

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ
FACULTAD DE CIENCIAS E INGENIERÍA



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DEL PERÚ

**ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE DATAMARTS
PARA LAS ÁREAS DE VENTAS Y RECURSOS HUMANOS DE
UNA EMPRESA DEDICADA A LA EXPORTACIÓN E
IMPORTACIÓN DE PRODUCTOS ALIMENTICIOS**

Tesis para optar por el Título de Ingeniero Informático, que presenta el bachiller:

Rolando Heli Moreno Reyes

ASESOR: Edric Ricardo Jugo Soma

Lima, junio del 2013

Resumen

Hoy en día en donde el tiempo es un factor muy importante para las empresas; muchas de estas se ven en desventaja frente a otras debido a la demora en la toma de decisiones de cara al negocio, esto es debido al retardo de procesamiento de datos; es decir, muchas veces los reportes sobre la venta y compra de productos, manejo de dinero, personal, entre otros, no son realizados a tiempo causando que las decisiones administrativas se realicen de manera tardía. Es por ello que las diversas áreas de una empresa están sujetas a mejoras y depende de los directores o gerentes priorizar sobre qué área llevar a cabo dichas mejoras de acuerdo a las necesidades del negocio.

El presente proyecto de fin de carrera ha tomado como estudio una empresa dedicada a la comercialización de productos alimenticios, que en los últimos años ha crecido considerablemente y ha obtenido grandes ganancias; sin embargo arrastra dos problemas claramente marcados relacionados a las áreas de ventas y recursos humanos.

Pese a que las ganancias de los últimos años han sido buenas, el sobre-stock de productos ha aumentado debido a la producción masiva de los mismos lo que ha llevado que en muchas épocas de año sean rematados e incluso eliminados de la comercialización. La empresa se encuentra entonces en la necesidad de equilibrar su producción para lo cual debe de considerar una tendencia de las ventas a realizar, así como las diversas zonas del Perú donde serán comercializado los productos, la época del año de dichas zonas y los clientes de las mismas a quienes van dirigidos los productos elaborados. Asimismo la empresa no lleva un control de las metas mensuales a las que deben llegar los vendedores de las distintas zonas donde se comercializa, ni tampoco un control mensual de las devoluciones que en muchas épocas del año han sido más de lo normal.

Por otro lado, se ha evidenciado que el personal administrativo no cumple con sus horas de trabajo y excede con el número de solicitudes de licencias, perjudicando así la productividad de la empresa. A esto se suma la gran cantidad de pagos realizados por horas extras ocasionando que el presupuesto destinado para cubrir la planilla y contratación de nuevo personal sea mayor a lo inicial. Para este caso la empresa se encuentra ante la necesidad de evaluar el horario de trabajo de su

personal; así como llevar un control del presupuesto que maneja teniendo en cuenta las áreas de la organización, los puestos y los roles del personal.

En base a lo anterior, se propone la construcción de una solución de Inteligencia de Negocios mediante la implementación de dos Datamarts para las áreas de ventas y recursos humanos que servirán de apoyo en el proceso de la toma de decisiones administrativas para una empresa dedicada a la exportación e importación de productos alimenticios.

El uso de los Datamarts de ventas y recursos humanos permitirá que se cuente con dimensiones (*conjunto de información acerca de un evento*) que puedan interrelacionarse entre sí y brindar información a detalle sobre una venta, personal o presupuesto de la organización en estudio; así por ejemplo podemos obtener la consulta sobre *las ventas realizadas en el segundo trimestre para los clientes del departamento de Ica*, o la consultas sobre *el personal nombrado del área de producción que no cumplió con sus horas obligatorias en el mes de diciembre del año 2006*. Además el uso de Datamarts permitirá tener resultados predefinidos en una o más tablas de hechos (*tablas que relaciona las dimensiones y los eventos o hechos*); como *el monto o la cantidad de ventas mensuales, horas diarias trabajadas del personal*, entre otros; los que junto a las dimensiones permitirán que las consultas sean mucho más óptimas y rápidas en comparación con las bases de datos transaccionales en donde las consultas complejas implica barrerse varias tablas del modelo.

Para la implementación del presente proyecto de fin de carrera se optó por usar un híbrido de herramientas siguiendo todos los pasos que implica una solución de Inteligencia de Negocios: análisis, diseño y construcción de los Datamarts, creación y programación de los procesos ETL, creación de los cubos, creación de los informes y finalmente la implementación de la plataforma BI (Web). Para el proceso ETL se usó la herramienta Kettle de la suite de Pentaho permitiendo de este modo ahorrar costos de licencia de software ya que esta es una herramienta que se encuentra bajo licencia GNU GPL. Para el diseño y explotación del cubo se usó la herramienta SQL Server 2008 de Microsoft (Analysis Services y Reporting Services) el cual permite explotar los reportes no sólo vía web sino también crear reportes adicionales que se necesite mediante tablas dinámicas en excel. Esta última es la herramienta con la cual los usuarios finales (personal administrativo) ya se

encuentran familiarizados ya que es su herramienta del día a día; lo cual se convierte en una ventaja al momento de realizar la explotación del cubo.

Resumiendo, el presente proyecto de fin de carrera consiste en implementar una solución de inteligencia de negocios en la empresa en estudio que permita tomar las mejores decisiones frente a los problemas presentados en las áreas de ventas y de recursos humanos; aprovechando, de este modo, todas las ventajas que una solución de BI brinda; como la granularidad de la información, uso de técnicas de explotación como *dice*, *slice*, *drill down*, consultas rápidas y cuyo objetivo es generar una mayor rentabilidad en la organización.



Dedicatoria

Dedico el presente trabajo de tesis a mis queridos padres Marcial Moreno y Julia Reyes que con su constante esfuerzo, dedicación y cariño lograron que yo pueda culminar mis estudios universitarios con éxito.

Agradecimientos

Deseo expresar mis sinceras muestras de agradecimiento a mi familia por sus constantes consejos y apoyo durante toda mi etapa universitaria; a Carla Basurto, quien me asesoró en una primera instancia de la tesis transmitiéndome toda su experiencia y conocimientos; al Sr. Alfonso Pinedo por haberme apoyado en la realización del presente proyecto; y finalmente a mi asesor y amigo Edric Jugo por haber aceptado desinteresadamente en apoyarme y asesorarme en la última etapa de mi tesis universitaria.

Índice

Introducción	1
1. CAPITULO 1: Generalidades	2
1.1. Identificación del Problema	2
1.2. Marco Conceptual	3
1.2.1. Conceptos: Inteligencia de negocios	4
1.2.2. Conceptos: Modelo Dimensional	11
1.2.3. Conceptos: Organización	13
1.3. Estado del Arte	17
1.3.1. Trabajos Realizados	17
1.3.2. Herramientas para la Implementación	21
1.4. Planificación	24
1.4.1. Análisis del Proyecto	24
1.4.2. Cronograma del Proyecto	26
1.5. Descripción y Sustentación de la Solución	28
2. CAPÍTULO 2: Análisis	31
2.1. Metodología	31
2.1.1. Ciclo de Vida de Un sistema de Soporte a Decisiones	31
2.1.2. Herramientas a utilizar	36
2.1.3. Enfoques	38
2.1.4. PMBOK	40
2.2. Identificación de los Requerimientos	41
2.2.1. Requerimientos Funcionales	41
2.2.2. Requerimientos No Funcionales	43
2.3. Análisis Dimensional	44
2.3.1. Datamart de Ventas	44
2.3.2. Datamart de Recursos Humanos	46
2.3.3. Facts	48
3. CAPÍTULO 3: Diseño	53
3.1. Arquitectura	53
3.2. Proceso de extracción y transformación	54
3.3. Diseño de explotación	57
3.3.1. Configuración del Reporte	57
3.3.2. Vistas y reportes	59
4. CAPÍTULO 4: Construcción y Pruebas	63

4.1.	Puesta en Producción.....	63
4.1.1.	Instalación del Pentaho.....	63
4.1.2.	Configuración de la Herramienta ETL.....	64
4.1.3.	Instalación de SQLServer 2008 R2	65
4.2.	Plan de Pruebas	65
4.2.1.	Objetivos	65
4.2.2.	Características a ser probadas.....	65
4.2.3.	Casos de pruebas.....	66
4.2.4.	Desarrollo de las pruebas	67
5.	CAPÍTULO 5: Observaciones, Conclusiones y Recomendaciones o Trabajos Futuros	78
5.1.	Observaciones	78
5.2.	Conclusiones	81
5.3.	Recomendaciones o Trabajos Futuros	82
6.	Referencias	84



Índice de Figuras

Figura 1.1 - OLAP Multidimensional [RMR 2013].....	7
Figura 1.2 - OLAP Relacional [RMR 2013].....	8
Figura 1.4 - Técnica Slice [TEC 2013].....	9
Figura 1.3 - OLAP Híbrido [RMR 2013].....	9
Figura 1.5 - Técnica Dice [TEC 2013]	10
Figura 1.6 - Técnica Drill Down/Up [TEC 2013]	10
Figura 1.7 - Técnica Pivot [TEC 2013]	11
Figura 1.8 - Cubo	12
Figura 1.9 - Modelo Estrella	13
Figura 1.10 - Organigrama de la empresa [APM 2012].....	14
Figura 1.11 - Diagrama de Flujo de la ejecución de un pedido [APM 2012]	17
Figura 1.12 - Cronograma de Actividades.....	27
Figura 1.13 - Diagrama de Gantt.....	27
Figura 1.14 - Diagrama WBS	28
Figura 2.1 - Ciclo de vida para un proyecto de BI [MOS 2003].....	35
Figura 2.2 - Ciclo de vida para un proyecto de BI [MOS 2003].....	36
Figura 2.3 - Datamart de Ventas	44
Figura 2.4 - Datamart de Recursos Humanos.....	46
Figura 3.1 - Arquitectura de la Solución	53
Figura 3.2 - Prototipo de Reporte Crosstab.....	58
Figura 3.3 - Prototipo de Reporte Gráfico	58
Figura 3.4 - Prototipo de Reporte Híbrido	59
Figura 4.1 - Pagina de Pentaho	64
Figura 4.2 - Pantalla de bienvenida de Kettle de Pentaho	64
Figura 4.3 - ETL de Ventas	71
Figura 4.4 - Log del Datamart de Ventas I	72
Figura 4.5 - Log del Datamart de Ventas II	72
Figura 4.6 - ETL de Recursos Humanos.....	76
Figura 4.7 - Log del Datamart de Recursos Humanos I.....	77
Figura 4.8 - Log del Datamart de Recursos Humanos II.....	77

Índice de Tablas

Tabla 1.1 - Proyectos de Fin de Carrera	21
Tabla 1.2 - Herramientas de ETL	23
Tabla 1.3 Herramientas de explotación.....	23
Tabla 2.1 - Requerimiento Funcionales para Ambos Datamarts.....	41
Tabla 2.2 - Requerimiento Funcionales el Datamart de Ventas.....	42
Tabla 2.3 - Requerimiento Funcionales para el Datamart de Recursos Humanos ..	43
Tabla 2.4 - Requerimiento No Funcionales	44
Tabla 2.5 - Fact V.S. Dimensiones.....	49
Tabla 3.1 - Bases de Datos de Extracción	54
Tabla 3.3 - Reportes V.S Datamart	59
Tabla 5.1 - Análisis TCO Para Una Herramienta ETL.....	79
Tabla 5.2 - Análisis TCO Para Una Herramienta BI.....	80



Introducción

Hoy en día las organizaciones se encuentran en constante toma de decisiones de manera que estas contribuyan a la mejora de sus procesos permitiendo así el desarrollo de la misma. Sin embargo llegar a tomar una decisión no es fácil, para ello se debe de realizar un estudio previo de todas las ventajas y desventajas que esta implica. En base a esta necesidad surgen los sistemas de apoyo a decisiones que permiten a los usuarios tomar las decisiones correctas en el momento correcto.

Muchas organizaciones cuentan ya con sistemas de información transaccionales los cuales almacenan los datos de sus procesos en las diferentes bases de datos OLTP existentes; sin embargo, no todas las empresas cuentan con un sistema de soporte a decisiones que permita integrar estos dato, explotarlos y convertirlos en información útil.

Lo que se busca es poder integrar estos datos trasladándolos desde las diferentes bases de datos relacionales hacia un nuevo modelo que permita explotarlos y convertirlos en información útil para la empresa. Es allí donde surge el concepto de Datamart el cual es la agrupación de datos específicos sobre un tema en particular de la organización cuya explotación ayudará a los usuarios a tomar las decisiones correctas.

El presente proyecto de fin de carrera tiene como objetivo la construcción de una solución de inteligencia de negocios mediante la implementación de dos Datamarts para las áreas de ventas y recursos humanos para una empresa cuyo rubro es la importación, exportación y ventas de productos alimenticios; de manera que permita a los usuarios finales tomar las decisiones adecuadas sobre las ventas y el personal de la organización.

1. CAPITULO 1: Generalidades

En este capítulo se definen los problemas que se desean resolver con los Datamarts a elaborar; además se resume cada uno de los conceptos utilizados al momento de elaborar una solución de inteligencia de negocios; así como también los conceptos propios de la organización en estudio. También se hace hincapié en algunos trabajos con éxito en diversas empresas y se realiza un plan de proyecto basado en entregables que permitirán gestionar el presente proyecto de fin de carrera. Finalmente se realiza una descripción y sustentación de la solución.

1.1. Identificación del Problema

Hoy en día en donde el tiempo es un factor muy importante para las empresas; muchas de estas se ven en desventaja frente a otras debido a la demora en la toma de decisiones de cara al negocio, esto es debido al retardo de procesamiento de datos; es decir, muchas veces los reportes sobre la venta y compra de productos, manejo de dinero, personal, entre otros, no son realizados a tiempo causando que las decisiones administrativas se realicen de manera tardía. Es por ello que las diversas áreas de una empresa están sujetas a mejoras y depende de los directores o gerentes priorizar sobre que área llevar a cabo dichas mejoras de acuerdo a las necesidades del negocio. [RMR 2013]

Según el personal entrevistado de la empresa en estudio el área de ventas es la encargada de realizar las operaciones de ventas de los productos alimenticios a los clientes con los que actualmente se cuenta y con ello generar rentabilidad para el negocio. Pese a que las ganancias de los últimos años han sido buenas, el sobrestock de productos ha aumentado debido a la producción masiva de los mismos lo que ha llevado que en muchas épocas de año sean rematados e incluso eliminados de la comercialización. La empresa se encuentra entonces en la necesidad de equilibrar su producción para lo cual debe de considerar una tendencia de las ventas a realizar, así como las diversas zonas del Perú donde serán comercializado los productos, la época del año de dichas zonas y los clientes de las mismas a quienes van dirigidos los productos elaborados. Asimismo la empresa no lleva un control de las metas mensuales a las que deben llegar los vendedores de las distintas zonas donde se comercializa, ni tampoco un control mensual de las devoluciones que en muchas épocas del año han sido más de lo normal. [APM 2012]

Por otro lado el personal entrevistado manifiesta que la empresa no cuenta con un control de las horas trabajadas y licencias del personal administrativo, que en muchas ocasiones no cumple con sus horas obligatorias de trabajo, perjudicando así la productividad de la empresa. A esto se suma la gran cantidad de pagos realizados por horas extras ocasionando que el presupuesto destinado para cubrir la planilla y contratación de nuevo personal sea mayor a lo inicial. Para este caso la empresa se encuentra ante la necesidad de evaluar el horario de trabajo de su personal; así como llevar un control del presupuesto que maneja teniendo en cuenta las áreas de la organización, los puestos y los roles del personal. [APM 2012]

1.2. Marco Conceptual

En esta sección se explicará los conceptos asociados que se requieren para llevar a cabo el presente proyecto de fin de carrera.

1.2.1. Conceptos: Inteligencia de negocios

Inteligencia de Negocios

Inteligencia de negocios (*Business Intelligence*) es una disciplina en la cual se hace centro en el análisis de la información para la correcta toma de decisiones que le permita a la organización cumplir con los objetivos de negocio. Inteligencia de negocios engloba a aquellos procesos, tecnologías y herramientas para transformar datos en información, información en conocimiento y conocimiento en planes que conduzcan una acción en beneficio del negocio. [KIM 2002]

Datawarehouse

El Datawarehouse es un repositorio de datos y eventos que puede basarse en estructuras relacionales o estructuras multidimensionales (cubos) en las que se almacenan la información calculando previamente todas las combinaciones de todos los niveles de todas las aperturas de análisis. Es como un producto cartesiano donde se almacena todas las combinaciones posibles. [KIM 2002]

El Datawarehouse es una colección de datos organizados orientados a un tema de que brinda apoyo a toma de decisiones. Estos son integrados, no volátiles, variantes en el tiempo. [INM 2003]

Un Datawarehouse es una copia de los datos transaccionales, específicamente diseñada para realizar consultas y análisis. [KIM 2002]

Datamart

Se denomina Datamart al conjunto de datos estructurados que provienen de las diferentes aplicaciones operacionales. Es un subconjunto de un Datawarehouse con un alcance de contenido limitado, el cual es usado sólo por un área específica o un problema particular de análisis dentro de la organización. [KIM 2002]

Sistemas Transaccionales - OLTP

Los sistemas transaccionales, cuyas siglas en inglés son OLTP (Online Transaction Process) son los sistemas que soportan las transacciones realizadas en una organización, los cuales se caracterizan porque son los encargados de soportar cada operación realizada en el negocio, por ello deben dar respuesta rápida a cada transacción realizada, además se encargan de recolectar los datos asociados a cada transacción y almacenar estos datos en bases de datos a través de sistemas

de información. Si un sistema OLTP deja de funcionar puede ocasionar grandes pérdidas en una compañía por el hecho de no poder realizar transacciones. [KIM 2002]

Sistemas de Soporte a Decisiones – DSS

Los sistemas de soporte a decisiones; cuyas siglas en inglés son DSS (Decision Support System); son sistemas que se han construido con el fin de proporcionar información acerca del desempeño organizacional o del negocio, es una forma de medir y ver el negocio o parte de él en un momento en el tiempo. La finalidad de ver, medir, evaluar el desempeño del negocio es lograr tomar las decisiones adecuadas que permitan superar algunas deficiencias, o aprovechar algunas oportunidades. [KIM 2002]

Gestión del Conocimiento

Es la disciplina que habilita la generación, retención, organización, distribución y mantenimiento del conocimiento de manera corporativa, con el fin de mejorar el proceso de toma de decisiones. [KIM 2002]

Extracción, Transformación y Carga de Datos – ETL

Al hablar de ETL relacionamos a los diferentes procesos que se concentran en el concepto de toma, transformación y carga de datos en un Datawarehouse. [KIM 2002]

El proceso de extracción consiste en extraer los datos desde los sistemas de origen y convertirlos en un formato previo para iniciar el proceso de transformación. [KIM 2002]

El proceso de transformación de datos se refiere a la conversión de los datos para que sean cargados aplicando una serie de reglas de negocio o funciones como la limpieza de datos, el cual consiste en la corrección de los datos escritos de forma errónea. [KIM 2002]

El proceso de carga consiste en la carga de los datos en el sistema de destino, es decir que todos los datos son cargados en el área de presentación del Datawarehouse. [KIM 2002]

Jerarquías de Una Dimensión

Son relaciones lógicas entre los niveles o elementos de una dimensión. Los atributos de una dimensión, permiten la navegación entre niveles, y una dimensión puede tener múltiples jerarquías. [KIM 2002]

Atributos en una dimensión

Son etiquetas descriptivas de un determinado elemento de la dimensión. [KIM 2002]

Tipos de Almacenamiento - OLAP

Los sistemas OLAP es una solución que se usa en la Inteligencia de Negocios con la finalidad de optimizar la consulta de grandes cantidades de datos, para lo cual utiliza los denominados cubos OLAP, que es un medio de almacenamiento lógico que contiene datos resumidos de grandes bases de datos o sistemas transaccionales. Los cubos se clasifican en tres tipos: [KIM 2002]

MOLAP

En el modo de almacenamiento MOLAP (**OLAP Multidimensional**) se requiere un preprocesamiento y almacenamiento de la información contenida en el cubo OLAP. Los datos, junto con sus agregaciones, son almacenados en una estructura multidimensional. [KIM 2002]

Entre las características del almacenamiento MOLAP tenemos:

- Provee excelente rendimiento y compresión de datos.
- Tiene mejor tiempo de respuesta, dependiendo solo del porcentaje de las agregaciones del cubo, las cuales suelen estar precalculadas e incluso el cálculo puede realizarse dentro del proceso ETL por lo que una consulta puede ser bastante rápida.
- Si el nivel de granularidad del cubo es muy fino o el número de dimensiones es alto, el cubo puede llegar a ser muy grande afectando de este modo su performance.
- En general este método, es muy apropiado para cubos con uso frecuente y que no tengan un alto nivel de granularidad debido a su rápida respuesta.

La Figura 1.1 muestra cómo se trabaja con un modo de almacenamiento MOLAP donde se observa que los datos son transferidos desde la base de datos OLTP y almacenados y preprocesados en una base de datos multidimensional desde donde serán explotados por los usuarios.

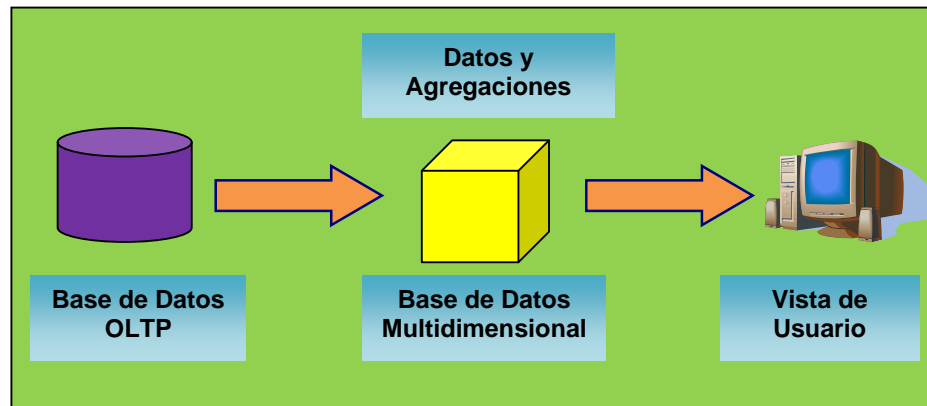


Figura 1.1 - OLAP Multidimensional [RMR 2013]

ROLAP

En un modelo ROLAP (**OLAP Relacional**) toda la información del cubo, sus datos, su agregación, sumas, etc., son almacenados en una base de datos relacional.

A diferencia del modo de almacenamiento MOLAP, ROLAP accede a las tablas de la base de datos relacional cuando necesita responder a las consultas. Generalmente es mucho más lenta que las otras estrategias de almacenamiento (MOLAP o HOLAP). [KIM 2002]

Entre las características del almacenamiento ROLAP tenemos:

- ROLAP se utiliza para ahorrar espacio de almacenamiento cuando se trabaja con grandes conjuntos de datos que se consultan con poca frecuencia; por ejemplo, datos exclusivamente históricos.
- ROLAP también es adecuado cuando se requiera llegar a mucho nivel de detalle.

