

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

FACULTAD DE CIENCIAS E INGENIERÍA



**PONTIFICIA
UNIVERSIDAD
CATÓLICA
DEL PERÚ**

**ANÁLISIS DISEÑO E IMPLEMENTACIÓN DE UNA SOLUCIÓN
DE INTELIGENCIA DE NEGOCIOS PARA EL ÁREA DE
COMPRAS Y VENTAS DE UNA EMPRESA
COMERCIALIZADORA DE ELECTRODOMÉSTICOS**

Tesis para optar por el Título de Ingeniero Informático, que presentan las
bachilleras:

Keller Gladys Rodríguez Cabanillas

Angela Lucia Mendoza Peña

ASESORA: Ing. Carla Shirley Basurto Figueroa

Lima, marzo del 2011

Resumen

Las pequeñas y medianas empresas comercializadoras de electrodomésticos crecen en el mercado peruano generando ingresos y empleo. El rápido avance de la tecnología permite a más familias acceder a productos que faciliten su trabajo diario en el hogar y en el trabajo. Esto obliga a dichas empresas a volverse más competitivas en cuanto a precios, promociones, publicidad, tecnología, infraestructura y recursos humanos. Las actividades principales de este tipo de empresas comercializadoras son la compra de electrodomésticos y negociación con los proveedores, así como la venta dirigida y el servicio brindado a sus clientes.

Para volverse más competitivas muchas empresas de este rubro toman decisiones a base de la experiencia y resultados anteriores.

Debido a que estas decisiones generalmente no se toman de manera estructurada, se plantea como solución el uso de una herramienta de inteligencia de negocios que permita en tiempo real a los gerentes y jefes de producto generar escenarios,

pronósticos y reportes que apoyen a la toma de decisiones en la compra y venta de electrodomésticos. El uso de esta herramienta se traduce en una ventaja competitiva y son muchas las empresas que se han beneficiado por la implementación de un sistema de inteligencia de negocios, además se pronostica que con el tiempo se convertirá en una necesidad de toda empresa. [Vit 2002]

Como solución de Inteligencia de Negocios se diseña un *Data Mart* de Compras y un *Data Mart* de Ventas, luego se realizan los procesos de extracción, transformación y carga de datos, para finalmente explotar los datos mediante reportes que permitan hacer el análisis de la información.

El proceso de extracción, transformación y carga (ETL) permite mover datos de diferentes fuentes, transformarlos y cargarlos a los *Data Marts*. El proceso de Explotación permite generar los reportes que el usuario final usa para el análisis de la información y para la toma de decisiones.

Se decide usar software libre para el desarrollo de la solución y se elige como herramienta la plataforma de Pentaho, la cual está escrita en Java y presenta una solución flexible para cubrir las necesidades de la empresa. Pentaho al ser una herramienta de software libre es accesible económicamente a las empresas comercializadoras de electrodomésticos, brindando así una ventaja diferencial frente a otras herramientas de inteligencia de negocios de precio elevado.

Pentaho permite la creación de cubos, la creación de informes e implementación de la plataforma BI (web) lo cual genera un nexo amigable entre la herramienta y los usuarios finales.

Dedicatorias

“A mis padres, por darme una carrera y apoyarme de forma incondicional a cumplir mis metas. A mi esposo y mi hija, quienes se han convertido en mi fuente de motivación para ser cada día mejor”

-Angela Mendoza Peña

“Dedico este esfuerzo a mis padres y familia por su amor incondicional y por su constante apoyo para hacer realidad cada proyecto en mi vida, y a Dios fuente de toda mi energía.”

-Keller Rodríguez Cabanillas.

Agradecimientos

Gracias a todas las personas que han hecho posible este proyecto. A la empresa comercializadora de electrodomésticos por la confianza depositada. A nuestra asesora por su permanente orientación y apoyo. A nuestros compañeros por compartir sus conocimientos y experiencias con nosotras. A nuestros familiares por animarnos a seguir adelante. A Dios por hacer posible la culminación de esta obra. Este trabajo es para todos ustedes.

Índice General

Resumen	1
Índice General	3
Índice de figuras	7
Índice de cuadros	9
1. Marco Conceptual.....	10
1.1. Definición del Problema.....	10
1.2. Conceptos Relacionados.....	12
1.3. Plan de Proyecto	20
1.4. Estado del Arte.....	23
1.5. Descripción y sustentación de la solución	31
1.5.1. Descripción actual de una empresa comercializadora de electrodomésticos.....	31
1.5.2. Sustentación de la solución	35
2. Análisis	39
2.1. Metodología.....	39
2.1.1. Marco Conceptual	41
2.1.2. Análisis.....	42
2.1.3. Diseño	43
2.1.4. Construcción	45
2.1.5. Implantación	48
2.2. Requerimientos funcionales para el área de compras	49

2.3.	Requerimientos no funcionales para el área de compras	51
2.4.	Requerimientos funcionales para el área de ventas	52
2.5.	Requerimientos no funcionales para el área de ventas	54
2.6.	Plan de Pruebas.....	55
2.6.1.	Plan de Pruebas para el área de compras.....	56
2.6.2.	Reporte de <i>slow moving</i>	57
2.6.3.	Reporte de stock de productos	58
2.6.4.	Reporte de compras a realizar	59
2.7.	Plan de Pruebas para el área de ventas.....	59
2.7.1.	Reporte de ventas realizadas	60
2.7.2.	Reporte de Precio de Venta al Público (PVP).....	60
2.7.3.	Reporte de Margen teórico vs. Margen real.....	61
2.7.4.	Reporte de ventas a 30 días.....	62
2.6.	Software a utilizar	63
2.7.5.	Modelador de datos.....	63
2.7.6.	Gestor de Base de Datos	64
2.7.7.	Plataforma BI.....	65
3.	Diseño	70
3.1.	Modelo Dimensional	70
3.1.1.	Diagrama del Modelo Dimensional	73
3.1.2.	Dimensiones.....	74
3.1.3.	Tablas de Hechos	76
3.1.4.	Estándares del Modelo – Nomenclatura	76

3.2.	Arquitectura	77
3.3.	Diseño de Extracción.....	81
3.3.1.	Carga de la Dimensión Producto	81
3.3.2.	Descripción de Tablas Fuentes	81
3.3.3.	Estandarización de Datos y Limpieza de Datos.....	82
3.3.4.	Fuente de Datos.....	83
3.3.5.	Tabla Destino	83
3.3.6.	Proceso	84
3.3.7.	Esquema de Extracción.....	85
3.4.	Diseño de explotación para el área de compras	85
3.4.1.	Reporte “Compras Realizadas”	86
3.5.	Diseño de explotación para el área de ventas	89
3.5.1.	Reporte “Ventas Realizadas”	89
4.	Construcción y Pruebas.....	92
4.1.	Construcción	92
4.1.1.	Instalación y configuración del software	92
4.1.2.	Construcción de Procesos de Carga	97
4.1.3.	Construcción de Reportes	99
4.2.	Pruebas.....	104
4.2.1.	Pruebas de Ejecución de Procesos de Carga.....	104
4.2.2.	Pruebas de Reportes.....	107
5.	Observaciones, Conclusiones y Recomendaciones.....	109
5.1.	Observaciones	109

5.2.	Conclusiones.....	110
5.3.	Recomendaciones.....	111
	Bibliografía	113



Índice de figuras

- Figura 1.1** Diagrama de WBS
- Figura 1.2** Diagrama de Gantt
- Figura 1.3** Demanda electrodomésticos en mercado nacional
- Figura 2.1** Metodología
- Figura 2.2** Arquitectura funcional de Pentaho
- Figura 2.3** Funcionamiento de Mondrian
- Figura 2.4** Ejemplos de reportes avanzados
- Figura 3.1** Diagrama del modelo dimensional
- Figura 3.2** Arquitectura conceptual básica
- Figura 3.3** Diseño de reporte de compras realizadas
- Figura 3.3** Diseño de reporte de ventas realizadas
- Figura 4.1** Configuración de PostgreSQL
- Figura 4.2** Instalación de Pentaho 1
- Figura 4.3** Instalación de Pentaho 2
- Figura 4.4** Configuración Spoon
- Figura 4.5** Ejecución de Cube Designer
- Figura 4.6** Diseño proceso de carga
- Figura 4.7** Proceso de creación de reportes
- Figura 4.8** Proceso de creación de reportes – Paso 1
- Figura 4.9** Proceso de creación de reportes – Paso 2
- Figura 4.10** Proceso de creación de reportes – Paso 3
- Figura 4.11** Proceso de creación de reportes – Paso 4
- Figura 4.12** Proceso de creación de reportes – Paso 5
- Figura 4.13** Proceso de creación de reportes
- Figura 4.14** Proceso de carga satisfactoria

Figura 4.15 Datos de OLTP tableproducto

Figura 4.16 Reporte de ventas realizadas

Figura 4.17 Reporte de ventas realizadas - tipo gráfico



Índice de cuadros

- Cuadro 1.1** Cuadro comparativo de herramientas de extracción
- Cuadro 1.2** Cuadro comparativo de herramientas de explotación
- Cuadro 1.3** Cuadro comparativo Kimball vs Inmon
- Cuadro 2.1** Cuadro de requerimientos funcionales para el área de compras
- Cuadro 2.2** Cuadro de requerimientos no funcionales para el área de compras
- Cuadro 2.3** Cuadro de requerimientos funcionales para el área de ventas
- Cuadro 2.4** Cuadro de requerimientos no funcionales para el área de ventas
- Cuadro 2.5** Resultado esperado para el reporte de compras realizadas
- Cuadro 2.6** Resultado esperado para el reporte de slow moving
- Cuadro 2.7** Resultado esperado para el reporte de stock de productos
- Cuadro 2.8** Resultado esperado para el reporte de compras a realizar
- Cuadro 2.9** Resultado esperado para el reporte de ventas realizadas
- Cuadro 2.10** Resultado esperado para el reporte de PVP
- Cuadro 2.11** Resultado esperado para el reporte de margen
- Cuadro 2.12** Resultado esperado para el reporte de venta a 30 días
- Cuadro 3.1** Cuadro comparativo modelo estrella vs modelo copo de nieve
- Cuadro 3.2** Cuadro comparativo arquitectura MOLAP vs arquitectura ROLAP

1. Marco Conceptual

Este primer capítulo describe el tema de tesis y el ambiente en que se desarrolla. Define el problema y la mejor solución de soporte de decisiones encontrada. También muestra los conceptos relacionados al trabajo de tesis.

1.1. Definición del Problema

Una empresa comercializadora de electrodomésticos tiene como funciones principales la compra mayorista y venta minorista de productos, siendo principalmente su margen de ganancia la diferencia entre el costo del producto y el precio de venta al público. Los productos que ofrece una empresa de este tipo varían constantemente en el tiempo, ya que se basa en tecnología y tendencias.

El área de compras es importante en este tipo de empresas por la negociación que involucra una compra mayorista y dado que aquí se empieza a definir el margen de ganancia. Así también el área de ventas es importante ya que se encarga de finalizar el proceso que hace efectivo el negocio. Las decisiones que se toman actualmente en estas áreas generalmente son en base a la experiencia y a los datos de las compras y ventas realizadas diariamente.

Los principales procesos identificados para el área de compras son:

- Decidir los productos que se van a comercializar en base a las ventas obtenidas en meses anteriores considerando la rotación de los productos.
- Negociar con los proveedores el precio de los productos en función a un margen de ganancia mínimo definido previamente considerando precios y costos de experiencias anteriores.
- Realizar un control continuo a los proveedores hasta que se haga efectiva la compra.
- Definir los productos y cantidades que se envían a cada tienda según su localidad y demanda.

Los principales procesos identificados para el área de ventas son:

- Definir el precio de venta en base al margen de ganancia de cada producto y precio de compra.
- Identificar los productos que menos se venden y a partir de ello incentivar su compra mediante campañas promocionales.
- Realizar un seguimiento de las ventas diarias y en base a ellas proyectar el total de ventas a 30 días.
- Medir el impacto de cada campaña o promoción realizada.

La inteligencia de negocios es una herramienta de información estratégica que ayuda a las empresas a la toma de decisiones en las áreas de marketing, finanzas, operaciones, logística, administración, recursos humanos, entre otras áreas, por medio del análisis de los datos, brindando información disponible y rápida, permitiendo detectar fallas en los procesos, descubriendo oportunidades de negocio y cuantificando relaciones con proveedores y clientes. Este tipo de solución se basa en la extracción de datos de diversas fuentes, transformándolas y almacenándolas en un repositorio, desde el cual se genera información mediante reportes para los usuarios finales.

Luego de identificado el problema de falta de información sólida para toma de decisiones en las áreas de compras y ventas en una empresa comercializadora, se plantea como solución la implementación de un *Data Mart* de compras y un *Data Mart* de ventas los cuales son herramientas de inteligencia de negocios que se desarrollan para el presente proyecto de tesis.

1.2. Conceptos Relacionados

A continuación se presentan algunos conceptos relacionados a Inteligencia de Negocios:

BI (*Bussines Intelligence*, Inteligencia de Negocios): Es un enfoque para la gestión empresarial que le permite a una organización definir qué información es útil y relevante para la toma de decisiones corporativas. Inteligencia de Negocios es un esquema polifacético que fortalece a las organizaciones para tomar mejores decisiones rápidamente, convertir los datos en información y usar una estrategia inteligente para la gestión empresarial. [Vit 2002]

Data Warehouse: Es un almacén o repositorio de datos que integra información de diferentes fuentes (base de datos, archivos de texto, hojas de cálculo, etc.) y permite un análisis para la toma de decisiones. Muchos expertos definen el *data warehouse* como un almacén de datos centralizados que introduce datos en un almacén de datos específico llamado *data mart*. Otros aceptan una amplia definición de *data warehouse*, como un conjunto integrado de *data marts*. [Vit 2002]

Data Mart: Es un conjunto de datos que son estructurados de una forma que facilite su posterior análisis. Un *data mart* contiene la información referente a un área, un tema o una función en particular, con datos relevantes que provienen de las diferentes aplicaciones operacionales. Los *data marts* pueden ser de diversas bases de datos OLAP dependiendo del tipo de análisis que se quiera desarrollar. [Vit 2002]

OLAP (*On Line Analytical Processing*, Procesamiento analítico en línea): La funcionalidad de los sistemas OLAP se caracteriza por ser un análisis multidimensional de datos corporativos, que soportan los análisis del usuario y posibilidades de navegación, seleccionando la información a obtener. En general, estos sistemas deben:

- Soportar requerimientos complejos de análisis.
- Analizar datos desde diferentes perspectivas.
- Soportar análisis complejos contra un volumen ingente de datos. [Vit 2002]

Existen diferentes tipos de OLAP: OLAP multidimensional (MOLAP) y OLAP relacional (ROLAP). OLAP híbrido (HOLAP) se refiere a las tecnologías que combinan MOLAP y ROLAP.

MOLAP: La arquitectura MOLAP usa unas bases de datos multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos multidimensionalmente.

ROLAP: La arquitectura ROLAP cree que las capacidades OLAP están perfectamente implantadas sobre bases de datos relacionales. La arquitectura ROLAP es capaz de usar datos pre calculados si estos están disponibles, o de generar dinámicamente los resultados desde los datos elementales si es preciso. Esta arquitectura accede directamente a los datos del *data warehouse*, y soporta técnicas de optimización de accesos para acelerar las consultas.

HOLAP: La tecnología HOLAP permite manejar lo mejor de ambos mundos. Para información sumariada, HOLAP utiliza tecnología multidimensional para un mejor desempeño. Cuando se necesita llegar a la información detallada, HOLAP utiliza técnicas de datos relacionales para llegar a ésta.

Extracción de datos: Sobre el *data warehouse* construido actúa la plataforma de extracción de datos, la información se genera en tiempo real y con acceso universal, vía Internet, intranet, extranet o dispositivos PDA. [Mic 2006]

Explotación de datos: La explotación de la información se realiza a través de un amplio conjunto de herramientas de consulta y análisis de la información. [Mic 2006]

Metodología de Bill Inmon: Define una metodología descendente *top-down* (hacia abajo) donde los *data marts* se crearán después de haber terminado el *data warehouse* completo de la organización.

Metodología de Ralph Kimball: Determina que un *data warehouse* es la unión de todos los *data marts* de una organización. Define la metodología *bottom-up* (hacia arriba).

A continuación se presentan algunos conceptos relacionados al proceso de Extracción de datos:

Staging area (Área de ensayo): Es el área donde se ejecutan los procesos ETL y que tiene un carácter volátil, es decir al finalizar el proceso, el área de ensayo debe de quedar vacía, sin contenido. [Bar 1998]

Procesos ETL (Extract, Transform and Load, Extraer, transformar y cargar): Son los procesos responsables del transporte e integración de datos desde uno o más sistema fuentes hacia uno o más sistemas de destino. [Bar 1998]

A continuación se presentan algunos conceptos relacionados al proceso de Explotación de datos:

Dashboard (Tablero de mando): Es una página desarrollada en base a tecnología web mediante la cual se despliega en tiempo real información de la empresa extraída de varias fuentes o bases de datos. Su característica de tiempo real otorga a los usuarios un conocimiento completo sobre la marcha de la empresa y permite hacer análisis instantáneos.

Scorecard (Pronosticador): Los *dashboards* nos proporcionan ventanas al pasado y al presente de la información disponible sobre nuestro negocio mientras que el

scorecard es una herramienta indicada para mirar hacia el futuro del rendimiento de nuestra empresa.

Data Mining (Minería de datos): Es el proceso de correr datos en algoritmos completamente sofisticados, relevando significantes patrones y correlaciones que pueden estar escondidos. Esto puede ser usado para ayudar a entender lo mejor para el negocio y explotar el rendimiento de éste en un futuro prediciendo completamente en el análisis.

A continuación se presentan algunos conceptos relacionados al Análisis Dimensional:

Métrica: Cuando utilizamos el término métrica nos referimos a una medida numérica directa, que representa un conjunto de datos de negocios en la relación a una o más dimensiones. [Pen 2008]

Jerarquía: Las jerarquías pueden existir en una dimensión en la cual sirven como navegaciones predefinidas y están compuestas de uno o más niveles. Una dimensión puede tener una o más jerarquías. [Pen 2008]

Dimensión: Llamada también *lookup table* (tabla de búsqueda). Es un elemento conformado por componentes que en su conjunto caracterizan a dicho elemento. Una dimensión puede ser el tiempo, el cual está compuesto por día, mes, año, semestre, etc. Una dimensión se representa por una tabla donde sus columnas indican cada uno de sus componentes.

Cubo: Las bases de datos multidimensionales dependen de estructuras llamadas cubos. Un cubo es una colección de medidas y dimensiones. Este puede tener “n” dimensiones, las medidas dentro de un cubo son evaluadas en la intersección de todas las “n” dimensiones. Los cubos permiten la agregación a través de jerarquías dimensionales y la navegación hacia arriba /abajo rápida, siendo así mucho más flexible que una construcción basada en tablas.

Tabla de variables: Llamada también tabla de hechos o *fact table*. Es la tabla central que representa datos numéricos en el contexto de las entidades del negocio.

Modelo entidad relación: Constituye una forma de representar conceptualmente la realidad mediante su abstracción en entidades y relaciones. De esta forma intentamos representar el mundo que nos rodea, los datos de nuestro problema mediante una serie de entidades que representan objetos o conceptos así como las relaciones que se dan entre ellos tales como su uso y composición.

Modelo estrella: Es una arquitectura de almacén de datos simple. En este diseño del almacén de datos la tabla de variables (hechos) está rodeada por dimensiones y juntos forman una estructura que permite implementar mecanismos básicos para poder utilizarla con una herramienta de consultas OLAP.

Modelo copo de nieve: En las bases de datos usadas para *data warehousing*, un esquema en copo de nieve es una estructura más compleja que el esquema en estrella. Se da cuando existen un gran número de tablas de hechos sin que sea factible reducir su número. Aunque puede reducir espacio, tiene la contrapartida de tener menor rendimiento al tener que crear más tablas de dimensiones y más relaciones entre las tablas.

Análisis Ad-hoc: Es el análisis flexible de datos sin formatos establecidos o predefinidos. Este análisis les permite a los usuarios la posibilidad de preguntar y obtener respuestas a una infinita variedad de preguntas rápidamente. [Bar 1998]

DSS (*Decision Support System*, Sistema de soporte de decisiones): Es el sistema que provee los mecanismos de acceso a los datos y el análisis de los mismos. [Mic03]

Slice and Dice: Son dos métodos complementarios para interactuar con los datos. *Slicing* (rebanar) significa aislar un miembro específico de una dimensión para hacer análisis. *Dicing* (dividir) o romper un conjunto de datos en pequeñas piezas para examinar cómo las medidas interceptan múltiples dimensiones. [Bar 1998]

Metadatos: Datos acerca de los datos. Información acerca de las propiedades de datos tales como lógica de negocios que definen la estructura y contenido de dimensiones y medidas. [Bar 1998]

A continuación se presentan algunos conceptos relacionados a Software Libre:

Open Source (Código abierto): Es el software que, una vez obtenido, puede ser usado, copiado, estudiado, modificado y redistribuido libremente. El software libre suele estar disponible gratuitamente, pero no hay que asociar software libre a software gratuito, o a precio del costo de la distribución a través de otros medios; sin embargo no es obligatorio que sea así y, aunque conserve su carácter de libre, puede ser vendido comercialmente. [Wik07]

OSBI (*Open Source Business Intelligence*, Inteligencia de negocios de software libre): Es una solución de Inteligencia de negocios basada en tecnologías *open source*. [Ari 2007]

Pentaho: Es un proyecto iniciado por una comunidad *open source*, provee una alternativa de soluciones de inteligencia de negocios en distintas áreas como en la arquitectura, soporte, funcionalidad e implantación. Estas soluciones al igual que su ambiente de implantación están basados en JAVA, haciéndolo flexible en cubrir amplias necesidades empresariales. A través de la integración funcional de diversos proyectos de *open source* permite ofrecer soluciones en áreas como: análisis de información, reportes, tableros de mando conocidos como *dashboards*, flujos de trabajo y minería de datos.

