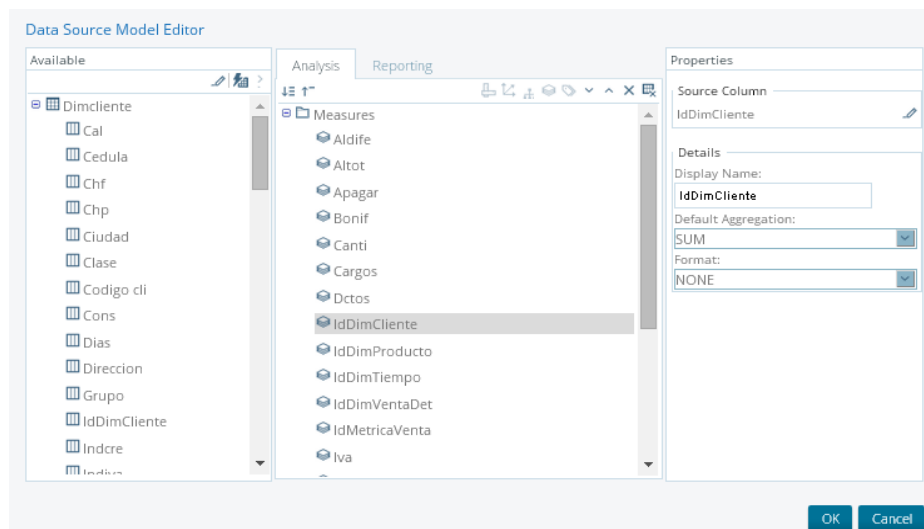




Figura 4.65: Diseño del Modelo de análisis.

Se debe crear la dimensión seleccionando la opción <  > y crear la jerarquía seleccionando la opción <  >.

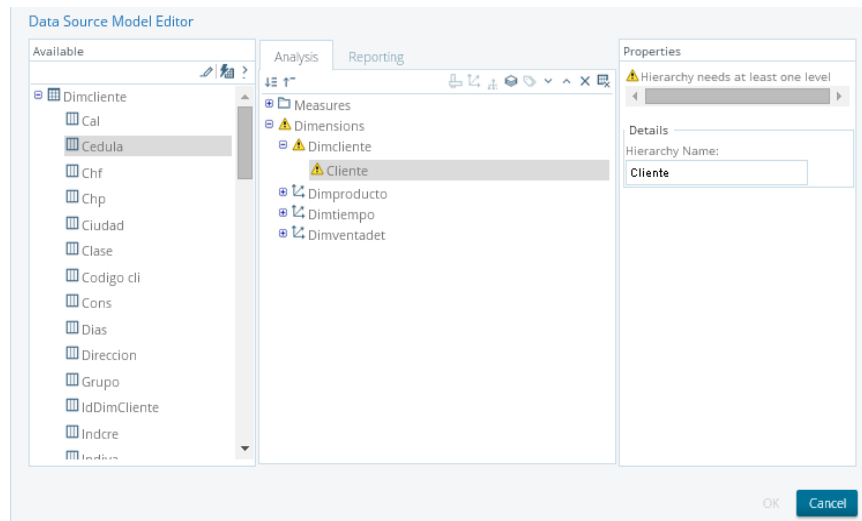



Figura 4.66 Creación de la Dimensión y Jerarquía.

Crear los niveles de cada jerarquía seleccionando la opción <  >.

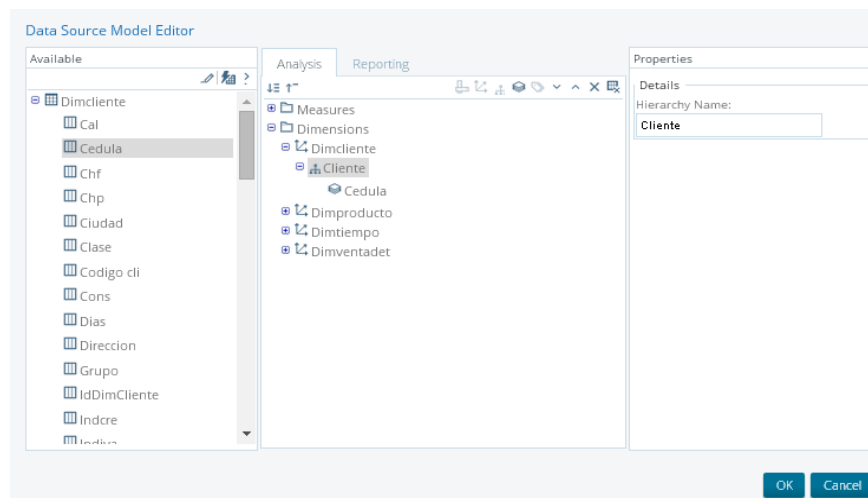


Figura 4.67: Creación del nivel.

Se debe realizar el mismo proceso para todas las dimensiones:

- Dimcliente
- Dimproducto
- Dimtiempo
- Dimventa

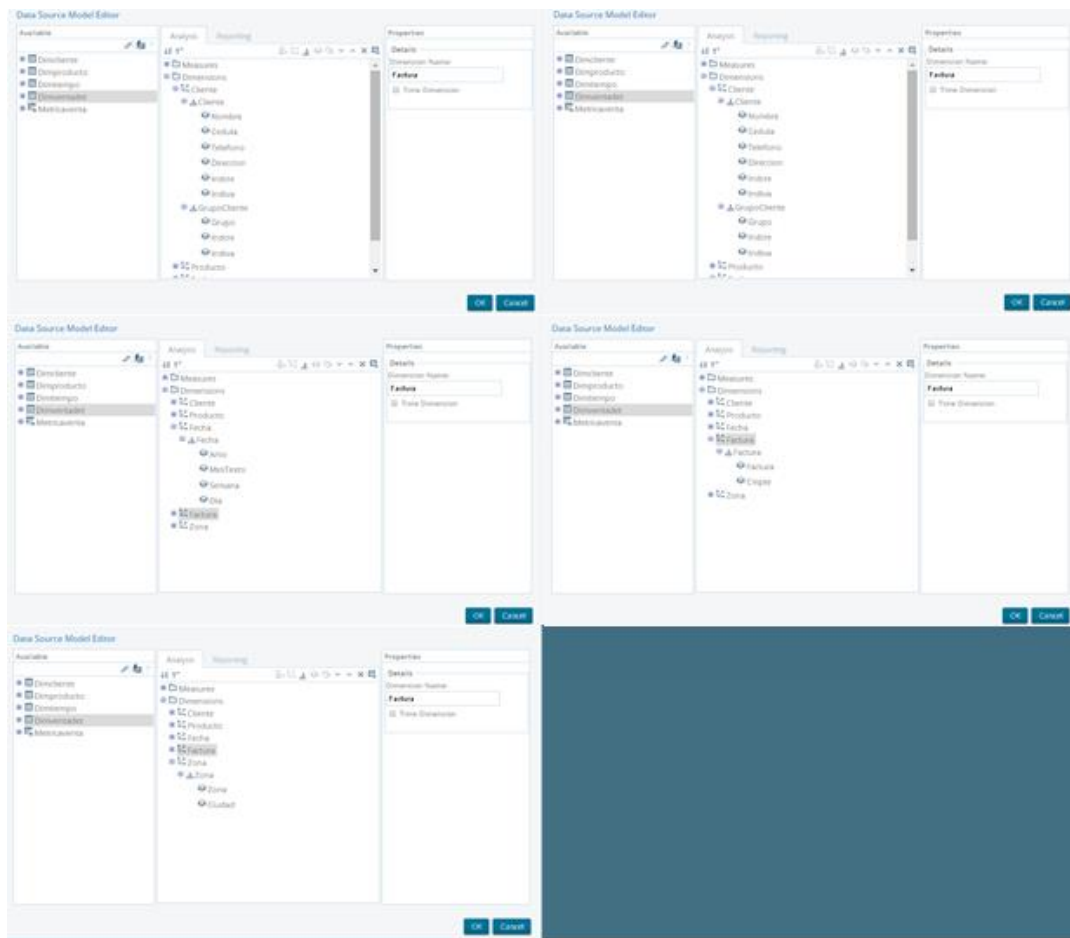


Figura 4.68: Resultado creación de dimensiones.

Crear las métricas y su agregación en la opción <Measures>, presionar la opción <Ok> una vez finalizado el modelo.