La Figura 1.2 muestra cómo se trabaja con un modo de almacenamiento ROLAP donde se observa que todos los datos se encuentran preprocesados y almacenados en una base de datos relacional desde donde serán explotados por los usuarios.

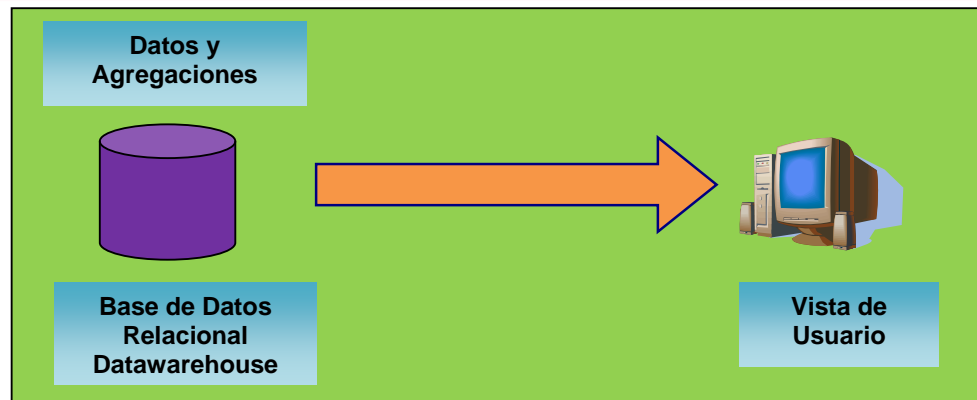


Figura 1.2 - OLAP Relacional [RMR 2013]

HOLAP

HOLAP (**OLAP híbrido**) combina atributos de MOLAP y ROLAP. Al igual que MOLAP, HOLAP hace que las agregaciones se almacenen en una estructura multidimensional, y los datos a nivel de detalle, en una base de datos relacional como lo hace el almacenamiento ROLAP. [KIM 2002]

Para procedimientos de búsqueda que acceden a los datos generales (sin tanto nivel de detalle), HOLAP es equivalente a MOLAP. Por el contrario, si los procesos de consultas accedieran a los máximos niveles de detalle, deberían recuperar los datos de la base de datos relacional y esto no sería tan rápido comparado con una estructura MOLAP. [KIM 2002]

Los cubos almacenados como HOLAP, son más pequeños que los MOLAP y responden más rápidos que los ROLAP. [KIM 2002]

Usos comunes de HOLAP:

- Cubos que requieren rápida respuesta.
- Solución de compromiso para bajar el espacio ocupado sin perjudicar totalmente el rendimiento de las consultas.

La Figura 1.3 muestra cómo se trabaja con un modo de almacenamiento HOLAP donde una parte de los datos están almacenados en una base de datos multidimensional y otra parte en una base de datos relacional.

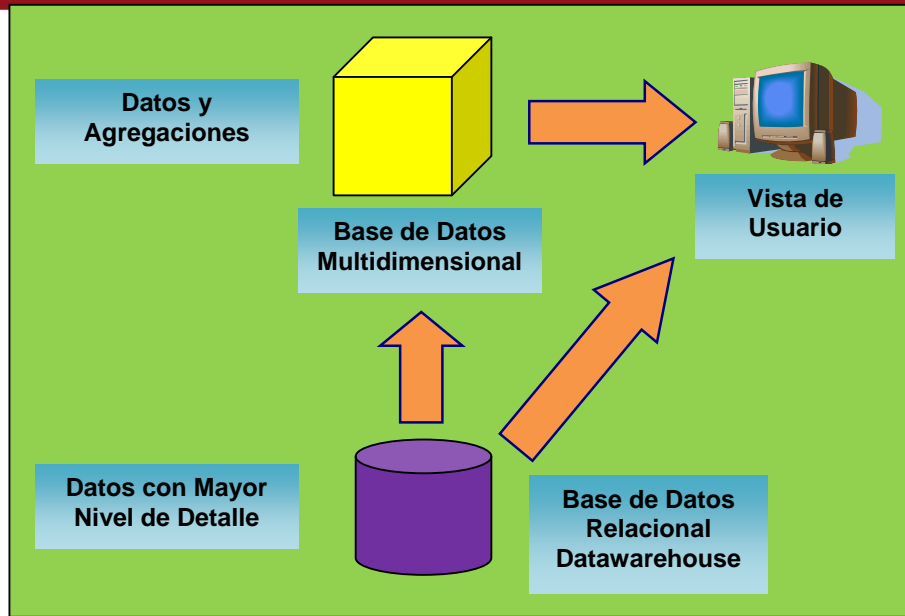


Figura 1.3 - OLAP Híbrido [RMR 2013]

Técnicas de Explotación

Las técnicas de explotación hacen referencia a la forma de cómo puede ser explotada la información en un reporte; algunas de ellas son: [RMR 2013]

SLICE

Es la técnica que permite escoger un subconjunto de un cubo eligiendo un único valor para una de sus dimensiones. [TEC 2013]

La Figura 1.4 ilustra un ejemplo de la técnica de explotación “slice” donde se muestra que se obtiene un subconjunto de las ventas de todos los productos y regiones de ventas eligiendo el valor 2004 de la dimensión tiempo.

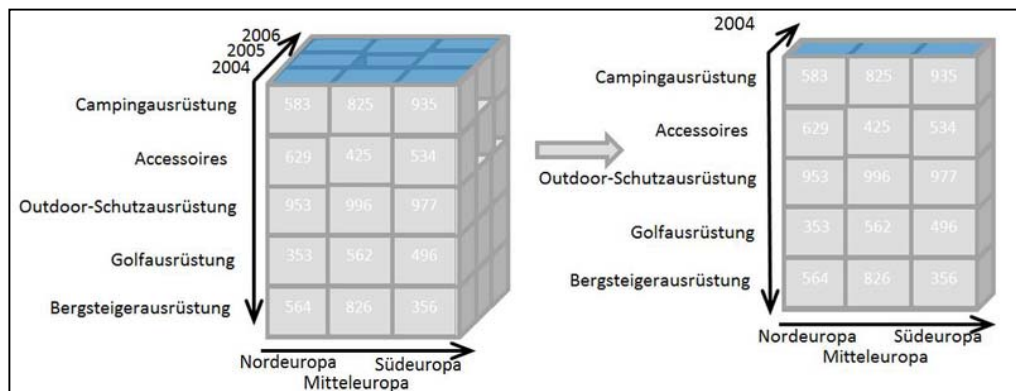


Figura 1.4 - Técnica Slice [TEC 2013]

DICE

Es la técnica que permite escoger un subconjunto de un cubo eligiendo valores específicos de múltiples dimensiones. [TEC 2013]

La Figura 1.5 ilustra un ejemplo de la técnica de explotación “dice” donde se muestra que se obtiene un subconjunto de las ventas para un grupo de productos quedando las dimensiones tiempo y región igual que el primer cubo.

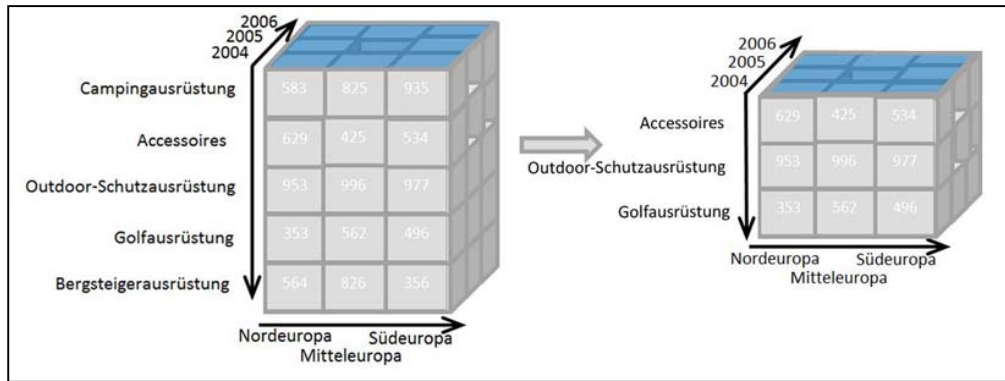


Figura 1.5 - Técnica Dice [TEC 2013]

DRILL DOWN/UP

Es la técnica que permite navegar desde niveles más resumidos (up) hacia niveles con más detallados (down). [TEC 2013]

La Figura 1.6 ilustra un ejemplo de la técnica de explotación “drill down/up” donde se muestra que se pasa de la categoría de producto “outdoor” a ver las ventas individuales de los productos pertenecientes a esta categoría.

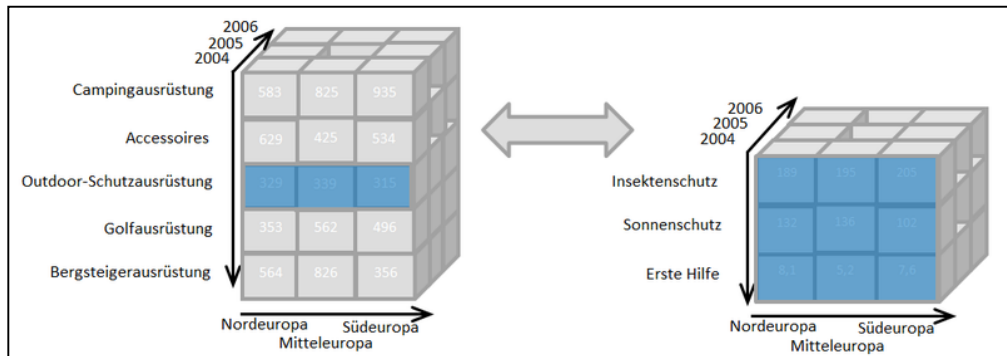


Figura 1.6 - Técnica Drill Down/Up [TEC 2013]

PIVOT

Es la técnica que permite rotar el cubo en el espacio para ver sus diferentes caras.
[TEC 2013]

La Figura 1.7 ilustra un ejemplo de la técnica “pivot” donde se muestra que el cubo es rotado quedando las regiones alineadas horizontalmente y el tiempo verticalmente teniéndose así otra perspectiva de la data.

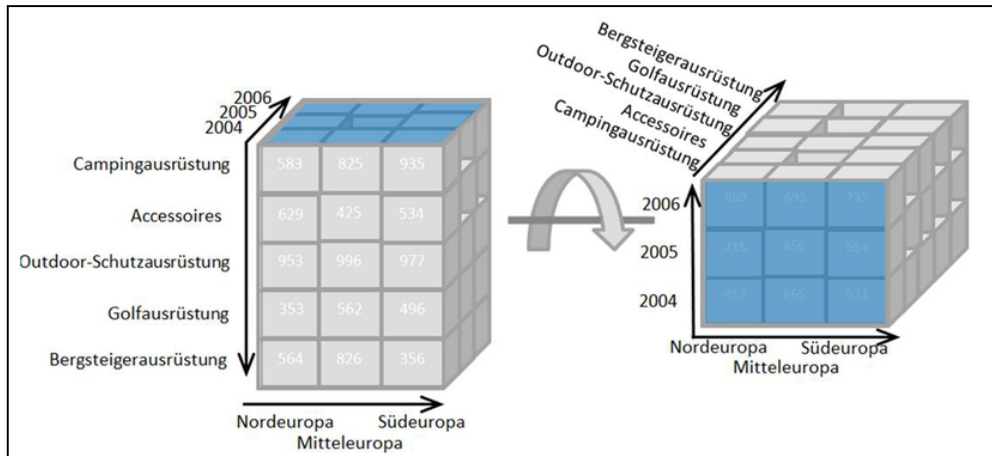


Figura 1.7 - Técnica Pivot [TEC 2013]

1.2.2. Conceptos: Modelo Dimensional

Modelo Dimensional

El modelo dimensional es la forma en que modelamos los datos para el análisis. En este modelo las perspectivas o dimensiones se combinan en una entidad denominada hechos, la cual contiene las medidas. Las medidas pueden analizarse desde cada una de las perspectivas o dimensiones que hayan definido. [KIM 2002]

La Figura 1.8 muestra un cubo de ventas con sus dimensiones Ubicación, Tiempo y Producto. Donde la ubicación sería el eje “y” el producto sería el eje “x” y el tiempo sería el eje “z” y cada punto en el espacio donde haya una intersección de las 3 dimensiones contendría la(s) medida(s) relacionada(s) a esa combinación.

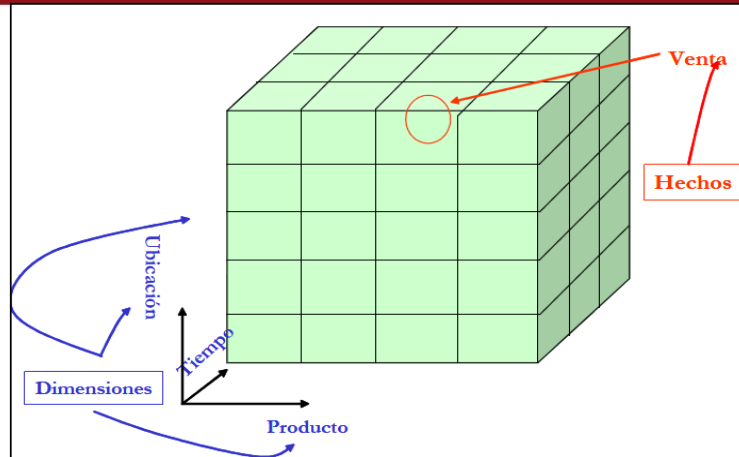


Figura 1.8 - Cubo

Tabla de Hechos (*Fact Table*)

La tabla de hechos es la tabla primaria del modelo dimensional y contiene los valores del negocio que se desea analizar. Cada tabla de hechos contiene las claves externas, que se relacionan con sus respectivas tablas de dimensiones, y las columnas con los valores que serán analizados. La tabla de hechos está constituida por medidas y por llaves primarias. Hay modelos que en lugar de llaves primarias se usan llaves foráneas. [KIM 2002]

Dimensiones

Las dimensiones son calificadores que dan sentido a la tabla de hechos, las cuales organizan los datos en función de un área de interés para los usuarios. Cada dimensión describe un aspecto del negocio y proporciona el acceso intuitivo y simple a datos. [KIM 2002]

Medidas o Métricas

Una medida es una columna cuantitativa, numérica, en la tabla de hechos. Las medidas representan los valores que son analizados, como cantidad de personas con horas extras o importe de ventas. [KIM 2002]

Granularidad

Es el mínimo nivel de detalle (agregación) de los datos en un modelo dimensional, está referido a los hechos y definido por el mínimo nivel de detalle de las dimensiones. [KIM 2002]

Esquema Estrella

El esquema estrella es una estructura compuesta por una tabla central (tabla de hechos) y un conjunto de tablas organizadas alrededor de esta (tablas de dimensiones). En las puntas de la estrella se encuentran las tablas de dimensión que contienen los atributos de las aperturas que interesan al negocio que se pueden utilizar como criterios de filtro y son relativamente pequeñas. Cada tabla de dimensión se vincula con la tabla de hechos por un identificador. La Figura 1.9 muestra el esquema del modelo estrella. [KIM 2002]

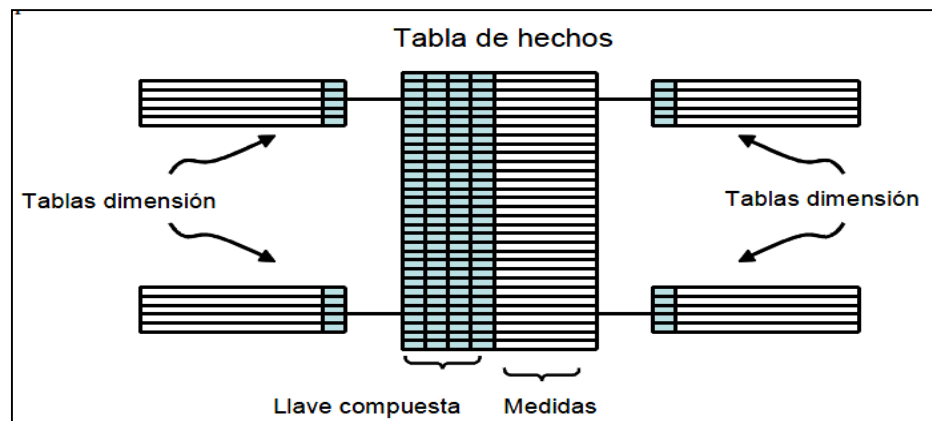


Figura 1.9 - Modelo Estrella

Agregaciones

Las agregaciones son resúmenes de datos pre-calculados que mejoran el tiempo de respuesta por el simple hecho de tener preparadas las respuestas antes de que se planteen las preguntas. [KIM 2002]

1.2.3. Conceptos: Organización

La organización en estudio es una empresa de prestigio cuyo rubro es la comercialización de productos alimenticios. Sus principales clientes son los supermercados Metro y Wong con los cuales trabaja todo el año; además cuenta con pequeños clientes con los que trabaja por temporadas. La empresa cuenta con dos sucursales una en Ica y otra en Lima. [APM 2012]

La Figura 1.10 muestra el organigrama de la empresa. [APM 2012]

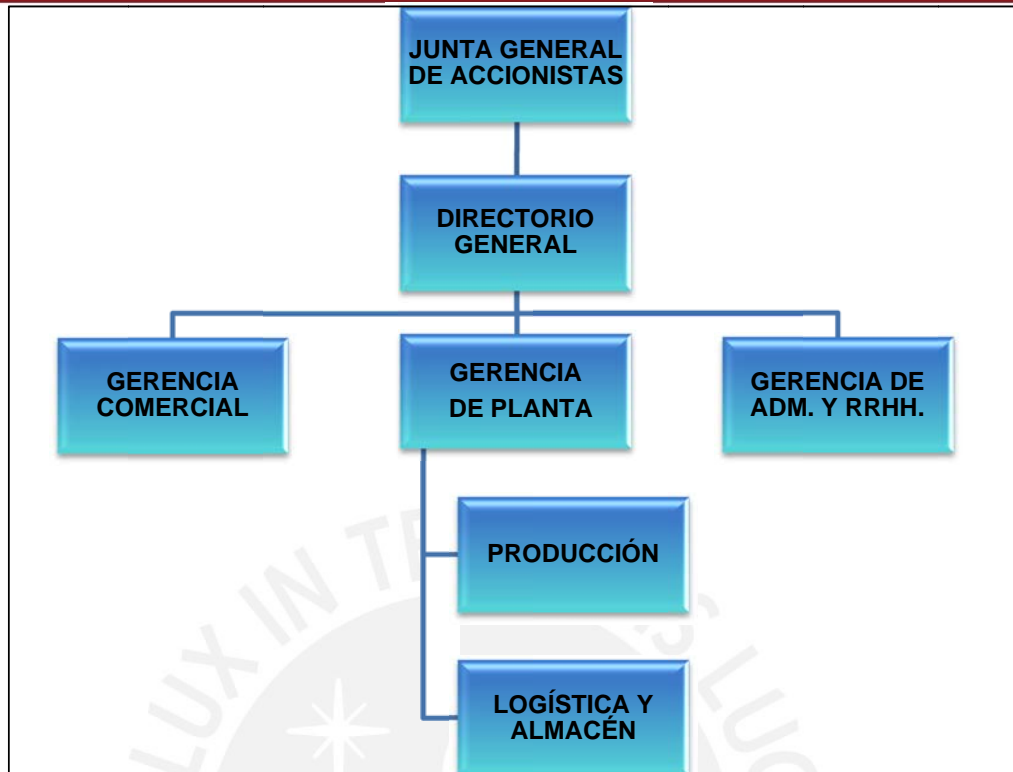


Figura 1.10 - Organigrama de la empresa [APM 2012]

Gerencia Comercial

Es el área encargada de contactar al cliente y de realizar las venta de productos. Los vendedores son los encargados de contactar con los supermercados y tiendas del medio peruano y realizar la lista de pedidos que ellos soliciten. Uno de los Datamarts a elaborar está centrado en esta área. [APM 2012]

Las principales funciones de la gerencia comercial son:

- Contactar y gestionar con el cliente la venta de productos.
- Coordinar con el área de almacén el pedido de productos que solicita el cliente.
- Realizar campañas de ventas de producto.

El Datamart de ventas permitirá lo siguiente:

- Lograr segmentar a los clientes que tienen un mayor índice de compras y ofrecerles beneficios como descuentos, ofertas, etc. Actualmente la empresa realiza ofertas a clientes sin un análisis previo, se podría decir por pura “intuición” dando preferencia a supermercados ubicados en la capital.
- Lograr segmentar en que temporadas se realizan un mayor índice de compras tanto por zonas como por productos vendidos. De esta forma se podrá controlar la cantidad de productos a elaborar, así como realizar diversas campañas de ventas en distintas temporadas en los diferentes lugares donde opera la organización. Actualmente la empresa no realiza campañas de ventas.
- Identificar quienes son los vendedores que realizan un mayor número de ventas de manera que se les pueda ofrecer bonos o beneficios de acuerdo a las ventas realizadas. Actualmente la empresa no cuenta con una administración de sus vendedores.

Gerencia de Planta

La gerencia de Planta es la encargada de elaborar los productos alimenticios de acuerdo al pedido que la gerencia comercial le solicite. Se dividen en dos áreas: Producción y Logística y Almacén. [APM 2012]

Logística y Almacén

Es el área encargada de recibir el pedido de la Gerencia Comercial y solicitar al área de producción la elaboración del mismo. En caso el área de producción no cuente con un stock de insumos y materia prima necesarios para elaborar los productos descritos en el pedido, el área de Logística y Almacén será la encargada de proporcionarlos. [APM 2012]

Producción

Es el área encargada procesar la materia prima y transformarlo en un producto terminado. Una vez realizado los productos estos son devueltos al área de Almacén. [APM 2012]

Gerencia de Administración y Recursos Humanos

La gerencia de Administración refiere básicamente al área contable, es la encargada de administrar la contabilidad de la empresa (sueldos al personal, presupuestos, estados financieros); mientras que el área de recursos humanos es la encargada del reclutamiento y capacitación del personal; así como el control de horarios de este. Uno de los Datamarts a elaborar esta centrado en estas dos áreas. [APM 2012]

Las principales funciones de la gerencia de administración y recursos humanos son:

- Administrar los estados financieros de la empresa.
- Administrar el presupuesto de la empresa.
- Contratar y capacitar al personal con el que cuenta la organización.
- Administrar los sueldos y horario de trabajo del personal de la empresa.

El Datamart de recursos humanos permitirá lo siguiente:

- Llevar un adecuado control de los movimientos de entrada y salida, licencias y vacaciones del personal de la organización; de manera que se pueda verificar si estos cumplen o no con sus horarios de trabajo. Actualmente la empresa no cuenta con un control de su personal.
- Administrar de una mejor manera el presupuesto que tiene la empresa. Con ello se podrá distribuir el dinero de una forma óptima para los contratos del personal, pago de horas extras, realizar nombramientos, capacitaciones, incentivos, bonos, etc.

La Figura 1.11 describe el diagrama de flujo desde la solicitud de un pedido hasta la entrega del mismo. La ampliación de esta figura se encuentra en el **Anexo L - Figuras**.