A continuación se presentan algunos conceptos relacionados a las áreas de compras y ventas:

Slow Moving (Movimiento lento): Es un indicador usado en el área de compras que muestra los días que un producto no tiene movimiento. Lo óptimo es minimizar el tiempo transcurrido entre la fecha de compra y venta de un producto.

PSI (*Purchase Stock Inventory*, Inventario de compras y stock): Es un valor que indica la cantidad de productos que se deben comprar para cumplir los objetivos de venta deseada.

Margen de ganancia: Es el porcentaje de ganancia que se debe tener por la venta de un producto. Este margen no siempre es igual al margen real del producto.

Venta a 30 días: muestra la proyección de ventas que se estima realizar en un mes.

1.3. Plan de Proyecto

El proyecto de tesis se divide en los siguientes capítulos principales: marco conceptual, análisis, diseño, construcción y pruebas, los cuales son tratados a detalle en el documento de tesis.

En Marco Conceptual, se describe el problema encontrado y la solución propuesta. Asimismo se examina brevemente el estado del arte relevante para el tema de tesis.

En Análisis, se detalla paso a paso la metodología usada para el proyecto, requerimientos y plan de pruebas.

En Diseño, se establece el modelo dimensional, arquitectura y procesos de extracción y explotación.

Finalmente en Construcción, se desarrolla el proceso ETL y reportes. El siguiente gráfico esquematiza los entregables a realizar en el proyecto, bajo la estructura de gestión de proyectos WBS (*Work Breakdown Structure*, Estructura de Descomposición del Trabajo).

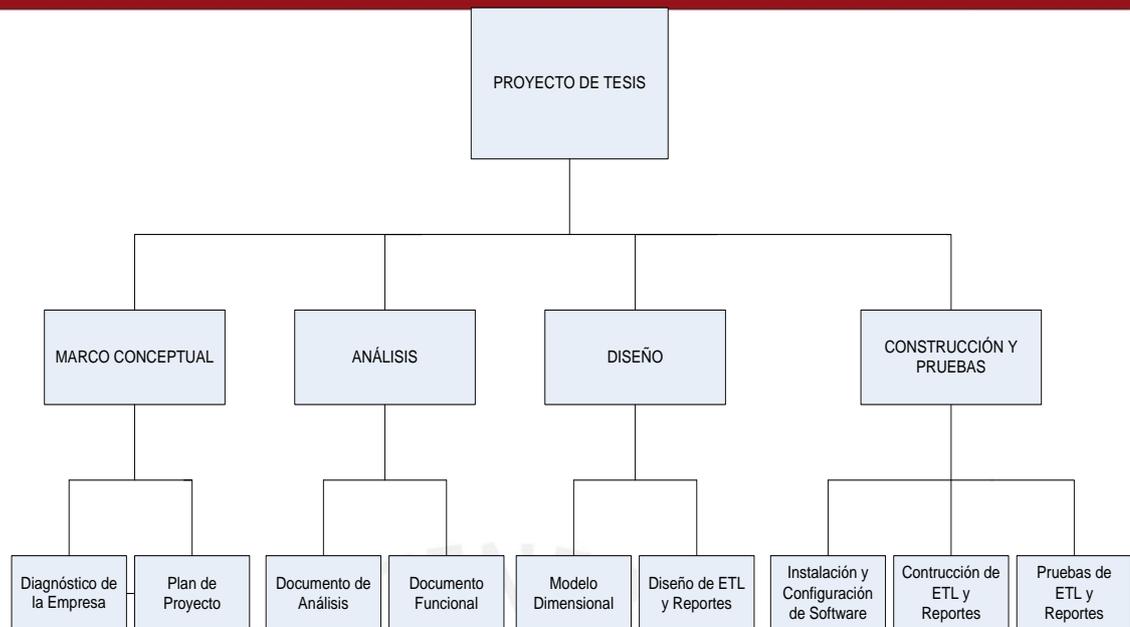


Figura 1.1 Diagrama de WBS. Fuente: Elaboración propia

El tiempo de desarrollo del proyecto es de 10 meses a tiempo completo por cada integrante del grupo de tesis.

La percepción del costo de una solución de BI es comúnmente denominado con el TCO (*Total Cost of Ownership*, Costo Total de Propiedad) al cual se le atribuyen ciertos factores que participan en el cálculo del mismo, por ejemplo:

Software

- Sistema Operativo
- Plataforma de BI
- Herramienta de BI (puede estar integrada en la plataforma o ser independiente)

Servicios

- Implementación y mantenimiento Hardware

- Implementación, mantenimiento y desarrollo de Software
- Capacitación de personal técnico y usuarios
- Consultoría en procesos de negocios

Hardware

- Servidores
- Conectividad
- Respaldos y mantenimiento
- Equipos Cliente

Cada uno de los rubros pudiera tener infinidad de elementos que lo comprenden, asimismo existen rubros de costo escondidos los cuales son difíciles de medir como lo son la cultura organizacional, el rechazo al cambio, el aprovechamiento correcto de la solución, entre otros, que pueden dejar fuera de la realidad dicha interpretación.

		Nombre de tarea	Duratio	Start	Finish
1		Definición de Asesor	7 days	Mon 8/20/07	Sun 8/26/07
2		Definición del Tema	30 days	Thu 8/23/07	Fri 9/21/07
3		Levantamiento de Información	30 days	Sun 9/2/07	Sat 10/20/07
4		Aprobación del Tema de Tesis	7 days	Mon 9/3/07	Fri 10/26/07
5		+ Plan de Tesis	32 days	Sat 9/8/07	Tue 10/9/07
16		+ Marco Conceptual	30 days	Wed 10/17/07	Thu 11/15/07
23		+ Análisis	39 days	Tue 3/18/08	Fri 4/25/08
33		+ Diseño	81 days	Sat 4/26/08	Tue 7/15/08
40		+ Construcción y Pruebas	21 days	Wed 7/16/08	Tue 8/5/08
46		+ Observaciones, Conclusiones y Recomendaciones	3 days	Wed 8/6/08	Fri 8/8/08
51		Revision General de la Monografía de Tesis	8 days	Wed 10/14/09	Wed 10/21/09
52		Revision Final de la Monografía de Tesis	7 days	Mon 9/13/10	Sun 9/19/10
53		Anexos del Documento	15 days	Mon 9/20/10	Mon 10/4/10
54		Construccion de Procesos de Carga al 100%	7 days	Wed 10/6/10	Tue 10/12/10
55		Construccion de Reportes al 100%	7 days	Wed 10/6/10	Tue 10/12/10

Figura 1.2 Diagrama de Gantt

El plan de proyecto a detalle según fechas, recursos y dependencias se encuentra en el Anexo A: Plan de Proyecto.

1.4. Estado del Arte

En el ambiente de los negocios se ha empezado a considerar las aplicaciones de inteligencia de negocios como una herramienta fundamental para la toma de decisiones. Para llevar ésto a cabo es necesario el uso de una plataforma que soporte la extracción y explotación de datos.

Dentro del ámbito peruano se encuentran disponibles herramientas de extracción de datos; entre los principales están:

Business Objects.- Cuenta con la herramienta *Business Objects Data Integrator*, la cual es una plataforma de integración de datos que puede abarcar a toda la organización. Permite explorar, extraer, transformar y distribuir datos para que los usuarios puedan disponer de información precisa, puntual y fiable. [Bus 2009]

Microsoft SQL Server.- Cuenta con una plataforma SSIS (*SQL Server Integration Services*, Servicios de integración de servidores); esta plataforma permite generar soluciones de integración de datos de alto rendimiento, entre las que se incluyen paquetes de extracción, transformación y carga para el almacenamiento de datos. [Mic 2009]

IBM DataStage.- Es una herramienta de IBM que permite crear y mantener fácil y rápidamente *data marts* y *data warehouses*. Soporta la extracción, integración y

transformación de altos volúmenes de datos desde estructuras simples hasta muy complejas. [Ibm 2009]

Pentaho.- Kettle.- programa ETL que incluye un conjunto de herramientas para realizar la extracción y transformación de datos. Uno de sus objetivos es que el proyecto ETL sea fácil de generar, mantener y desplegar. [Pen 2010]

Java - Clover.- Posee un *framework* llamado *CloverETL* que puede ser utilizado para la transformación y carga de datos. Esta herramienta se puede integrar con eclipse y permite trabajar con cualquier base de datos a través de JDBC (*Java Database Connectivity*, Conectividad a la base de datos JAVA). [Jav 2009]

Java - Octopus.- Es una herramienta ETL java que permite realizar transformaciones definidas en ficheros XML con cualquier base de datos para que exista JBDC. [Jav 2009]

El siguiente cuadro muestra las características de las herramientas de extracción mencionadas:

Características de la Herramienta					Java Clove	Java Octopus
¿Forma parte de una plataforma integrada de inteligencia de negocios?	Sí	No	Sí	Sí	No	No
¿La herramienta de extracción posee una interface grafica de uso?	Sí	No	Sí	Sí	No	Sí
¿Soporta diversos tipos de bases de datos?	No	Sí	No	Sí	Sí	Sí
¿Permite cargas desde ficheros excel, xml y planos?	Sí	No	Sí	Sí	Sí	Sí
¿Requiere una fácil instalación de la herramienta?	Sí	Sí	Sí	No	No	No
¿La plataforma posee una herramienta de explotación, herramientas de <i>reporting</i> , herramientas de consultas y análisis?	Sí	No	Sí	Sí	No	No
¿Se encuentra fácilmente consultoras de sistemas para la herramienta?	Sí	Sí	Sí	No	No	No

Cuadro 1.1 Cuadro comparativo de herramientas de extracción. **Fuente:** Elaboración propia

Luego de realizar un análisis y comparación entre las diversas herramientas de extracción se concluye que al cumplir Pentaho con la mayoría de características, es la herramienta más apropiada para el desarrollo de este proyecto.

Así también se tiene diversas herramientas de explotación tales como:

Oracle - Discoverer.- Es la herramienta de análisis y consulta de informes que proporciona un acceso inmediato a la información del *data warehouse* o *data mart*. La información en Oracle *Discoverer* se organiza de un modo similar que en Microsoft Excel. [Ora 2009]

Business Objects.- Los productos de consulta y análisis de *Business Objects* permiten al usuario formular preguntas, crear consultas, realizar cálculos y compartir información. [Bus 2009]

Los productos que ofrece *Business Objects* son:

- *BusinessObjects Web Intelligence*: Permite acceder a la información del negocio a través de *intranets* y *extranets*.
- *BusinessObjects Desktop Intelligence*: Permite a los usuarios controlar, entender y gestionar con facilidad todo el volumen de información almacenada en distintas fuentes de datos.

SAS.- Ofrece la herramienta de explotación, predice y optimiza su cadena entera del servicio para transformar operaciones *after-sales* de la ayuda de centros de costo desconectados en centros de beneficio integrados. El sistema de la solución permite a organizaciones ganar una ventaja competitiva conservando a clientes satisfechos y mejorando imagen de marca de fábrica. [Sas 2009]

Cognos – PowerPlay.- Cognos en el mercado BI presenta una amplia gama de funcionalidades orientadas al usuario, una mayor flexibilidad y mejoras técnicas. Esta nueva versión optimiza las capacidades de Cognos *Enterprise Business Intelligence Series* que integra a *PowerPlay*, *Cognos ReporNet* y *Metrics Manager* acercando los datos críticos de negocio a una extensa variedad de usuarios. Posee un rendimiento de misión crítica que incluye análisis *drag-and-drop* (*arrastrar y soltar*), además de estar integrado estrechamente con las últimas funcionalidades para cuadros de mando y *reporting*. [Cog 2009]

IBM - SurfAid para análisis.- Se emplea para atender a las demandas de gran complejidad y volumen para el rápido crecimiento del *web site*. Algunas herramientas y métricas adicionales incluyen acceso al banco de datos relacional, permitiendo consultas ad hoc e informes dinámicos. El *SurfAid* para análisis incluye acceso a los consultores para asistencia en la búsqueda de métodos de extracción de valor avanzado en base a los datos del cliente. [Ibm 2009]

Pentaho Análisis.- Con esta solución se puede navegar por la información con el uso de tablas dinámicas, cubos multidimensionales, *dashboards*, entre otros, permitiendo analizar de manera rápida el comportamiento de la empresa en sus diferentes rubros. Tanto los reportes como los análisis se pueden integrar en un *dashboard*, es posible integrar varios tipos de gráficos, tablas y velocímetros además que pueden ser convertidos en *portlets* (componentes modulares de la interfaz) para consultarlos dentro del portal web de Pentaho. [Pen 2010]

SpagoBI.- Es una plataforma *open source* para inteligencia de negocios. Fue creado teniendo como base la filosofía del software de código abierto, es capaz de cubrir todos sus aspectos funcionales, tales como: organización (estática y

dinámica) de los datos, investigación, información predictiva a través de la minería de datos, entre otros. [Spa 2009]

Java – BIRT (*Business Intelligence and Reporting Tools*, Herramienta de reporte de inteligencia de negocios) es un proyecto código abierto basado en ECLIPSE.

“*Birt*” es un sistema de reportes BI para aplicaciones web, basado en Java y J2EE. El proyecto “*Birt*” contempla una gran variedad de necesidades del reporting con una típica aplicación Java. Los principales componentes de “*Birt*” son: “*Report Designer*”, “*Report Engine*” y un completo sistema de Apis. [Jav 2009]

Java – JasperReports.- es una poderosa herramienta de creación de informes desarrollada en Java. Luego de la elaboración de reportes, éstos pueden ser exportados a ficheros PDF, HTML, XLS, CSV y XML. Puede ser usado en gran variedad de aplicaciones Java, incluyendo J2EE o aplicaciones web, para generar contenido dinámico.

Su propósito principal es ayudar a crear documentos de tipo páginas, preparados para imprimir en una forma simple y flexible. [Jav 2009]

El siguiente cuadro muestra las características de las herramientas de explotación anteriormente mencionadas:

Características de la Herramienta	ORACLE E-BUSINESS SUITE 11	Business Objects	SAS	COGNOS		pentaho	Spago BI	Java BIRT	Java Jasper Report
¿La plataforma integra los reportes web, reportes en cliente/servidor y archivos de Word, Excel, PowerPoint y PDF?	Sí	Sí	Sí	Sí	Sí	Sí	No	No	No
¿Los reportes permiten análisis gráficos pre-definidos, <i>templates</i> o motores analíticos (no reportes) con posibilidad de análisis a través de períodos?	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí
¿Los tableros de control incluyen una plantillas de análisis predictivo y contempla diversos algoritmos de predicción (logarítmico, exponencial, etc.)?	No	No	Sí	No	Sí	Sí	Sí	Sí	Sí
¿Permite el análisis de segmentaciones?	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí
¿Los tableros de control de la plataforma utilizan la misma herramienta de metadata que utiliza la herramienta de consultas y análisis por web?	Sí	Sí	Sí	Sí	Sí	Sí	No	No	No
¿Ofrece facilidad en la instalación de la herramienta?	Sí	Sí	No	Sí	Sí	No	No	No	No
¿La herramienta de metadata posee funciones adicionales a las provistas por la base de datos en uso?	Sí	Sí	Sí	Sí	Sí	Sí	No	No	No
¿Se cuenta con diversas consultoras que brinden asesoría en instalación y uso de la herramienta?	Sí	Sí	No	Sí	Sí	No	No	No	No
¿Facilidad de uso de la herramienta para la capacitación de usuarios?	Sí	Sí	No	Sí	No	Sí	No	No	No

Cuadro 1.2 Cuadro comparativo de herramientas de explotación. **Fuente:** Elaboración propia

A partir del cuadro se concluye que el uso de Pentaho es también recomendable en el proceso de explotación por cumplir con la mayoría de características necesarias, tales como: integración de reportes, algoritmos de predicción, análisis por web, entre otros.

Pentaho es una plataforma de Inteligencia de Negocios propuesta por la solución OSBI, es una herramienta integral que brinda soluciones para ambos procesos: extracción y explotación.

Algunas ventajas y beneficios del uso de *open source* son:

1. Al construir, integrar y desarrollar componentes *open source* dentro de una sola plataforma el coste de implementar soluciones Business Intelligence se reduce, beneficiando de esta manera principalmente a las pymes que no pueden acceder a la tecnología comercial actual.
2. Si se reduce la inversión en licencias, se puede destinar ese dinero a mejorar el análisis, el diseño, la formación de usuarios, la toma de decisiones.
3. Permite realizar prototipos de reportes de una forma más rápida y permanente en el tiempo.
4. El sistema es fácil y óptimamente escalable, pudiendo extender la misma entre diferentes servidores sin limitaciones del número de usuarios concurrentes, es decir, no existe un número limitado de licencias para su uso.
5. Se proporciona interfaces de usuarios muy intuitivos que son fácilmente customizables por los propios usuarios.

1.5. Descripción y sustentación de la solución

Se ha tomado como modelo una empresa comercializadora de electrodomésticos peruana.

En base al levantamiento de información obtenido en las diversas entrevistas y reuniones al área de Marketing y Sistemas, se ha realizado el análisis, diseño e implementación de este proyecto.

La solución abarca el modelo de negocio, fuentes de información, funcionalidades principales (compra y venta), modelamiento de datos y necesidades de información de una empresa comercializadora de electrodomésticos. Este levantamiento de información permite aplicar esta solución a otras empresas del mismo rubro, con la salvedad de que se tengan que realizar adecuaciones específicas propias de cada empresa.

1.5.1. Descripción actual de una empresa comercializadora de electrodomésticos

La empresa cuenta con una importante participación en el mercado peruano de electrodomésticos ya que una de sus fortalezas es la cobertura nacional a través de la red de tiendas ubicadas estratégicamente en el ámbito nacional. La empresa siempre cuenta con un plan de expansión y crecimiento para adecuarse a las necesidades de sus clientes.

Análisis de la competencia

Actualmente en el mercado de electrodomésticos peruano la competencia es mucho más dura en Lima. Sin embargo, la empresa comercializadora de electrodomésticos saca ventaja en sus tiendas ubicadas en provincias como: Arequipa, Cuzco, Tacna, Iquitos, Cerro de Pasco, Andahuaylas, Trujillo, entre otras.

Actualmente la empresa comercializadora de electrodomésticos cuenta con una participación del 15% a nivel Lima, y con 51% a nivel de provincias, lo que en conjunto hace que la empresa posea aproximadamente el 33% del mercado nacional, situándose como la cadena de mayor presencia y participación a nivel nacional gracias a la cobertura que posee.

El mercado de electrodomésticos en el Perú está subiendo exponencialmente. Actualmente supera los US\$ 850 millones como se muestra en el siguiente gráfico:

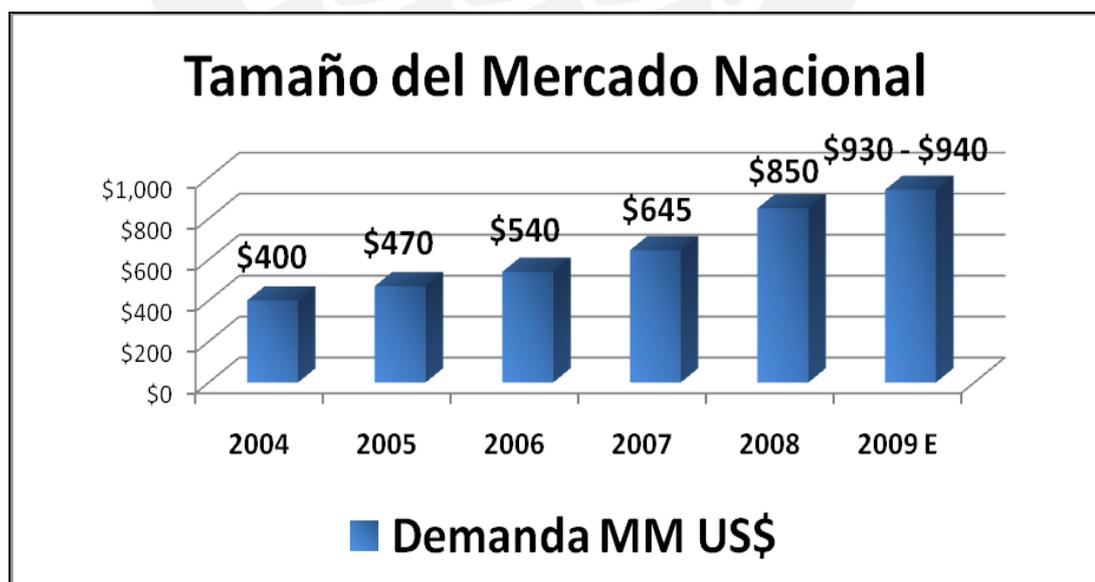


Figura 1.3 Demanda electrodomésticos en mercado nacional.

Fuente: Diario El Comercio.

A la fecha, 60% de las ventas de electrodomésticos en Lima se realiza a crédito, mientras que a nivel nacional este porcentaje se reduce a 45%, asociado a menores niveles de penetración del sector financiero en las regiones fuera de la capital. Asimismo, estuvo sustentado en la mayor diversidad de marcas de bajo costo y en constantes mejoras tecnológicas en los productos, especialmente aquellas referidas a comodidad y funcionalidad. La mayor oferta de electrodomésticos se explica por un mejor posicionamiento de la industria nacional, especialmente en línea blanca, así como por la diversificación de países de procedencia de las importaciones.

El mercado de electrodomésticos está compuesto por tres líneas: (i) línea blanca (refrigeradoras, lavadoras, cocinas, entre otros), que presenta la menor rotación de artículos por su condición de productos duraderos, cuyo consumo está determinado por el bajo poder adquisitivo promedio de la población, alcanzando una participación de 30% en las ventas totales de electrodomésticos; (ii) línea marrón (televisores, DVD, equipos de sonido, entre otros), que mantiene un crecimiento constante y muestra una alta dependencia por parte del público, lo que explica que muchos productos que ya se encuentran en la fase de maduración sigan mostrando altos niveles de comercialización; y (iii) artefactos menores (licuadora, tostadoras, entre otros), los cuales presentan altos índices de rotación y gran número de unidades comercializadas, participando del 10% de las ventas totales de electrodomésticos. La línea marrón sustentó parte importante de las mayores importaciones de electrodomésticos, al representar cerca del 60% de las ventas totales del segmento.