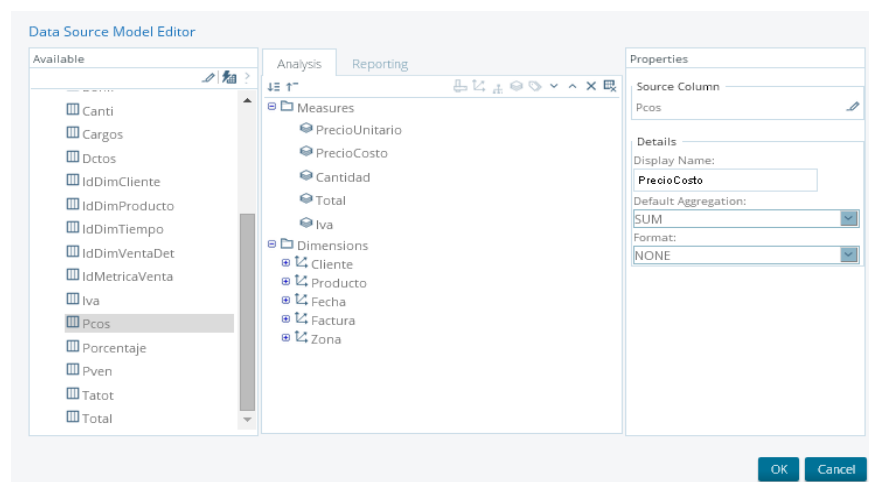


Figura 4.69: Creación de las métricas.

### 4.3.3 Creación de Herramientas de Análisis

En el menú principal de la plataforma Pentaho seleccionar la opción <New>, del submenú que se despliega seleccionar la opción <JPivot View>.

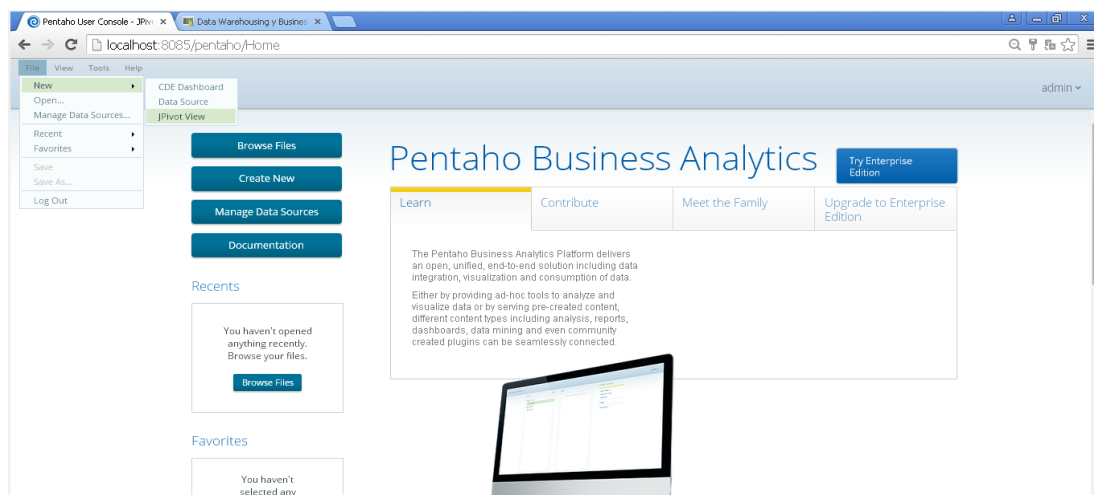


Figura 4.70: Creación de vista de análisis.

Seleccionar el esquema del que se realiza la vista de análisis y el cubo correspondiente a dicho esquema, una vez seleccionado el esquema y el cubo presionar la opción <Ok>.

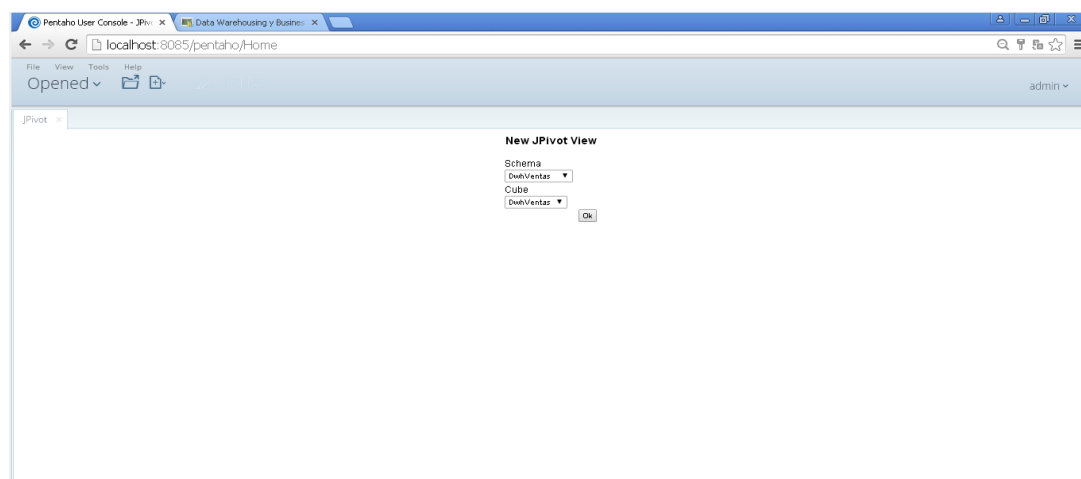


Figura 4.71: Selección del esquema y cubo.

Como resultado se tiene una vista de análisis del cubo de información, listo para el trabajo e interpretación.

Fecha	Cliente	Producto	Factura	Zona	Medidas		
					PrecioUnitario	Cantidad	Total
All Facturas	All Clientes	All Productos	All Facturas	All Zonas	35740	101229	7660739
2014	All Clientes	All Productos	All Facturas	All Zonas	26994	62489	5479062
Abril	All Clientes	All Productos	All Facturas	All Zonas	3791	9171	1012620
Agosto	All Clientes	All Productos	All Facturas	All Zonas	1700	3524	228302
Diciembre	All Clientes	All Productos	All Facturas	All Zonas	2273	4684	267981
	ALBARRAN AGUSTIN	All Productos	All Facturas	All Zonas	54	70	4160
		ACRILICO	All Facturas	All Zonas	54	70	4160
		MUJER	All Facturas	All Zonas	54	70	4160
		ABERTO BOTON N°40	All Facturas	All Zonas	25	40	2090
		PONCHOS	All Facturas	All Zonas	29	30	2090
		PAULINA CERRADO	All Facturas	All Zonas	14	10	1040
		RAYADO	All Facturas	All Zonas	14	20	1040
		FC000009482	All Facturas	All Zonas	14	20	1040
	ALFREDO CASTAÑEDA	All Productos	All Facturas	All Zonas	10	7	81
	ANDRES BELTRAN	All Productos	All Facturas	All Zonas	93	179	19473
	ANTONIO CANDI	All Productos	All Facturas	All Zonas	99	461	23543
	ANTONIO CASTAÑEDA	All Productos	All Facturas	All Zonas	101	47	2630

Figura 4.72: Resultado Vista de Análisis.

### 4.3.4 Creación de Tableros de Mando

Para la creación de un nuevo dashboard o tablero de mando, seleccionar en el menú *<File>*, *<New>*, *<CDEDashboard>*.

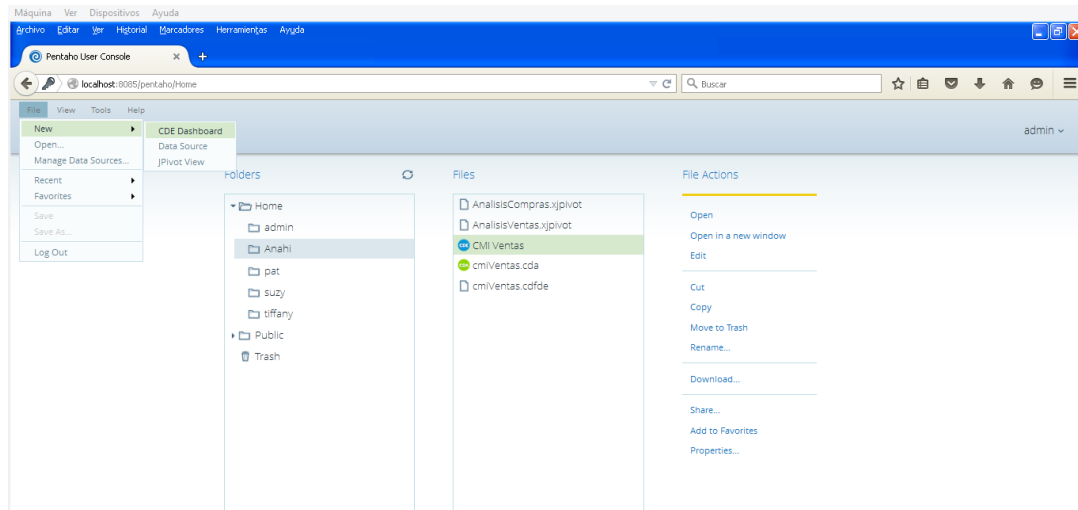


Figura 4.73: Creación nuevo tablero de mando.