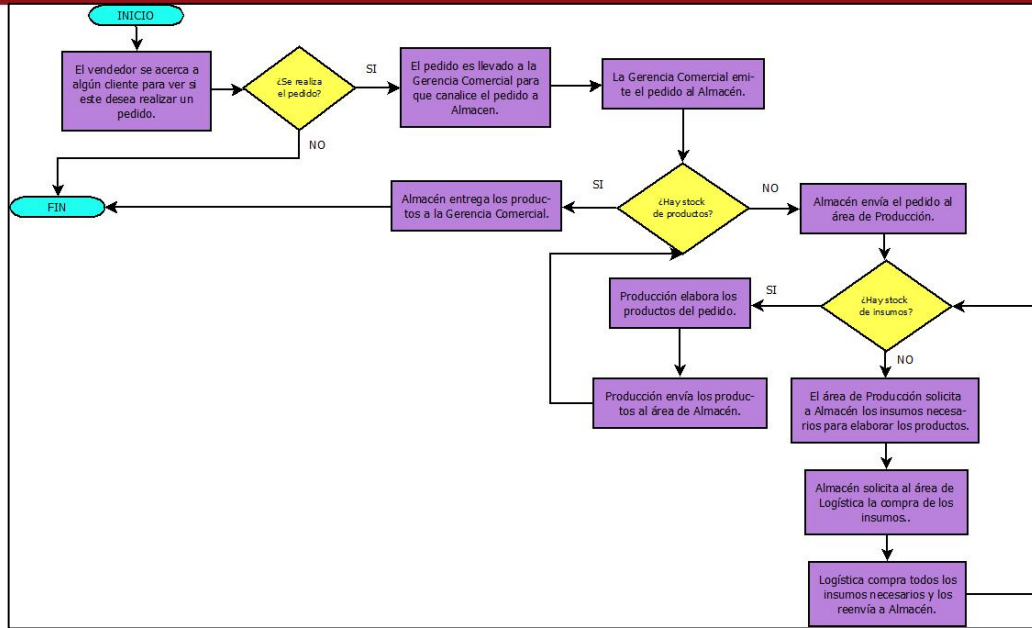


Figura 1.11 - Diagrama de Flujo de la ejecución de un pedido [APM 2012]

1.3. Estado del Arte

A continuación se muestra algunos trabajos realizados que atienden a los diversos problemas hallados en las distintas áreas de las empresas comercializadoras y/o productoras; así como las herramientas existentes en el proceso de toma decisiones.

1.3.1. Trabajos Realizados

Diseño e implantación de un Datamart para el área de facturación y ventas de la empresa Macro Comercializadora, S.A. Sede La Urbina. [GUE 2005]

Guédez y Zambrano han realizado el proyecto de fin de carrera denominado “Diseño e implantación de un Datamart para el área de facturación y ventas de la empresa Macro Comercializadora, S.A. Sede La Urbina” cuyo objetivo general es diseñar, construir e implantar un Datamart y una aplicación Web de Intranet para la construcción de reportes en el área de facturación y ventas de la empresa Macro Comercializadora, S.A. sede La Urbina, basándose en el enfoque de Ralph Kimball.

La metodología usada por Guédez y Zambrano abarca 3 etapas:

Etapa 1: Es la etapa de planificación que se resume en la entrega de un documento de la definición del proyecto y un plan detallado del mismo basado en entrevistas y consultas teóricas preliminares.

Etapa 2. Es la etapa de desarrollo donde se entrega un prototipo y un cronograma de la etapa de implantación, este prototipo está acompañado de los procesos que permiten obtener, transformar y cargar los datos correctos.

Etapa 3: Es la etapa de implantación cuyos resultados son un Datamart con data cargada y su documentación.

Los alcances que Guédez y Zambrano proponen son los siguientes:

- Implementar los procesos de ETL para poder extraer los datos referentes a las ventas de las diferentes tiendas y poder transformarlas en información válida para el Datamart. [GUE 2005]
- Diseño y construcción de un Datamart que contemple las dimensiones de Clientes, Artículos, Tiendas, Tiempo y Proveedores; incluyendo sus respectivas jerarquías y agrupaciones.
- Diseño, construcción e implantación de una herramienta web para la intranet de la empresa que permita hacer consultas sobre la información almacenada en la Datamart.

Desarrollo de un sistema de Business Intelligence para el área de compras de la empresa Salmofood S.A. [BAR 2005]

Barrientos denomina a su proyecto de fin de carrera “Desarrollo de un sistema de Business Intelligence para el área de compras de la empresa Salmofood S.A.”.

Barrientos opina que como se trabajará con un área específica entonces sólo basta con un Datamart para usarlo como repositorio de toda la información.

No toda la información existente respecto a Compras es analizada o es importante para el personal a cargo de la toma de decisiones. Muchas veces, las tablas existentes en los distintos sistemas de información poseen varios campos que no necesitan ser considerados para tomar una decisión, por lo que no tendría sentido copiar toda la información de un lado para otro. Es por este motivo que se usarán procesos ETL (*Extract Transform Load data*) que llevarán solamente los datos necesarios al Datamart de Compras. Usando los procesos ETL, se almacenará la información en un repositorio central, que usará como esquema de almacenamiento el modelo multidimensional de tipo estrella. [BAR 2005]

Sobre el Datamart de Compras se realizarán todas las consultas necesarias para extraer la información que se presentará a las personas que están encargadas del proceso de toma de decisiones relacionadas con el área de compras. Estas personas serán capaces de trabajar con esta información utilizando herramientas de análisis especialmente diseñadas para este fin y podrán llevar esta información a otras aplicaciones para que le puedan sacar el mayor provecho posible. [BAR 2005] Barrientos concluye en su tesis que luego de implementar el sistema de BI para el área de compras de la empresa se ha conseguido que los usuarios utilicen lo que se les ha desarrollado, lo cual indica un alto grado de conformidad respecto a lo que han solicitado. Se ha logrado que el tiempo que antes ocupaban para poder desarrollar algunas de sus tareas sea muchísimo menor, con lo que pueden ocupar el tiempo que les queda libre en desarrollar nuevos requerimientos para su área y luego poder implementarla, de forma que la gestión de la empresa iría mejorando cada vez. La importancia de que este grupo de usuarios utilice el sistema en el desarrollo de sus tareas diarias es que ayudará a que las demás áreas se interesen en ayudar al desarrollo de un sistema de BI para ellos, ya que sólo va en beneficio de ellos mismos. [BAR 2005]

Planeación operativa en el área de ventas basada en la metodología de Balanced Scorecard en una empresa dedicada a la fabricación y comercialización de repuestos de cauchos. [PAD 2010]

El trabajo de Padilla tiene como fundamento mejorar los márgenes de ventas mediante el diseño del Balanced Scorecard enfocado en el área de ventas de la empresa CAUCHOS INDUSTRIALES S.A., de la ciudad de Guayaquil. Cauchos Industriales S.A., es una empresa dedicada a la fabricación de piezas y accesorios automotrices de caucho y metal. Este diseño de planeación estratégica se llevara a

cabo mediante un proceso de medición de indicadores mediante una base de datos realizada en Access y luego importada a Excel 2007 para medir los objetivos estratégicos que cumplan con las metas propuestas por el departamento de ventas, contribuirá a ser más eficiente y a la optimización de los recursos de publicidad y propaganda. [PAD 2010]

Padilla propone la elaboración de un Datamart para el área de ventas y así poder hacer frente a los problemas que hasta ese entonces se venían suscitando en la organización. Entre estos problemas destacan los siguientes:

- Al realizar el análisis FODA se encontró que la empresa tiene carteras elevadas por cobrar; es decir, otorga mucho tiempo de crédito a los clientes.
- Se encontró un problema en la actualización del kardex; es decir, no se controla la entrada y salida de mercadería.
- El porcentaje promedio de reclamos y devoluciones a nivel general son del 28 y 33%.

Luego con la implementación de un Datamart para ventas se pudo medir el nivel de ventas por producto, por vendedor, por zona, así como también analizar el porcentaje de las devoluciones por zonas, reclamos por zona y el número de empleados que fueron capacitados en diferentes cursos. Este análisis se realiza mediante semáforos donde se establecieron valores para los límites máximos y mínimos donde se indica que el color rojo no alcanzó la meta, el verde que superó la meta y cuando está amarillo se encuentra en un nivel medio. [PAD 2010]

Análisis, diseño e implementación de un Datamart para la Dirección Financiera y Recursos Humanos de la Escuela Politécnica del Ejército para una toma de decisión efectiva [BEL 2007]

La elaboración de este proyecto de fin de carrera fue desarrollada para la institución educativa ecuatoriana “Escuela Politécnica del Ejército”(ESPE); específicamente para la unidades de Finanzas y de Recursos Humanos encargada de administrar los procesos de selección, capacitación, evaluación, reclasificación, ascensos, promoción y bienestar social del personal docente y administrativo de la organización. [BEL 2007]

Beltrán opina que el Datamart desarrollado para la Escuela Politécnica del Ejército permitirá poco a poco mejorar el manejo de la información plasmándose en decisiones efectivas y oportunas integrando las áreas de Finanzas y Recursos Humanos mejorando así la gestión administrativa financiera de la organización. [BEL 2007]

Cabe mencionar que la herramienta usada por Beltrán fue Oracle Discoverer 4.0 ya que es la herramienta con la que la empresa desarrolla sus aplicaciones de análisis de información. Beltrán concluye que el proyecto de Business Intelligence desarrollado permitirá a los directivos de la ESPE tener información precisa y confiable que apoye a una correcta toma de decisiones.

La Tabla 1.1 muestra un resumen de los proyectos de fin de carrera evaluados.

Proyectos de Tesis	Área a elaborar el Datamart	Navegación por Internet	Exportación a Excel	Manejo de Perfiles	Resultados Óptimos
[GUE 2005]	Facturación y Ventas	Si	No	No	Si
[BAR 2005]	Compras	No	No	No	Si
[PAD 2010]	Ventas	No	Si	No	Si
[BEL 2007]	Finanzas y Recursos Humanos	No	No	No	Si

Tabla 1.1 - Proyectos de Fin de Carrera

Ante los proyectos de fin de carrera presentados anteriormente, la implementación de un sistema de inteligencia trajo consigo resultados óptimos para cada caso, la mayor preocupación reside en que problema se debe solucionar y determinar si la aplicación de inteligencia de negocios es la mejor estrategia de solución para ello.

1.3.2. Herramientas para la Implementación

WebSphere Data Integrator (Data Stage)

Data Stage es una herramienta que apoya a la recolección, integración y transformación de grandes volúmenes de datos, con estructuras de datos que van desde simples a muy complejas. Permite a las empresas resolver problemas de negocio a gran escala a través del procesamiento masivo de datos. Además permite la conectividad entre cualquier fuente de datos y cualquier aplicación. [DAT 2010]

SSIS

Microsoft Integration Services es una plataforma para la creación de soluciones empresariales de transformaciones de datos e integración de datos. Integration Services sirve para resolver complejos problemas empresariales mediante la copia o descarga de archivos, el envío de mensajes de correo electrónico como respuesta a eventos, la actualización de almacenes de datos, la limpieza y minería de datos, y la administración de objetos y datos de SQL Server. Los paquetes pueden funcionar por separado o conjuntamente con otros paquetes para hacer frente a las complejas necesidades de la empresa. Integration Services puede extraer y transformar datos de muchos orígenes distintos, como archivos de datos XML, archivos planos y orígenes de datos relacionales, y, posteriormente, cargarlos en uno o varios destinos.[SQL 2010]

Sunopsis

Es un producto que posee un alto desempeño e integración efectiva. Trabaja con una arquitectura ELT (extraer, cargar y transformar) en lugar de la arquitectura tradicional ETL. Esto implica un mayor rendimiento y escalabilidad, junto con un mantenimiento más fácil para las soluciones de integración de datos. [SUN 2010]

Microstrategy

Es un software usado para el análisis de datos, informes y cuadros de mando, inteligencias móviles, minería de datos, la previsión, gestión de operaciones y toma de decisiones ejecutivas; permitiendo de esta forma mejorar y predecir el comportamiento del negocio. [MIC 2010]

Cognos

Software del grupo IBM cuyo uso permite comprender mejor el movimiento del negocio y tomar decisiones adecuadas. [COG 2010]

Business Objects

Plataforma de Inteligencia de negocios que ofrece una amplia gama de herramientas y aplicaciones diseñadas para ayudar a optimizar el rendimiento empresarial conectando personas, información y empresas a través de redes de negocios. [BUS 2010]

Octopus

Es una herramienta de extracción, transformación y carga (ETL) de Microsoft basada en Java. Puede conectarse a cualquier fuente de datos JDBC y realizar transformaciones definidas en un archivo XML. [OCT 2010]

Pentaho

Es una herramienta de código abierto y de uso libre muy completa, pues incluye elaboración de reportes, cubos, data mining, ETL y una plataforma BI. [PEN 2010]

La Tabla 1.2 y Tabla 1.3 muestran un cuadro comparativo para poder realizar la elección de la mejor herramienta de ETL y de explotación respectivamente.

Herramienta	Permite ETL	Uso libre (Open Source)	Soporte de la Herramienta	Soporta múltiples plataformas
Data Stage	Si	No	Si	Si
SQL Server	Si	No	Si	Si
Sunopsis	Si	No	-	-
Pentaho	Si	Si	No	Si
Octopus	Si	Si	-	-

Tabla 1.2 - Herramientas de ETL

Herramienta	Elaboración de Reportes	Uso libre (Open Source)	Análisis de la información en todas las dimensiones	Exportación a archivos Excel para visualizar la información	Drill and Down	Filtro y búsquedas personalizadas	Slice and Dice	Soporte de la Herramienta	Soporta múltiples plataformas
SQL Server	Si	No	Si	Si	Si	Si	Si	Si	Si
Microstrategy	Si	No	Si	Si	Si	Si	Si	Si	Si
Business Objects	Si	No	Si	Si	Si	No	Si	Si	Si
Cognos	Si	No	Si	Si	Si	Si	Si	Si	Si
Pentaho	Si	Si	Si	Si	Si	Si	Si	No	Si

Tabla 1.3 Herramientas de explotación

Se elegirá la herramienta ETL (Kettle) de Pentaho para la extracción, transformación y carga de datos debido a que es una herramienta muy potente en lo que ha extracción y transformación se refiere; y es Open Source lo que permitirá ahorrar costos a la organización. Además esta puede ser soportada en múltiples

plataformas de Windows, Linux, entre otras. Asimismo esta herramienta cuenta con una gama de funcionalidades que no tienen otras herramientas facilitando la elaboración de los jobs ETL; así por ejemplo cuenta con un stage de correos de alerta, stage visualización de logs, stage uso de scheduler; stage de estadísticas de redes sociales como facebook y twitter, entre otros. Si bien no tiene soporte por ser una herramienta Open Source, la página web brinda un foro de ayuda de la herramienta.

Además para el procesamiento de cubos y creación de reportes se empleará la herramienta SQLServer. De esta se empleará básicamente las funcionalidades de Analysis Services para el procesamiento de cubos OLAP y Reporting Services para la creación de reportes. Se usará SQLServer dado a que esta permite consultar los cubos no sólo por web, sino también de forma dinámica a través de Microsoft excel. Para ello se ha descargado la versión de prueba de SQLServer 2008 R2 que tiene agregadas ambas funcionalidades. Del mismo modo los usuarios cuentan con una mayor experiencia en Microsoft excel ya que es la herramienta donde todos los reportes de ventas de la organización son actualmente elaborados, lo que permitirá su fácil uso y la rápida adaptación de estos ante la nueva forma de explotar los datos.

Resumiendo, se empleará la herramienta ETL de Pentaho para la extracción, transformación, carga y limpieza de datos; la herramienta SQLServer para la creación y procesamiento de cubos y los reportes podrán ser explotados vía web o a través de Microsoft excel.

1.4. Planificación

La presente sección muestra los detalles sobre el proyecto de fin de carrera a realizar como la justificación y los periodos en que se realizará cada tarea.

1.4.1. Análisis del Proyecto

Proyecto:

Análisis, Diseño e Implementación de Datamarts para las áreas de Ventas y Recursos Humanos de una empresa dedicada a la exportación e importación de productos alimenticios.

Justificación del Proyecto:

La toma de decisiones en gerencia se realiza de manera tardía; es por esto que se requiere tener resultados inmediatos de manera que se sepa cuándo, cómo y dónde realizar las operaciones de venta de manera óptima y con ello mejorar los ingresos para la empresa en cuestión; así como equilibrar la relación entre lo que se produce y lo que se vende. Del mismo modo se desea saber los movimientos de entrada y salida, licencias y vacaciones del personal administrativo para determinar si estos cumplen o no con sus horas obligatorias dentro de la organización; así como poseer informes sobre la forma en que se administra el presupuesto, destinado al pago de los trabajadores de la organización.

En base a lo anterior se propone la implementación de dos Datamarts para las áreas de ventas y recursos humanos que permita tener integrado los datos de los distintos aplicativos o fuentes en un solo repositorio Datawarehouse, emplear instrucciones avanzadas de consultas SQL (*MDX*) y usar técnicas de navegación como *drill down/up*, *slice*, *dice*, *pivot*, entre otros; y con ello poder realizar una buena gestión de la información mejorando así los niveles de rentabilidad dentro de la organización.

Descripción del Proyecto:

El proyecto tiene como fin cubrir las principales necesidades de explotación de información de las áreas de Ventas y Recursos Humanos mediante el uso de reportes, gráficos web y selecciones dinámicas en excel que ayuden a la toma de decisiones para una empresa dedicada a la exportación e importación de productos alimenticios.

Actividades del Proyecto

Los objetivos del presente proyecto de fin de carrera son los siguientes:

- Realizar un análisis dimensional que sirva de soporte a las decisiones de los procesos de las áreas de ventas y recursos humanos.
- Realizar el diseño de la extracción, transformación y carga de los datos.
- Implementar los Datamarts de ventas y recursos humanos.
- Implementar los cubos de ventas y recursos humanos.
- Diseñar los reportes adecuados para las áreas de ventas y recursos humanos.
- Realizar los reportes para las áreas de ventas y recursos humanos.

1.4.2. Cronograma del Proyecto

El siguiente cronograma muestra cuánto tiempo durará cada una de las tareas que se requieren.

Trabajo: 900 horas

Las horas de trabajo se distribuyen en las fases del proyecto de la siguiente manera:

- Marco Conceptual: 88 horas.
- Análisis: 184 horas
- Diseño: 200 horas.
- Construcción y Pruebas: 832 horas.
- Observaciones, Conclusiones y Recomendaciones: 16 horas.

Roles Requeridos:

- Analista Funcional.
Quien será el encargado de levantar información y hacer los requerimientos.
- Diseñador del modelo.
Encargado de plasmar en un modelo las reglas de negocio en las cuales se basa la información.
- Analista Programador.
Encargado de codificar los requerimientos solicitados por medio de técnicas alcanzadas.
- Certificador:
Encargado de realizar los casos de prueba para validar que lo construido este de acuerdo a los requerimientos funcionales.
- Usuario.
Encargado de brindar las necesidades y sus detalles.

Cronograma de Actividades:

La Figura 1.12 ilustra el diagrama de actividades para el presente proyecto de fin de carrera. La ampliación de esta figura se encuentra en el **Anexo L - Figuras**.

ID	Nombre de tarea	Duración	Comienzo	Fin	Prede	Nombres de los recursos
1	Elaboración de los Datamarts de Ventas y Recursos Humanos	165 días	mar 01/03/11	lun 17/10/11		
2	Marco Conceptual	11 días	mar 01/03/11	mar 15/03/11		Analista Funcional
3	Definición del Problema	2 días	mar 01/03/11	mié 02/03/11		Analista Funcional
4	Definición de Conceptos	3 días	jue 03/03/11	lun 07/03/11	3	Analista Funcional
5	Definición de Áreas de la Organización	2 días	mar 08/03/11	mié 09/03/11	4	Analista Funcional
6	Estado del Arte	1 día	jue 10/03/11	jue 10/03/11	5	Analista Funcional
7	Planificación del Proyecto	3 días	vie 11/03/11	mar 15/03/11	6	Analista Funcional
8	Análisis	23 días	mié 16/03/11	vie 15/04/11	2	Diseñador del Modelo
9	Documento de Análisis	7 días	mié 16/03/11	jue 24/03/11	7	Diseñador del Modelo
10	Requerimientos	1 día	vie 25/03/11	vie 25/03/11	9	Diseñador del Modelo
11	Análisis Dimensional	15 días	lun 28/03/11	vie 15/04/11	10	Diseñador del Modelo
12	Diseño	25 días	lun 18/04/11	vie 20/05/11	8	Diseñador del Modelo, Analista Programador
13	Modelamiento Dimensional	10 días	lun 18/04/11	vie 29/04/11	11	Diseñador del Modelo
14	Especificaciones del ETL	15 días	lun 02/05/11	vie 20/05/11	13	Analista Programador
15	Construcción	104 días	lun 23/05/11	jue 13/10/11	12	Analista Programador
16	Instalación de Pentaho	1 día	lun 23/05/11	lun 23/05/11	14	Analista Programador
17	Instalación de SQL Server	1 día	mar 24/05/11	mar 24/05/11	16	Analista Programador
18	Construcción del ETL de Ventas	30 días	mié 25/05/11	mar 05/07/11	17	Analista Programador
19	Construcción del ETL DE Recursos Humanos	30 días	jue 04/08/11	mié 14/09/11	22	Analista Programador
20	Construcción del Cubo de Ventas	7 días	mié 06/07/11	jue 14/07/11	18	Analista Programador
21	Construcción del Cubo de Recursos Humanos	7 días	jue 15/09/11	vie 23/09/11	19	Analista Programador
22	Elaboración de los Reportes de Ventas	14 días	vie 15/07/11	mié 03/08/11	20	Analista Programador
23	Elaboración de los Reportes de Recursos Humanos	14 días	lun 26/09/11	jue 13/10/11	21	Analista Programador
24	Pruebas	104 días	lun 23/05/11	jue 13/10/11		Certificador
25	Elaboración del Plan de Pruebas	104 días	lun 23/05/11	jue 13/10/11	15CC	Certificador
26	Desarrollo de las Pruebas	104 días	lun 23/05/11	jue 13/10/11	15CC	Certificador
27	Observaciones, Conclusiones y Recomendaciones	2 días	vie 14/10/11	lun 17/10/11	24	Analista Funcional
28	Documento de Conclusiones	2 días	vie 14/10/11	lun 17/10/11	26	Analista Funcional

Figura 1.12 - Cronograma de Actividades

Diagrama de Gantt:

La Figura 1.13 ilustra cómo se genera el diagrama de Gantt de acuerdo a las actividades establecidas en el cuadro anterior: Además muestra los recursos empleados en cada actividad realizada. La ampliación de esta figura se encuentra en el **Anexo L - Figuras**.