Las principales funciones de la empresa son las compras y las ventas de los productos.

El área de Marketing es la encargada de realizar las compras a los proveedores y

colocar los productos en la tienda para su venta, por lo tanto la solución planteada está dirigida a esta área. El área de Marketing está constituida por las diferentes Líneas de Productos, cada línea cuenta con un jefe de línea y un asistente.

Las líneas de productos que encontramos en la empresa comercializadora de electrodomésticos son:

- Audio.- Comprende equipos de sonido, radiograbadoras, reproductor portátil, car-audio.
- Video.- Está comprendido por productos como televisores, DVD, filmadoras, cámaras fotográficas, juegos de video, accesorios de video.
- Línea Blanca.- Son los productos como refrigeradoras, cocinas, microondas, lavadoras, congeladoras, campanas.
- Pequeños Electrodomésticos.- Pequeños productos como licuadoras, batidoras, cafeteras, hervidor, exprimidor, sandwichera, tostadoras, wafieras, plancha.
- Cómputo y Telefonía.- Computadoras, impresoras, celulares, laptops, entre otros.
- Deportes.- Bicicletas, equipos gimnasio.
- Línea para el Hogar.- Muebles de sala, colchones, menaje, maquinas de coser.
- Motocicletas.- Llantas, productos de fuerza.

Los productos son colocados en las tiendas según la demanda de la zona. Es por ésto que se hace un análisis de las ventas por región.

1.5.2. Sustentación de la solución

Existen dos enfoques a seguir para implementar soluciones de inteligencia de negocios, las cuales son: Teoría de Ralph Kimball y Teoría de Bill Inmon. [Ple 2009]

Kimball vs. Inmon

Las principales diferencias se identifican en los siguientes puntos:

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

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

Cuadro 1.3 Cuadro comparativo Kimball vs Inmon. **Fuente:** Elaboración propia

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

En la actualidad, el almacén de datos en pequeñas y medianas empresas se acerca más a la idea de Ralph Kimball. Esto se debe a que en la mayoría de los casos los proyectos de inteligencia de negocios comienzan como un esfuerzo dentro de un área específica, la cual se inicia con la construcción de *data marts*. Así se contempla que en un futuro se pueda llegar a construir un *data warehouse* donde se almacene la totalidad de datos de la empresa.

Con lo explicado líneas arriba se ha llegado a la conclusión de seguir la metodología de Ralph Kimball, por ser la más apropiada a aplicar para la realidad actual de las empresas comercializadoras de electrodomésticos en el Perú.

Se va a construir dos *data marts*, uno para el área de compras y otro para el área de ventas. La construcción de ambos *data marts* representan una solución estratégica de inteligencia de negocios dentro de las empresas del mismo rubro, por abarcar las principales actividades de este tipo de negocio.

Para mejorar la toma de decisiones en una empresa comercializadora de electrodomésticos, en las áreas de compras y ventas se presenta una solución de soporte de decisiones en software libre que comprende el desarrollo de dos *data marts*:

Data mart de compras.- Este *data mart* contiene la información necesaria para el análisis de las compras realizadas a los proveedores, por marca, modelo, precio de

compra, producto, línea del producto, fecha, entre otros elementos propios del entorno.

Además, por medio de los reportes el usuario puede verificar que compras le falta realizar (predicción de compas), como distribuir sus productos por tienda y decidir con que proveedores trabajar en base a su cumplimiento.

Data mart de ventas.- Este *data mart* contiene la información de las ventas realizadas en diversas sucursales de la empresa comercializadora, permitiendo el análisis de la tendencia de ventas por producto, línea del producto, precio de venta al público, clientes, localidad, fecha así como el margen de ganancia. Los reportes permiten conocer qué productos son los que tienen mejor acogida en el mercado y cuáles se deben dejar de comprar. Así también, se puede conocer las tiendas top, productos top, y cuáles cumplen su objetivo de venta.

2. Análisis

Este Capítulo describe la metodología a utilizar para poner en práctica el proyecto de tesis. También, detalla los requerimientos y necesidades de los usuarios. Finalmente, el capítulo muestra el plan de pruebas, el cual mide la eficacia del proyecto.

2.1. Metodología

A continuación se presenta una metodología que tiene como base el ciclo de vida de los *data warehouses* definido por Ralph Kimball. El esquema presentado por Ralph Kimball con el nombre de BDL (*Business Dimensional Lifecycle*, Ciclo de vida dimensional del negocio) grafica las diferentes etapas por las que debe pasar todo proceso de *data warehousing*. Se ilustra la secuencialidad de tareas de alto nivel

requeridas para el efectivo diseño, desarrollo e implementación de un proyecto de inteligencia de negocios [Mic 2006].

Este enfoque de implementación se muestra en el siguiente diagrama:

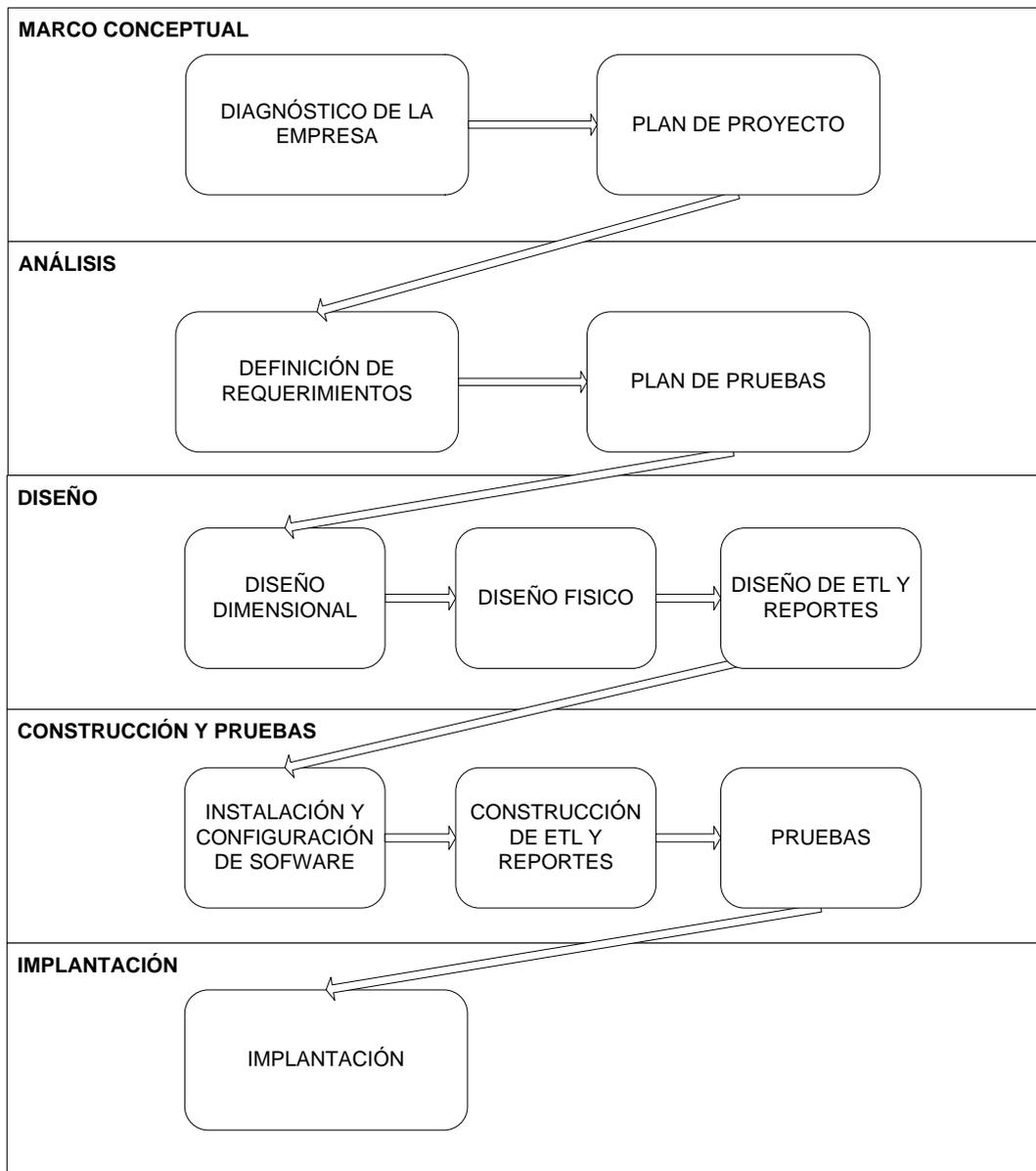


Figura 2.1 Metodología Fuente: Elaboración propia

El diagrama muestra una vista general de la ruta que sigue el proyecto en el cual cada rectángulo indica una etapa del proyecto y el orden a seguir.

Cualquiera que haya pasado por algún proyecto de *data warehousing* sabe que la magnitud de recursos y tiempo requerido no es igual para cada etapa. El BDL se focaliza en secuencialidad y concurrencia no en tiempos y plazos.

A continuación se describe cada una de las etapas en el presente proyecto de tesis:

2.1.1. Marco Conceptual

2.1.1.1 Diagnóstico de la empresa

Esta es la primera etapa de todo proyecto de Inteligencia de Negocios, en la cual se realiza el levantamiento de información de la empresa, se identifica sus necesidades, se diagnostica el problema y se plantea una solución.

El desarrollo de todas las etapas del proyecto se logra a partir de las diversas entrevistas y reuniones con los diferentes gerentes de línea del área de marketing y personal del área de sistemas.

2.1.1.2 Plan del proyecto

En esta etapa se definen las actividades, alcance y cronograma del proyecto incluyendo recursos y tiempo.

Desde la etapa de planificación del proyecto se establece una relación con los miembros del área de marketing, área de compras, área de ventas (usuarios finales) y del área de sistemas, quienes interactúan en conjunto con el presente equipo de tesis permitiendo el desarrollo del plan de proyecto y su seguimiento.

Según sentencia de Kimball, “Antes de comenzar un proyecto de data warehouse o data mart, hay que estar seguros si existe demanda y de dónde proviene. Si no se tiene un sólido usuario sponsor y no hay usuarios entusiasmados, posponga el proyecto”. Se da fe que el presente proyecto de tesis empieza con el entusiasmo referido por Kimball. [Kim 2008]

2.1.2. Análisis

2.1.2.1 Definición de requerimientos

La definición de los requerimientos del negocio es el punto de inicio del proyecto y a la vez es el cimiento para las siguientes etapas del proyecto, las cuales se enfocan en la tecnología, los datos y las aplicaciones. Esta definición se hace altamente crítica y es el centro de atención del BDL.

En esta etapa se realizan entrevistas a los usuarios finales (área de marketing, área de compras, área de ventas) lo cual permite entender los procesos del negocio, sus necesidades y requerimientos. Como estrategias para el relevamiento de información se ha manejado búsqueda de información pre-entrevista, selección de entrevistados y desarrollo de los cuestionarios para la correcta conducción de las entrevistas, las cuales ayudaron a descubrir eficazmente los requerimientos del negocio. Los requerimientos permitieron entender los factores claves que guían el negocio (compras y ventas) determinando la organización y alcance del proyecto.

Además, durante la etapa de levantamiento de información se identifica los sistemas transaccionales fuente, se conoce la arquitectura de los sistemas transaccionales, se realiza un inventario de éstos y se analiza la calidad de los datos.

2.1.2.2 Plan de pruebas

Se analizan las pruebas que permitan detectar errores para su corrección oportuna. El tipo de prueba utilizada es de caja negra, el cual consiste en el ingreso de datos que luego deben responder al resultado esperado sin importar el proceso que se realice con los datos. Estas pruebas se realizan sobre el producto final: reportes del área de compras y área de ventas.

2.1.3. Diseño

2.1.3.1 Diseño dimensional

Se ha usado como referente en esta etapa a Ralph Kimball. En el libro *“The Data Warehouse Toolkit”* (“El kit de herramientas data warehouse”), Kimball dedica tres capítulos a la etapa del modelado dimensional, donde describe las diferentes arquitecturas de construcción, esquemas de modelos y de tablas de hechos, definiendo el concepto de *data mart*.

Con la ayuda de la definición de los requerimientos del negocio se determinan los datos necesarios para cumplir los requerimientos analíticos de los usuarios finales. Para diseñar el modelo de datos que soporte estos análisis, se comienza identificando los correspondientes *data marts*, luego cada concepto del negocio (dimensión), se especifican los diferentes grados de detalle (atributos dentro de cada dimensión), como así también la granularidad de cada indicador (métrica) y las diferentes jerarquías dentro de cada dimensión que dan forma al modelo dimensional del negocio.

2.1.3.2 Diseño físico

El diseño físico de la base de datos o arquitectura de base de datos se focaliza sobre la selección de las estructuras de almacenamiento necesarias para soportar el diseño lógico, así como los métodos que garanticen un acceso eficiente a los datos.

Para llevar a cabo esta etapa, se decide cuál es el DBMS (*Data Base Management System*, Sistema administrador de base de datos) que se va a utilizar, ya que el esquema físico se adapta a él.

2.1.3.3 Diseño de ETL y reportes

Se siguen los siguientes pasos en el diseño de la base de datos para los *data marts*:

1. Obtener un conjunto de relaciones (tablas) y las restricciones que se deben cumplir sobre ellas siguiendo el modelo dimensional.
2. Determinar las estructuras de almacenamiento (variables, tipo, longitud, entre otros) y métodos de acceso a los datos.
3. Implementar la *meta data* basada en el diccionario de datos y tomando en cuenta la *meta data* de negocio, técnica y operacional.
4. Finalmente se diseña el modelo de seguridad del sistema.

Entre el diseño físico y el diseño lógico hay una retroalimentación, ya que algunas de las decisiones que se tomen durante el diseño físico pueden afectar a la estructura del esquema lógico y viceversa.

En esta fase también se diseñan los reportes analíticos según la información que necesita analizar el usuario final. Estos reportes se caracterizan por facilitar la

obtención de información en forma rápida y oportuna llegando incluso, si se quisiera, a un mayor detalle (*drill down* o uso de filtros).

2.1.4. Construcción

2.1.4.1 Configuración e instalación de software

En esta etapa se configura e instala la base de datos a utilizar: PostgreSQL, así como también la plataforma de Inteligencia de Negocios de software libre: Pentaho.

2.1.4.2 Construcción de ETL y reportes

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

1. Extracción: Durante esta sub etapa se siguen los procesos necesarios para obtener los datos que permiten efectuar la carga del modelo físico.
2. Transformación: Es esta sub etapa se siguen los procesos para convertir los datos fuente a fin de calcular las métricas y mantener un formato estándar de los datos.
3. Carga: Durante la carga de datos, se siguen los procesos necesarios para poblar los data marts.

La calidad de los datos es un factor determinante en el éxito de un proyecto de *data warehousing*. Es en esta etapa donde deben sanearse todos los inconvenientes relacionados con la calidad de los datos fuente.

Como advierte Kimball [Kim98] “el proceso de *Data Staging* es el *iceberg* de un proyecto de *data warehousing*”. En general esta es una de las etapas que siempre

termina tomando más tiempo del previsto. Ralph Kimball propone un plan de 10 ítems que ayudan a guiar esta etapa:

Plan:

1. Crear un diagrama de flujo fuente-destino esquemático.
2. Probar, elegir e implementar una herramienta de data staging.
3. Profundizar en detalle por tabla destino, gráficamente describir las reestructuraciones o transformaciones complejas. Desarrollo preliminar de la secuencialidad de los trabajos.

Carga de dimensiones:

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

Tabla de hechos y automatización:

1. Construir y probar la carga histórica de las fact tables (carga masiva de datos). Incluyendo búsqueda y sustitución de claves.
2. Construir y probar los procesos de cargas incrementales.
3. Construir y probar la generación de agregaciones.
4. Diseñar, construir y probar la automatización de los procesos.

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

Se decide descartar el uso de tablas temporales debido a que no se maneja una gran cantidad de datos que pudieran dañar la eficiencia del proceso *data staging*.

Asimismo está fuera del alcance del presente proyecto el proceso de limpieza de datos.

Para la construcción de reportes primero elaboramos con ayuda de la herramienta *Cube Designer* los cubos con la información a analizar.

El uso de esta herramienta es muy intuitivo y consta de tres pasos:

1. Elegir dimensiones.
2. Seleccionar los atributos de cada dimensión.
3. Seleccionar las métricas de cada dimensión.

Luego de realizados estos tres pasos, se genera el reporte en formato XML y puede ser visualizado vía web accediendo al siguiente url: <http://localhost:8080/pentaho>.

En esta interfaz se pueden seleccionar a los atributos como filtros, se puede hacer *drill down* y ver de manera grafica el reporte final.

2.1.4.3 Pruebas

Luego de la construcción del ETL y reportes se pone en práctica el plan de pruebas que permite identificar los errores originados en las etapas anteriores. Las pruebas son realizadas por los integrantes del presente proyecto utilizando diferentes parámetros de entrada.

2.1.5. Implantación

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

Existen varios factores extras que aseguran el correcto funcionamiento del producto, entre ellos se encuentran la capacitación, el soporte técnico, la comunicación y las estrategias de *feedback*. Todas estas tareas deben ser consideradas antes que cualquier usuario pueda tener acceso a los *data marts*.

Kimball en su libro “The Data Warehouse Toolkit” afirma que las organizaciones frecuentemente subestiman el esfuerzo y el tiempo requerido para esta etapa. El autor propone como solución seguir una lista de actividades de verificación antes de la implantación para un correcto funcionamiento del producto.

Esta lista planteada por Kimball incluye: configuración de hardware, conexión a las bases, acceso a intranet o internet, direcciones LAN (si no son dinámicamente asignadas), auditorías de tecnología sobre las configuraciones en las que se encuentran las PCs, proveer actualizaciones de hardware y software (determinando responsables, proyecto o área de usuario), verificaciones de seguridad, prueba de procedimientos de instalación, planificación de instalación, capacitación de usuarios, entre otros.

Finalmente Kimball propone un esquema de versionado. Primero se pasa por la versión Alpha, primera oportunidad para el grupo de trabajo de realizar una prueba del sistema. En esta versión todos los componentes del sistema deben ser testeados (infraestructura técnica, extracción, transformación, carga, procedimientos de calidad, *performance*, *templates*, entre otros)

Luego viene la versión Beta cuyo objetivo es conducir una prueba a nivel usuario. El grupo Beta está formado por los usuarios finales.

Una vez superadas estas dos versiones llegamos a un estado GA (*General Availability*, Disponibilidad general). Kimball sugiere que, todo cambio y/o modificación que se realice posteriormente pase internamente por un estado Alpha y Beta aunque externamente sea una nueva versión.

Esta etapa de implantación no está contemplada dentro del proyecto, sin embargo se considera importante la fase de capacitación del usuario, ya que es necesario que los usuarios conozcan el contenido de los *data marts* (los datos disponibles, qué significan, cómo se usan y para qué usarlos). Esta capacitación debe estar dirigida según el perfil del usuario (usuario final, *power user*). Además es importante que los usuarios cuenten con una documentación actualizada de los *data marts*, al cual puedan acceder en línea.

2.2. Requerimientos funcionales para el área de compras

Los requerimientos funcionales definen el comportamiento del software: cálculos, detalles técnicos, manipulación de datos y otras funcionalidades específicas. En base al levantamiento de información realizado los requerimientos funcionales para el área de compras son:

N°	Descripción	Dif	Prio	Ex
1	Reporte de compras realizadas. El reporte permite elegir la línea de producto y el periodo de tiempo a analizar. Este reporte muestra el proveedor y la forma de pago de una compra; así como la cantidad de productos comprados y el costo de estos. El reporte permite tomar decisiones sobre que proveedores elegir para la compra de productos y la mejor forma de pago de la compra. Además otorga un margen de negociación con el proveedor según las cantidades compradas en un periodo de tiempo para obtener tarifas rebajadas y formas de pago que beneficien a la empresa.	3	1	E
2	Reporte de Slow moving. El reporte muestra los días en que un producto no es vendido. Se elige una línea de producto específico. El reporte muestra un indicador denominado <i>slow moving</i> cuyo valor son los días transcurridos entre la última venta del producto y la fecha actual. El indicador es de color verde cuando la rotación del producto es favorable, de color amarillo cuando la rotación es aceptable, y de color rojo cuando ha sobrepasado el tiempo máximo permitido. El reporte refleja una alarma de los productos tienen bajas ventas y decidir si impulsarlos con una campaña o dejar de incluirlos en compras futuras. Este reporte permite decidir qué productos comprar nuevamente y que productos impulsar para incrementar las ventas.	2	2	E
3	Reporte del stock de productos por tienda. Este reporte muestra la cantidad total de productos en una tienda y cuántos de estos se encuentran en exhibición. El reporte permite la elección de una línea de producto y la tienda a analizar, mostrando información de cada producto. Este tipo de reportes permite la toma de decisiones referente a la mejor distribución por tienda, así como conocer que tiendas tienen mayor demanda de productos.	3	2	E
4	Reporte de compras a realizar de un producto. Este reporte muestra la cantidad de productos que se necesita para cubrir las ventas estimadas a 30 días. El reporte permite elegir la línea de producto y este muestra los productos, el stock, la cantidad de productos en exhibición y la venta a 30 días. Este reporte ayuda a decidir qué productos se deben comprar y en qué cantidad para cubrir la	2	1	E

	demanda de los mismos.			
5	Reporte de proveedores. Se elige un periodo de tiempo para el análisis. El reporte permite conocer los proveedores con los que trabaja la empresa en el periodo de tiempo y encontrar con quienes se tienen las mayores compras. Este reporte permite una mejor negociación con el proveedor buscando los mejores precios de compra y mejor margen de ganancia.	2	3	D
6	Reporte de estados de compra. Este reporte muestra el estado de las compras (pedidas, atendidas, etc.). Permite tener un seguimiento de las compras realizadas por proveedor, según una línea de producto elegido. El reporte permite conocer a que proveedores contactar para pedir agilizar el proceso de compra. Este reporte también ayuda a tener un control sobre el proceso de compra.	2	3	D
7	El sistema muestra en todos los reportes como marca de agua el logo de la empresa comercializadora de electrodomésticos.	2	3	D
8	Los usuarios tienen la opción de imprimir los reportes si lo requieren.	3	1	E
9	Se generan gráficos tipo columnas en base a los reportes.	1	2	E
10	El sistema presenta Dashboards para visualizar el estado de los indicadores claves de la empresa, como el semáforo dentro del reporte de <i>slow moving</i> .	3	2	D
11	Todos los reportes pueden ser exportados a otros archivos con formato: PDF, XLS, HTML y texto.	2	1	E
12	EL usuario define los campos y filtros que muestra el reporte.	1	1	E

Cuadro 2.1 Cuadro de requerimientos funcionales para el área de compras

2.3. Requerimientos no funcionales para el área de compras

Los requerimientos no funcionales para el área de compras son aquellos que se usan para todo el proyecto y durante todo su desarrollo.