El siguiente paso es la estructuración de Layout, en esta sección se define la distribución que tendrá el dashboard a nivel de sus campos. Para ello en el *Layout Panel* se crea nuevas filas y columnas según la estructura que se desee visualizar y se asigna un nombre a cada uno de los campos.

Los componentes principales a incluir son: Encabezado<Header>, Cuerpo <Body>, Pie <Foot>. Para el elemento seleccionado (Fila o columna) se muestra la tabla de propiedades, donde se puede configurar características como: largo y ancho, color, alineación del texto, entre otros.

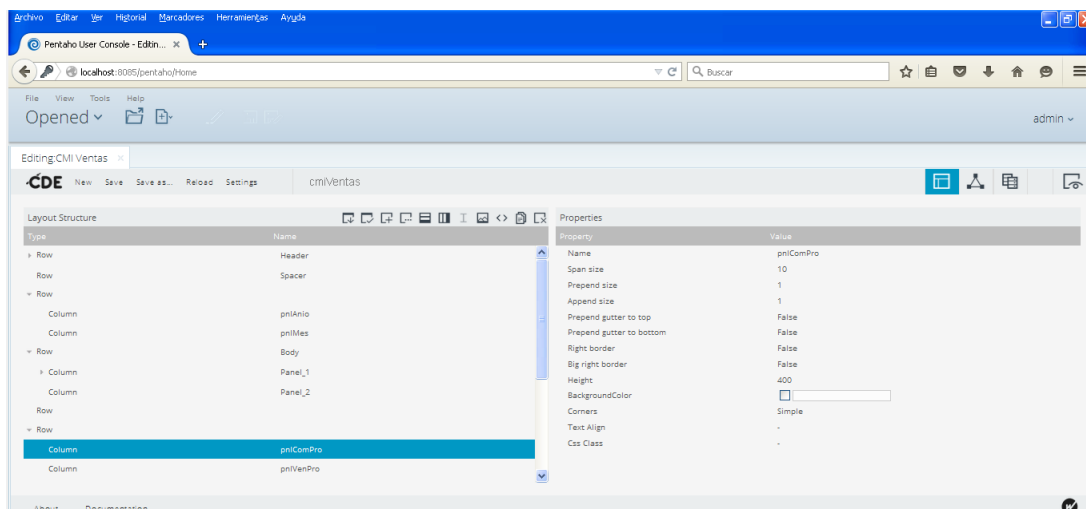


Figura 4.74: Estructura del Layout

Una vez estructurado el diseño del dashboard, se procede a incluir las fuentes de datos correspondientes; para ello se utiliza el panel *DataSources Panel*, ubicado en el lado derecho de la ventana.

En el menú principal, seleccionar <SQL Queries> y <SQL Over SQL Indi> para crear una nueva consulta SQL, parametrizable a los datos que se quiere mostrar.

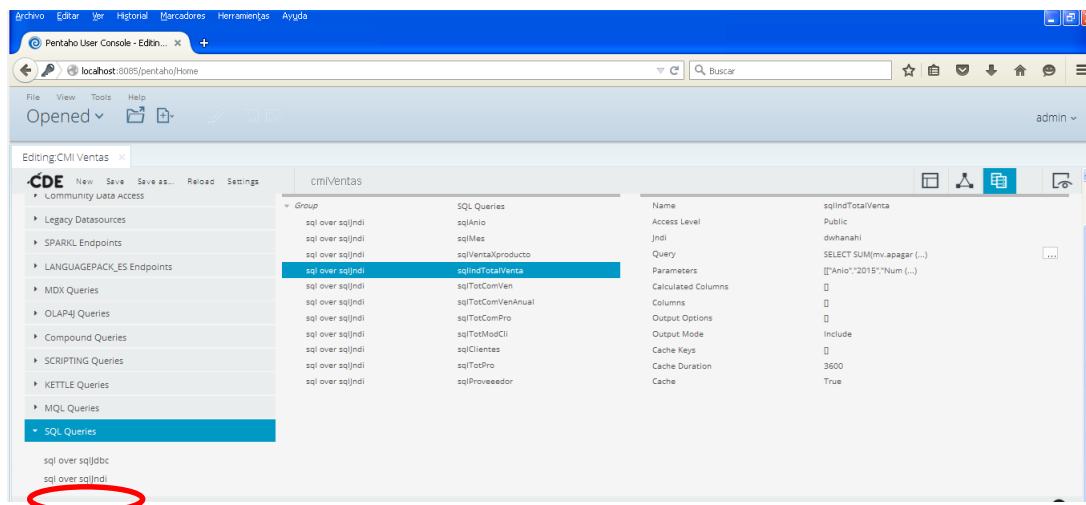


Figura 4.75: Consultas SQL para el dashboard.

La siguiente fase es la creación de componentes, aquí se crean los diferentes parámetros que van a ser utilizados:

- Parámetros Genéricos como Año, Mes, Cliente, Proveedor.
- Parámetros de Selección (Filtros).
- Parámetros de Tablas, donde se selecciona la forma visual de presentar los datos, por ejemplo los diagramas de barras.

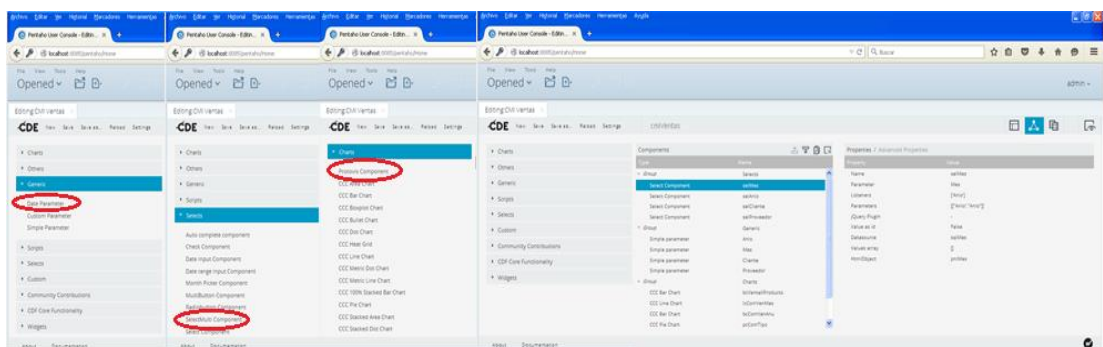


Figura 4.76: Creación de Componentes.

Antes de la publicación, es posible obtener una vista previa del dashboard, para verificar los resultados y su presentación.

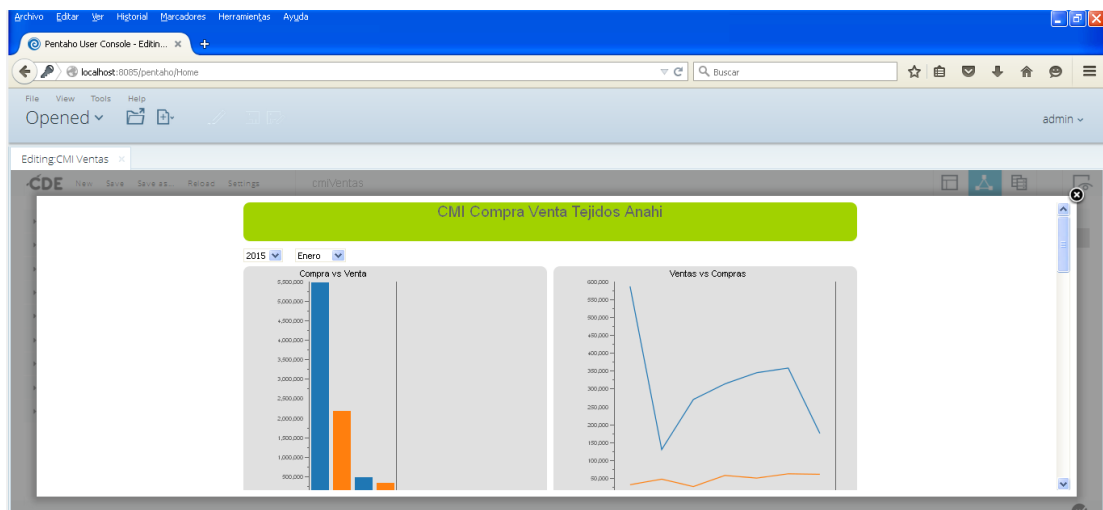


Figura 4.77: Vista Previa del Dashboard.