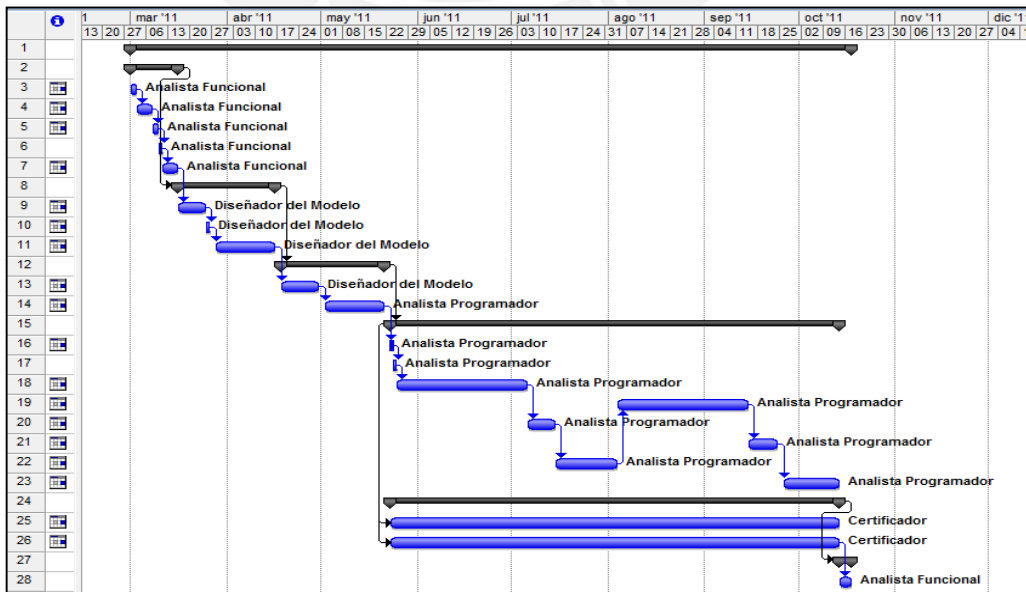


Figura 1.13 - Diagrama de Gantt

Diagrama WBS:

La Figura 1.14 ilustra las fases en las que se ha dividido el proyecto; así como las etapas en que cada una de estas se ha dividido. La ampliación de esta figura se encuentra en el **Anexo L - Figuras**.

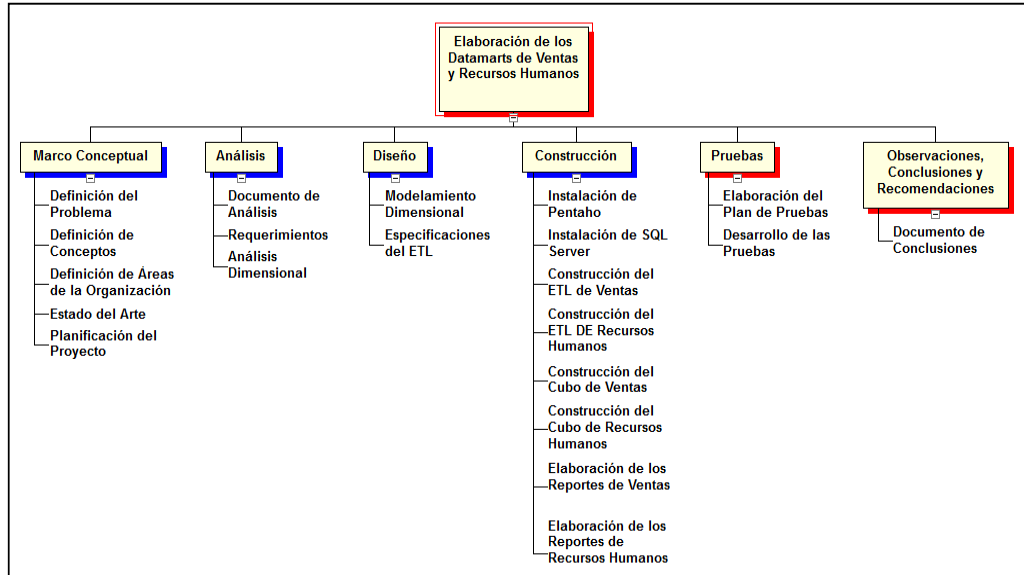


Figura 1.14 - Diagrama WBS

1.5. Descripción y Sustentación de la Solución

En las constantes entrevistas que se tuvo con los representantes de la organización en estudio se detectó que la problemática de la empresa, tal y como se detalla en el punto *1.1 Identificación del Problema*, radica en que no se explota información relacionada con las áreas de ventas y recursos humanos ocasionando que la gerencia no logre tomar las decisiones correctas en el momento correcto. Es por esto que se requiere tener resultados de las tendencias de las ventas en los diferentes puntos del litoral peruano de manera que no solo se pueda saber el momento preciso para realizar una venta, sino también lograr un equilibrio entre lo que se produce y lo que vende; además se desea llevar un control de las metas que logran los vendedores y un control de las devoluciones de productos que puedan generarse. Del mismo modo se desea contar con información que permita saber los movimientos de entrada y salida, licencias y vacaciones del personal administrativo para determinar si estos cumplen o no con sus horas obligatorias; así como poseer informes relacionados con el presupuesto con el que cuenta la organización.

La implementación de Datamarts para las áreas de ventas y recursos humanos es una buena solución para este tipo de situación ya que permitirá a los usuarios finales tomar decisiones, relacionadas a las ventas y al personal de la organización, de una manera rápida y óptima.

El uso de los Datamarts de ventas y recursos humanos permitirá que se cuente con dimensiones (*conjunto de información acerca de un evento*) que puedan interrelacionarse entre sí y brindar información a detalle sobre una venta, personal o presupuesto de la organización en estudio; así por ejemplo podemos obtener la consulta sobre *las ventas realizadas en el segundo trimestre para los clientes del departamento de Ica*, o la consultas sobre *el personal nombrado del área de producción que no cumplió con sus horas obligatorias en el mes de diciembre del año 2006*. Además el uso de Datamarts permitirá tener resultados predefinidos en una o más tablas de hechos (*tablas que relaciona las dimensiones y los eventos o hechos*); como *el monto o la cantidad de ventas mensuales, horas diarias trabajadas del personal*, entre otros; los que junto a las dimensiones permitirán que las consultas sean mucho más óptimas y rápidas en comparación con las bases de datos transaccionales en donde las consultas complejas implica barrerse varias tablas del modelo.

La propuesta de valor de la implementación de los Datamarts se basa en los siguientes puntos:

- Integrar los datos de los distintos aplicativos en un DWH respetando las reglas de negocio existentes.
- Optimización de los procesos a través de la gestión, monitoreo y mantenimiento de estos empleando instrucciones SQL avanzadas.
- Uso de técnicas como drill down/up, slice, dice, pivot, entre otras; cuyos conceptos se encuentran en el punto 1.2.1 *Conceptos: Inteligencia de negocios*; que permitan un mayor análisis de la información.

Por otro lado, el siguiente proyecto permitirá, en primera instancia, presentar reportes que permitan que los encargados de tomar las decisiones lo hagan de una forma rápida y correcta. El primer Datamart centra su análisis en las ventas que realiza la empresa mediante un historial que brinde información sobre cuáles son

sus principales clientes, los productos que tienen mayor demanda o rentabilidad, las zonas de ventas, entre otros. Una vez hecho el análisis de las ventas se podrán tomar acciones correctivas en cuanto al sobre-stock de productos alimenticios y tomar buenas decisiones en cuanto a la oferta y demanda de productos. Por otro lado; el segundo Datamart centra su análisis en el personal con el que cuenta la organización de manera que se tenga un mayor control del personal administrativo; así como, un buen manejo del presupuesto que posee la empresa.

El beneficio que traerá el presente proyecto se centra en las áreas de ventas y recursos humanos identificando oportunidades de negocio y reducción de costos para la empresa. Por un lado, se incrementará la utilidad ya que se podrá detectar cómo, cuándo y dónde poder realizar las ventas que proporcionen una mayor utilidad a la organización, así como poder reducir el sobre-stock de productos que ha originado pérdidas considerables en la organización. Por otro lado, se tendrá un mayor control del horario de trabajo del personal administrativo, así como una mejor administración del dinero con el que cuenta la empresa destinado al pago de los trabajadores.

En resumen, la solución de inteligencia de negocios planteada no solo brindará una serie de reportes con información resumida de los procesos de negocio; sino también permitirá hacer un análisis de la información a través de las diversas funcionalidades existentes como drill down, slice, etc. Ello permitirá que los encargados de tomar las decisiones de la empresa, lo hagan de una manera óptima en base a un análisis detallado y correcto.

2. CAPÍTULO 2: Análisis

En este capítulo se hace hincapié en la metodología a usar. Además se describe los requerimientos funcionales y no funcionales que deberá cumplir la solución de inteligencia de negocios. Por último se detalla el análisis dimensional de los Datamarts de ventas y de recursos humanos.

2.1. Metodología

Casi todo proyecto de ingeniería, estructural o de software, pasa a través de seis etapas desde la concepción hasta la implementación. La Figura 2.1 y la Figura 2.2 muestran el ciclo de vida para desarrollar un sistema de soporte de decisiones, tomando como base el ciclo que propone Moss Larissa. [MOS 2003]

2.1.1. Ciclo de Vida de Un sistema de Soporte a Decisiones

Justificación:

Toma como punto de partida la necesidad de la organización de un nuevo sistema de soporte a decisiones. [MOS 2003]

Evaluación de la Organización

El problema o la oportunidad de negocio es definida y una solución de Inteligencia de negocios es propuesta. La aplicación de Inteligencia de Negocios debe de ser justificada; planteando una solución que resuelva el problema de la organización, especificando las desventajas y beneficios que esta tendría. [MOS 2003]

Planeamiento:

En esta etapa se desarrolla las estrategias y planificaciones que permitirán saber cómo se logrará y desarrollará el proyecto. [MOS 2003]

Evaluación de la Infraestructura de la Organización

Estructura que soporta todas las aplicaciones de Inteligencia de Negocios de la organización y posee dos componentes: [MOS 2003]

- Infraestructura técnica: incluye hardware, software, sistemas de manejo de base de datos, sistemas operativos, sistema de red, repositorios de metadata, sistemas utilitarios. [MOS 2003]
- Infraestructura no técnica: estándares de metadata, modelo lógico del negocio, metodologías, procedimiento de pruebas, procesos del control de cambio. [MOS 2003]

Planeamiento del Proyecto

Los planeamientos de proyectos deben ser detallados, y el cumplimiento de este debe de ser seguido y reportado ya que los cambios en el alcance, el personal, el presupuesto, la tecnología, y los sponsors pueden afectar en gran medida el éxito del proyecto. [MOS 2003]

Análisis del Negocio:

En esta etapa se desarrolla un análisis del problema o de la oportunidad en la organización; y así poder entender en forma completa los requerimientos para una posible solución. [MOS 2003]

Definición de los Requerimientos del Proyecto

En esta etapa se debe de elegir los requerimientos de acuerdo a las posibilidades de cada entregable, sin crear falsas expectativas. [MOS 2003]

Análisis de los Datos

En esta etapa se evalúan los datos fuentes con la finalidad de obtener una mejor calidad de estos. [MOS 2003]

Prototipo de Aplicación

Esta etapa permite a los desarrolladores y a los involucrados ver el potencial y las limitaciones de la tecnología, y también brinda la oportunidad de ajustar los requerimientos del proyecto, y las expectativas del mismo. [MOS 2003]

Análisis del Repositorio de Metadata

La metadata debe ser guardada en un repositorio cuyos requerimientos deben ser documentados en un modelo lógico. [MOS 2003]

Diseño:

Esta etapa se concibe un producto que solucionará el problema de la organización; es decir se describirá cómo y de qué manera se realizará la aplicación de soporte a decisiones. [MOS 2003]

Diseño de la Base de Datos

El diseño de la base de datos debe estar acorde con los requerimientos planteados para acceder a la información con la que cuenta la organización. [MOS 2003]

Diseño ETL (Extracción, Transformación y Carga)

En esta etapa se realizarán las especificaciones del ETL y se fusionarán los datos que vienen de diversas plataformas en un formato para el Datawarehouse. [MOS 2003]

Diseño del Repositorio de Metadata

Si se construye el repositorio, se debe tomar la decisión si éste estará basado en entidad-relación u orientado a objetos. [MOS 2003]

Construcción:

En esta etapa se construye el producto en un margen de tiempo predeterminado. [MOS 2003]

Desarrollo ETL

En esta etapa se elige la herramienta ETL y se realiza el proceso de extracción, transformación y carga de datos. Para elegir qué herramienta de ETL conviene más se debe tomar en cuenta la limpieza de los datos, la transformación requerida para los datos, el análisis de los datos, el diseño, el costo y beneficios del ETL. [MOS 2003]

Desarrollo de la aplicación

Una vez que el prototipo se ha ajustado a los requerimientos de la empresa u organización, se dará inicio al análisis y el desarrollo de cómo se accederá a los datos. El desarrollo de la aplicación que se mostrará al usuario, se hará en forma paralela con el desarrollo del ETL y del repositorio de meta data. [MOS 2003]

Data Mining

Por medio del Data Mining se puede descubrir información en los datos que generalmente permanecen escondidas y con ello aprovechar el sistema de soporte a decisiones. Esta etapa no se tomará en cuenta para el presente proyecto de fin de carrera. [MOS 2003]

Instalación:

Se implementa el producto final y luego se mide si este cubre los requerimientos funcionales. [MOS 2003]

Implantación

Con la implantación se comienza a instalar el motor de la base de datos y la plataforma Web. Se programa entrenamiento para los usuarios. En esta etapa comienza las funciones de soporte, que incluye mantener la base de datos, programar y correr los jobs ETL, monitorear el comportamiento del sistema y afinar la base de datos. [MOS 2003]

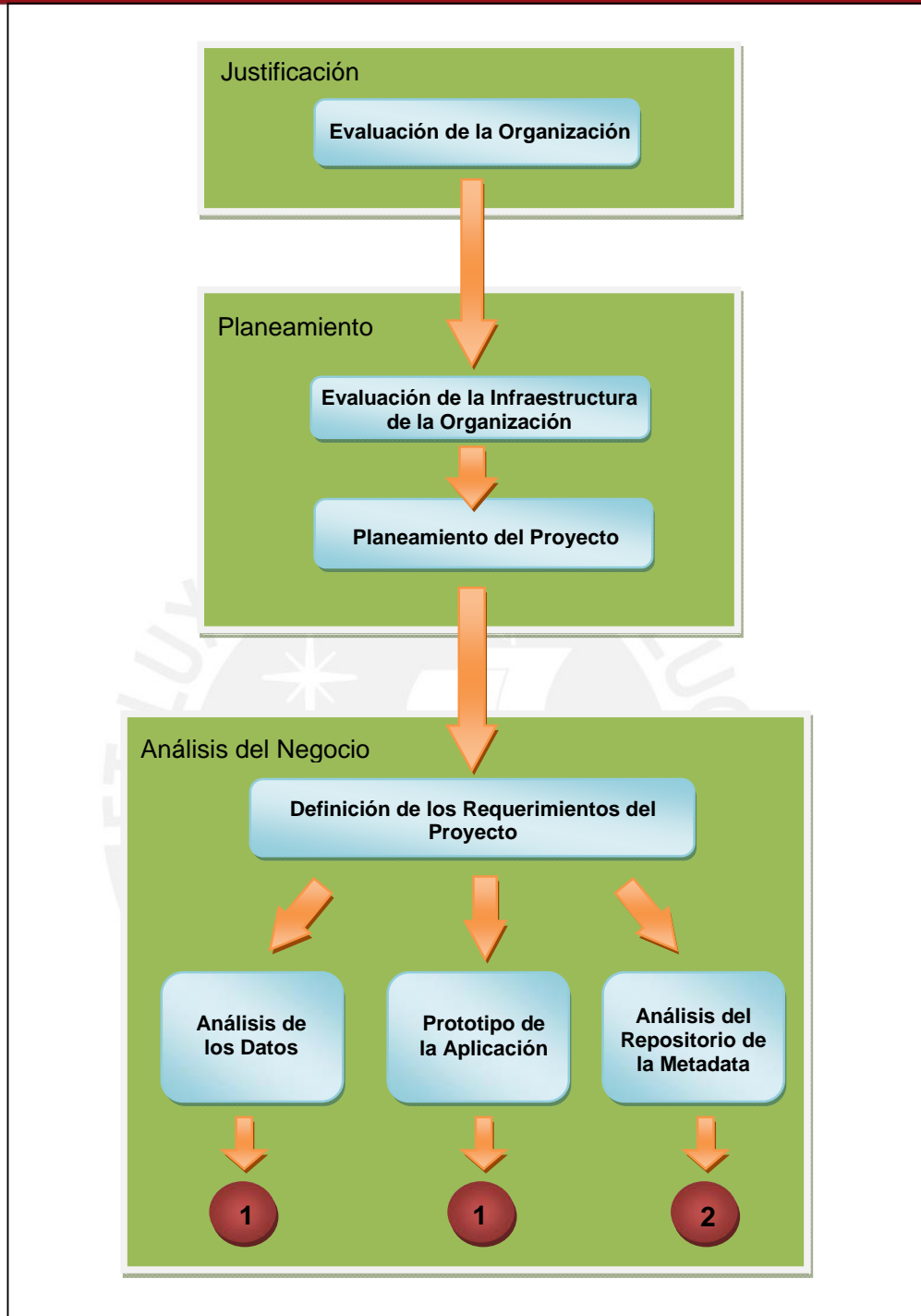


Figura 2.1 - Ciclo de vida para un proyecto de BI [MOS 2003]

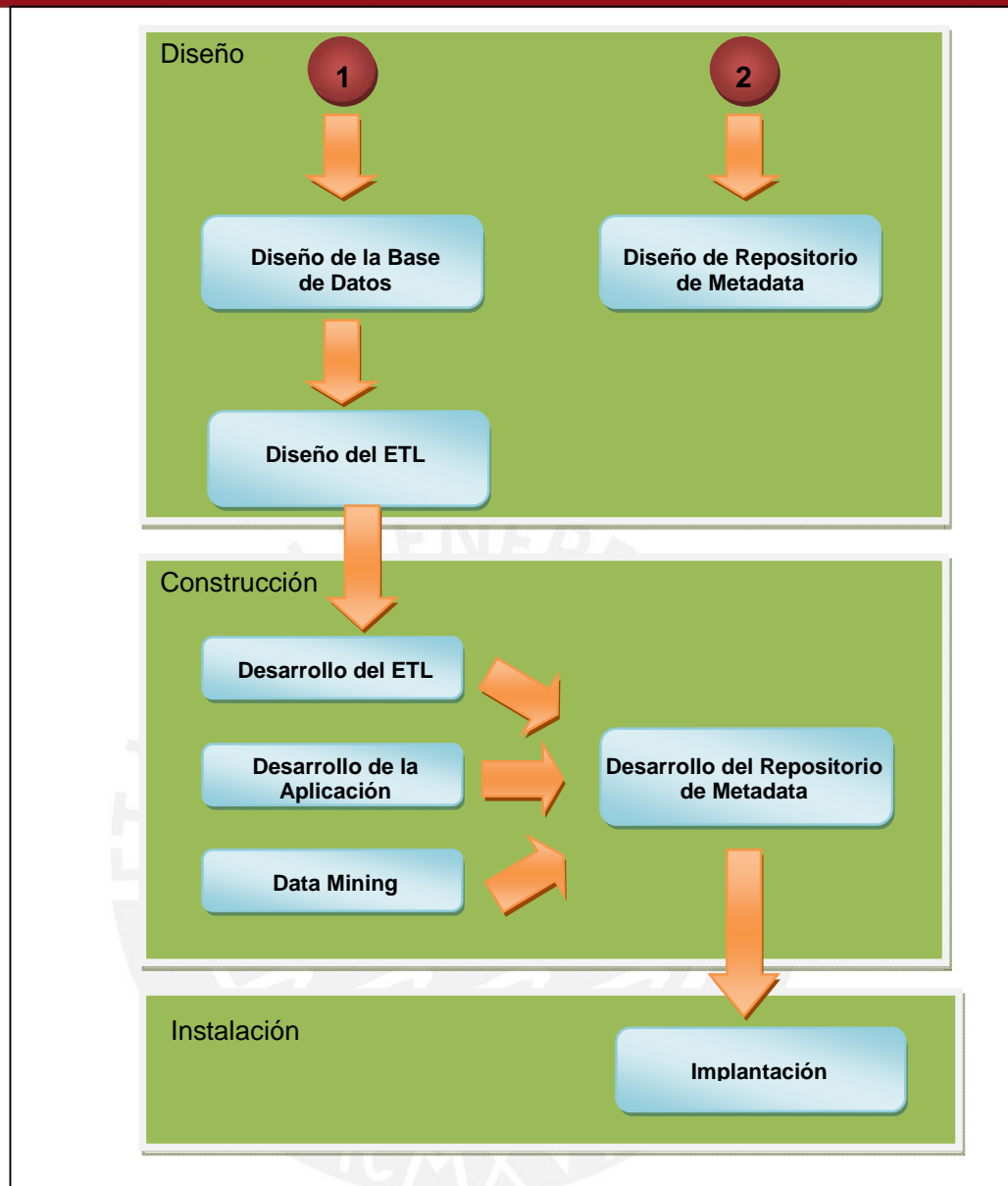


Figura 2.2 - Ciclo de vida para un proyecto de BI [MOS 2003]

2.1.2. Herramientas a utilizar

En cuanto a las herramientas de Inteligencia de Negocios a utilizar se ha elegido una de las que pertenecen al proyecto Pentaho, específicamente para el desarrollo del ETL, porque es una herramienta muy completa, de uso libre, muy bien documentada y que ha demostrado excelentes resultados en otros proyectos. Para el procesamiento de cubos y reportes se usará las funcionalidades de Analysis Services y Reporting Services del paquete de SQL Server 2008; así como la herramienta Excel para la explotación de los datos de forma dinámica.

Plataforma BI

Se optó por utilizar un híbrido de herramientas existentes. Por un lado, la herramienta Kettle de Pentaho debido a que es una herramienta de licencia libre lo cual permitirá ahorrar costos, facilitar el monitoreo de la ejecución de “jobs” y trabajar con una interfaz muy amigable que permite conectarse a diversas bases de datos; además posee gran documentación y una versión de demostración que permite aprender rápido el manejo de la herramienta. Por otro lado, se usará las funcionalidades de Analysis Services y Reporting Services del paquete de SQL Server 2008 para el procesamiento de cubos y reportes; esto es, debido a que es una de las herramientas con las que actualmente la empresa opera. Además, una funcionalidad importante de SQL Server es la explotación de los datos en excel, herramienta con la cual actualmente, los gerentes y jefes de la organización, analizan los cuadros del negocio; pudiendo facilitar, de este modo, el uso de la misma, pero esta vez orientado al análisis de los cubos BI.

ETL

Se utilizará Kettle Data Integration para el proceso de ETL. Kettle es una herramienta libre y forma parte del proyecto Pentaho. Kettle es una herramienta gráfica, lo que facilita su aprendizaje y su uso y puede trabajar con diversas fuentes de datos y conectarse a muchos motores de base de datos, tanto como fuente, como para destino. Además posee rutinas que facilitan el proceso de limpieza de datos.

Diseñador de Cubos

Se utilizará la funcionalidad de Analysis Services que viene junto a la herramienta de SQLServer 2008 de manera que los resultados, no sólo sean mostrados en reportes web, sino también puedan ser explotados en hojas de Excel y de forma dinámica permitiendo que su uso sea más fácil para las personas a quienes va dirigido el presente proyecto de fin de carrera. Debido al uso de excel dentro de la organización es que se opta por Analysis Services.