N°	Descripción	Dif	Prio	Ex
1	El sistema es elaborado en software libre.	1	1	E
2	El sistema tiene como motor de base de datos PostgreSQL, base de datos en software libre.	3	1	E
3	El sistema cuenta con una interfaz gráfica en web para mostrar los reportes para el área compras.	2	1	E
4	El sistema es de fácil manejo y presenta una interfaz amigable para el usuario final. proveer un sistema intuitivo y fácil de usar que permita a los usuarios finales generar sus propios reportes y análisis.	2	1	E
5	El tiempo de familiarización con el sistema no debe ser mayor a sesenta minutos con asistencia guiada.	1	2	D
6	El sistema es accesible 24 x 7 los 365 días.	3	2	D

Cuadro 2.2 Cuadro de requerimientos no funcionales para el área de compras

2.4. Requerimientos funcionales para el área de ventas

Los requerimientos funcionales para el área de ventas son:

N°	Descripción	Dif	Prio	Ex
1	Reporte de ventas realizadas. Este reporte permite analizar la cantidad de productos vendidos y los ingresos obtenidos, muestra que producto o línea de productos tienen mayor acogida por tienda para una fecha específica, controla el cumplimiento de objetivos de venta por tienda o línea de producto. Además en caso de baja rotación de un producto se impulsa mediante campañas comerciales.	3	1	E

N°	Descripción	Dif	Prio	Ex
2	Reporte de precio de venta al público de un producto (PVP). Este reporte ayuda a decidir el precio de venta de cada producto al público. El cálculo de PVP se apoya en el costo de compra y margen de ganancia, existe una variable más que influye en este cálculo: las campañas estacionales, que conllevan a la paquetización de precios.	3	1	E
3	Reporte del margen teórico vs margen real de venta de un producto. Este reporte muestra el margen de ganancia definido para cada producto y el margen de ganancia calculado a partir del precio de venta final del producto. Este reporte permite analizar los ingresos reales y compararlos con los objetivos de ingresos por línea de producto y tienda.	2	1	E
4	Reporte de ventas proyectadas a 30 días (Vta30días). Este reporte muestra la cantidad de productos vendidos en el mes consultado y las ventas proyectadas al cierre del mes. En caso la proyección de ventas este por debajo de la meta trazada se toma acciones a tiempo que ayuden a elevar el número de ventas para cumplir la meta.	2	1	E
5	Reporte de impacto de Campañas. Este reporte muestra la efectividad de las campañas calculado a partir del número de ventas, clientes e ingresos por tienda. Esta medición permite analizar si la venta de un producto o paquete de productos tuvo acogida para un segmento en particular dentro de un periodo definido y si se llegaron a los resultados esperados.	3	1	D
6	Reporte de medición de devoluciones. Este reporte ayuda a calcular el valor real de ventas realizadas, además permite conocer las razones de desistimiento de compra de los clientes como la identificación de errores de fabricación de un producto.	2	2	D
7	El sistema muestra en todos los reportes como marca de agua el logo de la empresa comercializadora de electrodomésticos.	2	3	D
8	Los usuarios tienen la opción de imprimir los reportes si lo requieren.	3	1	E
9	Se genera gráficos tipo columnas en base a los reportes.	1	2	E

N°	Descripción	Dif	Prio	Ex
10	El sistema presenta Dashboards para visualizar el estado de los indicadores claves de la empresa, como el semáforo dentro del reporte de <i>slow moving</i> .	3	2	D
11	Todos los reportes pueden ser exportados a otros archivos con formato: PDF, XLS, HTML y texto.	2	1	E
12	EL usuario define los campos y filtros que muestra el reporte.	1	1	E

Cuadro 2.3 Cuadro de requerimientos funcionales para el área de ventas

2.5. Requerimientos no funcionales para el área de ventas

Los requerimientos no funcionales son aquellos que se usan para todo el proyecto y durante todo su desarrollo.

N°	Descripción	Dif	Prio	Ex
1	El sistema es elaborado en software libre.	1	1	E
2	El sistema tiene como motor de base de datos PostgreSQL, base de datos en software libre.	3	1	E
3	El sistema cuenta con una interfaz gráfica en web para mostrar los reportes para el área compras.	2	1	E
4	El sistema es de fácil manejo y presenta una interfaz amigable para el usuario final.	2	1	E
5	El tiempo de familiarización con el sistema no debe ser mayor a sesenta minutos con asistencia guiada.	1	2	D
6	El sistema es accesible 24 x 7 los 365 días.	3	2	D

Cuadro 2.4 Cuadro de requerimientos no funcionales para el área de ventas

Dif: Dificultad

Valores	Descripción
1	Alta
2	Media
3	Baja

Imp: Importancia/Prioridad

Valores	Descripción
1	Alta
2	Media
3	Baja

Ex: Exigibilidad

Valores	Descripción
E	Exigible
D	Deseable

2.6. Plan de Pruebas

La fase de pruebas añade valor al producto final y tiene como objetivo detectar errores en el momento oportuno, así como brindar calidad al proyecto.

Para esta fase se prueba el resultado de los reportes de compras y ventas realizadas usando el tipo de prueba de caja negra o de entrada/salida.

Las pruebas de caja negra se centran en lo que se espera de un proceso, es decir, intentan encontrar casos en que el proceso no se atiene a su especificación. Por ello también se denominan pruebas funcionales, y el probador se limita a suministrarle datos como entrada y estudiar la salida, sin preocuparse de lo que pueda estar haciendo el proceso por dentro.

Las pruebas de caja negra se usan especialmente en el producto final que va a ser interfaz con el usuario (en este caso, los reportes).

Las pruebas de caja negra se apoyan en la especificación de requisitos. De hecho, se habla de "cobertura de especificación" para dar una medida del número de requisitos que se han probado.

Luego de revisar los requisitos, se sigue una técnica algebraica conocida como "clases de equivalencia". Esta técnica trata cada parámetro como un modelo algebraico donde unos datos son equivalentes a otros.

Para identificar clases de equivalencia se toma en cuenta los siguientes puntos:

Un parámetro de entrada debe estar comprendido en un cierto rango.

Una entrada requiere un valor concreto.

Si una entrada es booleana, hay 2 clases: sí o no.

Los mismos criterios se aplican a las salidas esperadas.

Una vez identificadas las clases de equivalencia significativas, se procede a coger un valor de cada clase, que no esté justamente al límite de la clase. Este valor aleatorio, hace las veces de cualquier valor normal que se le pueda pasar en la ejecución real.

La data utilizada para el Plan de Pruebas se especifica en el Anexo B: Data Input del Plan de Pruebas.

2.6.1. Plan de Pruebas para el área de compras

El plan de pruebas mide la eficacia de los reportes. El plan de pruebas se detalla para cada reporte del área de compras de la siguiente manera:

2.6.1.1 Reporte de compras realizadas

Este reporte ayuda a los jefes de línea y gerentes de áreas a conocer las diferentes formas de pago con que se trabaja con los proveedores. Este reporte también es un soporte para la negociación con el proveedor.

Para efecto de prueba se ingresa como parámetro de inicio la línea: BLANCA y el año: 2008. El resultado esperado teniendo la data del transaccional especificado en el Anexo H Pruebas de carga, es:

Proveedor	Forma Pago	Cantidad	Costo (S/.)
HP	CONTADO	40	20,000
	CREDITO	80	24,560
LG	CREDITO	66	19,398
SAMSUNG	CONTADO	25	17,500
	CREDITO	256	96,098

Cuadro 2.5 Resultado esperado para el reporte de compras realizadas

2.6.2. Reporte de *slow moving*

El reporte de *slow moving* ayuda a medir el grado de movimiento que tiene un producto en la empresa. Es un indicador que avisa al usuario si un producto tiene poca circulación y por lo tanto puede llevar al jefe de línea a no seguir adquiriendo dicho producto.

Este reporte muestra el cálculo del *slow moving* como indicador, según la fórmula: **fecha actual - fecha de última compra**. Dicho indicador, si tiene un valor de 0 a 30 días se muestra de color verde (Bueno), de 30 a 90 se muestra de color ámbar (Aceptable) y de 90 a más se muestra de color rojo (Crítico). Para la prueba se

considera como fecha actual el día 3 de diciembre del 2009 y como parámetro de ingreso la línea de producto: AUDIO. El resultado esperado teniendo la data del transaccional especificado en el Anexo H Pruebas de carga, es:

Producto	Fecha Última Venta	Slow Moving
MP3 EQUIPO AUDIO BASICO PHILLIPS 34	02/12/2009	1
MP3 EQUIPO AUDIO BASICO PHILLIPS 67	02/12/2008	360

Cuadro 2.6 Resultado esperado para el reporte de slow moving

2.6.3. Reporte de stock de productos

Dicho reporte muestra la cantidad de productos por tienda y ayuda a los jefes de línea a colocar mejor los productos después de su compra, además puede ayudar a decidir si se deja de comprar o no un producto en caso la demanda sea baja. El reporte de stock de productos es de tipo matriz o tabla dinámica, se muestra el stock por tienda y por producto. Se ingresa como parámetro inicial la línea del producto: COMPUTO. El reporte muestra la cantidad total de productos y la cantidad de productos en exhibición por tienda. El resultado esperado teniendo la data del transaccional especificado en el Anexo H Pruebas de carga, es:

Producto	BARRANCA		CAÑETE	
	Exhibición	Total	Exhibición	Total
LAP TOP COMPUTADORA COMPUTO PUNTUAL HP 52	1	43		
PC COMPUTADORA COMPUTO BASICO HP 1234	-	-	1	50

Cuadro 2.7 Resultado esperado para el reporte de stock de productos

2.6.4. Reporte de compras a realizar

Este reporte es de gran ayuda al jefe de producto para decidir la cantidad de productos a comprar, esto según las ventas y la cantidad de productos que se encuentran en stock y en exhibición.

Se ingresa como parámetro de inicio la línea de producto a analizar: COMPUTO. El campo "Cantidad" muestra la cantidad de productos que se deben comprar, según la fórmula: **(Vta30días + Exhibición) - Stock**. Si la fórmula tiene como resultado un valor negativo, se debe mostrar el valor de 0. El resultado esperado teniendo la data del transaccional especificado en el Anexo H Pruebas de carga, es:

Producto	Cantidad	Exhibición	Total	Venta 30 días
LAP TOP COMPUTADORA COMPUTO PUNTUAL HP 52	123	1	43	165
PC COMPUTADORA COMPUTO BASICO HP 1234	0	1	50	45

Cuadro 2.8 Resultado esperado para el reporte de compras a realizar

2.7. Plan de Pruebas para el área de ventas

El plan de pruebas se detalla para cada reporte del área de ventas de la siguiente manera:

2.7.1. Reporte de ventas realizadas

Este reporte ayuda a los jefes y gerentes de línea a medir las ventas realizadas por tienda y a partir de ello medir el cumplimiento de sus objetivos de venta. Con este reporte se visualiza los resultados de las ventas y los ingresos producto de estas ventas.

A continuación se verifica las ventas realizadas, para ello se ingresan como parámetros de inicio: un grupo de tiendas, la línea CÓMPUTO y el año 2008.

El resultado esperado teniendo la data del transaccional especificado en el Anexo H Pruebas de carga, es:

Tienda	Cantidad vendida	Ventas realizadas S/.
Abancay	25	38,800
Ancash	43	27,000
Andahuaylas	11	18,100
Arequipa	7	7,200
Ate	57	49,400
Ayacucho	17	11,900

Cuadro 2.9 Resultado esperado para el reporte de ventas realizadas

2.7.2. Reporte de Precio de Venta al Público (PVP)

Este reporte ayuda a estimar el precio de venta de cada producto al público. El cálculo de PVP se apoya en el costo de compra y margen de ganancia.

Se ingresa como parámetro de inicio la línea: BLANCA. Para obtener el resultado del PVP, se sigue la siguiente fórmula: Precio de compra + Margen de venta + IGV (el IGV se calcula luego de calcular el margen de venta sobre el precio de compra), sin embargo este resultado puede ser modificado por ejemplo durante las campañas promocionales.

El resultado esperado teniendo la data del transaccional especificado en el Anexo H Pruebas de carga, es:

Producto	Precio de Compra	PVP
AUTOMATICA LAVADORA BLANCA BASICO LG 234	500	800
DORADOR MICROONDAS BLANCA BASICO MABE 564	100	300
ELECTRICA COCINA BLANCA ESPECIAL MABE 32	550	1100
SEMIAUTOMATICA LAVADORA BLANCA ESPECIAL MABE 678	600	1200

Cuadro 2.10 Resultado esperado para el reporte de PVP

2.7.3. Reporte de Margen teórico vs. Margen real

Este reporte sirve de apoyo para consultar el margen teórico (definido por el jefe de línea en base a la ganancia esperada) vs el margen de ganancia real por producto (que resulta del precio final de venta para cada producto).

Se ingresa como parámetro de inicio la línea: COMPUTO y se muestra el PVP, margen teórico y margen real obtenido después de la venta. Este margen se calcula según la siguiente fórmula: $(\text{Precio Venta Promedio} - \text{PVP}) / \text{Precio Venta Promedio}$.

El resultado esperado teniendo la data del transaccional especificado en el Anexo H Pruebas de carga, es:

Producto	PVP	Margen Teórico (%)	Margen Real (%)
LAP TOP COMPUTADORA COMPUTO MANUAL HP 52	1,500	50	114
LASER IMPRESORA COMPUTO ESPECIAL HP 78	600	50	100
MATRICIAL IMPRESORA COMPUTO PUNTUAL HP 85	500	50	100
PC COMPUTADORA COMPUTO BASICO HP 1234	2,000	50	67

Cuadro 2.11 Resultado esperado para el reporte de margen

2.7.4. Reporte de ventas a 30 días

Este reporte ayuda a proyectar el cierre de ventas mensual de cada producto en base a las ventas realizadas a la fecha de consulta.

Se ingresa como parámetros de inicio la línea de producto: BLANCA y tienda: Chiclayo Real. Se calcula el proyectado de ventas a 30 días a partir de la fecha actual: 02 de Noviembre del 2009, se muestra la cantidad vendida durante el mes a esta fecha y las ventas proyectadas al cierre del mes.

El resultado esperado teniendo la información del transaccional especificado en el Anexo H Pruebas de carga, es:

Producto	Cantidad vendida	Venta a 30 días
AUTOMATICA LAVADORA BLANCA BASICO LG 234	9	135
DORADOR MICROONDAS BLANCA BASICO LG 45	31	465
ELECTRICA COCINA BLANCA ESPECIAL MABE 32	27	405
SEMIAUTOMATICA LAVADORA BLANCA ESPECIAL MABE 678	47	705

Cuadro 2.12 Resultado esperado para el reporte de venta a 30 días

2.6. Software a utilizar

Se describe el software que se emplea durante el ciclo de vida de desarrollo del presente proyecto. Comenzando por el modelado de datos, la elección del repositorio usado como base de datos, extracción y explotación de los datos y finalmente elaboración de reportes.

A continuación se describe el uso de cada una de las herramientas utilizadas.

El detalle de instalación y configuración de cada una se encuentra en el Capítulo 4.

2.7.5. Modelador de datos

El uso de un modelador de datos sirve de base principalmente para el proceso de análisis y desarrollo de este proyecto. Ha permitido que los diagramas puedan ser fácilmente creados y modificados. Para facilitar el diseño de la base de datos usada como repositorio se ha utilizado ERwin 4.0

ERwin es una herramienta para el diseño de base de datos, que brinda productividad en su diseño, generación, y mantenimiento de aplicaciones, desde un modelo lógico (mostrando todas las entidades, atributos, relaciones, y llaves importantes) de los requerimientos de información, hasta el modelo físico perfeccionado para las características específicas de la base de datos diseñada. Además ERwin permite visualizar la estructura, los elementos importantes, y optimizar el diseño de la base de datos. Genera automáticamente las tablas y miles de líneas de *stored procedure* y *triggers* para los principales tipos de base de datos. La migración automática garantiza la integridad referencial de la base de datos. ERwin establece una conexión entre una base de datos diseñada y una base de datos ya existente, permitiendo transferencia entre ambas y la aplicación de

ingeniería reversa. Usando esta conexión, ERwin genera automáticamente tablas, vistas, índices, reglas de integridad referencial (llaves primarias, llaves foráneas), valores por defecto y restricciones de campos y dominios.

ERwin soporta principalmente bases de datos relacionales SQL y bases de datos que incluyen Oracle, Microsoft SQL Server, Sybase. El mismo modelo puede ser usado para generar múltiples bases de datos, o convertir una aplicación de una plataforma de base de datos a otra.

2.7.6. Gestor de Base de Datos

Para el presente proyecto de Tesis, se decide utilizar un sistema de gestión de bases de datos relacional en software libre, y así complementar la solución *OpenSource Business Intelligence* (OSBI). Se elige PostgreSQL, ya que su diseño permite soportar una gran carga de forma muy eficiente. Es por esto que es ideal para la creación de un *Data Mart* o *Data Warehouse*.

PostgreSQL es el gestor más usado en el mundo del software libre, debido a su gran rapidez y facilidad de uso. Esta gran aceptación es debida a que existen infinidad de librerías y otras herramientas que permiten su uso a través de gran cantidad de lenguajes de programación, además de su fácil instalación y configuración.

Las principales características de este gestor de bases de datos son las siguientes:

- Aprovecha la potencia de sistemas multiprocesador.
- Soporta gran cantidad de tipos de datos para las columnas.
- Dispone de API's en gran cantidad de lenguajes (C, C++, Java, PHP, entre otros).
- Gran portabilidad entre sistemas.

- Gestión de usuarios y *passwords*, manteniendo un muy buen nivel de seguridad en los datos.

2.7.7. Plataforma BI

Compuesta por componentes *OpenSource* provee la arquitectura y la Infraestructura a la vez. Forma un proceso centralizado bajo un marco de trabajo orientado a la solución de problemas de soporte de decisión empleando componentes de BI y permitiendo desarrollos completos para atender a soluciones de Inteligencia de Negocios.

En la siguiente ilustración se muestra la arquitectura funcional de Pentaho:

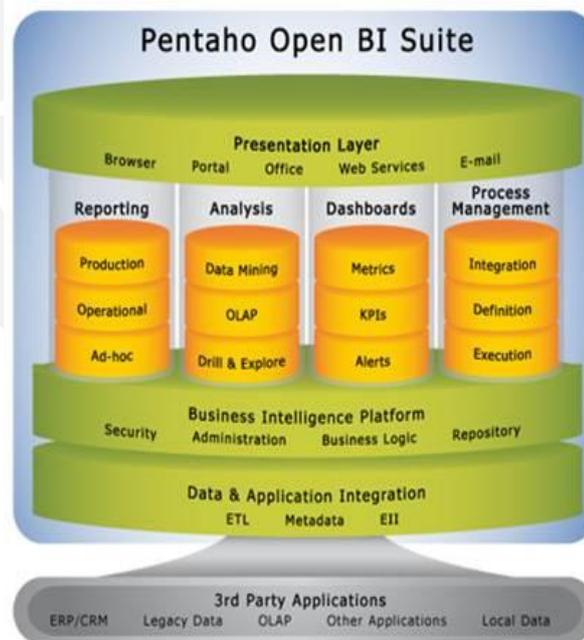


Figura 2.2 Arquitectura funcional de Pentaho **Fuente:** website Pentaho BI

En resumen la Plataforma BI integra componentes *OpenSource* mostrando una combinación de flujos de trabajo y administración de procesos.

A continuación se muestra los componentes de la plataforma Pentaho que se utilizan en el desarrollo del proyecto:

Pentaho Data Integration



Llamado por sus siglas en inglés “*K Extraction Transformation Transportation Load E*”.

Usa una interface grafica “SPOON” que permite diseñar “*jobs*” de transformación de datos para ser procesados bajo las herramientas de Kettle (PAN y KITCHEN).

PAN es un motor de transformación de datos permitiendo la lectura y escritura sobre diversas fuentes de datos. KITCHEN es una aplicación que permite ejecutar “*Jobs*” planificados en modo “*batch*” (serie, lote) para correr automáticamente en intervalos de tiempo definidos.

Pentaho Analysis Services



Ahora bautizado como “Pentaho Analysis Service” forma parte del motor OLAP integrado en el Suite BI de Pentaho. Un ejemplo rápido sobre el flujo de datos que utiliza esta herramienta es:

1. El cliente manda una solicitud de consulta bajo la interfaz web JPivot
2. Mondrian recibe la solicitud y bajo el esquema de metadatos que definen sus conceptos multidimensionales busca si ya tiene los datos en cache respondiendo rápidamente a la petición.
3. Si los datos no se encontraron en cache ejecuta las sentencias SQL para generar los datos.

4. Se almacenan los datos recibidos en cache para agilizar posteriores consultas.
5. Y finalmente se devuelve el resultado al usuario cliente a través de la interfaz.