Luego de finalizar el diseño, se establece el directorio donde se almacenará y se lo guarda.

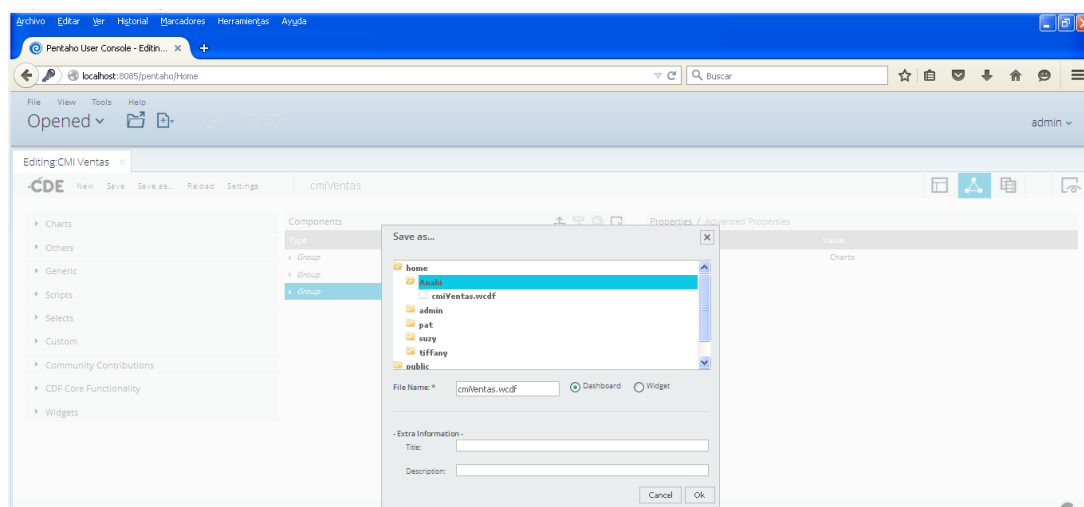


Figura 4.78: Directorio para almacenar el Dashboard.

Para visualizar el dashboard definitivo, se accede a la plataforma Pentaho, se busca el archivo en el directorio guardado, se lo abre dando un clic en la opción <Open in a new window>.

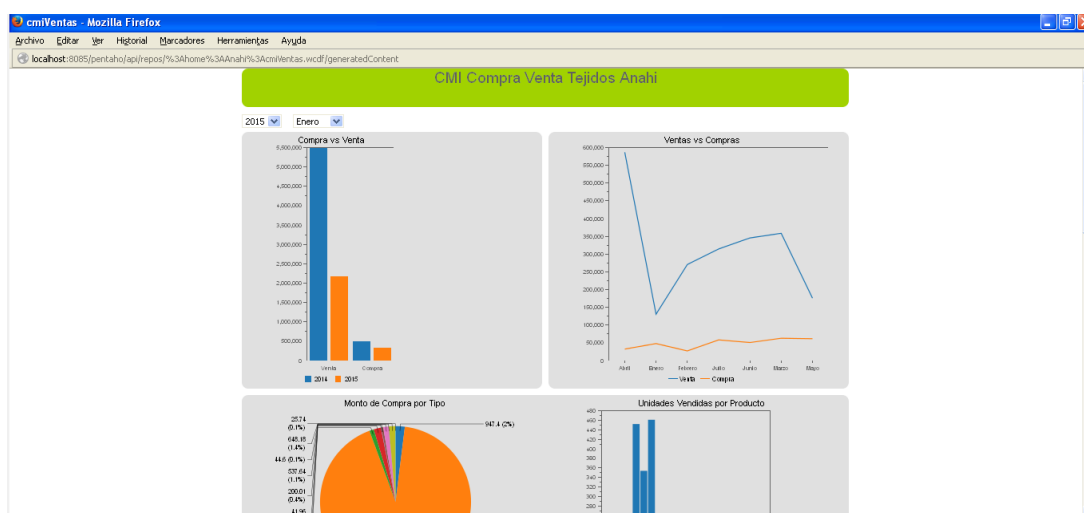


Figura 4.79: Dashboard Compra Venta Tejidos Anahi.

### 4.3.5 Creación de Reportes

En el directorio de Pentaho en la ruta <C:\Pentaho\repot-designer>, ejecutar el archivo batch <report-designer>.



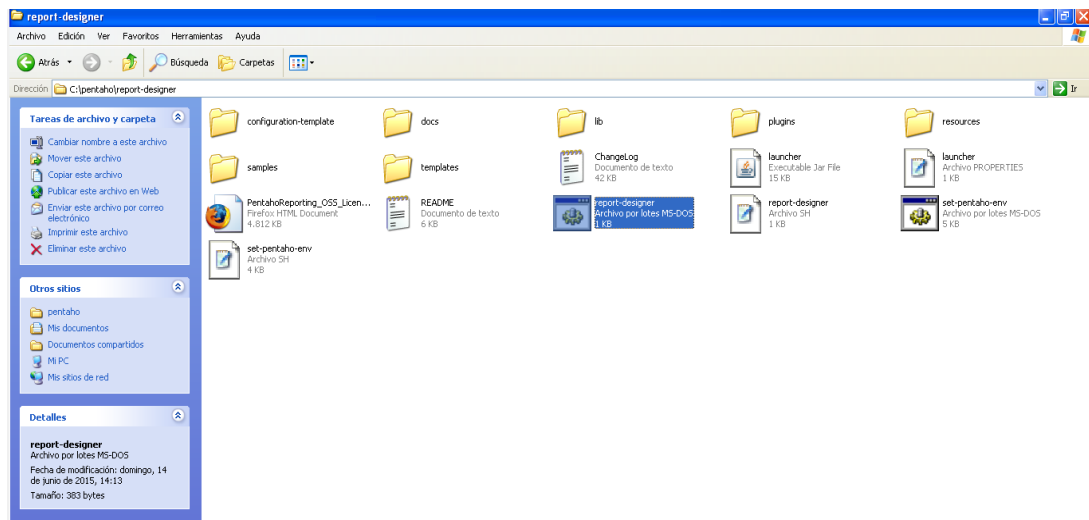


Figura 4.80: Directorio Report Designer.



Figura 4.81: Carga de la herramienta Report Designer.

En la pantalla de inicio seleccionar la opción <New Report>.

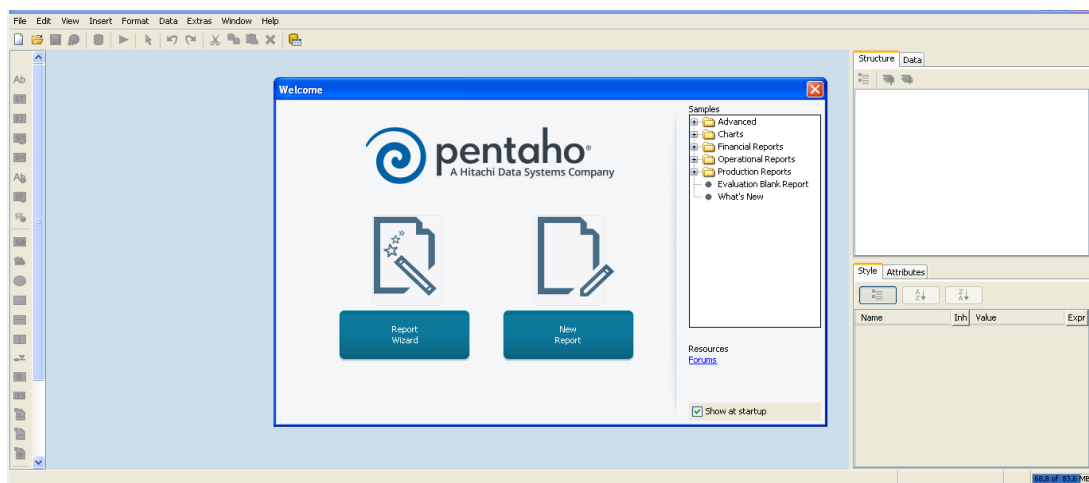


Figura 4.82: Pantalla de inicio de Report Designer

Para la creación de la fuente de información que va a manejar el reporte. En el menú superior seleccionar la opción <Datasource>.