Diseñador de Informes

Para los reportes, como se menciona anteriormente, se utilizará reportes en web y además la herramienta Excel. En esta última se podrá generar cuadros dinámicos y gráficos que ayuden a tomar las mejores decisiones.

2.1.3. Enfoques

Básicamente se tienen dos enfoques claramente marcados:

Enfoque de W.H Inmon

Según W.H. Inmon uno de los mayores problemas que presenta la “arquitectura telaraña” de sistemas de información, es la dificultad en obtener información de los datos que estos sistemas han generado. [INM 2003]

W.H. Inmon plantea un “ambiente arquitecturado” con 4 niveles de datos:

1. Operacional
2. Atómico/Datawarehouse
3. Departamental
4. Individual

El nivel operacional contiene los datos primitivos generados por los sistemas OLTP, son muy volátiles (normalmente se tiene el último dato actualizado) y están orientados a las aplicaciones. [INM 2003]

En el nivel de Datawarehouse, los datos están integrados, orientados a un tema, son detallados como los primitivos, no pueden ser actualizados por transacciones, puede contener algunos datos derivados y guardan la historia. [INM 2003]

En el nivel departamental o nivel de Datamart los datos son derivados y están modelados de acuerdo los requerimientos de análisis del usuario departamental específico. Finalmente el nivel individual es aquel en el que se ha hecho un análisis sobre los datos. [INM 2003]

El enfoque de W.H. Inmon también es conocido como el enfoque “Top-down” porque plantea la arquitectura de capas donde hay una capa integradora, luego a partir de esta capa se obtienen los datos a nivel agregado de Datamart y luego a partir del Datamart un nivel aún mas agregado que es a nivel individual. [INM 2003]

Enfoque de Ralph Kimball

El enfoque de Ralph Kimball también es conocido como el enfoque “Bottom-up” porque parte de la premisa que el Datawarehouse se construye a partir de un Datamart que está orientado a un proceso de negocio y normalmente está referido a un tema puntual. Luego se desarrollan otros Datamarts haciendo que las dimensiones sean comunes y sean las fuentes de integración. [KIM 2002]

El enfoque “Bottom-up” producirá resultados rápidos, un enfoque “Top-down” tomará más tiempo dependiendo del tamaño del proceso de negocio específico que se haya modelado en el Datamart. [KIM 2002]

Así mismo plantea que el Datawarehouse debe tener los siguientes componentes:

1. Sistemas operacionales fuente
2. Data staging área
3. Datos de presentación
4. Herramientas de acceso a datos

Los sistemas operacionales fuente son los que capturan los registros producidos en cada transacción de negocio, las consultas se caracterizan porque actualizan o leen un registro a la vez. [KIM 2002]

El segundo componente llamado “Data staging área” es el sistema que permanece entre las fuentes y el Datawarehouse o Datamarts cuyo objetivo es facilitar la extracción, limpieza y calidad de datos mediante los procesos ETL. [KIM 2002]

El componente de presentación es donde los datos están almacenados, organizados, y listos para ser consultados y analizados, este componente es un conjunto de Datamarts integrados a través de sus dimensiones comunes, también llamadas dimensiones conformes, y un Datamart presenta la información sobre un proceso de negocio. [KIM 2002]

La capa de herramientas de acceso que son el conjunto de herramientas que consultan los datos, también se denomina la capa de presentación, estas herramientas pueden ser tan simples como una herramienta para hacer consultas (Query tool) o tan sofisticadas como herramientas de minería de datos o

modelamiento pasando por las herramientas analíticas o herramientas OLAP que sería las un nivel de complejidad intermedio. [KIM 2002]

Para el proyecto de fin de carrera se tomara el enfoque de Kimball ya que de acuerdo a las necesidades del negocio conviene la implementación de Datamarts que permitan dar solución a los problemas que actualmente existen y poco a poco crecer y formar un Datawarehouse; además este enfoque permite que los resultados sean obtenidos de manera rápida.

2.1.4. PMBOK

Las áreas de conocimiento de PMBOK relacionadas para el presente proyecto de fin de carrera son las siguientes.

Gestión de Alcance

El presente proyecto de fin de carrera tiene un límite propuesto, el cual es la implementación de los Datamarts para las áreas de ventas y recursos humanos.

Gestión de Tiempo.

El presente proyecto de fin de carrera está conformado por etapas de desarrollo, cada una de las cuales comprende un espacio de tiempo determinado, los cuales deben de cumplirse.

Gestión de Calidad

El presente proyecto de fin de carrera debe de tener un rango de calidad, tanto en los datos para que se pueda realizar un adecuado manejo de estos, así como también, calidad en los resultados con los cuales se tomarán las decisiones de la organización.

Gestión de Recursos Humanos

El presente proyecto de fin de carrera será elaborado por personal externo a la organización desde su análisis hasta su implementación.

2.2. Identificación de los Requerimientos

En este apartado se detalla los requerimientos funcionales y no funcionales que son necesarios para la elaboración del presente proyecto de fin de carrera.

2.2.1. Requerimientos Funcionales

Los requerimientos funcionales son tomados a partir de la metodología anteriormente planteada teniendo en cuenta las necesidades de los usuarios en la empresa. A continuación se muestra un cuadro con los requerimientos funcionales:

Para ambos Datamarts

La Tabla 2.1 muestra los requerimientos funcionales para ambos Datamarts.

Número	Requerimiento	Nivel de Prioridad
1	Se extraerán los datos de los diversos sistemas de OLTP que existen en la empresa.	1
2	Se estandarizará los datos obtenidos de las fuentes de OLTP	2
3	Se realizará la limpieza de los datos según los criterios de la empresa, de manera que se pueda verificar la validez y calidad de estos.	2
4	Se creará los Datamarts de ventas y recursos humanos.	1
5	Se automatizará el proceso ETL para cada Datamart, según la necesidad de actualización de cada uno de estos	3
7	Se creará ETL totales para las cargas absolutas y ETL incrementales para las cargas periódicas.	3
8	Se implementará los cubos para las áreas de ventas y recursos humanos.	1

Tabla 2.1 - Requerimiento Funcionales para Ambos Datamarts

Para el Datamart de Ventas

La Tabla 2.2 muestra los requerimientos funcionales para el Datamart de Ventas:

Número	Requerimiento	Nivel de Prioridad
1	Se implementará un reporte que muestre para cada tipo de productos los montos de ventas, cantidad de ventas realizadas, cantidad de devoluciones, el monto devuelto y el factor devuelto (porcentaje de devoluciones del total de la cantidad vendida) por zona, cliente y producto. Cuando el factor devuelto sea mayor a 12% se mostrarán unos semáforos de alerta. Se usarán las dimensiones Ubicación, Cliente Producto y Tiempo.	2

2	Se implementará un reporte que muestre el monto de las ventas por productos para determinados periodos de tiempo. Se manejará una jerarquía para la dimensión Producto (tipos de producto, producto) y para la dimensión Tiempo (año, trimestre y mes). También se usará la dimensión Ubicación.	1
3	Se implementará un reporte que muestre para cada zona el monto de ventas que realizaron los vendedores en un determinado periodo de tiempo. Se manejarán semáforos que indique quienes son los vendedores que cumplieron o no con las ventas establecidas por la organización; las dimensiones involucradas son Vendedor, Ubicación y Tiempo. Además el reporte permitirá hacer un salto a otro reporte donde se muestra a detalle cada venta de un vendedor en particular	1
4	Se implementará un reporte que muestre las ventas a detalles de cada vendedor en particular. Servirá como respaldo para el reporte N° 3 cada vez que se desee ver a detalle las ventas de algún vendedor en particular. Se usarán las dimensiones Ubicación, Tiempo, Vendedor, Producto y Cliente.	3
5	Se implementará un gráfico de líneas que muestre los montos de ventas para un periodo de tiempo. Se usarán las dimensiones Tiempo, Producto y Ubicación.	1
6	Se implementará un gráfico de barras que muestre los montos de ventas de cada línea de producto para cada zona. Las dimensiones involucradas son Ubicación, Tiempo y Producto.	1
7	Se implementará un gráfico de pipes con subreportes que muestre los porcentajes de ventas para cada zona. Las dimensiones a usar son Ubicación, Tiempo y Producto.	1
8	Se implementará un reporte donde se muestre lo montos de ventas para cada zona en un determinado periodo de tiempo. Se manejará una jerarquía para la dimensión Ubicación y otra para la dimensión Tiempo.	1
9	Se implementará un reporte donde se muestre la cantidad de productos devueltos y el monto de los productos que han sido devueltos en un determinado periodo de tiempo. Las dimensiones a usar son Ubicación, Tiempo y Producto. Adicionalmente se trabajará con semáforos que indiquen si un producto determinado ha pasado el porcentaje de devolución permitida por la empresa. (12% de la cantidad de productos vendidos)	1

Tabla 2.2 - Requerimiento Funcionales el Datamart de Ventas

Para el Datamart de Recursos Humanos

La Tabla 2.3 muestra los requerimientos funcionales para el Datamart de Recursos Humanos:

Número	Requerimiento	Nivel de Prioridad
1	Se implementará un reporte donde se muestre las horas obligatorias, las horas trabajadas y el porcentaje de cumplimiento de trabajo en el mes por cada empleado administrativo. Las dimensiones a usar son Oficina, Personal, Tiempo y Puesto.	1
2	Se implementará un gráfico de líneas que muestre la historia del porcentaje de horas trabajadas por oficina en un determinado periodo de tiempo. Las dimensiones involucradas son Oficina y Tiempo.	1

3	Se implementará un reporte donde se muestre el total de horas de licencia que tiene cada empleado administrativo por tipo de licencias. Las dimensiones involucradas son Oficina, Personal, Puesto, Licencia y Tiempo.	3
4	Se implementará un reporte donde se muestre por cada persona, el número de horas extras trabajadas y el importe que estas generaron en la empresa para un determinado periodo de tiempo. Se manejará una jerarquía para la dimensión Tiempo. Las dimensiones a usar son Persona, Puesto, Tiempo y Oficina.	1
5	Se implementará un reporte donde se muestre por cada rol, la cantidad de personas y el importe total de sus sueldos brutos en un determinado periodo de tiempo. Se manejará una jerarquía para la dimensión Tiempo. Las dimensiones a usar son Rol, Tiempo y Oficina.	3
6	Se implementará un reporte que muestre la cantidad de personas contratadas y nombradas por cada área, así como el total de importe en planilla de cada una de ellas en un periodo determinado de tiempo. Se usarán semáforos para determinar cuáles son las áreas de la organización que sobrepasan o no el número límite de personas nombradas. Las dimensiones a usar son Oficina, Tiempo y Vínculo.	1
8	Se implementará un reporte que muestre el resumen de importe de pagos y descuentos en el mes por cada rol en un determinado periodo de tiempo. Se podrá navegar desde cada Rol listando los empleados que lo conforman. Las dimensiones a emplear son Personal, Tiempo, Vínculo y Rol.	1

Tabla 2.3 - Requerimiento Funcionales para el Datamart de Recursos Humanos

2.2.2. Requerimientos No Funcionales

Los requerimientos no funcionales son aquellos requerimientos relacionados directamente con la performance y rendimiento de los Datamarts; del mismo modo están relacionados con la elección de las herramientas a emplear para la elaboración de estos. La Tabla 2.4 muestra los requerimientos no funcionales:

Número	Requerimiento	Nivel de Prioridad
1	La extracción, transformación, carga y limpieza de datos se realizará con la herramienta Kettle de Pentaho, que es una herramienta Open Source.	1
2	La extracción de la data se realizara a mensualmente debido a que los reportes son analizados los primeros días de cada mes.	1
3	La creación y procesamiento de cubos y la creación de reportes se desarrollaran con la herramienta SQLServer 2008. Para ello se deberá descargar la versión de prueba SQLServer 2008 R2, la cual trae todas las funcionalidades que se empleará.	1
5	La interfaz de BI debe de ser fácil de usar con la finalidad que los usuarios no especializados en cuestiones informáticas pueda hacer uso de los reportes.	1
6	Los cubos podrán ser explotados dinámicamente mediante la herramienta Microsoft Excel 2007.	1

7	El sistema debe de realizar los reportes en un tiempo no mayor de 10 segundos.	1
8	La disponibilidad del aplicativo es durante las 24 horas del día.	1

Tabla 2.4 - Requerimiento No Funcionales

Leyenda de Prioridad:

1: Alta 2: Media 3: Baja

2.3. Análisis Dimensional

La presente sección muestra el análisis de las dimensiones tanto para el Datamart de ventas como para el Datamart de recursos humanos. Asimismo muestra el detalle de las fact tables y de las medidas correspondientes a cada una de éstas.

2.3.1. Datamart de Ventas

La Figura 2.3 ilustra el Datamart de ventas. La ampliación de esta figura se encuentra en el **Anexo L - Figuras**.

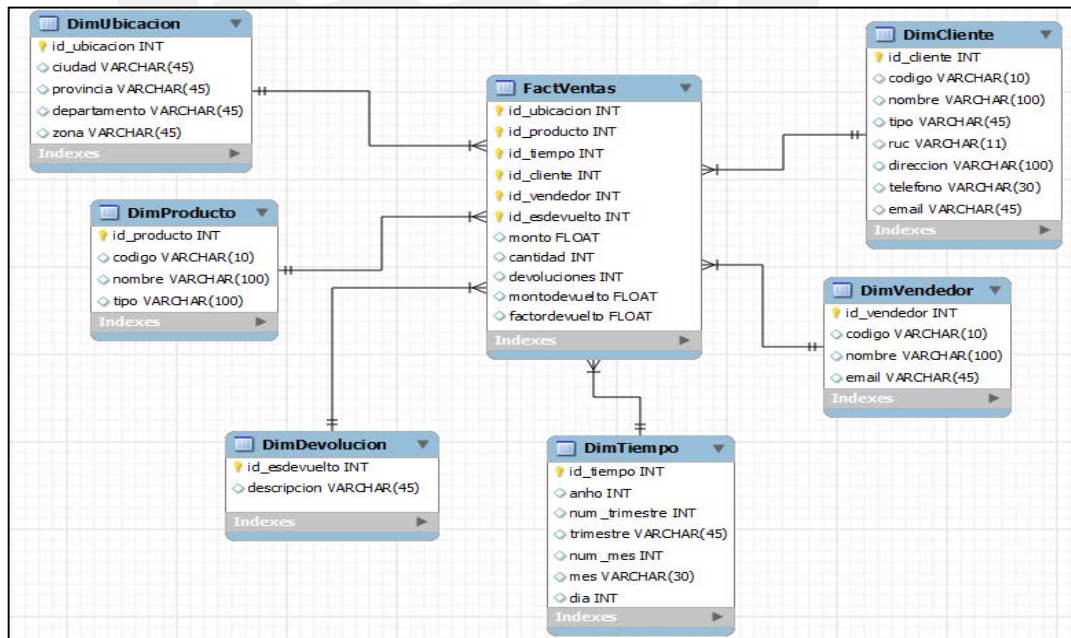


Figura 2.3 - Datamart de Ventas

El Datamart de ventas posee las dimensiones y la fact que permitirán dar soporte a los reportes de ventas que se plantearan en el siguiente capítulo. Estas dimensiones son obtenidas desde la base de datos de ventas de la organización.

Dimensiones

Las dimensiones que conforman el Datamart de Ventas son:

1. Dimensión Producto
2. Dimensión Cliente
3. Dimensión Tiempo
4. Dimensión Ubicación
5. Dimensión Vendedor
6. Dimensión Devolución

A continuación se describirá una dimensión del Datamart de Ventas, para un mayor detalle observar el **Anexo A - Dimensiones**.

Dimensión Producto

Descripción

La dimensión Producto es aquella dimensión que posee las características del producto con el que se trabaja en la organización. Son los productos a vender.

Jerarquías

Jerarquía 1

Representa a qué nivel de detalle se desea observar el producto (por línea o por producto)

Nivel	Atributos
Nivel 1	Tipo
Nivel 2	Producto

Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
id_producto	Código numérico del producto. Es un correlativo.	Numérico	Ninguno
codigo	Código que identifica al producto.	Cadena de caracteres	Ninguno
nombre	Nombre del producto.	Cadena de caracteres	Ninguno
tipo	Tipo al que pertenece el producto.	Cadena de caracteres	Ninguno

2.3.2. Datamart de Recursos Humanos

La Figura 2.4Figura 2.3 ilustra el Datamart de Recursos Humanos. La ampliación de esta figura se encuentra en el **Anexo L - Figuras**.

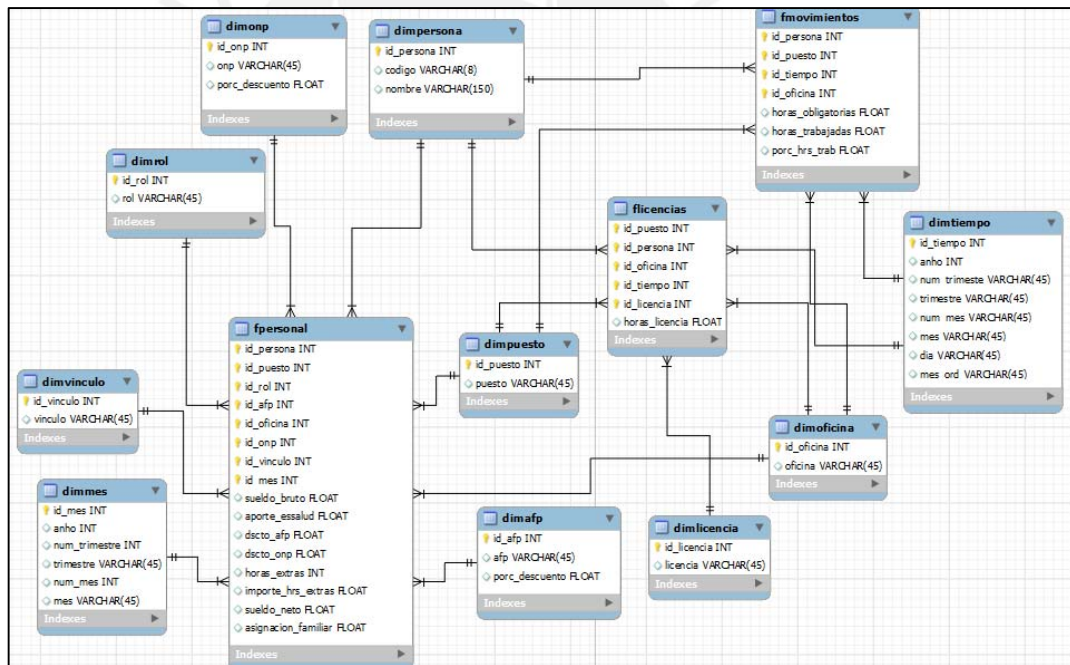


Figura 2.4 - Datamart de Recursos Humanos

El Datamart de recursos humanos posee las dimensiones y facts que permitirán dar soporte a los reportes de recursos humanos que serán planteados en el siguiente capítulo. Estas dimensiones son obtenidas de la base de datos de recursos humanos de la organización.

Dimensiones

Las dimensiones que conforman el Datamart de Recursos Humanos son:

1. Dimensión Mes
2. Dimensión Vínculo
3. Dimensión Rol
4. Dimensión ONP
5. Dimensión AFP
6. Dimensión Puesto
7. Dimensión Persona
8. Dimensión Oficina
9. Dimensión Licencia
10. Dimensión Tiempo

A continuación se describirá una dimensión del Datamart de Recursos Humanos, para un mayor detalle observar el **Anexo A - Dimensiones**.

Dimensión Mes

Descripción

La dimensión Mes es aquella dimensión que posee todos los meses respectivos a partir del año 2006.

Jerarquías

Jerarquía 1

Representa a qué nivel de detalle se desea observar las consultas (mes, trimestre, año).

Nivel	Atributos
Nivel 1	Año
Nivel 2	Trimestre
Nivel 3	Mes

Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
id_mes	Código numérico del mes. Es un correlativo.	Numérico	Ninguno
año	Es el año a que corresponde el mes.	Numérico	Ninguno
num_trimestre	Numero de trimestre al que corresponde el mes. De 1 a 4.	Numérico	Ninguno
trimestre	Nombre del trimestre al que corresponde el mes.	Cadena de caracteres	Ninguno
num_mes	Número del mes. De 1 a 12.	Numérico	Ninguno
mes	Nombre del mes.	Cadena de caracteres	

2.3.3. Facts

Datamart de Ventas

- 1) Fact de Ventas

Datamart de Recursos Humanos

- 1) Fact de Personal
- 2) Fact de Movimientos
- 3) Fact de Licencias

Facts vs. Dimensiones

La Tabla 2.5 muestra la relación entre las facts y las dimensiones a utilizar en los Datamarts de ventas y de recursos humanos.

Dimensiones	Facts			
	Fact de Ventas	Fact de Personal	Fact de Movimientos	Fact de Licencias
Dimensión Producto	✓			
Dimensión Cliente	✓			
Dimensión Tiempo	✓		✓	✓
Dimensión Ubicación	✓			
Dimensión Vendedor	✓			
Dimensión Devolución	✓			
Dimensión Mes		✓		
Dimensión Vínculo		✓		
Dimensión Rol		✓		
Dimensión ONP		✓		
Dimensión AFP		✓		
Dimensión Puesto		✓	✓	✓
Dimensión Persona		✓	✓	✓
Dimensión Oficina		✓	✓	✓
Dimensión Licencia				✓

Tabla 2.5 - Fact V.S. Dimensiones

Fact de Ventas

a) *Descripción*

Esta fact abarca todas las ventas realizadas. La fact permitirá cumplir con la elaboración de los reportes que se verán en el capítulo siguiente.

b) *Granularidad*

N°	Nombre de la Dimensión	Descripción	Llave Primaria
1	dimcliente	Esta dimensión comprende la información sobre el cliente. Mediante esta se podrá obtener los clientes que compraron determinados productos.	Si
2	dimubicación	Esta dimensión comprende toda información acerca del lugar donde se realiza una venta.	Si
3	dimproducto	Esta dimensión comprende toda información acerca de los productos que fueron vendidos.	Si
4	dimtiempo	Esta dimensión comprende toda información acerca de cómo se organiza el tiempo en la empresa para una venta determinada. Esta dimensión puede llegar hasta un nivel de detalle de días.	Si
5	dimvendedor	Esta dimensión comprende la información sobre los vendedores de la empresa. A través de esta dimensión se puede obtener los vendedores que ejecutaron alguna venta.	Si

6	dimdevolucion	Esta dimensión indica si, durante una venta, un producto particular ha tenido devoluciones o no.	Si
---	---------------	--	----

c) *Medidas*

N°	Nombre	Descripción
1	Monto	Mide el monto en soles que se obtuvo de una determinada venta.
2	Cantidad	Mide la cantidad de productos que fueron vendidos en un proceso de venta.
3	Devoluciones	Mide la cantidad de productos que fueron devueltos en un proceso de venta.
4	Monto devuelto	Mide el equivalente en soles de los productos que han sido devueltos.
5	Factor devuelto	Es el porcentaje de productos devueltos sobre el total de productos vendidos para un determinado tipo de producto.