La siguiente imagen muestra el funcionamiento de Mondrian:

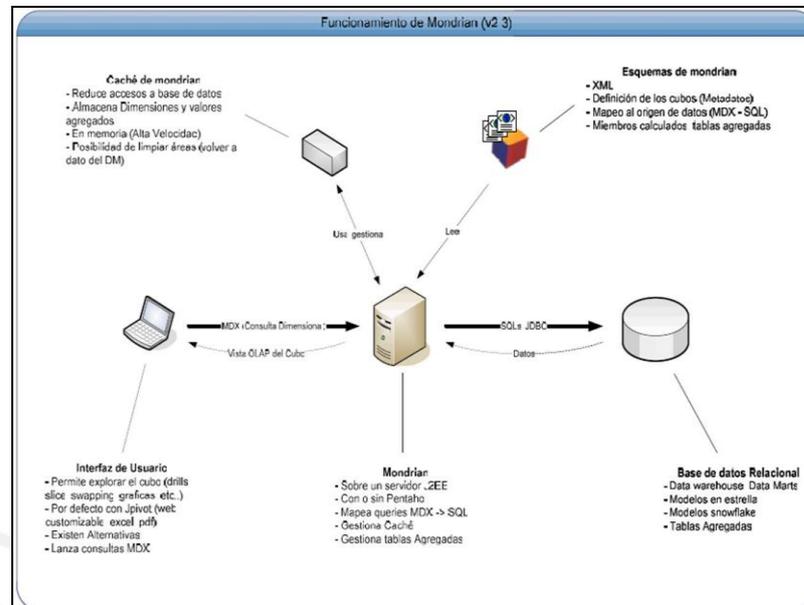


Figura 2.3 Funcionamiento de Mondrian **Fuente:** website Pentaho BI

JFreeReport / Pentaho Reporting



Herramienta de reportes flexibles y con clase empresarial, de escritorios o basados en Web. La herramienta de reportes Pentaho permite comenzar desde sencillos reportes iniciales hasta formar complejos reportes ajustados a las necesidades de negocio.

La siguiente imagen muestra ejemplos de reportes avanzados que se pueden realizar con la herramienta:

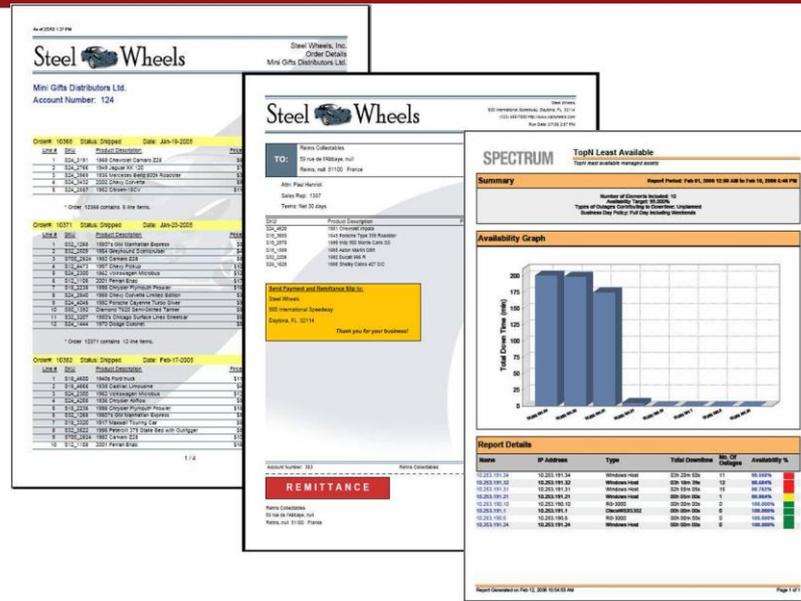


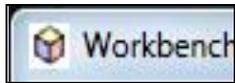
Figura 2.4 Ejemplos de reportes avanzados. Fuente: website Pentaho BI

Pentaho Cube Designer



Cube Designer es el fichero del esquema Cubo Mondrian. En él se definen dimensiones, niveles de jerarquía de las dimensiones, hechos y conexión a la base de datos relacional, los cuales sirven como datos para diseñar el cubo OLAP. El fichero del esquema Cubo Mondrian es de tipo XML que se puede crear con cualquier editor de texto, el problema es que puede resultar muy difícil y consumir mucho tiempo, por eso se aconseja usar la aplicación OpenSource Pentaho Cube Designer para la creación de cubos. La aplicación Cube Designer es un buen apoyo en creación de los cubos OLAP. Sin embargo en la versión actual del Pentaho Cube Designer (0.7.2.0) faltan algunas opciones avanzadas, por ejemplo creación de agregados, dimensiones compartidas o cubos múltiples. Se recomienda usar Cube Designer para crear modelos simples, no complejos y añadir las opciones avanzadas a mano editando directamente al fichero XML.

Workbench



Workbench es parte del proyecto Mondrian de la suite Pentaho. Su finalidad es la de construir los cubos OLAP y administrar el lenguaje de consulta ROLAP llamado MDX (*Multi-Dimensional Expression*, Expresión Multidimensional).

Esta herramienta ayuda a crear un cubo más complejo; para ello se modifica el archivo XML que corresponda al cubo creado. Dentro de la carpeta donde se instale Workbench, se ejecuta el archivo workbench.bat y se visualiza una aplicación donde se puede modificar los archivos XML (los cuales contienen a los cubos) pudiendo agregar, eliminar o editar dimensiones y métricas.

3. Diseño

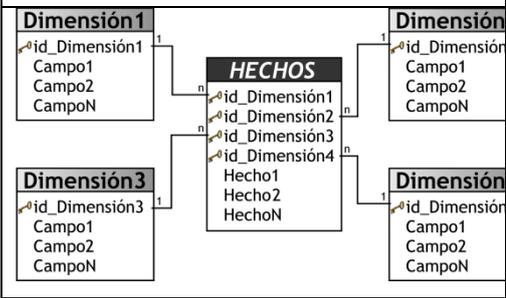
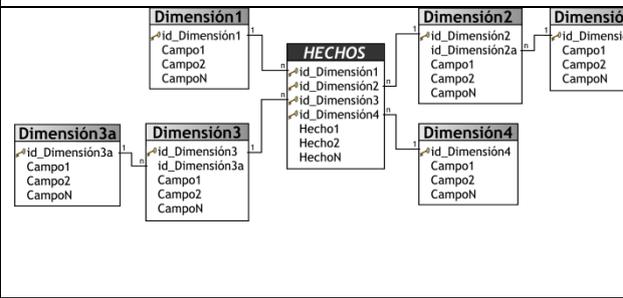
Este capítulo describe el diseño de los data marts, el cual incluye el modelo dimensional, la arquitectura y los procesos de extracción y explotación del proyecto.

3.1. Modelo Dimensional

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

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

Las diferencias entre ambos modelos son:

iniESTRELLA	COPO DE NIEVE
	
Modelo más simple	Modelo más complejo
Desnormalizado	Forma normalizada de las dimensiones (solo las dimensiones primarias están enlazadas con la tabla de hechos)
Habilidad para análisis dimensional	Rompe el análisis dimensional
Se usa con una cantidad de datos moderada	Se usa con gran cantidad de datos
Optimiza el tiempo de respuesta de base de datos	La consulta se realiza en más tiempo ya que se hace referencia a más tablas
Sentencia SQL simple	Sentencia SQL más compleja

Cuadro 3.1 Cuadro comparativo modelo estrella vs modelo copo de nieve

Con el modelo estrella la consulta se realiza accediendo a un menor número de tablas lo cual optimiza el tiempo de respuesta. La clave primaria de una tabla de hechos está formada principalmente por las llaves primarias de las dimensiones. Lo característico de la arquitectura de estrella es que sólo existe una tabla de dimensiones para cada dimensión y esta tabla representa la segunda forma normal.

El modelo copo de nieve es una variación del modelo estrella y usualmente se usa cuando no se puede implementar el modelo estrella.

Luego de realizar la comparación entre ambos modelos se decide usar el modelo estrella dado que la solución está dirigida a pequeñas y medianas empresas las cuales se caracterizan por manejar una menor cantidad de datos. Además con el uso del modelo estrella se optimiza el tiempo de respuesta de los reportes en comparación del modelo copo de nieve debido a que resulta ser un modelo más complejo.



3.1.1. Diagrama del Modelo Dimensional

El modelo dimensional está conformado por 8 dimensiones, 3 tablas de hechos y sus relaciones. A continuación se muestra el diagrama:

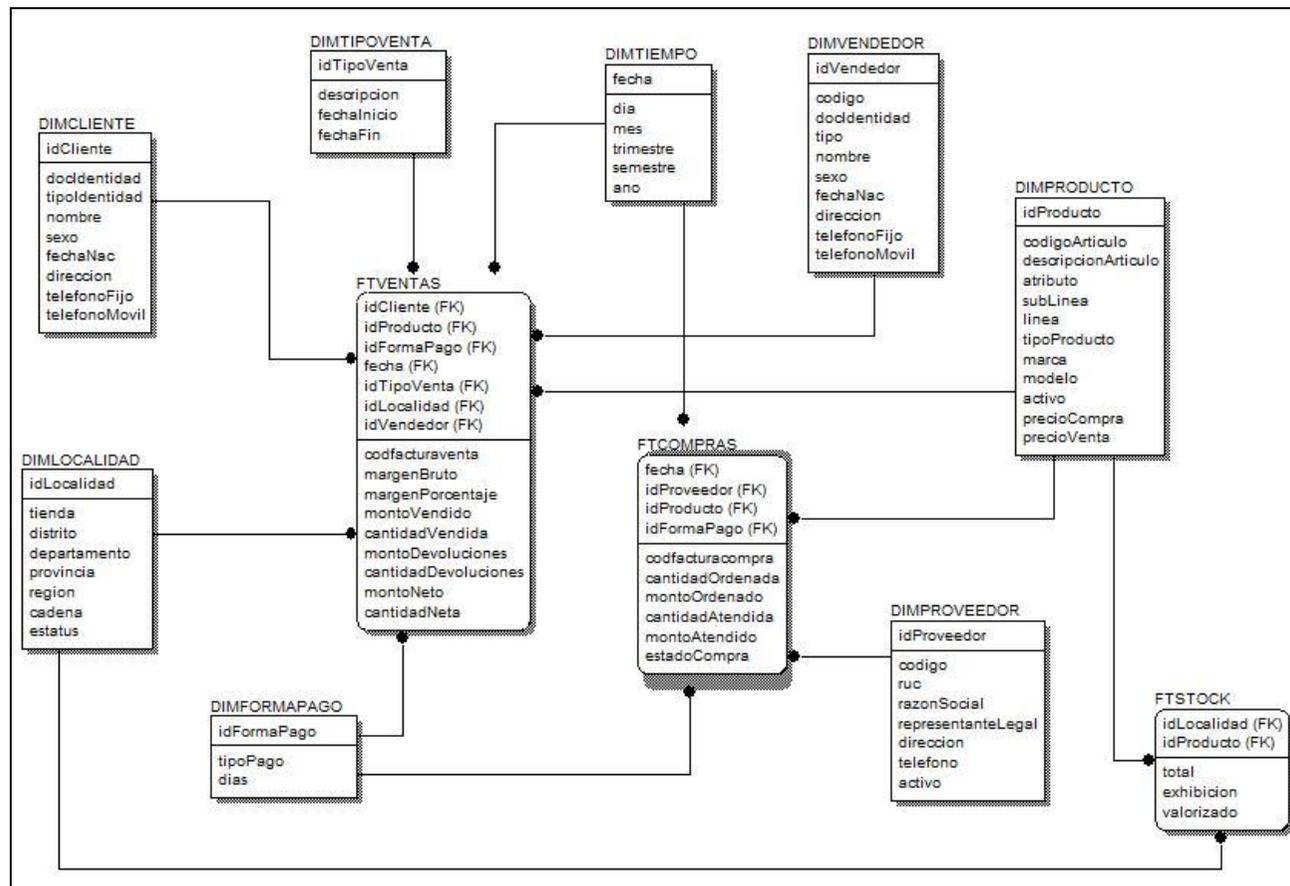


Figura 3.1 Diagrama del Modelo Dimensional

3.1.2. Dimensiones

El modelo dimensional del proyecto está compuesto por las siguientes dimensiones:

Dimensión Localidad.- Comprende la tabla Localidad y se refiere a la red de tiendas ubicadas en el ámbito nacional; en este caso se crea una jerarquía con la dimensión Localidad debido a que se tiene varios niveles de asociación de datos.

A continuación se muestra la jerarquía de esta dimensión:

- Tienda
 - Distrito
 - Departamento
 - Provincia
 - Región
 - Cadena

Dimensión Tiempo.- Comprende la tabla Tiempo, se ha creado niveles jerárquicos de tiempo como se muestra a continuación:

- Día
 - Mes
 - Trimestre
 - Semestre
 - Año

Dimensión Producto.- La dimensión de Producto contiene las características de los artículos que ofrece la empresa. Cada tipo de producto pertenece a una sub línea y esta a su vez a una línea, por ejemplo la línea Blanca tiene como sub líneas:

refrigeradoras, lavadoras, microondas, entre otras. A continuación se muestra la jerarquía para esta dimensión:

- Artículo
 - Sub línea
 - Línea

Dimensión Cliente.- Es la dimensión que contiene las características de un cliente, como: nombre y apellidos, documento de identidad, fecha de nacimiento, sexo, dirección, teléfono, entre otros.

Dimensión FormaPago.- La dimensión contiene las diferentes formas de pago para la compra y venta de los productos, siendo una característica principal la cantidad de días que se tiene como plazo para pagar a los proveedores. Por ejemplo crédito a 30 días.

Dimensión Proveedor.- Contiene las características de la empresa que provee de electrodomésticos a la comercializadora. La dimensión proveedor contiene las siguientes características: RUC, razón social, representante legal, entre otros.

Dimensión Vendedor.- Es la dimensión que almacena a los vendedores que trabajan en la empresa comercializadora. Contiene las siguientes características: nombre y apellidos, documento de identidad, fecha de nacimiento, entre otros.

Dimensión TipoVenta.- La tabla contiene los tipos de venta que se realizan en la empresa. Por ejemplo si se trata de una promoción o campaña por periodo. Por ejemplo: promoción del día de la madre, promoción del día del padre, entre otras.

3.1.3. Tablas de Hechos

El modelo dimensional tiene las siguientes tablas de hechos:

Tabla de Hechos FTStock.- Comprende la cantidad de productos disponibles por cada localidad de la empresa. Esta tabla de hechos tiene como métrica el monto en que está valorizado un producto. La información contenida en esta tabla permite la toma de decisiones para la compra y venta de electrodomésticos.

Tabla de Hechos FTCompras.- Contiene las características de las compras realizadas por la empresa comercializadora, cuantificándolas por monto y unidades. Cada compra realizada pasa por una serie de estados hasta que se concrete el proceso de compra. Se indica el tipo de compra realizado el cual depende de las ofertas realizadas por los proveedores, entre ellos: campaña navideña, fiestas patrias, campaña escolar, día de la madre.

Tabla de Hechos FTVentas.- Contiene las características de las ventas realizadas por la empresa comercializadora, cuantificándolas por monto y unidades. En esta tabla se registra el número de ventas y devoluciones efectuadas, con lo cual se calcula el número de ventas netas realizadas. La información de ventas permite conocer qué productos tienen mayor rotación por localidad, estacionalidad y mayor demanda.

3.1.4. Estándares del Modelo – Nomenclatura

Dimensiones: Las tablas que representan las dimensiones llevan la nomenclatura DIM + NOMBRE DE DIMENSION

Tablas de Hechos: Estas tablas tienen como nomenclatura: FT + NOMBRE DE TABLA DE HECHOS

Llave Primaria: Las llaves primarias de las dimensiones llevan la nomenclatura: id + Nombre de la dimensión. A excepción de la tabla DIMTIEMPO, la cual usa el campo fecha como llave primaria.

3.2. Arquitectura

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

La funcionalidad de los sistemas OLAP se caracteriza por ser un análisis multidimensional de datos corporativos, que soportan los análisis del usuario y unas posibilidades de navegación, seleccionando la información a obtener.

A continuación se realiza una comparación entre dos tipos de arquitectura OLAP, arquitectura MOLAP y arquitectura ROLAP:

	MOLAP	ROLAP
Ventajas	Excelente rendimiento: los cubos MOLAP son contruidos para tener una rápida recuperación de datos y esta optimizado para operaciones de <i>slicing</i> & <i>dicing</i> (rebanar y dividir).	Puede almacenar grandes cantidades de datos. La limitante de tamaño en la tecnología ROLAP es la limitante de la base de datos relacional, es decir ROLAP en sí misma no está limitada.

	MOLAP	ROLAP
Ventajas	Puede realizar cálculos complejos: ya que todos los cálculos han sido pre generados cuando el cubo se crea. Por lo tanto los cálculos complejos se almacenan y regresan su resultado rápidamente.	Puede cubrir funcionalidad inherente a las bases de datos relacionales, las cuales ya vienen con un set de funciones que son heredables ya que esta tecnología se monta sobre esta base de datos.
Desventajas	Limitado en la cantidad de datos a ser manejados. Porque todos los cálculos son construidos cuando se genera el cubo, no es posible incluir grandes cantidades de datos en el cubo en sí mismo. Esto no quiere decir que los datos del cubo no deriven de una gran cantidad de datos. Sí es posible, pero en este caso, sólo la información de alto nivel puede ser incluida en este.	Bajo rendimiento. Ya que cada informe ROLAP es esencialmente una o múltiples consultas en la base de datos relacional, el tiempo de respuesta se alarga si el tamaño de la base de datos es mayor. Está limitada a la funcionalidad SQL ya que la tecnología ROLAP utiliza básicamente sentencias SQL o <i>queries</i> (consultas) de la base de datos relacional, y SQL no aporta todas las necesidades de consultas multidimensionales.

Cuadro 3.2 Cuadro comparativo arquitectura MOLAP vs arquitectura ROLAP

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

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

Se utiliza como herramienta el motor ROLAP de la plataforma Pentaho denominado Mondrian. Las expresiones MDX (compatibles con Mondrian) pre compiladas permiten al servidor Mondrian optimizar consultas, seleccionando entre evaluar una expresión en memoria o enviar la consulta a la base de datos relacional.

Además se utiliza un asistente para diseño de cubos basado en eclipse: Cube Designer. Este ayuda en el proceso de creación de cubos facilitando la definición de las dimensiones, los niveles de jerarquía, las tablas de hechos, y la conexión a la base de datos relacional en un fichero XML.

La arquitectura conceptual básica comprende los componentes para la extracción y explotación de datos, necesarios para la construcción de los Data Marts de Compras y Ventas.

En el siguiente gráfico se muestran los componentes de dicha arquitectura:

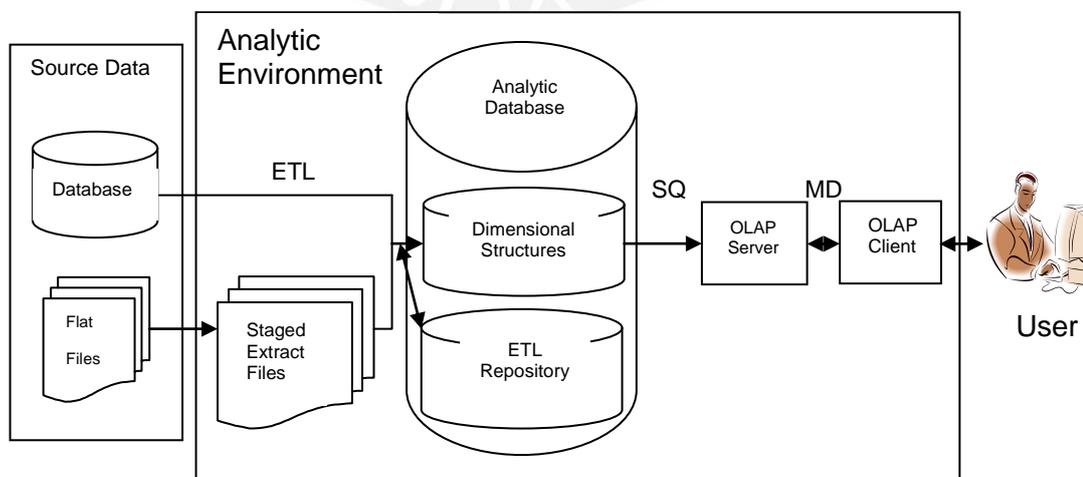


Figura 3.2 Arquitectura conceptual básica. **Fuente:** Elaboración propia

A continuación se definen cada uno de los componentes de la arquitectura:

- **Datos Fuentes (Source Data).**- Son las fuentes de origen de datos, dentro de las cuales tenemos el sistema transaccional de la empresa, archivos de texto y hojas de cálculo. Son los datos necesarios para el análisis. Los datos fuentes pueden ser extraídos directamente o en lote. Se cuenta con las siguientes fuentes de datos para el desarrollo del proyecto:

Fuente de Datos	Tipo	Conexión
Database	Base de datos relacional	OLDB
Flat Files	Hoja de cálculo	

- **ETL.**- “Extraer” datos de fuentes. “Transformar” los datos a través de limpieza, formateo e integración. “Cargar” (*Load* en inglés) los datos en una base de datos analítica optimizada.
- **Base de Datos Analítica (Analytic Database).**- RDBMS en la cual reside las estructuras de datos dimensionales (también conocido como diseño estrella). Optimizada para acceso OLAP.
- **Servidor OLAP (OLAP Server).**- Procesa consultas MDX, retorna resultados multidimensionales. Puede emitir consultas SQL al RDBMS de base de datos analíticas. Manejo de Cachés optimizado y construcción de agregados para mejorar rendimiento. Manejo de estados para sesiones de usuario final.

- **Cliente OLAP (OLAP Client).**- Herramienta de usuario final que provee segmentación, navegación y pivoteo wysiwyg (*what you see is what you get*). Genera MDX dinámicamente a partir de las peticiones de los usuarios. Generalmente provee una metáfora de “hoja de cálculo” para manipulación de datos y tablas rudimentarias.

3.3. Diseño de Extracción

El diseño de extracción de datos comprende la carga de los data marts desde las tablas fuente transaccionales.

Se muestra la extracción y carga de la dimensión Producto a modo de ejemplo:

3.3.1. Carga de la Dimensión Producto

La dimensión producto contiene los atributos y jerarquías de los productos para la compra y venta de la empresa comercializadora de electrodomésticos.

En esta etapa surgen problemas dada la falta de integración y estructura consistente (alineada) entre los sistemas fuentes.