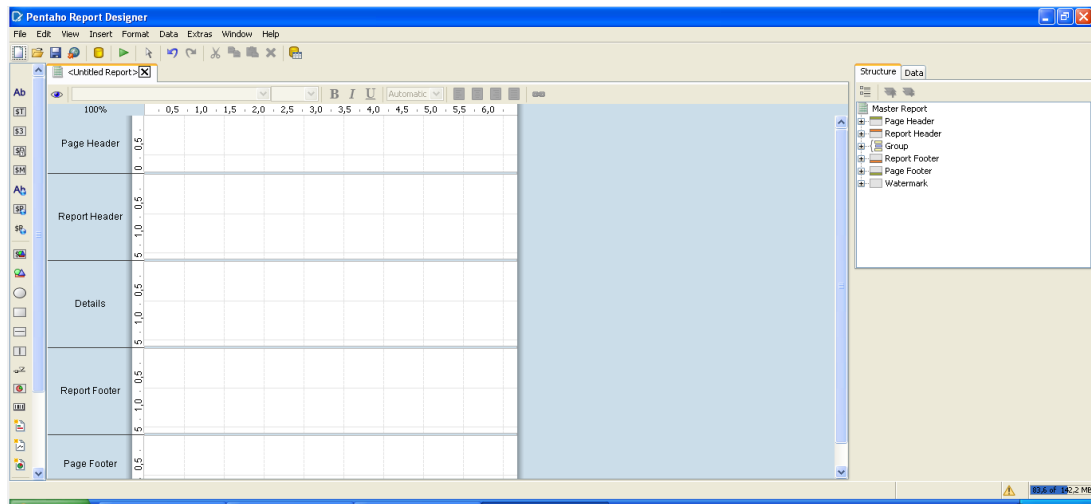


Figura 4.83: Creación del DataSource

Seleccionar la opción <+> para agregar una nueva fuente de datos.

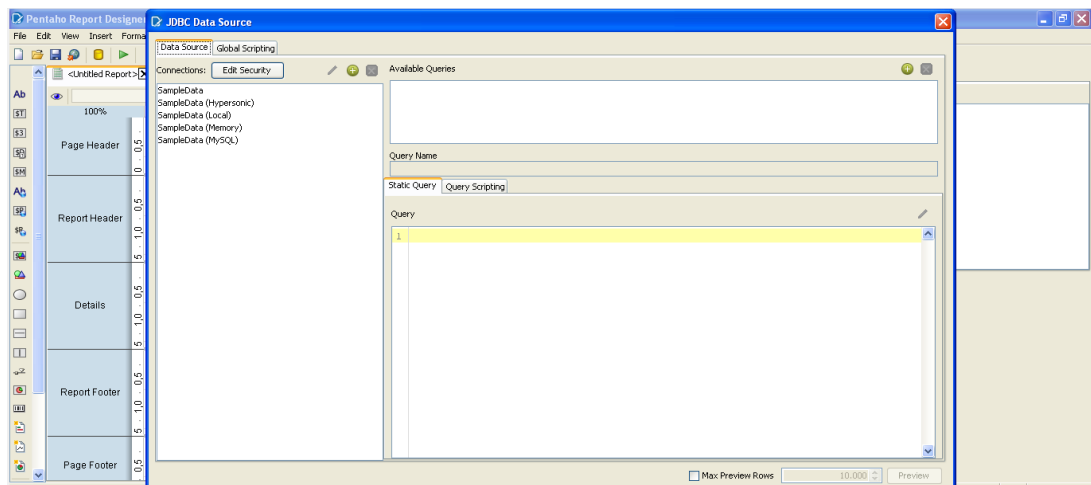


Figura 4.84: Creación nueva de fuente de datos.

Establecer el nombre y los parámetros de conexión a la base de datos que contiene el DW.

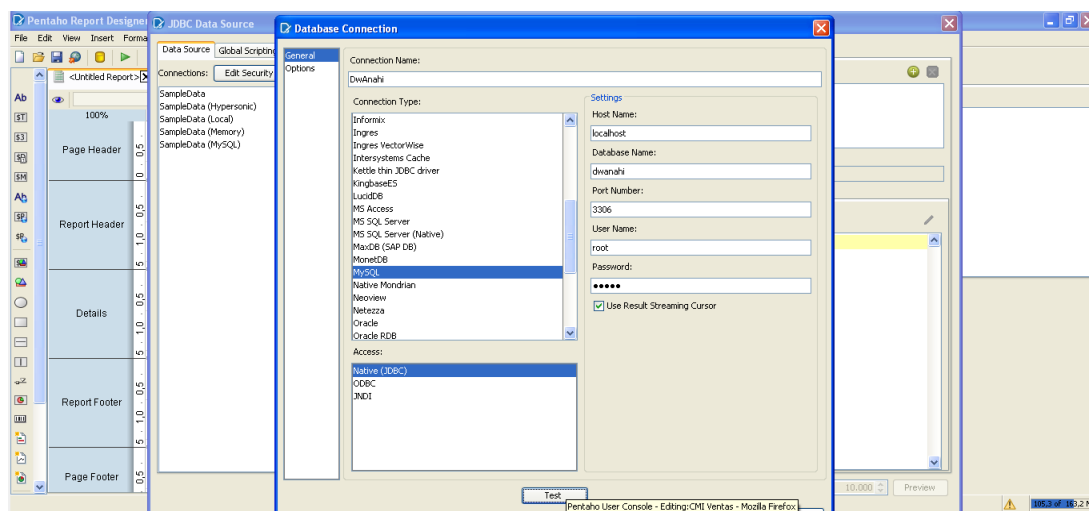


Figura 4.85: Configuración fuente de datos.

Una vez ingresados los parámetros de conexión, probar la misma presionando en la opción <Test>, si la prueba es satisfactoria presionar la opción <Ok>.

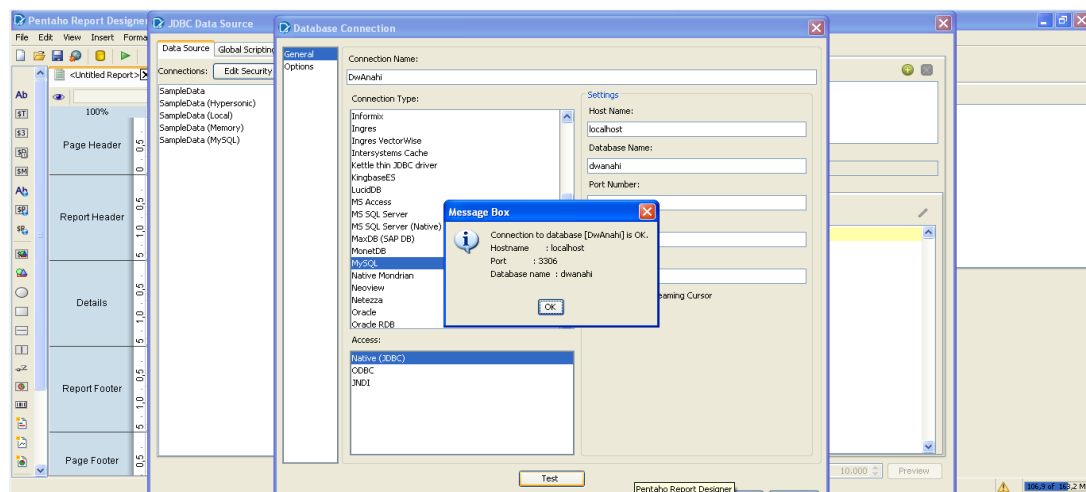


Figura 4.86: Resultado de la configuración fuente de datos.

Crear las sentencias SQL necesarias para obtener los campos de la información a presentarse en el reporte. Estas sentencias deben ser las de información descriptiva como a su vez las sentencias que se usara para los parámetros.

Seleccionar la opción <+> para agregar una instrucción SQL, establecer el nombre y la sentencia de la misma. Se debe tener en consideración que los parámetros van entre <\${Parametro}>.