Facts de Personal

a) *Descripción*

Esta fact almacena todo lo referente a la planilla del personal de la empresa y servirá como base para poder elaborar los reportes que se verán en el siguiente capítulo.

b) *Granularidad*

N°	Nombre de la Dimensión	Descripción	Llave Primaria
1	dimmes	Esta dimensión comprende los datos referentes a los meses desde el año 2006.	Si
2	dimvinculo	Esta dimensión indica si es que un personal es contratado o nombrado.	Si
3	dimrol	Esta dimensión comprende información sobre el rol del personal que labora en la empresa.	Si
4	dimONP	Esta dimensión indica si en personal que labora en la empresa está afiliado o no a la ONP.	Si
5	dimAFP	Esta dimensión comprende toda la información de las AFP a las cuales están afiliados los trabajadores de la empresa.	Si
6	dimpuesto	Esta dimensión comprende los puestos que existen en la organización.	Si
7	dimpersona	Esta dimensión comprende todas las características del personal de la empresa.	Si

8	dimoficina	Esta dimensión comprende toda la información de las distintas áreas que hay en la empresa.	Si
---	------------	--	----

c) *Medidas*

N°	Nombre	Descripción
1	sueldo_bruto	Sueldo del empleado sin aplicar los descuentos respectivos.
2	aporte_essalud	Es el aporte que mensualmente hace la empresa.
3	dscto_afp	Son los descuentos a los que está sujeto el sueldo del empleado que está suscrito a una AFP.
4	dscto_onp	Es el descuento que se aplica al sueldo de un empleado que está suscrito a la ONP.
5	horas_extras	Cantidad de horas extras hechas en el mes
6	importe_hrs_extras	El equivalente en soles por las horas extras trabajadas en el mes.
7	dscto_renta	Es el descuento correspondiente al impuesto a la renta.
8	sueldo_netto	Es la diferencia del sueldo en bruto mas los ingresos extras y menos los descuentos aplicados.
9	asignación_familiar	Es el 10% de la remuneración mínima que se otorga mensualmente a cada trabajador

Facts de Movimientos

a) *Descripción*

Esta fact almacena todo lo referente a los movimientos de entrada y salida que realiza el personal de la empresa.

b) *Granularidad*

N°	Nombre de la Dimensión	Descripción	Llave Primaria
1	dimtiempo	Esta dimensión comprende toda información acerca de cómo se organiza el tiempo en la empresa. Esta dimensión puede llegar hasta un nivel de detalle de días.	Si
2	dimpuesto	Esta dimensión comprende toda la información sobre los puestos que existen en la organización.	Si
3	dimpersona	Esta dimensión comprende todas las características del personal de la empresa.	Si
4	dimoficina	Esta dimensión comprende toda la información de las distintas áreas que hay en la empresa.	Si

c) *Medidas*

N°	Nombre	Descripción
1	horas_obligatorias	Son las horas obligatorias que debe cumplir el empleado durante un mes.
2	horas_trabajadas	Son las horas que ha trabajado el empleado durante un mes.
3	%horas_trabajadas	Es el porcentaje de las horas trabajadas sobre las horas obligatorias.

Facts de Licenciasa) *Descripción*

Esta fact almacena todo lo referente a las licencias que se otorgan al personal de la empresa.

b) *Granularidad*

N°	Nombre de la Dimensión	Descripción	Llave Primaria
1	dimtiempo	Esta dimensión comprende toda información acerca de cómo se organiza el tiempo en la empresa. Esta dimensión puede llegar hasta un nivel de detalle de días.	Si
2	dimpuesto	Esta dimensión comprende toda la información sobre los puestos que existen en la organización.	Si
3	dimpersona	Esta dimensión comprende todas las características del personal de la empresa.	Si
4	dimoficina	Esta dimensión comprende toda la información de las distintas áreas que hay en la empresa.	Si
5	dimlicencia	Esta dimensión comprende toda la información sobre las licencias otorgadas al personal de la empresa.	Si

c) *Medidas*

N°	Nombre	Descripción
1	horas_licencia	Es el total de horas de licencias acumuladas durante un mes.

3. CAPÍTULO 3: Diseño

En este capítulo se describe la arquitectura que tiene la solución de inteligencia de negocios; así como los procesos de ETL (extracción, transformación y carga) involucrados.

3.1. Arquitectura

La Figura 3.1 ilustra la arquitectura usada para la elaboración del presente proyecto de fin de carrera.

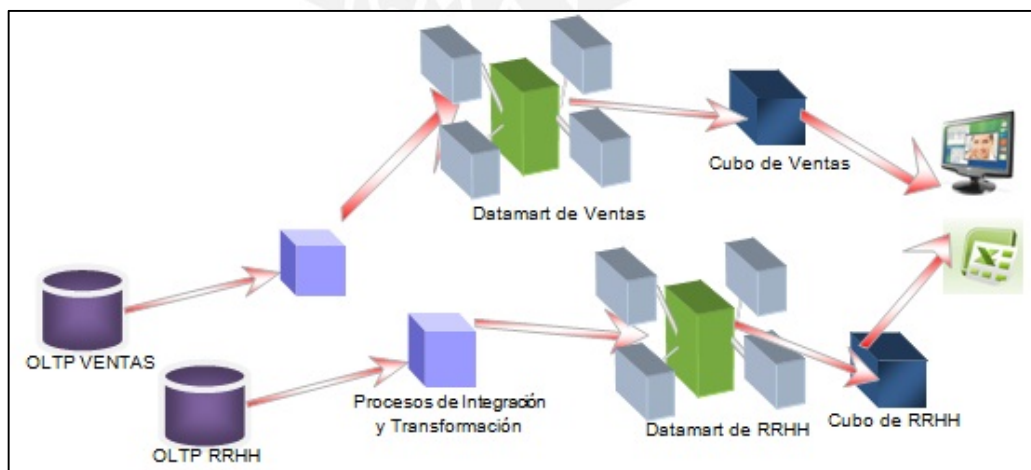


Figura 3.1 - Arquitectura de la Solución

En el gráfico se observa que la extracción de los datos se realiza desde las bases de datos OLTP propias de la empresa y estos serán transformados y cargados a los Datamarts respectivos desde donde se podrá realizar el procesamiento de los cubos OLAP y con ello la explotación de la información. La Tabla 3.1 muestra algunas características de las bases de datos de extracción.

Fuente	Tipo	Conexión
Base de datos de ventas.	MYSQL	ODBC MYSQL
Base de datos de recursos humanos.	MYSQL	ODBC MYSQL

Tabla 3.1 - Bases de Datos de Extracción

3.2. Proceso de extracción y transformación

En esta sección se muestra el proceso de extracción y transformación de los datos para el Datamart de ventas y el Datamart de recursos humanos. Para tener un mayor detalle observar el **Anexo B - Diseño de Extracción**.

DATAMART DE VENTAS

Carga de la dimensión Producto

a) Descripción

Es el proceso que consiste en realizar la carga de la dimensión “dimproducto” desde las tablas t0002_producto y t0003_linea que se encuentran en la base de datos de ventas.

b) Descripción de las tablas fuentes

Tipo de Fuente	Nombre de Tabla	Descripción
Base de Datos de la Empresa (MySQL)	t0002_producto	La tabla contiene todas las características de los productos con los que trabaja la organización.
Base de Datos de la Empresa (MySQL)	t0003_linea	La tabla contiene la clasificación de los diversos productos que vende la empresa.

c) Estandarización de Datos y Limpieza de Datos

Nombre	Llave	Tipo	Formato	Limpieza	Valor por Defecto
id_producto	PK	Integer	Numérico	No debe ser nulo	NO TIENE
codigo		Varchar(10)	Texto		NO TIENE
nombre		Varchar(100)	Texto		NO TIENE
tipo		Varchar(100)	Texto		NO TIENE

d) Fuentes de Datos

Tabla: t0002_producto				
Nombre	Llave	Tipo	Formato	Consideración Importante
id_producto	PK	Integer	Numérico	NO TIENE
codigo_producto		Varchar(10)	Texto	NO TIENE
nombre_producto		Varchar(100)	Texto	Espacios en Blanco

Tabla: t0003_linea				
Nombre	Llave	Tipo	Formato	Consideración Importante
descripcion_linea		Varchar(100)	Texto	NO TIENE

e) Tabla destino

Tabla: dimproducto		
Campo	Tipo	Mapeo
id_producto	Integer	t0002_producto.id_producto
codigo	Varchar(10)	t0002_producto.codigo_producto
nombre	Varchar(100)	t0002_producto.nombre_producto
tipo	Varchar(100)	t0003_linea.descripcion_linea

f) Proceso

i. Carga de registros en un archivo intermedio

Se extrajo los datos de la tabla t0002_producto y t0003_linea de acuerdo al mapeo y se cargó en un archivo de texto intermedio "Producto".

ii. Carga de la Dimensión

Se extrajo los valores del archivo “Producto” y se cargó en la dimensión “dimproducto”. En caso los registros hagan referencia a nuevos productos se insertan las nuevas filas en la tabla; caso contrario, si los registros hacen referencia a antiguos productos actualizar el código, nombre y tipo del producto.

iii. Borrar el archivo intermedio

Se borró el archivo intermedio “Producto”.

DATA MART DE RECURSOS HUMANOS

Carga de la dimensión Vínculo

a) *Descripción*

Es el proceso que consiste en realizar la carga de la dimensión “dimvinculo” desde la tabla t009_vinculo que se encuentra en la base de datos de recursos humanos.

b) *Descripción de las tablas fuentes*

Tipo de Fuente	Nombre de Tabla	Descripción
Base de Datos de la Empresa (MySQL)	t009_vinculo	La tabla muestra el vínculo que tiene un trabajador con la empresa: Contratado o Nombrado.

c) *Estandarización de Datos y Limpieza de Datos*

Nombre	Llave	Tipo	Formato	Limpieza	Valor por Defecto
id_vinculo	PK	Integer	Numérico	No debe ser nulo	NO TIENE
vinculo		Varchar(45)	Texto		NO TIENE

d) *Fuentes de Datos*

Tabla: t009_vinculo				
Nombre	Llave	Tipo	Formato	Consideración Importante
id_vinculo	PK	Integer	Numérico	NO TIENE

Tabla: t009_vinculo				
Nombre	Llave	Tipo	Formato	Consideración Importante
vinculo		Varchar(45)	Texto	NO TIENE

e) *Tabla destino*

Tabla: dimvinculo			
Campo	Tipo	Mapeo	
id_vinculo	Integer	t009_vinculo.id_vinculo	
vinculo	Varchar(45)	t009_vinculo vinculo	

f) *Proceso*

i. Carga de registros en un archivo intermedio

Se extrajo los datos de la tabla t009_vinculo de acuerdo al mapeo y se cargó en un archivo de texto intermedio "Vinculo".

ii. Carga de la Dimensión

Se extrajo los valores del archivo "Vinculo" y se cargó en la dimensión "dimvinculo".

iii. Borrar el archivo intermedio

Se borró el archivo intermedio "Vinculo".

3.3. Diseño de explotación

La explotación se llevara a cabo a través de los reportes los cuales serán elaborados en coordinación con el usuario (empleado de la empresa). Para un mayor detalle observar el **Anexo C - Diseño de Explotación**.

3.3.1. Configuración del Reporte

Se detalla la cabecera y pie de página de los reportes, así como los márgenes y los prototipos de estos. Para un mayor detalle observar el **Anexo C - Diseño de Explotación**.

Cuerpo del reporte

A continuación se presenta los prototipos de reportes usados en los informes desarrollados en el presente trabajo de fin de carrera.

Reporte de tipo crosstab:

La Figura 3.2 muestra el prototipo de reporte crosstab.

Filtros				
<Nombre del reporte>			<Logo>	
	Columna 1	Columna 2	Columna 3	Columna 4
Fila 1	0.00	0.00	0.00	0.00
Fila 2	0.00	0.00	0.00	0.00
Fila 3	0.00	0.00	0.00	0.00
Fila 4	0.00	0.00	0.00	0.00
Fila 5	0.00	0.00	0.00	0.00
Fila 6	0.00	0.00	0.00	0.00
Total	0.00	0.00	0.00	0.00

Figura 3.2 - Prototipo de Reporte Crosstab

Reporte de tipo gráfico:

Se debe considerar que hay diferentes tipos reportes de gráficos. La Figura 3.3 ilustra el prototipo de reporte gráfico.

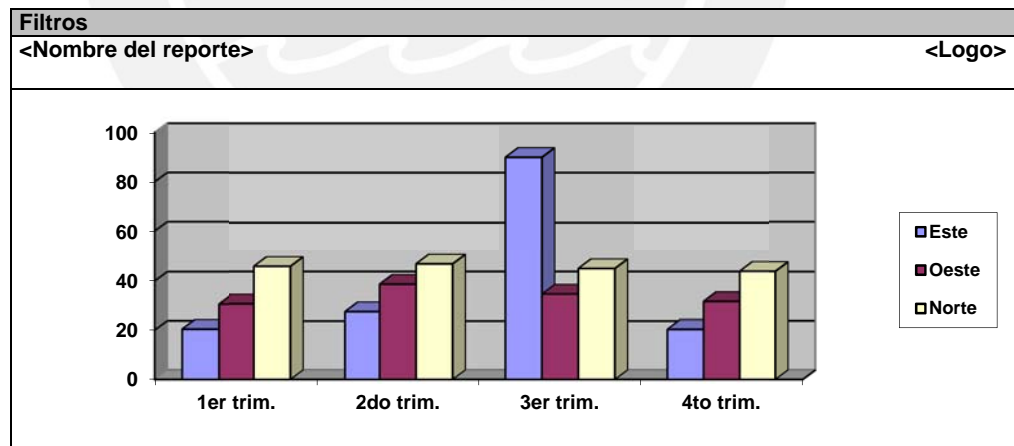


Figura 3.3 - Prototipo de Reporte Gráfico

Reporte de híbrido:

El reporte de tipo híbrido es aquel formado por un gráfico y un reporte crosstab. Tener en cuenta que no necesariamente el gráfico va arriba del reporte crosstab, el orden puede variar. La Figura 3.4 ilustra el prototipo de reporte híbrido.

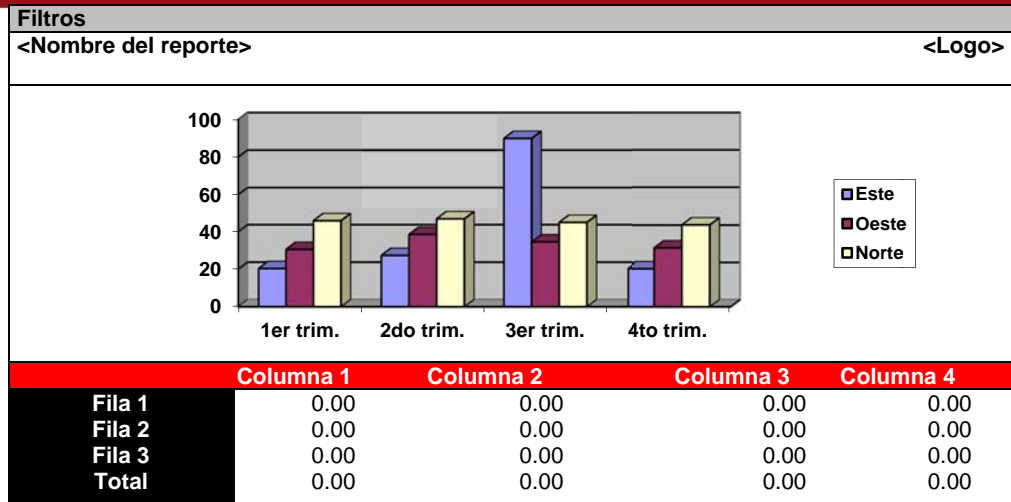


Figura 3.4 - Prototipo de Reporte Híbrido

3.3.2. Vistas y reportes

La Tabla 3.2 resume los reportes de cada Datamart.

Tema	Reporte
Análisis de Ventas	Reporte de Indicadores de Ventas para cada Línea Producto por Clientes por Producto - Tiempo - Ubicación.
	Reporte de Ventas por Producto por Tiempo - Tiempo - Año - Ubicación.
	Reporte para cada Zona del Total de Ventas de Vendedores por Tiempo - Año - Mes.
	Reporte de Indicadores para cada Vendedor.
	Gráfico de Ventas por Línea Producto por Mes - Año - Zona.
	Gráfico de Ventas por Línea Producto por Zona - Año - Mes.
	Gráfico de Porcentaje de Ventas por Tipo de Producto por Zona - Año - Mes.
	Reporte de Ventas por Zona por Tiempo - Año - Mes - Ubicación.
	Reporte de Indicadores de Devoluciones.
Análisis de Recursos Humanos	Reporte de Control de Horarios del Personal - Tiempo - Oficina.
	Gráfico de Historial del Porcentaje de Horas Trabajadas por Oficina por Mes.
	Reporte de Horas de Licencia del Personal por Tipo de Licencia - Oficina - Tiempo.
	Reporte de Indicadores de Horas Extras por Persona por Tiempo - Oficina.
	Reporte de Sueldos del Personal por Rol por Tiempo - Oficina.
	Reportes de Sueldos por Oficina por Vínculo Laboral y Tiempo.
	Reporte de Pagos y Descuentos por Rol por Persona por Tiempo - Vínculo

Tabla 3.2 - Reportes V.S Datamart

ANÁLISIS DE VENTAS

Reporte de Indicadores de Ventas para cada Línea Producto por Clientes por Producto - Tiempo - Ubicación

El informe muestra para cada tipo de productos la cantidad de ventas, monto de ventas, cantidad de devoluciones, monto de equivalente a las devoluciones, y el factor devuelto de las ventas realizadas por zona, cliente y producto.

El informe permite evaluar lo siguiente:

- i. La cantidad de productos vendidos a cada cliente.
- ii. El total del monto de las ventas realizadas a cada cliente.
- iii. La cantidad de devoluciones y el monto devuelto de cada cliente por producto.
- iv. Que tan rentable ha sido las ventas respecto a los productos devueltos.
- v. Tomar las mejores decisiones para reducir el índice de productos devueltos

a) Diseño

<Tiempo>	<Ubicación>	Producto				
Reporte de Indicadores de Ventas para cada Línea Producto por Clientes por Producto - Tiempo - Ubicación						
		Cantidad	Monto	Devoluciones	Monto devuelto	Factor devuelto
Ubicación 1	Cliente 1	###	0.00	###	0.00	0.00 %
Ubicación 2	Cliente 2	###	0.00	###	0.00	0.00 %
Ubicación 3	Cliente 3	###	0.00	###	0.00	0.00 %
Ubicación 4	Cliente 4	###	0.00	###	0.00	0.00 %
Ubicación 5	Cliente 5	###	0.00	###	0.00	0.00 %
Total		###	0.00	###	0.00	0.00 %

Tipo: crosstab

b) Filas

No.	Dimensión	Nivel / Categoría
1	Ubicación	Zona
2	Ubicación	Departamento
3	Cliente	Nombre

c) Columnas

No.	Dimensión	Nivel / Categoría
1	Producto	Nombre

d) *Medida*

No.	Medida	Formato
1	Cantidad	Numérico
2	Monto	Soles
3	Devoluciones	Numérico
4	Monto devuelto	Soles
5	Factor devuelto	Numérico

e) *Filtro*

No.	Operación
1	Tiempo = "Año", "Mes"
2	Ubicacion = "Zona", "Departamento", "Provincia", "Ciudad"

ANÁLISIS DE RECURSOS HUMANOS

Reporte de Control de Horarios del Personal - Tiempo - Oficina

El informe muestra las horas obligatorias, las horas trabajadas y el porcentaje de cumplimiento de trabajo en el mes por cada empleado administrativo. El informe contará con semáforos que indiquen si el personal cumplió o no con el porcentaje de horas obligatorias por mes.

El informe permite evaluar lo siguiente:

- i. Las horas obligatorias que tiene cada trabajador en el mes.
- ii. Las horas que realmente trabaja un empleado de la empresa.

a) *Diseño*

<Oficina> <Tiempo>			
Reporte de Control de Horarios del Personal por Mes - Oficina – Puesto			
	Horas Obligatorias	Horas trabajadas	% Horas Trabajadas
Empleado 1 Puesto 1	###	###	0.00 %
Empleado 2 Puesto 2	###	###	0.00 %
Empleado 3 Puesto 3	###	###	0.00 %
Empleado 4 Puesto 4	###	###	0.00 %

Tipo: crosstab

b) *Filas*

No.	Dimensión	Nivel / Categoría
1	Persona	Nombre
1	Puesto	Puesto

c) *Columnas*

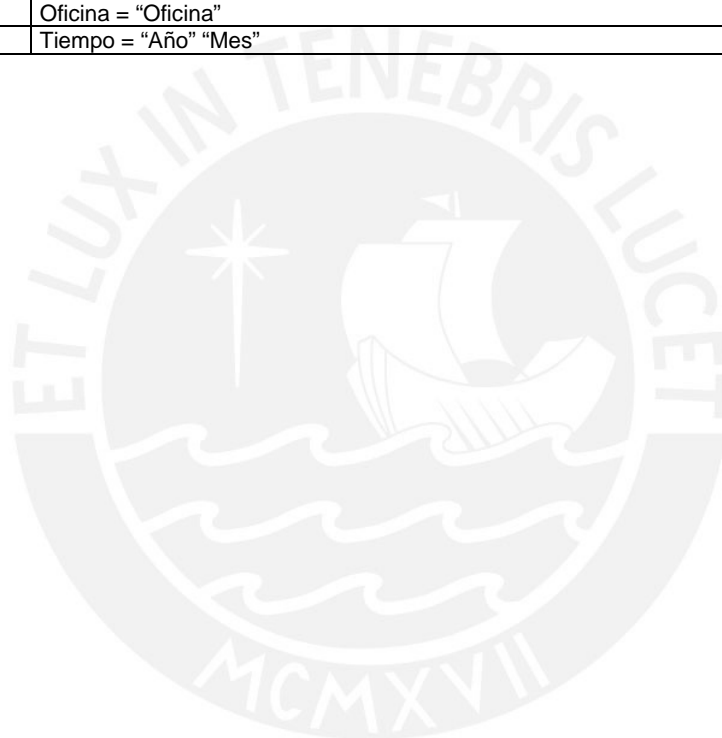
No.	Dimensión	Nivel / Categoría
-	-	-

d) *Medida*

No.	Medida	Formato
1	Horas obligatorias	Numérico
2	Horas trabajadas	Numérico
3	% Horas trabajadas	Porcentaje

e) *Filtro*

No.	Operación
1	Oficina = "Oficina"
2	Tiempo = "Año" "Mes"



4. CAPÍTULO 4: Construcción y Pruebas

Este capítulo hace hincapié en cómo se ha de instalar las herramientas de Pentaho BI (Kettle) y SQLServer 2008 para la elaboración del ETL y cubos respectivamente. Asimismo junto con la suite de Pentaho BI será necesaria la instalación de MySQL y así poder tener integrada la base de datos de prueba. Por otro lado, se realizarán las pruebas de verificación de data de manera que los datos de entrada en los procesos ETL sean reflejados en los reportes finales.