3.3.2. Descripción de Tablas Fuentes

Son las fuentes desde donde se extraen los datos de los productos para alimentar la dimensión producto. Adicionalmente para cargar los nuevos productos que lanza la empresa al mercado se usa una hoja de cálculo.

Tipo de Fuente	Nombre de Fuente	Descripción
Base de datos transaccional	Tableproducto	Esta tabla contiene el listado de productos que utiliza el sistema para realizar compras y ventas de la empresa comercializadora.
Hoja de cálculo	Productos	Es el archivo plano recibido periódicamente con el listado de nuevos productos que ingresan al mercado.

3.3.3. Estandarización de Datos y Limpieza de Datos

Especifica la nomenclatura y transformación que deben tener los datos antes de almacenarlos en los data marts. El ETL verifica los datos que ingresan al datamart, realizando una limpieza de estos si se requiere.

Nombre	Llave	Tipo	Formato	Limpieza	Valor por Defecto
idProducto	PK	Integer	Número entero	No Nulo	0
Línea		Varchar(50)	Texto		
Sublínea		Varchar(50)	Texto		
Atributo		Varchar(50)	Texto		
Codigoarticulo		Varchar(50)	Texto	No Nulo	
descripcionarticulo		Varchar(100)	Texto		
Tipoproducto		Varchar(50)	Texto		
Marca		Varchar(50)	Texto		
Modelo		Varchar(50)	Texto		
Activo		Char(1)	Carácter	Sólo puede ser "S" o "N"	S
Precio compra		Numeric(20,5)	Número real	>0	
Precio venta		Numeric(20,5)	Número real	>0	

3.3.4. Fuente de Datos

Se muestra las características de la tabla fuente tableproducto que se encuentra en la base de datos transaccional. La consideración importante indicada en el siguiente cuadro se toma en cuenta para la transformación de los datos.

Tabla:		Tableproducto		
Nombre	Llave	Tipo	Formato	Consideración Importante
Codigoarticulo	PK	Varchar(50)	Texto	Valor Numérico, borrar espacios en blanco
Descripcionarticulo		Varchar(100)	Texto	
Atributo		Varchar(50)	Texto	
Marca		Varchar(50)	Texto	
Modelo		Varchar(50)	Texto	
Línea		Varchar(50)	Texto	
Sublínea		Varchar(50)	Texto	
Activo		Varchar(50)	Texto	Sólo puede contener "S" o "N", de lo contrario asumir "S"
Tipoproducto		Varchar(50)	Texto	
Preciocompra		Numeric(20,5)	Número real	Número entero positivo, considerar 2 decimales
Precioventa		Numeric(20,5)	Número real	Número entero positivo, considerar 2 decimales

3.3.5. Tabla Destino

El mapeo de los datos se hace partiendo de los datos una vez "limpios". Se transforma los datos de acuerdo con las reglas de negocio y los estándares establecidos. La transformación incluye: cambios de formato, sustitución de códigos, valores derivados y agregados. Los agregados, como por ejemplo la suma

de las venta, normalmente se precalcular y se almacenan para conseguir mayores rendimientos cuando lanzamos las consultas que requieren el cálculo de totales al datamart de ventas. En el siguiente cuadro se muestra las características de la tabla destino Dimproducto que se utiliza en los data marts de compras y ventas.

Tabla:	Dimproducto	
Campo	Tipo	Mapeo
Idproducto	Integer	Seqproducto
Línea	Varchar(50)	tableproducto.linea
Sublínea	Varchar(50)	tableproducto.sublínea
Atributo	Varchar(50)	tableproducto.atributo
Codigoarticulo	Varchar(50)	tableproducto.codigoarticulo
descripcionarticulo	Varchar(100)	tableproducto.descripcionarticulo
Tipoproducto	Varchar(50)	tableproducto.tipo
Marca	Varchar(50)	tableproducto.marca
Modelo	Varchar(50)	tableproducto.modelo
Activo	Char(1)	tableproducto.activo
Preciocompra	Numeric(20,5)	tableproducto.preciocompra
Precioventa	Numeric(20,5)	tableproducto.precioventa

3.3.6. Proceso

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

Paso1: Extracción de los datos de la tabla Producto

Se toma los datos de la tabla OLTP tableproducto y se realiza la transformación de acuerdo al mapeo indicado.

Paso 2: Extracción de datos del archivo Excel de nuevos productos

Para ingresar nuevos productos que aún no se encuentran en el OLTP, se usa una hoja de cálculo. Se extraen los datos del archivo.

Paso 3: Carga de la Dimensión Producto

Se transforma e ingresan los datos en la dimensión Producto. En caso que sean nuevos productos insertarlos, en caso que sean antiguos productos sólo se actualizan los datos.

3.3.7. Esquema de Extracción

El esquema de extracción permite conocer el orden en que se realiza la carga de las dimensiones y tablas de hecho. El esquema de extracción que se usa para la dimensión producto es:

N°	Proceso	Dependencia
1	Carga Tabla Producto	
2	Carga Nuevos Productos	1
3	Fact Compras	2
4	Fact Ventas	2
5	Fact Stock	2

El detalle de la carga de las demás dimensiones se encuentra en el Anexo C: Diseño de extracción.

3.4. Diseño de explotación para el área de compras

El diseño de explotación difiere para cada una de las áreas por la información mostrada. Para el área de compras se tiene 4 reportes:

Tema	Reporte
Análisis de Compras	Compras realizadas
	Slow moving
	Stock de productos
	Compras a realizar

A modo de ejemplo se detalla el reporte de compras realizadas:

3.4.1. Reporte “Compras Realizadas”

Reporte de compras realizadas en cantidad y monto en soles. Este reporte permite usar filtros por periodo (mes, año) y línea de producto. La información mostrada es a nivel forma de pago.

3.4.1.1. Diseño

El siguiente reporte permite decidir con que proveedor trabajar y cuál es la mejor forma de pago. Esta información ayuda en la negociación con el proveedor al momento de una compra. A continuación se muestra el diseño del reporte de compras realizadas:

COMPRAS REALIZADAS			
Proveedor	Forma Pago	Cantidad	Costo (S/.)
HP	CONTADO	40	20,000
	CREDITO	80	24,560
LG	CREDITO	66	19,398
SAMSUNG	CONTADO	25	17,500
	CREDITO	256	96,098

Filtros Línea: Blanca, Año: 2008

Figura 3.3 Diseño de reporte de compras realizadas

3.4.1.2. Tipo

El tipo del reporte es tabla.

3.4.1.3. Filas

La tabla no contiene cabeceras por fila.

3.4.1.4. Columnas

Las columnas que se muestran en el reporte son:

No.	Dimensión	Nivel / Categoría
1	DimProveedor	Proveedor
2	DimFormaPago	Forma Pago

3.4.1.5. Medidas

Los cálculos que se realizan en el reporte son:

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

3.4.1.6. Filtros

Los filtros usados en el reporte son:

No.	Operación
1	Línea de producto = <Línea de producto seleccionado>
2	Año = <Año ingresado>

El detalle de los demás reportes de compras se encuentra en el Anexo D: Diseño de explotación para el área de compras.

3.5. Diseño de explotación para el área de ventas

Para el área de ventas se tienen los siguientes reportes:

Tema	Reporte
Análisis de Ventas	Ventas realizadas
	Precio de venta al público
	Margen teórico vs. margen real
	Ventas a 30 días

A modo de ejemplo se detalla el reporte de ventas realizadas.

3.5.1. Reporte “Ventas Realizadas”

Reporte de ventas realizadas en unidades y monto en soles. Este reporte permite usar filtros por periodo (mes, año) y línea de producto. La información mostrada es a nivel producto y año.

3.5.1.1. Diseño

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

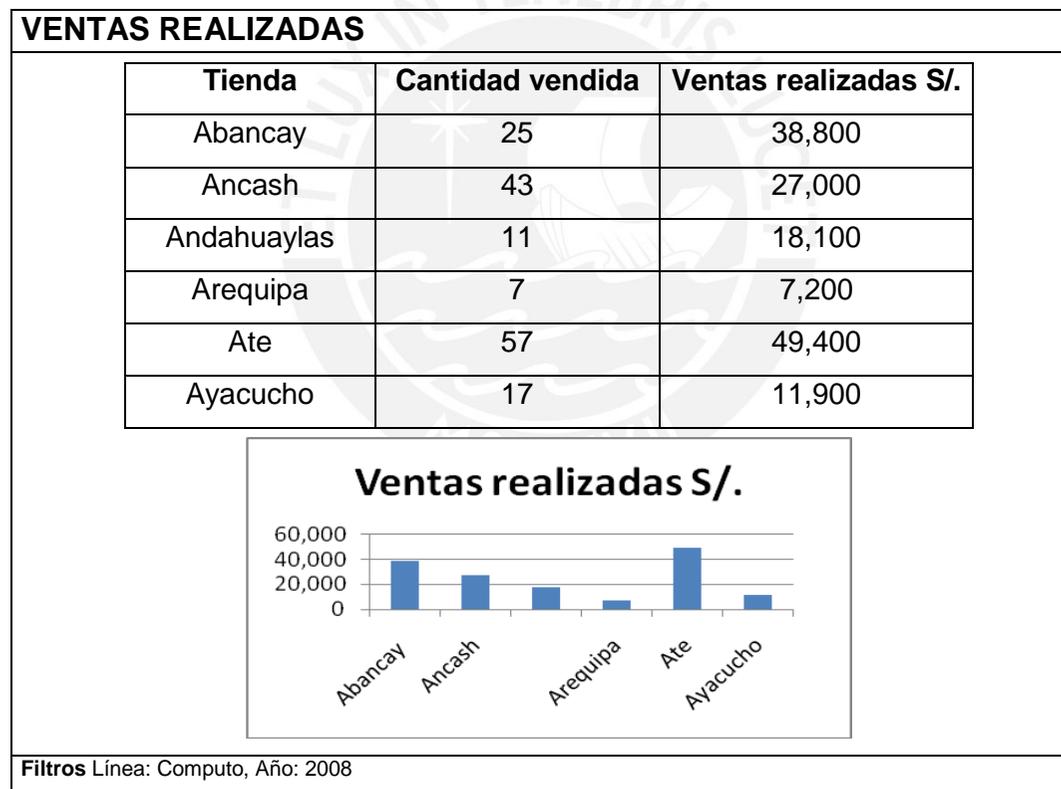


Figura 3.3 Diseño de reporte de ventas realizadas

3.5.1.2. Tipo

El tipo del reporte es gráfico y tabla.

3.5.1.3. Filas

La tabla no contiene cabeceras por fila.

3.5.1.4. Columnas

Las columnas que tiene el reporte de ventas realizadas:

No.	Dimensión	Nivel / Categoría
1	DimLocalidad	Tienda

3.5.1.5. Medidas

Los cálculos que se realizan para el reporte:

No.	Medida	Formato
1	Cantidad vendida	Numérico
2	Monto vendido	Moneda (S/.)

3.5.1.6. Filtros

Los filtros usados en el reporte:

No.	Operación
1	Línea de producto = <Línea de producto seleccionado>
2	Año = <Año ingresado>

El detalle de los demás reportes de ventas se encuentra en el Anexo E: Diseño de explotación para el área de ventas.

4. Construcción y Pruebas

Este capítulo describe la fase de construcción y prueba de los *Data Marts*, lo cual incluye configuración de todo el software utilizado, construcción de procesos de carga, construcción de reportes y ejecución de pruebas del proyecto.

4.1. Construcción

4.1.1. Instalación y configuración del software

Aquí se muestran las instrucciones a tomar en cuenta para la instalación y configuración de la herramienta de inteligencia de negocios que se usa para la implementación del proyecto. Para el proyecto se usa *Open Source*. Como repositorio del *Data Mart* se usa PostgreSQL y como herramienta de análisis se tiene la plataforma Pentaho Open Source Business Intelligence.

4.1.1.1. Instalación y configuración de PostgreSQL

El gestor de base de datos PostgreSQL es de software libre, para obtenerlo se descarga desde la página web: <http://www.postgresql.org/download>

El proceso de instalación y configuración es amigable ya que es realizado por un instalador Wizard.

Para configurar PostgreSQL primero se crea una cuenta de servicio donde se ingresan los datos mostrados en la siguiente figura:

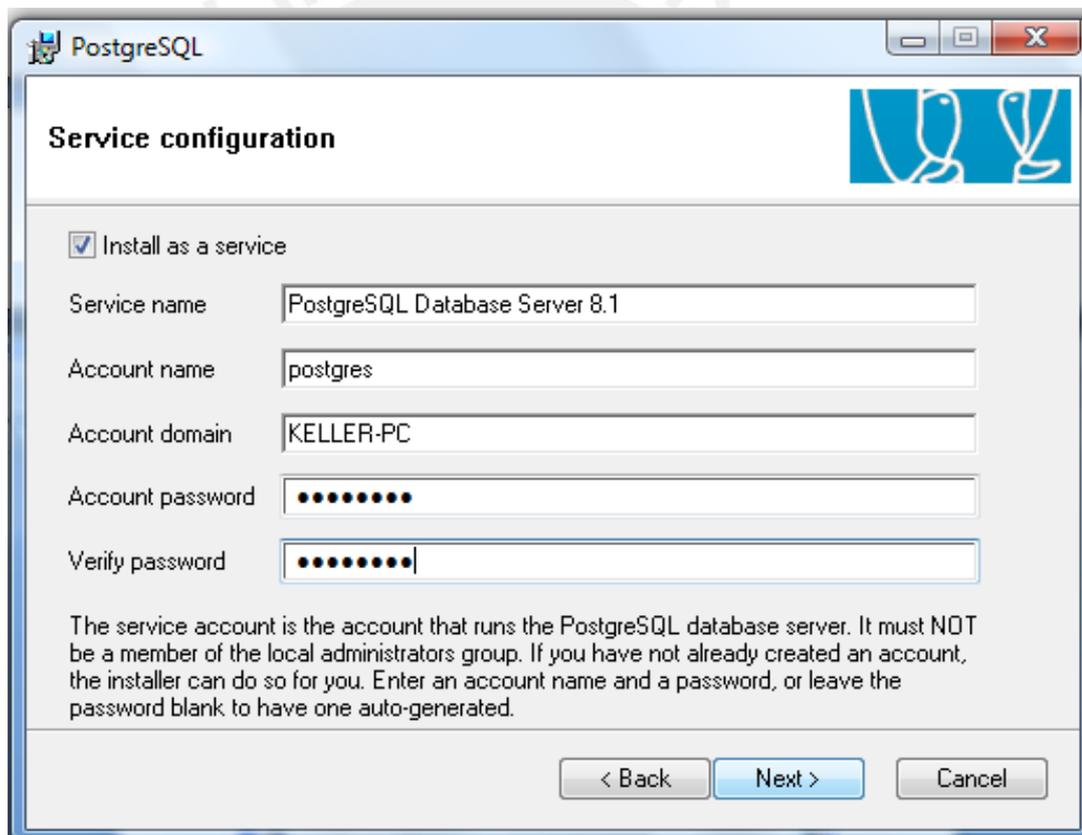


Figura 4.1 Configuración de PostgreSQL

La forma de instalación y configuración de PostgreSQL paso a paso se encuentra en el Anexo F: Instalación y configuración de PostgreSQL.

4.1.1.2. Instalación y configuración de BI Suit Pentaho

Se empieza ejecutando el instalador de BI Suite Pentaho, luego se configura la base de datos ingresando el ID Usuario y el *password* del repositorio solución tal como se muestra en la siguiente figura:



Figura 4.2 Instalación de Pentaho 1

Se valida la información mostrada acerca de la configuración y finalmente se muestra un mensaje donde se indica que está todo listo para iniciar la configuración de Pentaho BI Suite con lo que se concluye la instalación. Dicho mensaje se muestra en la siguiente figura:

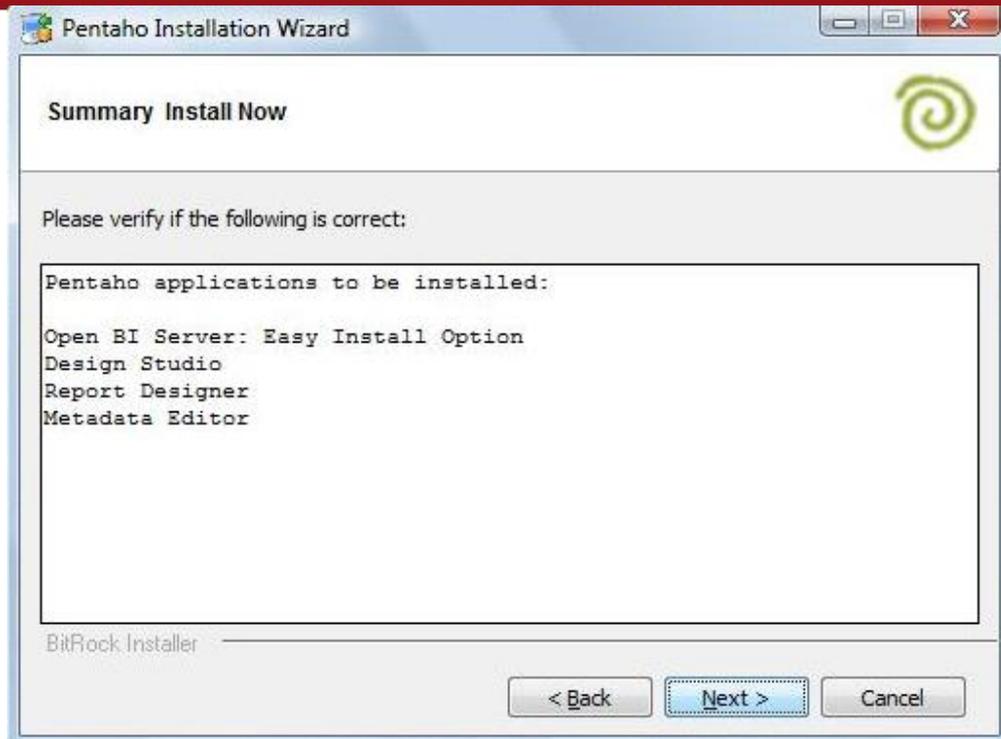


Figura 4.3 Instalación de Pentaho 2

4.1.1.3. Instalación y configuración del servidor Mondrian Pentaho

Mondrian es una de las aplicaciones más importantes de la plataforma Pentaho BI. Se instala en conjunto con la plataforma BI Suit Pentaho. Mondrian es un servidor OLAP de software libre que comunica la base de datos fuente y la aplicación OLAP (escrita en Java).

4.1.1.4. Instalación y configuración del Kettle

Los procesos de carga se realizan con la herramienta Kettle (*K Extraction Transformation Transportation Load E*) que usa la interfaz gráfica Spoon, la cual es un componente de Pentaho Data Integration.

La herramienta Spoon se obtiene de manera gratuita de la página web de Pentaho: www.pentaho.com/download y se copia en la PC del desarrollador para su uso.

Para empezar a usar la herramienta sólo se ejecuta el archivo “spoon.bat”.

Es necesario crear una conexión al repositorio que se encuentra en la base de datos PostgreSQL como se muestra en la siguiente figura:

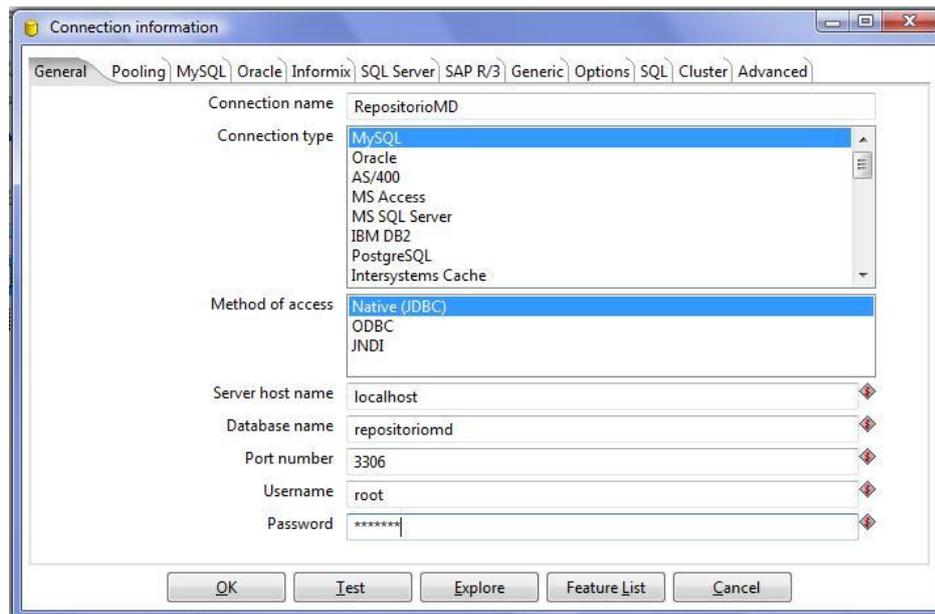


Figura 4.4 Configuración Spoon

En el Anexo G: Instalación y configuración de Kettle se encuentra la instalación y configuración paso a paso de Kettle.

4.1.1.5. Instalación y configuración de Cube Designer

Luego de haberse descargado e instalado el Pentaho BI Suite se procede a la instalación de Cube Designer. Esta herramienta se obtiene de manera gratuita de la página web de Pentaho: <http://www.pentaho.org/download/latest>.