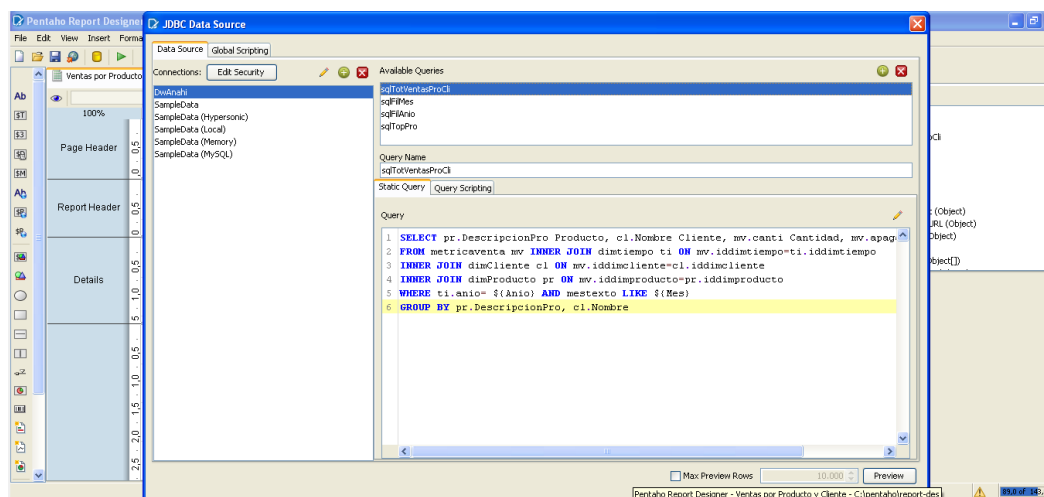


Figura 4.87: Sentencias SQL del reporte.

En la parte derecha seleccionar la pestaña <Data>, en esta sección se encuentra las sentencias anteriormente creadas, Para la elaboración de reporte basta con alar los campos generados en la sentencia SQL hacia el diseño del reporte en la sección de <Details>.

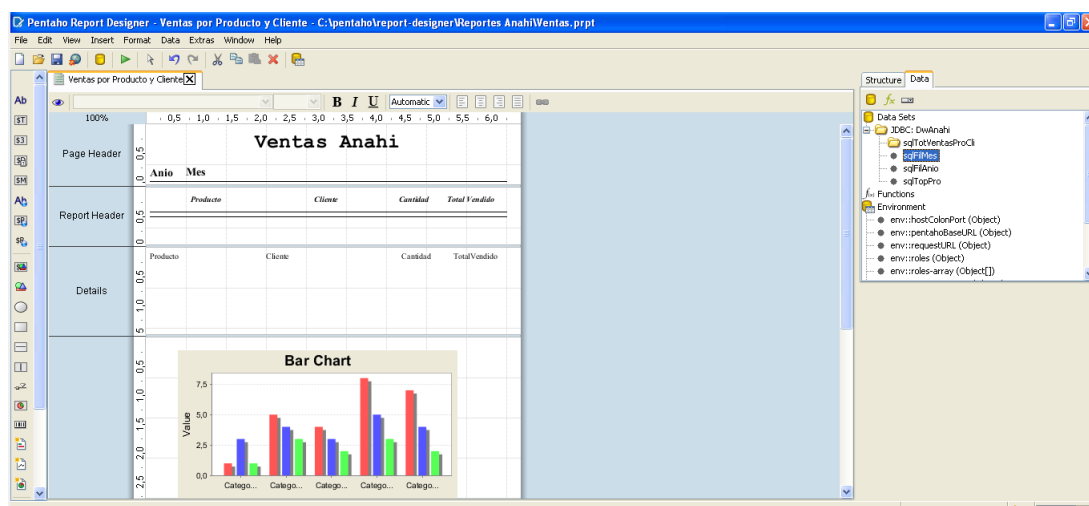


Figura 4.88: Diseño del reporte

Para la publicación del reporte en la plataforma Pentaho, en el menú principal seleccionar la opción publish < > y establecer los parámetros de conexión.

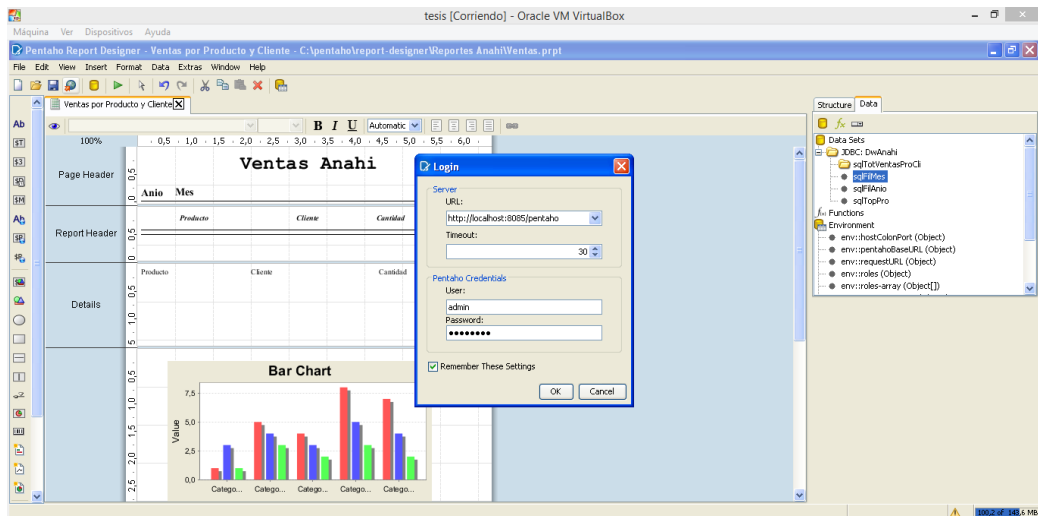


Figura 4.89: Publicación del reporte.

Es necesario establecer la ruta, nombre, título y el tipo de visualización del reporte.

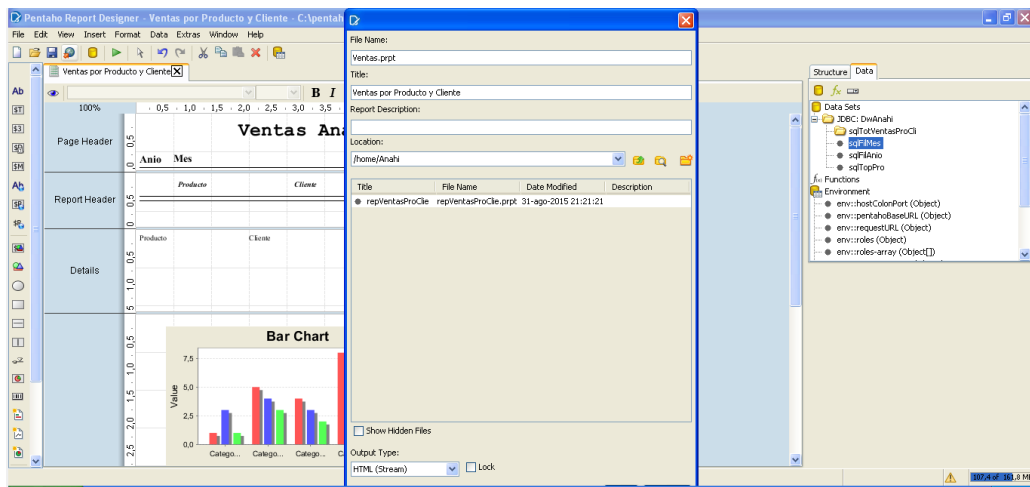


Figura 4.90: Nombre y ruta del reporte.

### 4.3.6 Creación de Usuarios

Ingresa en la plataforma Pentaho, Presionar la opción <File> en el menú principal, seleccionar la opción <Administration>.

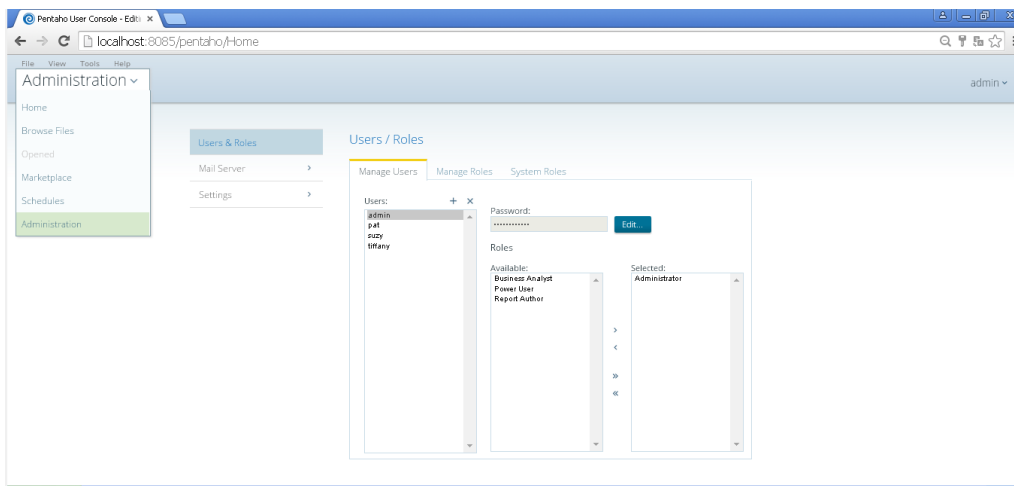


Figura 4.91: Creación de Usuarios.