4.1. Puesta en Producción

Esta sección resume todo lo referente a la instalación y configuración de las herramientas de Pentaho y de SQL Server 2008, necesarias para la elaboración de los procesos ETL y cubos OLAP respectivamente.

4.1.1. Instalación del Pentaho

La descarga del software Pentaho BI se hizo desde la siguiente página:
<http://kettle.pentaho.com/>. [KET 2012]

Se descargará solo el paquete Data Integration de Pentaho (Kettle) versión 4.2.1 necesario para desarrollar los procesos ETL del presente proyecto de fin de carrera. La Figura 4.1 ilustra la página de Pentaho desde donde se realizó la descarga.



Figura 4.1 - Pagina de Pentaho

Para un mayor detalle de la instalación observar el **Anexo D - Instalación Pentaho**.

4.1.2. Configuración de la Herramienta ETL

Para iniciar la instalación solo es necesario hacer clic en el ícono denominado Spoon del directorio “C:\pentaho\pdi-ce-4.2.1-stable\data-Integration”. La Figura 4.2 muestra la pantalla de bienvenida una vez se haya hecho clic en el ícono Spoon.



Figura 4.2 - Pantalla de bienvenida de Kettle de Pentaho

Los detalles acerca de la configuración de la herramienta ETL se encuentran en el **Anexo E - Configuración de la Herramienta ETL.**

4.1.3. Instalación de SQLServer 2008 R2

La descarga de SQLServer 2008 R2 se hizo desde la siguiente página web:
<http://www.microsoft.com/downloads/details.aspx?familyid=265f08bc-1874-4c81-83d8-0d48dbce6297&displaylang=es>. [SQL 2011]

Para un mayor detalle de la instalación de SQLServer 2008 R2 observar el **Anexo F - Instalación SQLServer 2008.**

4.2. Plan de Pruebas

La siguiente sección resume los casos de prueba y el desarrollo de estos de manera que se pueda comprobar que los datos ingresados se vean reflejados en los informes elaborados.

4.2.1. Objetivos

La finalidad del plan de pruebas es verificar que los resultados finales sean coherentes con los datos de entrada.

4.2.2. Características a ser probadas

Las características a probar son las siguientes:

- Se verificará que las extracciones sean realizadas correctamente.
- Se verificará que las transformaciones sean realizadas de manera correcta.
- Se verificará que la carga de data sea realizada de manera correcta.
- Se verificará que los reportes muestren la data actualizada.
- Se verificará que los datos de los reportes sean coherentes con los filtros aplicados; en caso existan.

4.2.3. Casos de pruebas

En esta sección se detallarán los casos de prueba para cada uno de los reportes del Datamart de ventas y del Datamart de recursos humanos. Para un mayor detalle observar el **Anexo G - Casos de Prueba**.

Datamart de Ventas

Caso de prueba para el informe “Indicadores de Ventas para cada Línea Producto por Clientes por Producto - Tiempo - Ubicación”

Esta prueba tiene como objetivo verificar que las tablas necesarias para mostrar el informe en mención sean correctamente cargadas.

Tablas OLTP / Archivos	Tablas Datamart	Resultados Finales Informe
t0001_cliente	dimcliente	El monto de ventas y la cantidad de productos vendidos para cada Línea por Clientes de acuerdo a los filtros de Tiempo y Ubicación.
t0002_producto	dimproducto	
t0003_linea	dimubicacion	
t0004_ciudad	dimtiempo	
t0005_provincia	factventas	
t0006_dpto		
t0007_zona		
devolucionestotal.txt		

Datamart de Recursos Humanos

Caso de prueba para el informe “Reporte de Control de Horarios del Personal - Tiempo - Oficina”

Esta prueba tiene como objetivo verificar que las tablas necesarias para mostrar el informe en mención sean correctamente cargadas.

Tablas OLTP	Tablas Datamart	Resultados Finales Informe
t002_oficina	dimoficina	El número de horas obligatorias, número de horas trabajadas y el porcentaje de horas
t004_persona	dimpersona	

Tablas OLTP	Tablas Datamart	Resultados Finales Informe
t003_puesto	dimpuesto	trabajadas de los empleados por oficina y tiempo.
t006_personal_mes	dintempo	
t007_mes	fmovimientos	
t008_movimiento		

4.2.4. Desarrollo de las pruebas

En esta sección se detalla cómo se realizaron las diversas pruebas y además se presenta los resultados de los reportes. Debido a la gran cantidad de registros de algunas tablas solo se muestran algunos de ellos. Además solo se mostrarán los campos más relevantes de cada tabla involucrada en el caso de prueba. Para un mayor detalle observar el **Anexo H - Desarrollo de las Pruebas**.

Datamart de Ventas

Caso de prueba para el informe “Indicadores de Ventas para cada Línea Producto por Clientes por Producto - Tiempo - Ubicación”

Datos de entrada:

- Tabla t0001_cliente: Solo se muestra la tabla con las columnas más relevantes para cargar el informe; sin embargo se podría obtener una mayor información desde los otros campos.

id_cliente	id_ciudad	codigo_cliente	nombre_cliente
1	1	cli001	METRO SAN ISIDRO
2	2	cli002	METRO JR. DE LA UNIÓN

- Tabla t0002_producto: Solo se muestra la tabla con las columnas más relevantes para cargar el informe; sin embargo se podría obtener una mayor información desde los otros campos.

id_prod	id_linea	cod_prod	nombre_producto
1	1	pis001	Pisco mosto verde acholado
2	2	mol001	Ajos molidos
3	2	mol002	Culantro molido

- Tabla t0003_linea: Solo se muestra la tabla con las columnas más relevantes para cargar el informe.

id_linea	codigo_linea	descripcion_linea
1	lin001	línea de piscos
2	lin002	línea de molidos

- Tabla t0004_ciudad: Solo se muestra la tabla con una parte de los registros que cargarán el informe.

id_ciudad	id_provincia	descripcion_ciudad
1	1	SAN ISIDRO
2	1	CENTRO DE LIMA

- Tabla t0005_provincia: Solo se muestra la tabla con una parte de los registros que cargarán el informe.

id_provincia	id_dpto	descripcion_provincia
1	1	LIMA

- Tabla t0006_dpto: Solo se muestra la tabla con una parte de los registros que cargarán el informe.

id_dpto	id_zona	descripcion_dpto
5	2	LIMA

- Tabla t0007_zona: Solo se muestra la tabla con una parte de los registros que cargarán el informe.

id_zona	descripcion_zona
2	CENTRO

- Archivo devolucionestotal.xls: Solo se muestra el archivo con una parte de los registros que cargarán el informe.

cliente	producto	fecha	cant	devoluciones
cli001	pis001	2006-01-08	18	5
cli002	pis001	2006-01-21	19	5

Resultados Intermedios:

- Dimensión dimcliente: Solo se muestran los campos de la dimensión que pueden ser cargados con los registros de prueba extraídos de la tabla t0001_cliente. Algunos campos no son tomados en cuenta.

id_cliente	codigo	nombre
1	cli001	METRO SAN ISIDRO
2	cli002	METRO JR. DE LA UNIÓN

- Dimensión dimproducto: Solo se muestran los campos de la dimensión que pueden ser cargados con los registros de prueba extraídos de las tablas t0002_producto y t0003_linea.

id_producto	codigo	nombre	tipo
1	pis001	Pisco mosto verde acholado	línea de piscos
2	mol001	Ajos molidos	línea de molidos
3	mol002	Culantro molido	línea de molidos

- Dimensión dimubicacion: Solo se muestran los campos de la dimensión que pueden ser cargados con los registros de prueba extraídos de las tablas t0004_ciudad, t0005_provincia, t0006_dpto y t0007_zona.

id_ubicacion	ciudad	provincia	departamento	zona
1	SAN ISIDRO	LIMA	LIMA	CENTRO
2	CENTRO DE LIMA	LIMA	LIMA	CENTRO

- Dimensión dimtiempo: Solo se muestran algunos registros y campos de la dimensión tiempo necesarios para cargar el informe.

id_tiempo	año	num_mes	mes
20060108	2006	1	Enero
20060102	2006	1	Enero
20060121	2006	1	Enero

- Fact factventas: Solo se muestran algunos registros y campos de la fact de ventas tomados para realizar la prueba.

id_ubicacion	id_producto	id_tiempo	id_cliente	monto	cantidad	devoluciones	montodevuelto	factordevuelto
1	1	20060108	1	4320	18	5	1200	27
1	2	20060108	1	384	1	0	0	0
1	3	20060108	1	768	2	0	0	0
2	1	20060121	2	460	19	5	1200	26
2	2	20060121	2	768	2	0	0	0

Ahora verificaremos si los productos analizados han tenido devoluciones para los clientes y periodos de tiempo analizados. Para esto bastará verificar en la hoja de cálculo excel desde donde se extrae la información de devolución de productos.

Luego de realizar los filtros necesarios para los clientes y productos estudiados se verifica que para Enero del 2006 el único de los productos ha tenido devoluciones es el pisco.

	A	B	C	D	E
1	cliente	producto	fecha	cant	devoluciones
3	cli001	pis001	2006-01-08	18	5
17	cli002	pis001	2006-01-21	19	5

De lo anterior se puede constatar que el pisco ha tenido 5 devoluciones para el cliente “cli001” (Metro San Isidro) y 5 devoluciones para el cliente “cli002” (Metro Jr. De la Unión) sobre un total de ventas de 18 y 19 piscos respectivamente. Si cada caja de pisco cuesta 240 soles entonces el monto por la devolución de 5 cajas es 1200 soles. Por otro lado el porcentaje de devoluciones se calcula sobre el total de productos vendidos; para Metro de San Isidro es $(5/18*100)$ y para el Metro de Jr. De la Unión es $(5/19*100)$.

Resultado Final:

A continuación se muestra el reporte final que muestra la cantidad y monto de ventas para los valores probados. La ampliación de esta figura se encuentra en el **Anexo L - Figuras.**

ventas de fila	línea de molidos					línea de pisos				
	Cantidad	Monto	Devoluciones	Montodevuelto	Factordevuelto	Cantidad	Monto	Devoluciones	Montodevuelto	Factordevuelto
METRO LIMA JR DE LA UNION	2	S/ 768,00	0	S/ 0,00	0,00%	1	S/ 384,00	0	0,00%	0,00%
METRO SAN ISIDRO	1	S/ 384,00	0	S/ 0,00	0,00%	2	S/ 768,00	0	0,00%	0,00%

Este resultado es verificado con un query realizado a la base de datos relacional. Se observa que tanto en el informe y en el query los montos de ventas son iguales para los clientes y productos evaluados para enero 2006.

```

select c.nombre_cliente, p.nombre_producto, date_format(fecha, 'MM-YY') as fecha,
       sum(cantidad_venta) as cantidad, sum(subtotal_venta) as monto
from nutreinventas.t0009_ventas v, nutreinventas.t0010_devventas d, nutreinventas.t0001_cliente c,
     nutreinventas.t0002_producto p
where
  v.id_venta=d.id_venta and
  v.id_cliente=c.id_cliente and
  d.id_producto=p.id_producto and
  v.id_cliente in (1,2) and
  d.id_producto in (1,2,3) and
  date_format(fecha, 'MM-YY') = 'January-2006'
group by c.nombre_cliente, p.nombre_producto, date_format(fecha, 'MM-YY');
    
```

nombre_cliente	nombre_producto	fecha	cantidad	monto
METRO LIMA JR DE LA UNION	AJOS MOLIDOS 212 ML TPV 063MM HOJA REDONDA	January-2006	2	768
METRO LIMA JR DE LA UNION	CULANTRO MOLIDO 212 ML TPV 063MM H.R.	January-2006	1	384
METRO LIMA JR DE LA UNION	pisco mosto verde acholado	January-2006	19	4560
METRO SAN ISIDRO	AJOS MOLIDOS 212 ML TPV 063MM HOJA REDONDA	January-2006	1	384
METRO SAN ISIDRO	CULANTRO MOLIDO 212 ML TPV 063MM H.R.	January-2006	2	768
METRO SAN ISIDRO	pisco mosto verde acholado	January-2006	18	4320

Ahora se presenta el proceso de Extracción, Transformación y Carga para el Datamart de Ventas mediante la herramienta de Pentaho Kettle, para tener un mayor detalle de la funcionalidad de la herramienta ver el **Anexo K - Funcionalidad de Kettle Pentaho**. La Figura 4.3 ilustra el ETL principal que alimenta el Datamart de Ventas.

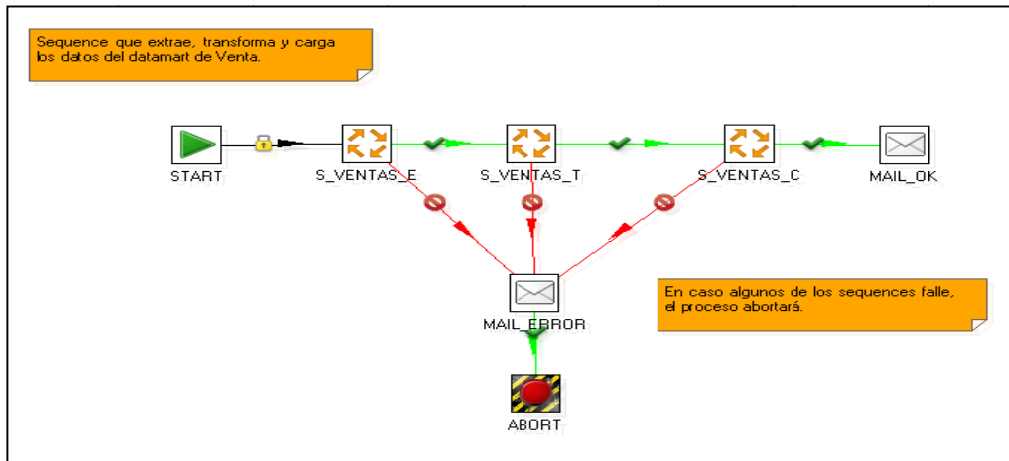


Figura 4.3 - ETL de Ventas

Este proceso muestra el job principal que extrae, transforma y carga la data a las tablas del Datamart de Ventas. En caso el proceso este correcto se envía un mail de satisfacción; de lo contrario se envía un mail indicando que el proceso ha fallado y aborta. Este es un proceso total; es decir todos los datos son extraídos, transformados y cargados en cada ejecución.

Entrada: Se ingresaron parámetros de entrada que permitirán conectarse a las tablas de la base de datos “nutreinaventas” desde donde se desea obtener los datos (MySQL); así como conectarse a la base de datos dimensional cuyas tablas se desean cargar (SQLServer).

Salida: Las dimensiones y la fact de ventas son cargadas correctamente.

Ejecución y Prueba: La Figura 4.4 y la Figura 4.5 ilustran los log del proceso ETL de Ventas.

```

2012/05/07 00:30:53 - F_VENTAS.0 - Running on slave server #0/1.
2012/05/07 00:30:53 - conexiónFact - New database connection defined
2012/05/07 00:30:53 - conexiónFact - Connected to database.
2012/05/07 00:30:53 - F_VENTAS.0 - Connected to database [conexiónFact] (commit=5000)
2012/05/07 00:30:53 - conexiónFact - Auto commit off
2012/05/07 00:30:53 - FVENTAS_C - Se ha podido inicializar el paso [FACT_VENTAS.0] correctamente.
2012/05/07 00:30:53 - FVENTAS_C - Se ha podido inicializar el paso [F_VENTAS.0] correctaments.
2012/05/07 00:30:53 - FVENTAS_C - Transformación ha preparado 2 hilos y 1 conjuntos de files.
2012/05/07 00:30:53 - FACT_VENTAS.0 - Iniciando ejecución...
2012/05/07 00:30:53 - FACT_VENTAS.0 - Opening file: C:\ROLANDO\TESIS\ETL\TRANSFORMACION_VENTAS\ventas.txt
2012/05/07 00:30:53 - F_VENTAS.0 - Iniciando ejecución...
2012/05/07 00:30:53 - F_VENTAS.0 - Prepared statement : INSERT INTO factventas (id_ubicacion, id_producto, id_cliente, id_vendedor, monto, cantidad) VALUES ( ?, ?, ?, ?, ?, ?
2012/05/07 00:31:56 - FACT_VENTAS.0 - Procesamiento finalizado (I=36149, O=0, R=0, W=36148, U=1, E=0
2012/05/07 00:32:13 - conexiónFact - Connection to database closed!
2012/05/07 00:32:13 - F_VENTAS.0 - Procesamiento finalizado (I=0, O=38148, R=38148, W=38148, U=0, E=0
2012/05/07 00:32:14 - S_VENTAS_C - Starting entry [MAIL_OK]
2012/05/07 00:32:14 - S_VENTAS_C - exec(10, 0, MAIL_OK.0)
2012/05/07 00:32:14 - MAIL_OK - Starting job entry
2012/05/07 00:32:16 - S_VENTAS_C - Finished job entry [MAIL_OK] (result={true})
2012/05/07 00:32:16 - S_VENTAS_C - Finished job entry [VENDEDOR_C] (result={true})
2012/05/07 00:32:16 - S_VENTAS_C - Finished job entry [PRODUCTO_C] (result={true})
2012/05/07 00:32:16 - S_VENTAS_C - Finished job entry [UBICACION_C] (result={true})
2012/05/07 00:32:16 - S_VENTAS_C - Finished job entry [CLIENTE_C] (result={true})
2012/05/07 00:32:16 - S_VENTAS_ETC - Starting entry [MAIL_OK]
2012/05/07 00:32:16 - S_VENTAS_ETC - exec(4, 0, MAIL_OK.0)
2012/05/07 00:32:16 - MAIL_OK - Starting job entry
2012/05/07 00:32:19 - S_VENTAS_ETC - Finished job entry [MAIL_OK] (result={true})
2012/05/07 00:32:19 - S_VENTAS_ETC - Finished job entry [S_VENTAS_C] (result={true})
2012/05/07 00:32:19 - S_VENTAS_ETC - Finished job entry [S_VENTAS_T] (result={true})
2012/05/07 00:32:19 - S_VENTAS_ETC - Finished job entry [S_VENTAS_E] (result={true})
2012/05/07 00:32:19 - S_VENTAS_ETC - Job execution finished
2012/05/07 00:32:19 - Spoon - Trabajo ha terminado.
    
```

Figura 4.4 - Log del Datamart de Ventas I

Trabajo / Entrada de Trabajo	Comentario	Resultado	Razón	Nombre Fichero
S_VENTAS_ETC				
Trabajo: S_VENTAS_ETC	Start of job execution	start		
START	Start of job execution	start		
START	Job execution finished	Exito		
S_VENTAS_E	Start of job execution		Followed unconditional link	file:///C:/ROLANDO/TESIS/ETL/EXTR...
Trabajo: S_VENTAS_E				
S_VENTAS_E	Job execution finished	Exito		file:///C:/ROLANDO/TESIS/ETL/EXTR...
S_VENTAS_T	Start of job execution		Followed link after success	file:///C:/ROLANDO/TESIS/ETL/TRAN...
Trabajo: S_VENTAS_T				
S_VENTAS_T	Job execution finished	Exito		file:///C:/ROLANDO/TESIS/ETL/TRAN...
S_VENTAS_C	Start of job execution		Followed link after success	file:///C:/ROLANDO/TESIS/ETL/CARG...
Trabajo: S_VENTAS_C				
S_VENTAS_C	Job execution finished	Exito		file:///C:/ROLANDO/TESIS/ETL/CARG...
MAIL_OK	Start of job execution		Followed link after success	
MAIL_OK	Job execution finished	Exito		
Trabajo: S_VENTAS_ETC	Job execution finished	Exito	finished	

Figura 4.5 - Log del Datamart de Ventas II

Notar que en la primera imagen el log muestra el mensaje “el proceso terminó con éxito”. En la segunda imagen se muestra a detalle que cada job que compone el proceso principal termina con éxito.

Para un mayor detalle de la ejecución y pruebas del ETL para el Datamart de ventas observar el **Anexo I - Ejecución y Pruebas del ETL**.

Datamart de Recursos Humanos

Caso de prueba para el informe “Reporte de Control de Horarios del Personal - Tiempo - Oficina”

Datos de entrada:

- Tabla t002_oficina: Solo se muestra la tabla con una parte de los registros que cargarán el informe.

id_oficina	oficina
2	ADMINISTRACIÓN Y RECURSOS HUMANOS

- Tabla t004_persona: Solo se muestra la tabla con las columnas más relevantes para cargar el informe; sin embargo se podría obtener una mayor información desde los otros campos.

id_persona	nombre
127	Carmen Hutz Yanac

- Tabla t003_puesto: Solo se muestra la tabla con una parte de los registros que cargarán el informe.

id_puesto	puesto
16	SECRETARIA

- Tabla t006_personal_mes: Solo se muestra la tabla con las columnas más relevantes para cargar el informe; sin embargo se podría obtener una mayor información desde los otros campos.

id_persona	id_mes	id_puesto	id_oficina
127	1	16	2

- Tabla t007_mes: Solo se muestra la tabla con las columnas más relevantes para cargar el informe; sin embargo se podría obtener una mayor información desde los otros campos.

id_mes	mes	año	num_mes
1	ENERO	2006	1

- Tabla t008_movimiento: Solo se muestra la tabla con algunos registros de enero 2006.

fecha	id_persona	id_mes	enthor	entmin	salhor	salmin
2006-01-02	127	1	7	56	5	1
2006-01-03	127	1	7	59	5	30
2006-01-04	127	1	7	40	5	0

Resultados Intermedios:

- Dimensión dimoficina: Solo se muestran los campos de la dimensión que pueden ser cargados con los registros de prueba extraídos de la tabla t002_oficina.

id_oficina	oficina
2	ADMINISTRACIÓN Y RECURSOS HUMANOS

- Dimensión dimpersona: Solo se muestran los campos de la dimensión que pueden ser cargados con los registros de prueba extraídos de la tabla t004_persona. Algunos campos no son tomados en cuenta.

id_persona	nombre
127	Carmen Hutz Yanac

- Dimensión dimpuesto: Solo se muestran los campos de la dimensión que pueden ser cargados con los registros de prueba extraídos de la tabla t003_puesto.

id_puesto	puesto
16	SECRETARIA

- Dimensión dimtiempo: Solo se muestran algunos registros y campos de la dimensión tiempo necesarios para cargar el informe.

id_tiempo	año	num_mes	mes
20060102	2006	1	Enero
20060103	2006	1	Enero
20060104	2006	1	Enero

- Fact fmovimientos: Solo se muestran algunos registros y campos de la fact de movimientos tomados para realizar la prueba.

id_persona	id_puesto	id_tiempo	id_oficina	horas_oblig	horas_trab	porc_hrs_trab
127	16	20060102	1	8	8,1	101 %
127	16	20060103	1	8	8,5	106.5 %
127	16	20060104	1	8	9,3	116.7 %

Resultado Final:

A continuación se muestra el reporte final que muestra las horas obligatorias, horas trabajadas y porcentaje de horas trabajadas para el personal evaluado.