Para instalarla se ejecuta el archivo “CubeDesigner.exe” y se muestra la siguiente figura:

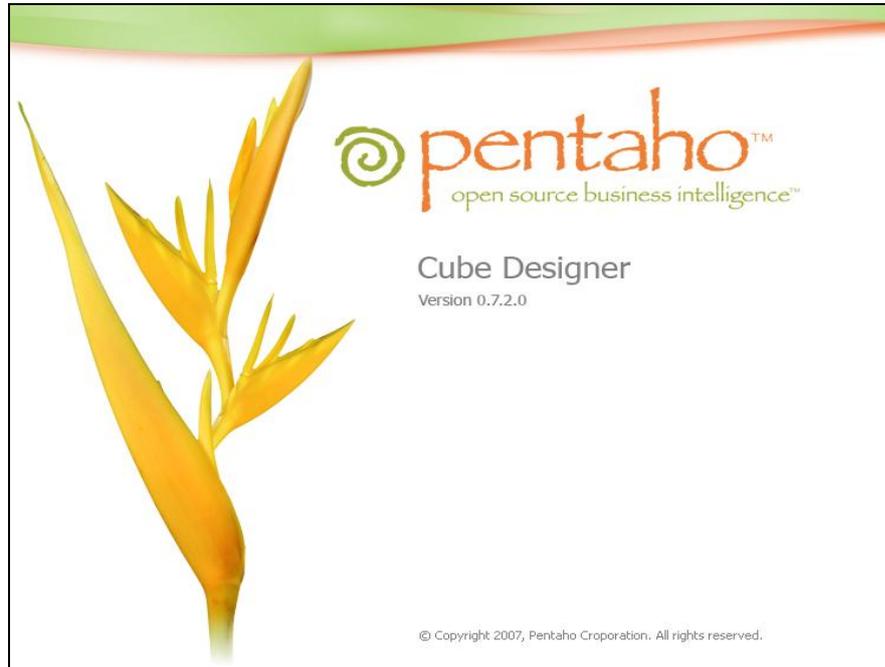


Figura 4.5 Ejecución de Cube Designer

4.1.2. Construcción de Procesos de Carga

A modo de ejemplo se crea el proceso de carga de la dimensión Producto.

Primero se arrastran los elementos al área de trabajo y se relacionan entre sí mediante un objeto llamado Hop como se muestra en la siguiente figura:

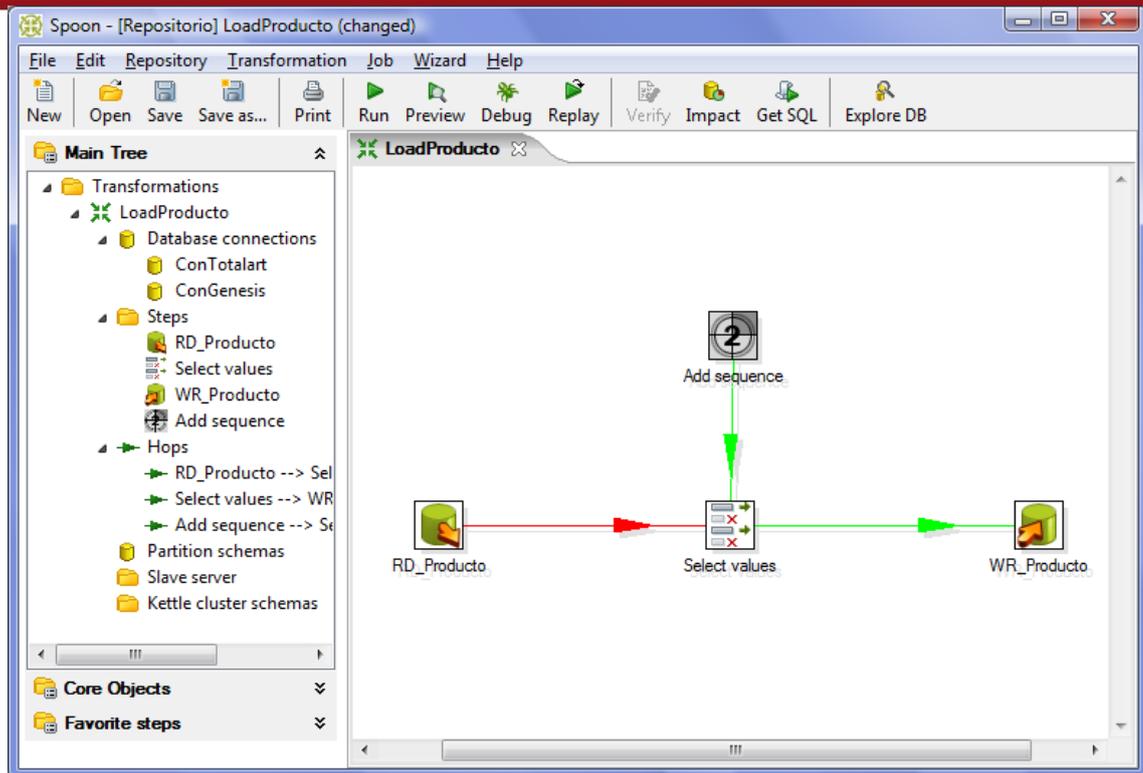


Figura 4.6 Diseño proceso de carga

Los elementos usados para el proceso de carga son los siguientes:

- Table Input: a este elemento se le asigna el nombre RD_Producto el cual contiene la sentencia SQL donde se consulta la tabla origen.
- Sequence_Producto: este elemento se encarga de generar la secuencia correlativa del id_producto (llave primaria de la tabla producto).
- Select values: en este elemento se realiza el mapeo de cada valor que ingresa al data Mart.
- Table Output: a este elemento se le asigna el nombre WR_Producto el cual contiene la sentencia SQL donde se consulta la tabla destino (*Data Mart*).

Este proceso de carga se guarda con el nombre LoadProducto.

4.1.3. Construcción de Reportes

Consiste en el desarrollo de reportes corporativos utilizando la herramienta Cube Designer. Para ello se toma la información contenida en el Datamart, se construyen cubos acorde a los requerimientos de la empresa y como producto final se construyen los reportes. Luego de este desarrollo, se ejecuta las pruebas y publicación de reportes.

Para iniciar el desarrollo de un reporte se ejecuta el archivo “CubeDesigner.exe” y dentro de la opción “File” se selecciona “New Cube Schema” tal como se muestra en la siguiente figura:

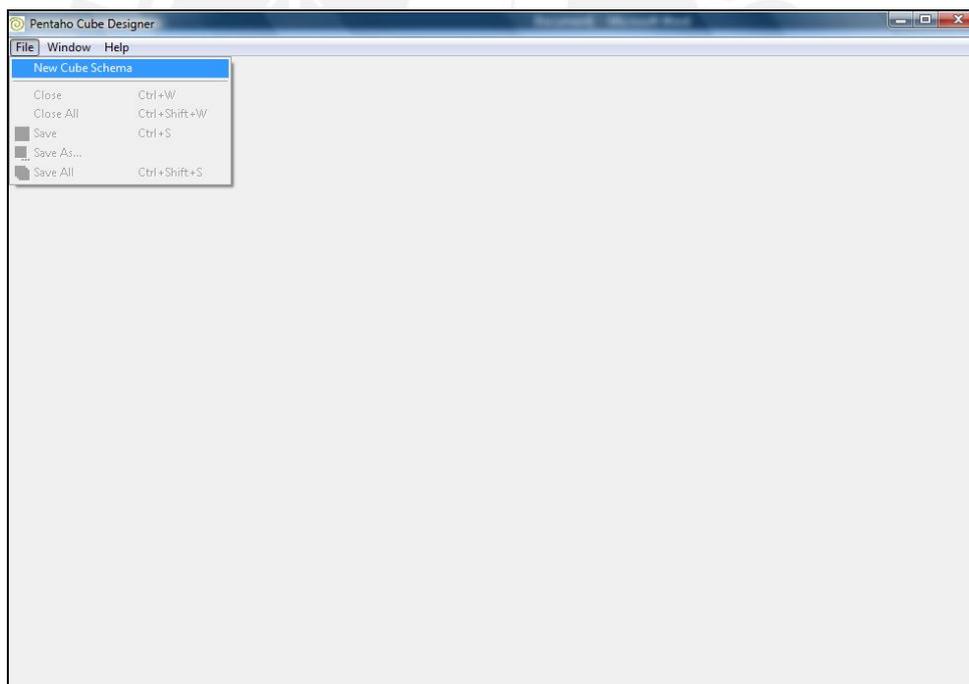


Figura 4.7 Proceso de creación de reportes

La creación del reporte consta de 5 pasos, los cuales se pasan a detallar:

Paso 1:

Se comienza ingresando el nombre y descripción del cubo a diseñar. Con el objetivo de podernos conectar con el repositorio transaccional se ingresan algunos datos como: nombre JNDI, *driver*, cadena de conexión, usuario y clave de la base de datos como se muestra en la siguiente figura:

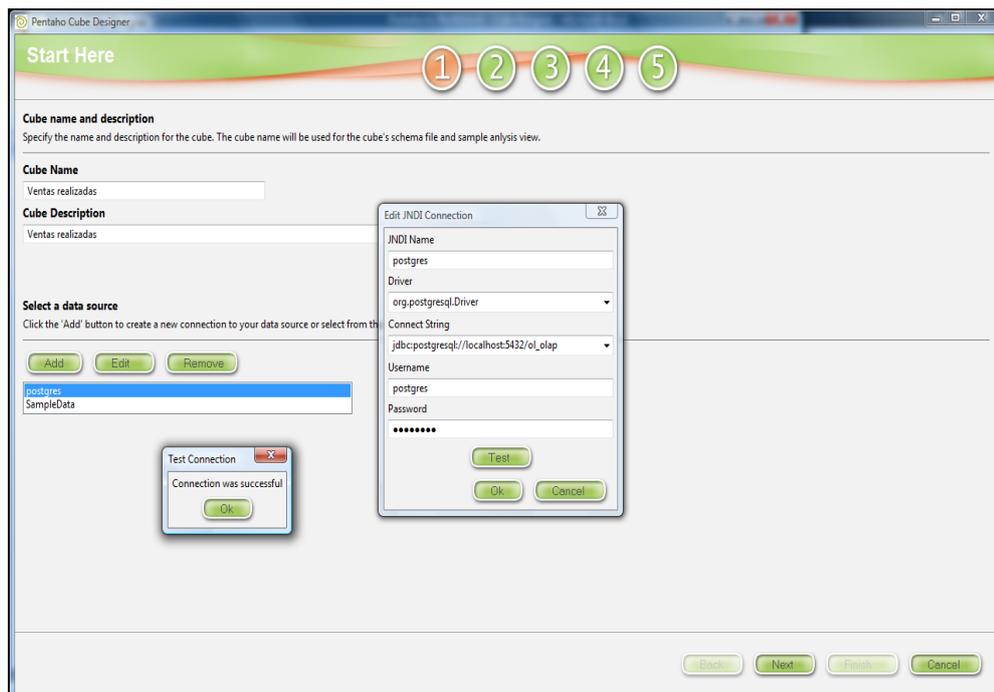


Figura 4.8 Proceso de creación de reportes – Paso 1

Paso 2:

Según el análisis que se desee realizar se eligen las dimensiones y *fact tables* involucradas, para luego seleccionar sus atributos como se indica en la siguiente figura:

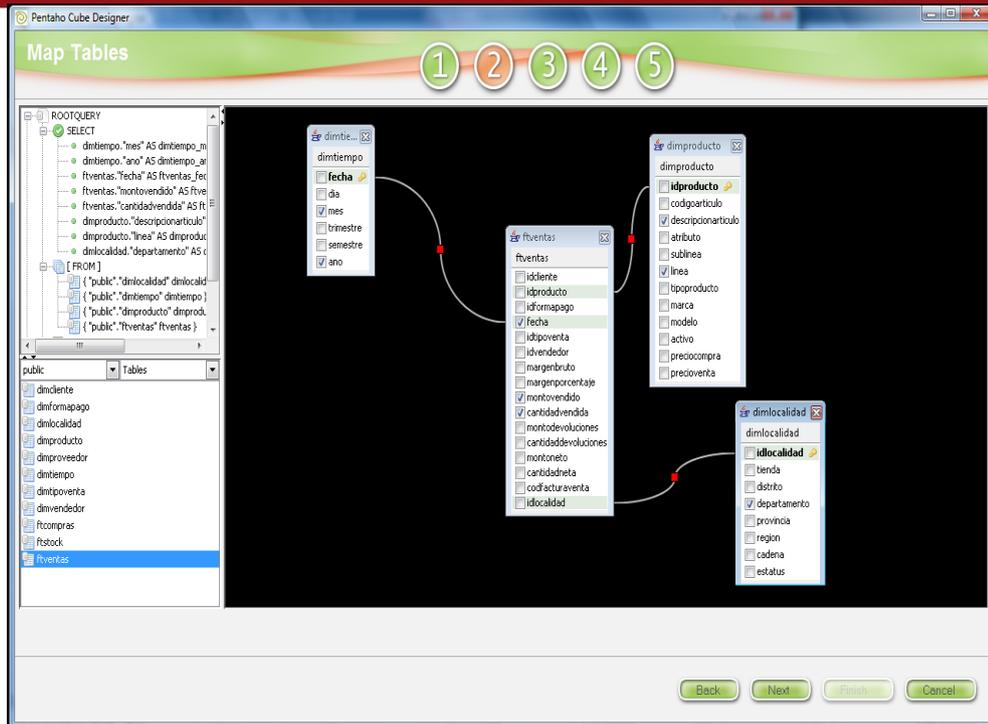


Figura 4.9 Proceso de creación de reportes – Paso 2

Paso 3:

En este paso se indican las métricas que contiene cada *fact table*, para ello primero se elige el *fact table* y se seleccionan las métricas pasándolas de la columna del *fact table* a la columna de Métricas tal como se muestra en la siguiente figura:

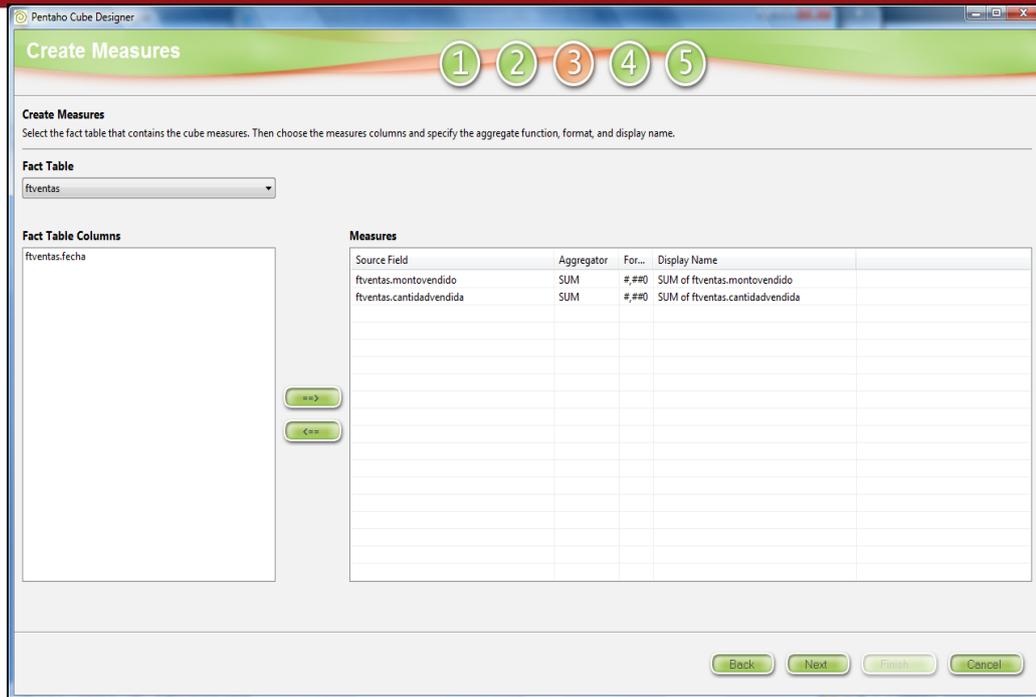


Figura 4.10 Proceso de creación de reportes – Paso 3

Paso 4:

En este paso se crean las dimensiones, se agregan una a una con sus respectivos atributos tal como lo muestra la siguiente figura:

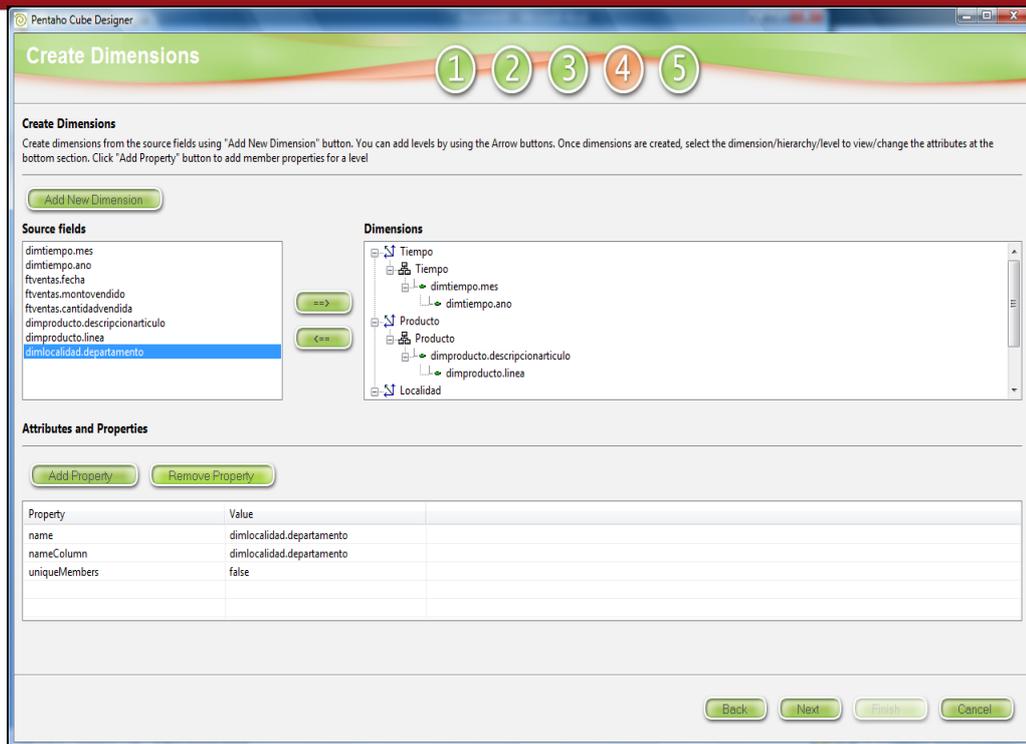


Figura 4.11 Proceso de creación de reportes – Paso 4

Paso 5:

Finalmente en este paso se publica el reporte en la interfaz web; para ello se ingresan los siguientes datos: carpeta donde se guarde el reporte, *publish password*, usuario y clave del servidor web como se muestra en la siguiente figura:

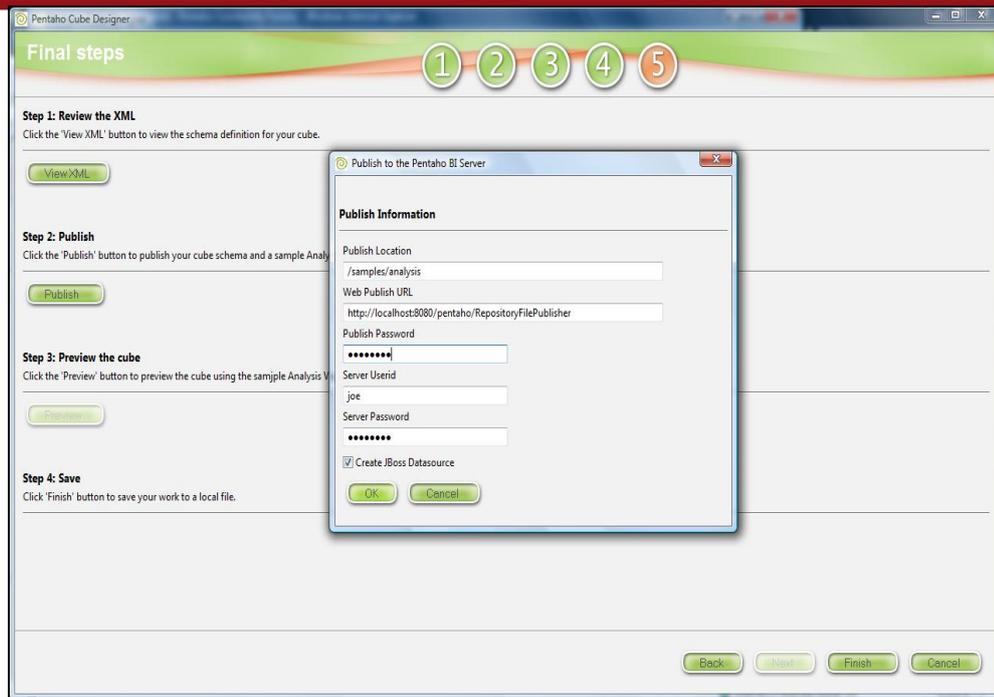


Figura 4.12 Proceso de creación de reportes – Paso 5

4.2. Pruebas

4.2.1. Pruebas de Ejecución de Procesos de Carga

Luego de la construcción de los procesos de carga se procede a su ejecución

mediante la opción  de la herramienta Spoon.