Seleccionar la pestaña <Manager users> en la opción <Users / Roles>, presionar en opción <+> para la creación de un nuevo usuario. Ingresar el nombre de usuario y contraseña, al finalizar presionar <OK>.

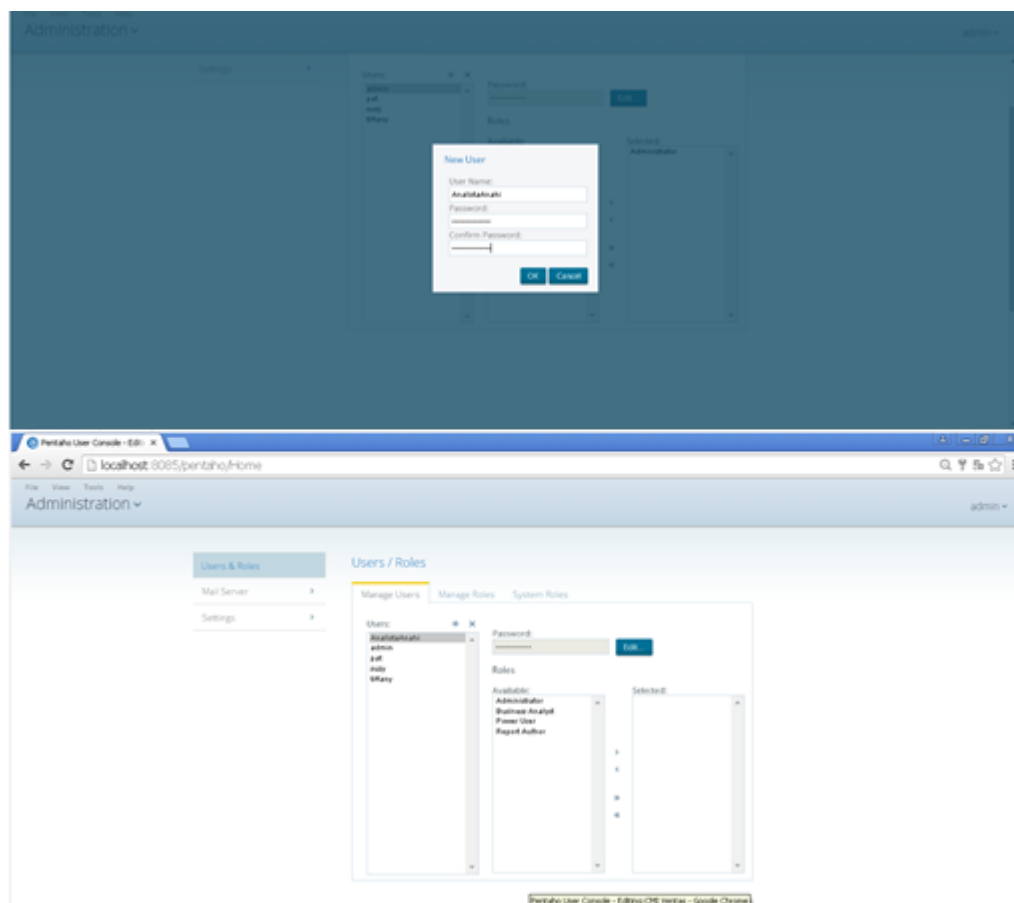


Figura 4.92: Resultado de la creación de usuarios.

## **CAPÍTULO 5.**

### **CONCLUSIONES Y RECOMENDACIONES**

#### **5.1 CONCLUSIONES**

La metodología Hefesto gracias a su fácil estructura y escalabilidad permite la implementación de una solución BI dentro de una PYME de forma correcta, rápida y estructurada.

Los datos extraídos de los procesos de compra y venta de la compañía permiten la identificación de los indicadores necesarios para el correcto diseño del datawarehouse y el análisis de la información, estos ayudarán a los procesos de toma de decisiones.

La correcta aplicación de la metodología Hefesto, permite la identificación de las fuentes de información, en este caso la base de datos de un ERP Financiero; también facilita el diseño de la estructura del Datawarehouse para los cubos de compra y venta, para finalmente implementarlo con las herramientas Pentaho Community.

Las herramientas Business Analytics Platform, Pentaho Report Designer y Community Dashboard Editor permiten el diseño de la navegabilidad, implementación de reportes y cuadros de mando respectivamente, para los cubos de Compra y Venta del datawarehouse.

#### **5.2 RECOMENDACIONES**

Se recomienda utilizar la metodología Hefesto para la implementación de soluciones BI dentro de una PYME por sus facilidades y rapidez en el proceso de diseño y construcción.

En el caso de las PYMES, se recomienda indagar minuciosamente los datos involucrados en cada proceso clave, para ampliar el margen de análisis de cada uno dentro de la compañía.

Una vez identificadas las fuentes de información, se recomienda seguir alimentando la base de datos para ampliar el contenido del datawarehouse e incrementar los elementos de análisis a todos los procesos de la compañía.

Se recomienda administrar periódicamente los niveles de acceso a los usuarios a las diferentes herramientas de visualización y explotación para mantener un acceso seguro a la información de la empresa.



## Bibliografía

- Bernabeu, R. (2010). HEFESTO DATA WAREHOUSING: Investigación y Sistematización de Conceptos – HEFESTO: Metodología para la Construcción de un Data Warehouse.
- BI-Business Intelligence Blog. (2009). Metadata. Recuperado de: <http://bi-businessintelligence.blogspot.com/2009/02/metadata.html>.
- Bouman, R., Dongen, V. (2009). Pentaho Solutions: Business Intelligence and Data Warehousing with Pentaho and MySQL.
- Cano, J. (2007). Business Intelligence: Competir con Información. Recuperado de [http://itemsweb.esade.edu/biblioteca/archivo/Business\\_Intelligence\\_competir\\_con\\_informacion.pdf](http://itemsweb.esade.edu/biblioteca/archivo/Business_Intelligence_competir_con_informacion.pdf).
- CEPAL/ONU. (2007). El desarrollo de las Pymes mediante el uso de tecnologías de la información y de las comunicaciones (TIC): las opciones para América Latina y el Caribe. Boletín FAL No. 254. Santiago de Chile. Recuperado de: <http://repositorio.cepal.org/handle/11362/36145>.
- Component One Studio. (2013). Working with OLAP Cubes. Recuperado de: <http://our.componentone.com/2014/11/13/working-with-olap-cubes/>.
- Data Runs the World. (2012). Advanved Visualizations. Recuperado de: <http://dataruntheworld.com/category/advanved-visualizations/>
- Durá, S. (2011). Sistemas de Información BI: Estado Actual y Herramientas de Software Libre. Recuperado de: [http://openaccess.uoc.edu/webapps/o2/bitstream/10609/8175/1/Sduras\\_TFM\\_0611.pdf](http://openaccess.uoc.edu/webapps/o2/bitstream/10609/8175/1/Sduras_TFM_0611.pdf).
- Eckerson, W. (2003). Evaluating ETL and Data Integration Platforms. Recuperado de [http://download.101com.com/tdwi/research\\_report/2003ETLReport.pdf](http://download.101com.com/tdwi/research_report/2003ETLReport.pdf)
- Espinosa, R. (2010). El Rincón del BI: Dashboard. Recuperado de: <https://churriwifi.wordpress.com/2010/02/23/14-8-2-dashboard-y-cuadros-de-mando-en-microstrategy-9-utilizando-documentos-de-report-services-ii/>
- Espinosa, R. (2010). El Rincón del BI: ETL. Recuperado de: <http://churriwifi.wordpress.com/category/etl/>
- Eckerson, W. (2005). Enterprise Business Intelligence: Strategies and Technologies for Deploying BI on an Enterprise Scale.
- ETL-Tools.Info. (2010-2015). Business Intelligence – Almacén de Datos – ETL. Recuperado de: [http://etl-tools.info/es/bi/almacendedatos\\_arquitectura.htm](http://etl-tools.info/es/bi/almacendedatos_arquitectura.htm).

Ferraro, C., Stumpo, G., Compiladores. (2010). Políticas de Apoyo a las pymes en América Latina: Entre avances innovadores y desafíos institucionales.