A modo de prueba se ejecuta el proceso LoadProducto y al terminar de correr el proceso la herramienta indica que este finaliza de manera satisfactoria como se muestra en la siguiente figura:

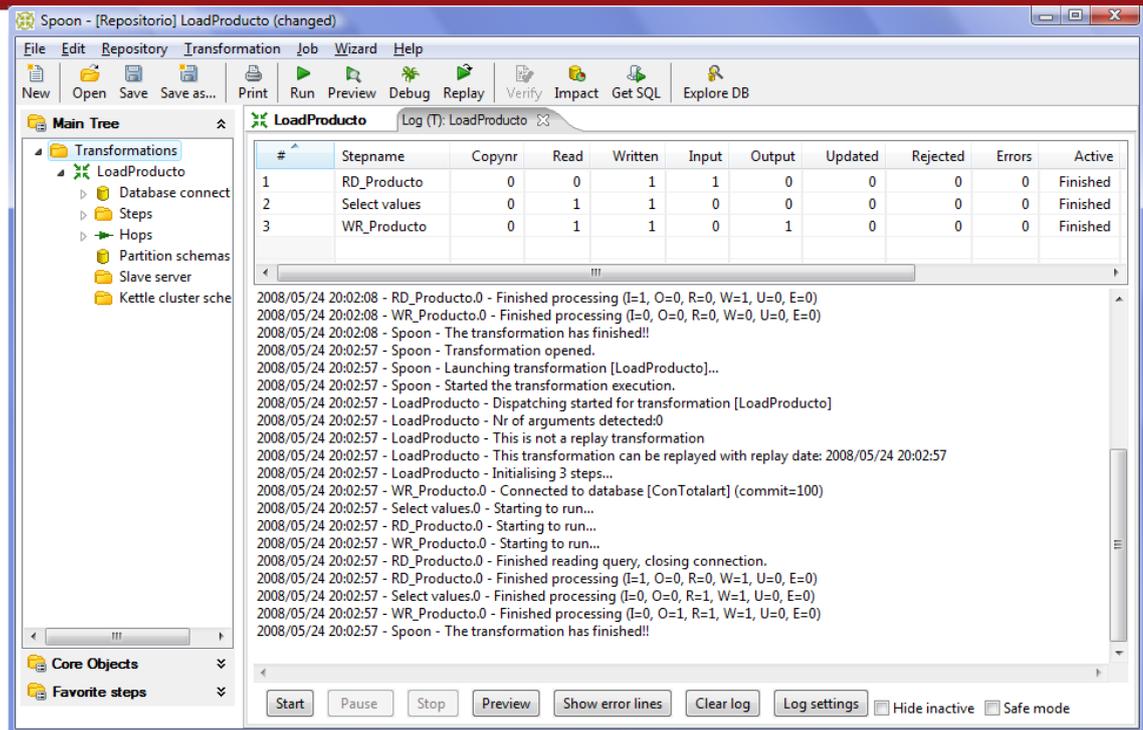
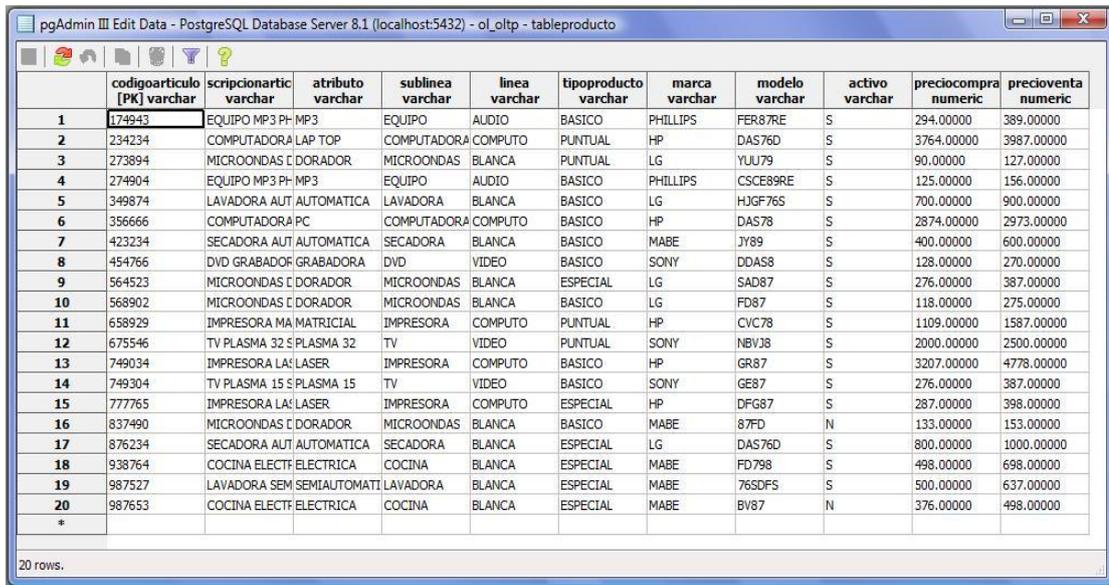


Figura 4.13 Proceso de carga satisfactoria

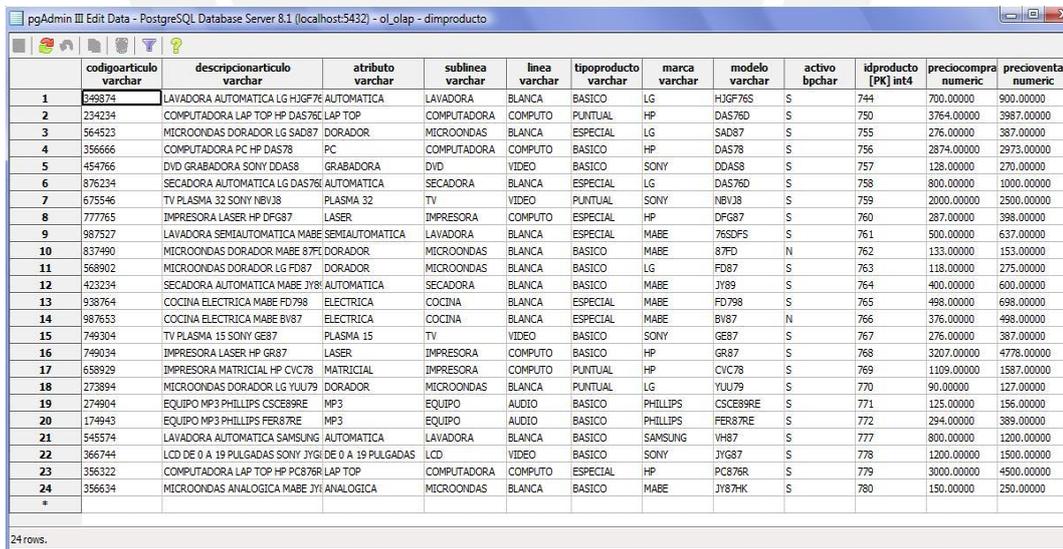
Para realizar la prueba de carga, primero se verifican los datos que se encuentran en la base fuente tableproductos de la base de datos OLTP. A continuación se muestran los datos contenidos en la OLTP fuente como se muestra en la siguiente figura:



	codigoarticulo [PK] varchar	descripcionarticulo varchar	atributo varchar	sublinea varchar	linea varchar	tipoproducto varchar	marca varchar	modelo varchar	activo varchar	preciocompra numeric	precioventa numeric
1	174943	EQUIPO MP3 PH MP3		EQUIPO	AUDIO	BASICO	PHILLIPS	FER87RE	S	294.00000	389.00000
2	234234	COMPUTADORA LAP TOP		COMPUTADORA	COMPUTO	PUNTUAL	HP	DAS76D	S	3764.00000	3987.00000
3	273894	MICROONDAS C DORADOR		MICROONDAS	BLANCA	PUNTUAL	LG	YUU79	S	90.00000	127.00000
4	274904	EQUIPO MP3 PH MP3		EQUIPO	AUDIO	BASICO	PHILLIPS	CSC89RE	S	125.00000	156.00000
5	349874	LAVADORA AUT AUTOMATICA		LAVADORA	BLANCA	BASICO	LG	HJGF76S	S	700.00000	900.00000
6	356666	COMPUTADORA PC		COMPUTADORA	COMPUTO	BASICO	HP	DAS78	S	2874.00000	2973.00000
7	423234	SECADORA AUT AUTOMATICA		SECADORA	BLANCA	BASICO	MABE	JY89	S	400.00000	600.00000
8	454766	DVD GRABADOR GRABADORA		DVD	VIDEO	BASICO	SONY	DDAS8	S	128.00000	270.00000
9	564523	MICROONDAS C DORADOR		MICROONDAS	BLANCA	ESPECIAL	LG	SAD87	S	276.00000	387.00000
10	568902	MICROONDAS C DORADOR		MICROONDAS	BLANCA	BASICO	LG	FD87	S	118.00000	275.00000
11	658929	IMPRESORA MA MATRICIAL		IMPRESORA	COMPUTO	PUNTUAL	HP	CVC78	S	1109.00000	1587.00000
12	675546	TV PLASMA 32 S PLASMA 32		TV	VIDEO	PUNTUAL	SONY	NBVJ8	S	2000.00000	2500.00000
13	749034	IMPRESORA LA LASER		IMPRESORA	COMPUTO	BASICO	HP	GR87	S	3207.00000	4778.00000
14	749304	TV PLASMA 15 S PLASMA 15		TV	VIDEO	BASICO	SONY	GE87	S	276.00000	387.00000
15	777765	IMPRESORA LA LASER		IMPRESORA	COMPUTO	ESPECIAL	HP	DFG87	S	287.00000	398.00000
16	837490	MICROONDAS C DORADOR		MICROONDAS	BLANCA	BASICO	MABE	87FD	N	133.00000	153.00000
17	876234	SECADORA AUT AUTOMATICA		SECADORA	BLANCA	ESPECIAL	LG	DAS76D	S	800.00000	1000.00000
18	938764	COCINA ELECTRICA		COCINA	BLANCA	ESPECIAL	MABE	FD798	S	498.00000	698.00000
19	987527	LAVADORA SEMI AUTOMATI		LAVADORA	BLANCA	ESPECIAL	MABE	76SDFS	S	500.00000	637.00000
20	987653	COCINA ELECTRICA		COCINA	BLANCA	ESPECIAL	MABE	BV87	N	376.00000	498.00000
*											

Figura 4.14 Datos del OLTP tableproducto

Luego de la ejecución del proceso de carga, se verifica que la tabla del OLAP dimproducto sean los datos que se encuentran en la base de datos OLTP:



	codigoarticulo varchar	descripcionarticulo varchar	atributo varchar	sublinea varchar	linea varchar	tipoproducto varchar	marca varchar	modelo varchar	activo bpchar	idproducto [PK] int4	preciocompra numeric	precioventa numeric
1	174943	EQUIPO MP3 PH MP3		EQUIPO	AUDIO	BASICO	PHILLIPS	FER87RE	S	744	294.00000	389.00000
2	234234	COMPUTADORA LAP TOP		COMPUTADORA	COMPUTO	PUNTUAL	HP	DAS76D	S	750	3764.00000	3987.00000
3	273894	MICROONDAS C DORADOR		MICROONDAS	BLANCA	ESPECIAL	LG	SAD87	S	755	276.00000	387.00000
4	274904	EQUIPO MP3 PH MP3		EQUIPO	AUDIO	BASICO	PHILLIPS	CSC89RE	S	756	125.00000	156.00000
5	349874	LAVADORA AUT AUTOMATICA		LAVADORA	BLANCA	BASICO	LG	HJGF76S	S	757	700.00000	900.00000
6	356666	COMPUTADORA PC		COMPUTADORA	COMPUTO	BASICO	HP	DAS78	S	758	2874.00000	2973.00000
7	423234	SECADORA AUT AUTOMATICA		SECADORA	BLANCA	BASICO	MABE	JY89	S	759	400.00000	600.00000
8	454766	DVD GRABADOR GRABADORA		DVD	VIDEO	BASICO	SONY	DDAS8	S	760	128.00000	270.00000
9	564523	MICROONDAS C DORADOR		MICROONDAS	BLANCA	ESPECIAL	LG	SAD87	S	761	276.00000	387.00000
10	568902	MICROONDAS C DORADOR		MICROONDAS	BLANCA	BASICO	LG	FD87	S	762	118.00000	275.00000
11	658929	IMPRESORA MA MATRICIAL		IMPRESORA	COMPUTO	PUNTUAL	HP	CVC78	S	763	1109.00000	1587.00000
12	675546	TV PLASMA 32 S PLASMA 32		TV	VIDEO	PUNTUAL	SONY	NBVJ8	S	764	2000.00000	2500.00000
13	749034	IMPRESORA LA LASER		IMPRESORA	COMPUTO	BASICO	HP	GR87	S	765	3207.00000	4778.00000
14	749304	TV PLASMA 15 S PLASMA 15		TV	VIDEO	BASICO	SONY	GE87	S	766	276.00000	387.00000
15	777765	IMPRESORA LA LASER		IMPRESORA	COMPUTO	ESPECIAL	HP	DFG87	S	767	287.00000	398.00000
16	837490	MICROONDAS C DORADOR		MICROONDAS	BLANCA	BASICO	MABE	87FD	N	768	133.00000	153.00000
17	876234	SECADORA AUT AUTOMATICA		SECADORA	BLANCA	ESPECIAL	LG	DAS76D	S	769	800.00000	1000.00000
18	938764	COCINA ELECTRICA		COCINA	BLANCA	ESPECIAL	MABE	FD798	S	770	498.00000	698.00000
19	987527	LAVADORA SEMI AUTOMATI		LAVADORA	BLANCA	ESPECIAL	MABE	76SDFS	S	771	500.00000	637.00000
20	987653	COCINA ELECTRICA		COCINA	BLANCA	ESPECIAL	MABE	BV87	N	772	376.00000	498.00000
*												

Figura 4.15 Datos de OLAP dimproducto

EL resto de pruebas de carga se encuentran a detalle el Anexo H: Pruebas de Carga.

4.2.2. Pruebas de Reportes

Como resultado final se obtienen los reportes en sí. A continuación se muestra el reporte de ventas realizadas en un grupo de tiendas en forma de tabla y gráfico.





Figura 4.16 Reporte de ventas realizadas

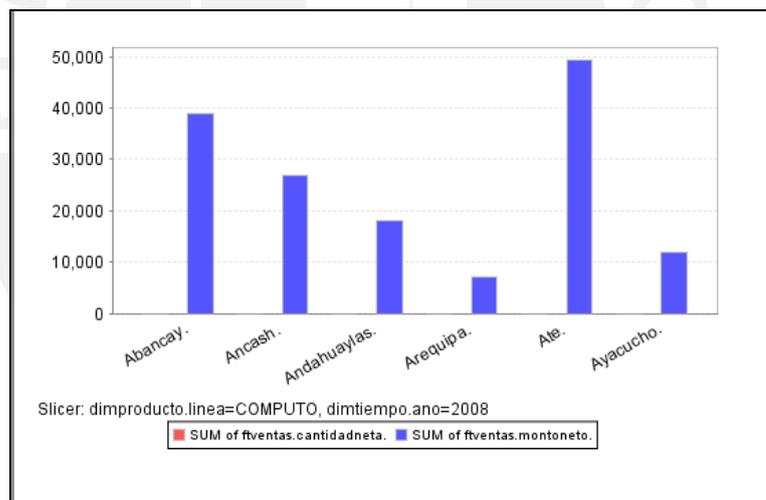


Figura 4.17 Reporte de ventas realizadas - tipo gráfico

A partir de estos resultados se verifica que coinciden con lo esperado especificado en el Anexo B: Data Input del Plan de Pruebas. El resto de las pruebas de reportes se encuentran en el Anexo I: Pruebas de reportes para el área de compras y en el Anexo J: Pruebas de reportes para el área de ventas

5. Observaciones, Conclusiones y Recomendaciones

El quinto y último capítulo del documento muestra las observaciones, conclusiones y recomendaciones del presente proyecto de tesis.

5.1. Observaciones

- La herramienta de BI es solo una parte de la solución de Business Intelligence pero no la más importante. Un buen modelamiento de datos acorde a las necesidades del negocio es eje principal para el éxito de una solución BI.
- El uso de Pentaho como herramienta *Open Source* permite el fácil acceso en costo de una solución BI para cualquier empresa.

- La configuración y uso de las herramientas *Open Source* tienen mayor complejidad.
- Para la elección de una adecuada base de datos se requiere investigación y comparación entre las diferentes opciones del mercado. Esto permite una mejor configuración con la plataforma Pentaho BI Suite.
- El constante contacto con los usuarios finales permite un mayor acercamiento a las necesidades de la empresa.

5.2. Conclusiones

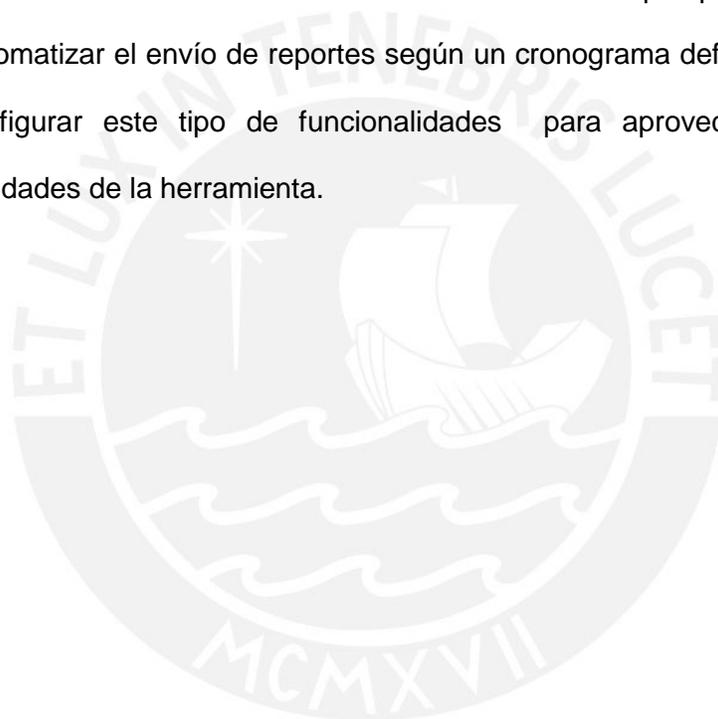
- El correcto levantamiento de información permite identificar los problemas y necesidades del área de compras y ventas con mayor acierto.
- Se decide usar las herramientas de *software libre* para base de datos y para los procesos de extracción y explotación por sus bajos costos.
- En base a las características propias de la empresa el uso de la metodología de Ralph Kimball o “Bottom-Up” resulta una solución eficaz en tiempo y recursos debido a que abarca la solución al problema en un corto plazo. Esta solución sirve de base para la futura construcción de un Data Warehouse.
- Se diseña un modelo dimensional adecuado según la cantidad y profundidad de datos que posee cada Data Mart.

- El desarrollo de los procesos de extracción, transformación y carga son los apropiados según la información requerida por el área.
- Se toma en cuenta que los reportes satisfagan las necesidades de los usuarios para una adecuada toma de decisiones. Además de ello, les ayuda a reducir tiempos de respuesta en el procesamiento y análisis de información, lo que se traduce en que lleguen a ser empresas sostenibles en el tiempo bajo un entorno competitivo.
- El uso de una interfaz BI de reportes permite un manejo intuitivo y sencillo a los usuarios finales para generar sus propios reportes y análisis acorde a las necesidades del negocio en comparación del uso de hojas de cálculo.

5.3. Recomendaciones

- Los datos que contiene el *data mart* permiten el diseño de otra serie de reportes para toma de decisiones en las áreas de compras y ventas.
- El modelo permite la creación de nuevas empresas que contenga el grupo empresarial en el futuro.
- Los *data marts* creados pueden originar la creación de un *data warehouse*, según la metodología de Ralph Kimball.

- Esta solución puede servir de base para cualquier empresa comercializadora de electrodomésticos peruana.
- El utilizar software libre como herramienta para empresas medianas y pequeñas es recomendable debido a que estas empresas no tienen recursos para invertir en software propietarios.
- Pentaho cuenta con funcionalidades adicionales que por ejemplo permiten automatizar el envío de reportes según un cronograma definido, se recomienda configurar este tipo de funcionalidades para aprovechar al máximo las bondades de la herramienta.



Bibliografía

- [Ari 2007] Emilio Arias, “Conferencia Internacional de Software Libre 3.0 (Todo BI)”, Badajoz 2007
- [Bar 1998] Ramón Barquin, “Notas de presentación de estrategia de Data warehousing”, MIGE 1998
- [Bus 2009] <http://www.businessobjects.com>, SAP – SAP Business Objects: Business Intelligence Software and Enterprise Performance Solutions, SAP, 04/09/2010
- [Cog 2009] <http://www.cognos.com>, IBM – Cognos Business Intelligence and Performance Management Software, IBM, 25/02/2009
- [Ibm 2009] <http://www.ibm.com/software>, IBM Software, IBM, 15/05/2009
- [Inm 1992] William H. Inmon, “Building the Data Warehouse”, 1992
- [Jav 2009] <http://java-source.net>, Open Source Business Intelligence – Open Source Software in Java, Java – Source.net, 03/02/2009
- [Kim 2002] Ralph Kimball, “The Data Warehouse Toolkit”, 2002

- [Kim 2008] Ralph Kimball, “The Data Warehouse Lifecycle Toolkit”, 2008
- [Mic 2006] MicroStrategy LATAM South, “Teoría sobre Business Intelligence”, Tercera Edición 2006
- [Mic 2009] <http://technet.microsoft.com>, Microsoft TechNet: Recursos para Profesionales de TI, Microsoft, 25/08/2009
- [Ora 2009] <http://www.oracle.com>, Oracle | Hardware and Software, Engineered to Work Together, Oracle, 19/08/2009
- [Pen 2008] Matrix CPM Solutions C.A., “Building Analytic Solutions using Pentaho”, 2008
- [Pen 2010] <http://www.pentaho.com>, Open Source Business Intelligence – Open Source Reporting, ETL & Data Integration and OLAP | Pentaho, Pentaho, 23/04/2010
- [Pmi 2004] PMI, “Guía de los fundamentos de la dirección de proyectos” (Guía del PMBOK), Tercera edición 2004
- [Ple 2009] http://www.pleroforea.com/Kimball_vs_Inmon.htm, Pleroforea – Kimball vs Inmon, Pleroforea, 25/03/2009

- [Sas 2009] <http://www.sas.com>, SAS | Business Analytics and Business Intelligence Software, SAS, 26/04/2009
- [Spa 2009] <http://www.spagoworld.org>, SpagoWorld, SpagoWorld, 15/05/2009
- [Vit 2002] Elizabeth Vitt, Michael Luckevich y Stacia Misner, “Business Intelligence Técnicas de análisis para la toma de decisiones estratégicas”, España 2002
- [Cur 2008] Matrix CPM Solutions, Curso “Construyendo Soluciones Analíticas con Pentaho”, Santiago Chile 2008.
- [Tes 2008] Villanueva Ojeda Álvaro, “Análisis, diseño e implementación de un datawarehouse de soporte de decisiones para un hospital del sistema de salud público / Álvaro Villanueva Ojeda”, PUCP 2008.
- [Tes 2004] Chipana Peceros Estanislao, “Análisis, diseño e implementación de un data mart del área de riesgos para la banca personal de una institución financiera”, PUCP 2004
- [Tes 2003] Rodas Cervantes Gilberto, “Análisis, diseño e implementación de un data mart de las áreas de inventario y compras para una empresa comercial e industrial ”, PUCP 2003