Guzmán, E. (2010). Inteligencia de Negocios. Recuperado de <http://dis.unal.edu.co/profesores/eleon/cursos/tebd/presentaciones/IntroBI.pdf>

Inmon, W. (1992). Building the datawarehouse.

Intellego. (2015). Business Intelligence. Recuperado de <http://www.grupointellego.com/business-intelligence/>

Open Business Intelligence. (2015). Analisis View de Pentaho. Recuperado de: <http://www.redopenbi.com/>

Paz, J. (2010). La Importancia de la Inteligencia de Negocios Aplicada a Empresas Medianas. Recuperado de <http://www.ibm.com/developerworks/ssa/local/data/dm-bi-pymes/#authorN1001A>.

Sinnexus. (2007-2012). Qué es Business Intelligence?. Recuperado de [http://www.sinnexus.com/business\\_intelligence](http://www.sinnexus.com/business_intelligence)

Sinnexus. (2007-2015). Datawarehouse. Recuperado de: [http://www.sinnexus.com/business\\_intelligence/datawarehouse.aspx](http://www.sinnexus.com/business_intelligence/datawarehouse.aspx).

Sivsa. (2015). Pentaho Data Integration. Recuperado de: <https://www.sivsa.com/site/es/productos-y-servicios/soluciones/business-intelligence/item/23?start=4>

StrateBI. (2011). STDashboard. Recuperado de: <http://www.stratebi.com/stdashboard>

## Glosario de Términos

- BI:** Business Intelligence.
- CDE:** Community Dashboard Editor.
- CRM:** Customer Relationship Management.
- DASHBOARD:** Tablero de control.
- DW:** Datawarehouse.
- ERP:** Enterprise Resource Planning.
- ETL:** Extract Transformation Load (Extracción, Transformación y Carga).
- FK:** Foreign Key.
- HOLAP:** Hybrid On Line Analytical Processing.
- JDBC:** Java Database Connectivity.
- JDNI:** Java Naming and Directory Interface.
- KPI:** Key Performance Indicator.
- MOLAP:** Multidimensional On Line Analytical Processing.
- OLAP:** On Line Analytical Processing.
- OLTP:** On Line Transaction Processing.
- PDI:** Pentaho Data Integration.
- PK:** Primary Key.
- PSW:** Pentaho Schema Workbench.
- PRD:** Pentaho Report Designer.
- RDBMS:** Sistemas de Gestión de Bases de Datos Relacionales.
- ROLAP:** On Line Analytical Processing Relational.
- SQL:** Structured Query Language
- TIC:** Tecnologías de la Información y la Comunicación.
- UTIC:** Unidad de Tecnología de Información y Comunicación.
- XML:** Extensible Markup Language.

## Curriculum Vitae

Alex Patricio Dávila Montúfar  
**Ingeniería en Sistemas**  
Ph. 0992659552  
alexdavila90@hotmail.com

### INFORMACION PERSONAL

**Fecha de Nacimiento:** 22 de Enero de 1990  
**Estado Civil:** Divorciado  
**Edad:** 25 años  
**Nacionalidad:** Ecuatoriano  
**Cédula de Ciudadanía:** 100295852-6  
**Dirección:** La Rábida N25-47 y Av. Colón  
Quito, Ecuador.



### OBJETIVO

Trabajar para mi desarrollo personal, profesional y el progreso de nuestra sociedad; además, proporcionar beneficios que nos incentiven y conlleven a una mutua superación para obtener un futuro mejor.

### ESTUDIOS REALIZADOS

**SECUNDARIA:** Colegio Fisco-Misional “San Francisco”  
Bachiller en Físico-Matemáticas  
**SUPERIOR:** Escuela Politécnica del Ejército  
Ingeniero en Sistemas.

### EXPERIENCIA LABORAL

**Empresa:** Sermatick Desarrollo de Sistemas  
**Área:** Departamento de Sistemas  
**Cargo:** Coordinador Tecnologías de la Información.  
**Jefe Encargado:** **Nombre:** Ing. Francisco Lara  
Gerente-Propietario.  
**Teléfono:** 0992 948 585  
**Duración:** 2 años

**Empresa:** Negrete In.G Ingeniería Global  
**Área:** Departamento de Sistemas  
**Cargo:** Coordinador Tecnologías de la Información  
**Jefe Encargado:** **Nombre:** Ing. Alberto Negrete Salas  
Gerente-Propietario  
**Teléfono:** 022 454 228  
**Duración:** 1 año 3 meses

**Empresa:** Actuaría Consultores Cía. Ltda.  
**Área:** Departamento de Sistemas  
**Cargo:** Analista de Sistemas  
**Jefe Encargado:** **Nombre:** Dr. Rodrigo Ibarra  
Presidente Ejecutivo  
**Teléfono:** 022 501 001  
**Duración:** 8 meses

#### REFERENCIAS PERSONALES

- Ing. Cecilia Montúfar  
Docente Universidad Técnica del Norte  
0991800156
- Ing. Francisco Lara  
Gerente Propietario de Sermatick Desarrollo de Sistemas  
0992948585
- Ing. Fernando Valenzuela  
Director de la Escuela de Contabilidad y Auditoría de la Universidad Técnica del Norte.  
0999531830
- Ing. Santiago Aulestia  
Director de Proyectos en Amaenergy Service S.A.  
0984435045

Gonzalo Wladimir Villares Dávila

## Ingeniería en Sistemas

Ph. 0969 628 363

gwillares@gmail.com

### INFORMACION PERSONAL

**Fecha de Nacimiento:** 2 de Agosto de 1989  
**Estado Civil:** Soltero  
**Edad:** 26 años  
**Nacionalidad:** Ecuatoriano  
**Cédula de Ciudadanía:** 172097376-5  
**Dirección:** Valle de los Chillos, Av Ilalo  
 Oe8-327. Quito, Ecuador.



### OBJETIVO

Desarrollarme como persona y profesional. Manifestando que soy sumamente emprendedor, investigador, responsable y honesto, de gran iniciativa y con muchos deseos de superación, me gustan los retos y poseo facilidad para desenvolverme en cualquier ambiente de trabajo.

### ESTUDIOS REALIZADOS

**SECUNDARIA:** Instituto Técnico Vida Nueva  
 Bachiller Técnico en Informática.  
**SUPERIOR:** Escuela Politécnica del Ejército  
 Ingeniero en Sistemas.

### PRACTICAS EMPRESARIALES

**Empresa:** Claro  
**Área:** Departamento de Sistemas.  
**Jefe Encargado:** **Nombre:** Ing. Omar Gómez  
**Teléfono:** 099449902  
**Fecha:** 1 de Diciembre del 2011 - 29 de Septiembre del 2012

**Empresa:** Banco General Rumiñahui  
**Área:** Departamento de Inteligencia de Negocios.

**Jefe Encargado:** **Nombre:** Diego Rolando Maldonado Guerrero  
Subgerente de Inteligencia de Negocios  
**Teléfono:** (593-2) 2509929, ext 4503

**Fecha:** 10 de Septiembre del 2012 - 07 de Marzo del 2013

**Empresa:** NOUX C.A  
**Área:** Inteligencia de Negocios  
**Cargo:** Especialista de EIM

**Jefe Encargado:** **Nombre:** María Fernanda Jaramillo.  
Sub Gerente de EIM  
**Teléfono:** (593-9)95027032

**Fecha:** 12 de Marzo del 2013 - 8 de Octubre del 2013

**Empresa:** TATA CONSULTING SERVICES  
**Área:** Inteligencia de Negocios  
**Cargo:** Analista IT

**Jefe Encargado:** **Nombre:** David Vivanco.  
Sub Gerente de Inteligencia de Negocios  
**Teléfono:** (593-9)984551854

**Fecha:** 14 de Octubre del 2013 - 28 Febrero del 2014

**Empresa:** DWCONSULWARE  
**Área:** Inteligencia de Negocios  
**Cargo:** Consultor BI

**Jefe Encargado:** **Nombre:** Marcia Marcalla.  
Gerente de Proyectos  
**Teléfono:** (593-9)96275514

**Fecha:** 14 de Octubre del 2014 - Presente

## REFERENCIAS PERSONALES

Ing. Magaly Villares  
Asistente de Gerencia Financiera  
DANEC  
Teléfono: 098 906541

Ing. Washington Morejón  
Gerente General  
Proveedor de Internet Airtel  
Teléfono: 095 494373

Ing. Lorena Llumiquinga  
Analista IT  
TATA TCS Ecuador  
Teléfono: 0998142809