Jerarquía de Año	2006		
Jerarquía de Mes	01. ENERO		
Jerarquía Area	ADMINISTRACION Y RECURSOS HUMANOS		
Valores			
Rótulos de fila	Horas Obligatorias	Horas Trabajadas	Porc
Carmen Huzt Yanac	176	183,8	104,4318182
Total general	176	183,8	104,4318182

Este resultado es verificado con un query realizado a la base de datos relacional. Se observa que tanto en el informe y en el query la persona evaluada tiene la misma cantidad de horas obligatorias, horas trabajadas y porcentaje de horas trabajadas.

```

SELECT A.año,A.mes,A.nombre,B.horas_obligatorias,A.horas_trabajadas,
(horas_trabajadas/horas_obligatorias)*100 AS PORC
FROM
(select year(fecha) as año,month(fecha) as mes,p.id_persona,p.nombre as nombre,
sum((((12+salhor)*60)+salmin)-((enthor*60)+entmin)-60)/60) as horas_trabajadas
from nutreinarrrh.t008_movimiento m,nutreinarrrh.t004_persona p
where
p.id_persona=m.id_persona and
p.id_persona=127 and
date_format(fecha,'%M-%Y') ='January-2006') A,
(select year(fecha) as año,month(fecha) as mes, p.id_persona,count(*)*8 as horas_obligatorias
from nutreinarrrh.t008_movimiento m,nutreinarrrh.t006_personal_mes p
where
p.id_persona=m.id_persona and
p.id_mes=m.id_mes and
p.id_persona=127 and
year(fecha) =2006 and
month(fecha) in (1) ) B
WHERE
A.año=B.año AND
A.mes=B.mes AND
A.id_persona=B.id_persona;
    
```

año	mes	nombre	horas_obligatorias	horas_trabajadas	PORC
2006	1	Carmen Huat Yanak	176	184.0833	104.59279400

Ahora se presenta el proceso de Extracción, Transformación y Carga para el Datamart de Recursos Humanos mediante la herramienta de Pentaho Kettle, para tener un mayor detalle de la funcionalidad de la herramienta ver el **Anexo K - Funcionalidad de Kettle Pentaho**. La Figura 4.6 ilustra el ETL que alimenta el Datamart de Recursos Humanos.

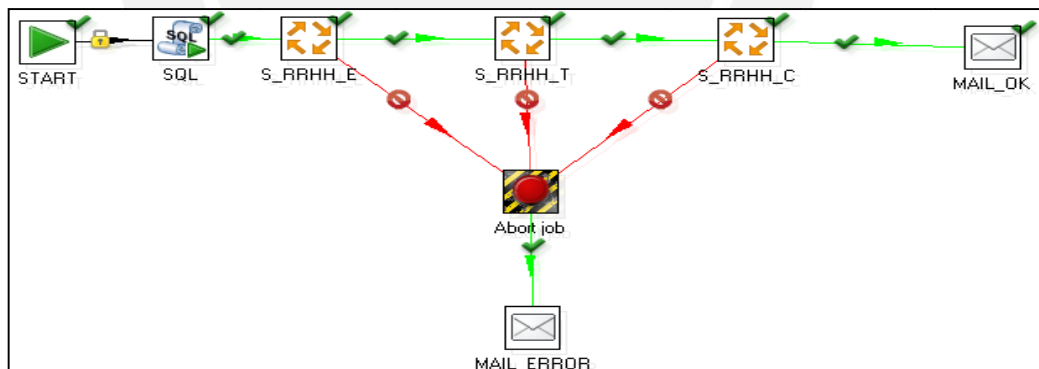


Figura 4.6 - ETL de Recursos Humanos

Este proceso muestra el job principal que extrae, transforma y carga la data a las tablas del Datamart de Recursos Humanos. En caso el proceso este correcto se envía un mail de satisfacción; de lo contrario se envía un mail indicando que el proceso ha fallado y aborta. Este es un proceso incremental; es decir, los datos son extraídos, transformados y cargados o actualizados periódicamente.

Entrada: Se ingresaron parámetros de entrada que permitirán conectarse a las tablas de la base de datos “nutreinarrhh” desde donde se desea obtener los datos (MySQL); así como conectarse a la base de datos dimensional cuyas tablas se desean cargar (SQLServer).

Salida: Las dimensiones y la fact de recursos humanos son cargadas correctamente.

Ejecución y Prueba: La Figura 4.7 y la Figura 4.8 ilustran los log del proceso ETL de Recursos Humanos.

```

2012/05/07 02:34:49 - Mail - Starting job entry
2012/05/07 02:34:51 - S_RRHH_C - Finished job entry [Mail] (result=[true])
2012/05/07 02:34:51 - S_RRHH_C - Finished job entry [F_MVTO_C] (result=[true])
2012/05/07 02:34:51 - S_RRHH_C - Finished job entry [F_PERSONAL_C] (result=[true])
2012/05/07 02:34:51 - S_RRHH_C - Finished job entry [F_LICENCIA_C] (result=[true])
2012/05/07 02:34:51 - S_RRHH_C - Finished job entry [MES_C] (result=[true])
2012/05/07 02:34:51 - S_RRHH_C - Finished job entry [VINCULO_C] (result=[true])
2012/05/07 02:34:51 - S_RRHH_C - Finished job entry [PUESTO_C] (result=[true])
2012/05/07 02:34:51 - S_RRHH_C - Finished job entry [PERSONA_C] (result=[true])
2012/05/07 02:34:51 - S_RRHH_C - Finished job entry [ONP_C] (result=[true])
2012/05/07 02:34:51 - S_RRHH_C - Finished job entry [LICENCIA_C] (result=[true])
2012/05/07 02:34:51 - S_RRHH_C - Finished job entry [CLASELABOR_C] (result=[true])
2012/05/07 02:34:51 - S_RRHH_C - Finished job entry [AREA_C] (result=[true])
2012/05/07 02:34:51 - S_RRHH_C - Finished job entry [AFP_C] (result=[true])
2012/05/07 02:34:51 - S_RRHH_ETC - Starting entry [MAIL_OK]
2012/05/07 02:34:51 - S_RRHH_ETC - exec(S, 0, MAIL_OK.0)
2012/05/07 02:34:51 - MAIL_OK - Starting job entry
2012/05/07 02:34:55 - S_RRHH_ETC - Finished job entry [MAIL_OK] (result=[true])
2012/05/07 02:34:55 - S_RRHH_ETC - Finished job entry [S_RRHH_C] (result=[true])
2012/05/07 02:34:55 - S_RRHH_ETC - Finished job entry [S_RRHH_T] (result=[true])
2012/05/07 02:34:55 - S_RRHH_ETC - Finished job entry [S_RRHH_E] (result=[true])
2012/05/07 02:34:55 - S_RRHH_ETC - Finished job entry [SQL] (result=[true])
2012/05/07 02:34:55 - S_RRHH_ETC - Job execution finished
2012/05/07 02:34:55 - Spoon - Trabajo ha terminado.
    
```

Figura 4.7 - Log del Datamart de Recursos Humanos I

Trabajo / Entrada de Trabajo	Comentario	Resultado	Razón	Nombre Fichero
S_RRHH_ETC				
Trabajo: S_RRHH_ETC	Start of job execution		start	
START	Start of job execution		start	
START	Job execution finished	Exito		
SQL	Start of job execution		Followed unconditional link	
SQL	Job execution finished	Exito		
S_RRHH_E	Start of job execution		Followed link after success	C:\ROLANDO\TESIS\ETL\EXTRACCIO...
Trabajo: S_RRHH_E	Job execution finished	Exito		C:\ROLANDO\TESIS\ETL\EXTRACCIO...
S_RRHH_T	Start of job execution		Followed link after success	C:\ROLANDO\TESIS\ETL\TRANSPOR...
Trabajo: S_RRHH_T	Job execution finished	Exito		C:\ROLANDO\TESIS\ETL\TRANSPOR...
S_RRHH_C	Start of job execution		Followed link after success	C:\ROLANDO\TESIS\ETL\CARGA_RR...
Trabajo: S_RRHH_C	Job execution finished	Exito		C:\ROLANDO\TESIS\ETL\CARGA_RR...
MAIL_OK	Start of job execution		Followed link after success	
MAIL_OK	Job execution finished	Exito		
Trabajo: S_RRHH_ETC	Job execution finished	Exito	finished	

Figura 4.8 - Log del Datamart de Recursos Humanos II

Notar que en la primera imagen el log muestra “el proceso terminó con éxito”. En la segunda imagen se muestra a detalle que cada job que compone el proceso principal termina con éxito.

Para un mayor detalle de la ejecución y pruebas del ETL para el Datamart de recursos humanos observar el **Anexo I - Ejecución y Pruebas del ETL**.

5. CAPÍTULO 5: Observaciones, Conclusiones y Recomendaciones o Trabajos Futuros

Este capítulo hace hincapié a todas las conclusiones que se han formado después de elaborar el presente proyecto de fin de carrera. Asimismo se muestra una lista de observaciones que se deben tomar en cuenta y una serie de recomendaciones para quienes deseen implementar el presente proyecto de fin de carrera en cuestión o utilizar las herramientas usadas en la elaboración de la misma.

5.1. Observaciones

- Para la elaboración del presente proyecto de fin de carrera se ha realizado un análisis de TCO (*costo total de propiedad*) para la herramienta ETL y para la herramienta de explotación de información SQL Server.

La Tabla 5.1 ilustra una comparación entre el cálculo del TCO para la herramienta ETL de software libre y el cálculo del TCO para la herramienta ETL de software propietario para el periodo de un año.

Características	Software ETL		
	TCO de Pentaho	TCO de otro software propietario	Ahorro
Licencia y soporte	NO	S/. 374 394	100 %
Diseño y desarrollo	S/. 60 000	S/. 60 000	0 %
Capacitación del Personal	S/. 10 000	S/. 10 000	0 %
Total TCO	S/. 70 000	S/. 444 394	84 %

Tabla 5.1 - Análisis TCO Para Una Herramienta ETL

Analizando la tabla anterior se puede observar que, para la herramienta ETL, el uso de software libre permite el ahorro de costos en licencias, soporte y migraciones a nuevas versiones. Con Pentaho Data Integration la organización puede ahorrar entre un 60 % y 90 % del costo por proyecto de Inteligencia de Negocios que si es que se empleara una herramienta ETL propietaria. [IBM 2013]

Para un mayor detalle observar el **Anexo J - Costo Total de Propiedad**.

En base a lo anterior es que se elegirá la herramienta Pentaho para la elaboración de los procesos ETL.

Por otro lado la empresa ya contaba con las licencias de SQL Server para trabajar con la elaboración de aplicaciones web. Aprovechando esta licencia se escogió la herramienta de SQL Server para la realización y explotación de los cubos BI y poder aprovechar sus múltiples funcionalidades que como software propietario ofrece; así como su integración con hojas de cálculo en excel, el envío de correos de los reportes en tiempo real; entre otros.

Para esta herramienta también se ha hecho un análisis de TCO con otros productos propietarios competidores como IBM u Oracle dando como resultado un ahorro de hasta un 48 %. La Tabla 5.2 muestra las características de un análisis de TCO para la herramienta SQL Server y otras herramientas de BI para el periodo de un año. [TCO 2012]

Características	Herramienta de BI		
	TCO de SQL Server	TCO de otro producto competidor	Porcentaje de ahorro
Licencia	S/. 13 000	S/. 39 000	67 %
Diseño y desarrollo	S/. 32 477	S/. 42 704	24 %
Capacitación del Personal	S/. 1 456	S/. 3 157	54 %
Mantenimiento	S/. 11 180	S/. 25 155	56 %
Total TCO	S/. 58 113	S/. 110 016	48 %

Tabla 5.2 - Análisis TCO Para Una Herramienta BI

Del cuadro anterior se puede concluir que dentro de las herramientas propietarias de BI usar SQL Server permitirá ahorrar costos por los diferentes beneficios que ofrece.

Para un mayor detalle observar el **Anexo J - Costo Total de Propiedad**.

- Las desventajas de usar software libre es que no se tiene el soporte ni la documentación necesaria. A pesar de ello las herramientas de software libre se encuentran en constantes crecimientos y su avance ofrece un mejor uso para su desarrollo y mantenimiento.
- El análisis de los Datamarts de Ventas y Recursos Humanos puede ser usado para cualquier empresa industrial que cuente con dichas áreas de acuerdo a sus necesidades. Los Datamarts de Ventas y Recursos Humanos para el presente proyecto de fin de carrera se hicieron tomando como base la empresa en estudio.
- El presente proyecto de fin de carrera es una alternativa para la empresa con la que se está trabajando, dado que puede implementarla y así poder satisfacer sus necesidades de gestión, análisis y toma de decisiones.
- El uso de software libre usado para el presente proyecto de fin de carrera solo se empleó bajo una sola premisa: Economía. Sin embargo existe en el mercado muchos software propietarios; con mayores funcionalidades de ETL y mucho más fáciles de usar; que no fueron empleados por falta de licencias. Además el software propietario cuenta con documentación y soporte que ayuda a dar un mantenimiento rápido del mismo.

- El presente proyecto de fin de carrera tuvo como alcance la elaboración de Datamarts para las áreas de ventas y recursos humanos de una empresa industrial dedicada a la importación y exportación de productos alimenticios. Para ello se implantó una plataforma de BI para la extracción, transformación y carga de datos, además se hizo la creación y procesamiento de los cubos con Analysis Services, la cual es una funcionalidad de SQLServer 2008. Se optó por este proceso dado que la empresa, a la cual está dirigido el presente proyecto de fin de carrera, cuenta con las licencias de SQLServer 2008.
- Los contadores y gerentes, que son las personas que toman las decisiones de la empresa, son los encargados de tomar las decisiones en base a los reportes que la plataforma de Bi proporcionará. Ellos actualmente usan la herramienta de Microsoft Excel 2007 para observar sus reportes estáticos como Balance General, Estado de Ganancias y Pérdidas, Plan de Ventas, entre otros. Asimismo uno de los requerimientos fue seguir prevaleciendo el uso de Microsoft Excel dado que es una herramienta sobre la cual ellos se sienten más familiarizados. Es por esto que se concluye que, además de los reportes en web, otra opción es que las personas encargadas de tomar las decisiones en la empresa lo hagan bajo la herramienta Microsoft Excel 2007. Además, con esta herramienta se podrá observar las dimensiones de manera dinámica pudiendo ellos tomar las mejores decisiones.

5.2. Conclusiones

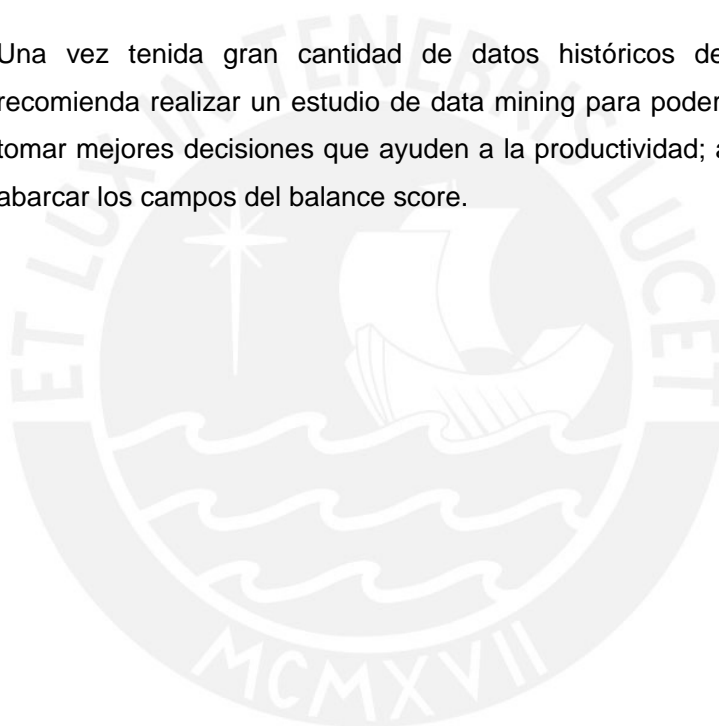
- La mejor posibilidad para desarrollar el presente proyecto de fin de carrera es el esquema de Kimball dado que no es necesario la creación de un Datawarehouse, simplemente se debe de extraer la data de las diferentes bases de datos existentes de la organización y con ello ir armando los Datamarts para las diferentes áreas del negocio.
- Es importante realizar en forma correcta el proceso de análisis ya que de no hacerlo implica que a lo largo del proyecto se tenga que hacer una reestructuración de los procesos, mapeos o reportes de los mismos. Para ello es de vital importancia entender los requerimientos que desea el cliente.

- Es importante realizar un prototipo de reportes junto con el usuario final de manera que estos no sufra grandes modificaciones una vez que hayan sido implementados.
- Para realizar una óptima implementación de un sistema de soporte a decisiones es muy importante las reuniones con el usuario final. Como mínimo deben de existir tres reuniones: una primera reunión donde se detalle las necesidades del usuario; una segunda reunión donde se fije las dimensiones junto con los indicadores y medidas necesarias y una tercera reunión donde se fije los prototipos de reportes a implementar.

5.3. Recomendaciones o Trabajos Futuros

- Para realizar una buena implementación, implantación y capacitación del presente proyecto de fin de carrera se recomienda documentarse bien sobre la herramienta Pentaho que es aquella con las que se hará la extracción, transformación y carga. Asimismo se recomienda documentarse sobre la funcionalidad de Analysis Services de SQLServer para realizar el procesamiento de los cubos y Reporting Services para la elaboración de informes web.
- A veces la documentación suele ser difícil de encontrar; para ello se recomienda participar en foros de las herramientas a utilizar o navegar en las páginas webs de Internet relacionadas donde los participantes brindan información en base a sus experiencias.
- Para tener la certeza que los procesos se ejecutarán de forma correcta en el ambiente de producción se recomienda realizar distintas pruebas en el ambiente de desarrollo de tal forma que si se encuentra alguna inconsistencia esta pueda ser resuelta antes del “pase a producción”.
- Para validar que la información que utilizará el usuario sea la correcta, se recomienda revisar varios reportes que hagan referencia al mismo Datamart de manera que los datos sean congruentes e idénticos entre ellos.

- Para dar solución a las problemática de otras áreas de la organización se recomienda la creación de Datamarts para todas estas en base a un análisis detallado de manera que la plataforma BI se expanda por toda la organización y ayude a la toma de decisiones de cada área.
- A medida que vayan creciendo la implementación de Datamarts en la organización se hará necesario el reúso de dimensiones o fact tables en caso corresponda; por lo que se recomienda tener mapeado las múltiples dimensiones y facts existentes en una base de datos de manera que se facilite su uso.
- Una vez tenida gran cantidad de datos históricos de las ventas, se recomienda realizar un estudio de data mining para poder predecirlas y así tomar mejores decisiones que ayuden a la productividad; así como también abarcar los campos del balance score.



6. Referencias

[APM 2012] Alfonso Pinedo Mirano. Entrevista. 17 junio 2012.

[BAR 2005] Mauricio Andrés Barrientos. Desarrollo de un sistema de Business Intelligence para el área de compras de la empresa Salmofood S.A. Universidad Austral de Chile, Facultad de Ciencias de la Ingeniería, Escuela de Ingeniería Civil en Informática. 2005.

<http://cybertesis.uach.cl/tesis/uach/2005/bmfcib275d/doc/bmfcib275d.pdf>.

Último acceso: mayo 2010.

[BEL 2007] Carlos López Beltrán. Análisis, diseño e implementación de un Datamart para la Dirección Financiera y Recursos Humanos de la Escuela Politécnica del Ejército para una toma de decisión efectiva. 2007.

<http://www3.espe.edu.ec:8700/handle/21000/400>.

Último acceso: noviembre 2011.

[BUS 2010] Business Objects Web Page

<http://www.sap.com/solutions/sapbusinessobjects/index.epx>

Último acceso: mayo 2010

[COG 2010] Cognos 8 Business Intelligence Web Page

<http://www-01.ibm.com/software/data/cognos/>

Último acceso: mayo 2010

[DAT 2010] Data Stage Web Page

<http://www-01.ibm.com/software/data/infosphere/datastage/>.

Último acceso: mayo 2010

[GUE 2005] Régulo Guédez, Yaranais Zambrano. Diseño e implantación de un Datamart para el área de facturación y ventas de la empresa Macro Comercializadora, S.A. Sede La Urbina. Universidad Católica Andrés Bello. Facultad de Ingeniería. Escuela de Ingeniería Informática. 2005.

http://200.2.12.152/wwwisis/anexos/marc/texto/AAQ4739_1.pdf.

Último acceso: mayo 2010

[IBM 2013] IBM Passport Advantage Express

[https://www-](https://www-112.ibm.com/software/howtobuy/buyingtools/paexpress/Express?part_number=D5982LL%2CD5988LL%2CD5987LL%2CD599ULL%2CD598CLL%2CD5975LL%2CD597FLL%2CD598PLL%2CD5978LL%2CD59AXLL%2CD59PDLL%2CD59PQLL%2CD59PSLL%2CD59PFL%2CD59PULL%2CD59PELL%2CD03TVLL%2CD03TTLL%2CD03SPLL%2CD03SYLL%2CD03SGLL%2CD03U6LL%2CD03T8LL%2CD03SQLL%2CD03U5LL%2CD03P8LL%2CD03PFL%2CD03PHLL%2CD03P9LL%2CD03PNLL%2CD03SELL%2CD03PDLL%2CD0IAALL%2CD0IAELL&catalogLocale=es_ES&Locale=null&country=ESP&PT=html&TACTICS=%26S_TACT%3D%26S_CMP%3D%26brand%3D&ibm-submit=Ver+precios+y+comprar)

[112.ibm.com/software/howtobuy/buyingtools/paexpress/Express?part_number=D5982LL%2CD5988LL%2CD5987LL%2CD599ULL%2CD598CLL%2CD5975LL%2CD597FLL%2CD598PLL%2CD5978LL%2CD59AXLL%2CD59PDLL%2CD59PQLL%2CD59PSLL%2CD59PFL%2CD59PULL%2CD59PELL%2CD03TVLL%2CD03TTLL%2CD03SPLL%2CD03SYLL%2CD03SGLL%2CD03U6LL%2CD03T8LL%2CD03SQLL%2CD03U5LL%2CD03P8LL%2CD03PFL%2CD03PHLL%2CD03P9LL%2CD03PNLL%2CD03SELL%2CD03PDLL%2CD0IAALL%2CD0IAELL&catalogLocale=es_ES&Locale=null&country=ESP&PT=html&TACTICS=%26S_TACT%3D%26S_CMP%3D%26brand%3D&ibm-submit=Ver+precios+y+comprar](https://www-112.ibm.com/software/howtobuy/buyingtools/paexpress/Express?part_number=D5982LL%2CD5988LL%2CD5987LL%2CD599ULL%2CD598CLL%2CD5975LL%2CD597FLL%2CD598PLL%2CD5978LL%2CD59AXLL%2CD59PDLL%2CD59PQLL%2CD59PSLL%2CD59PFL%2CD59PULL%2CD59PELL%2CD03TVLL%2CD03TTLL%2CD03SPLL%2CD03SYLL%2CD03SGLL%2CD03U6LL%2CD03T8LL%2CD03SQLL%2CD03U5LL%2CD03P8LL%2CD03PFL%2CD03PHLL%2CD03P9LL%2CD03PNLL%2CD03SELL%2CD03PDLL%2CD0IAALL%2CD0IAELL&catalogLocale=es_ES&Locale=null&country=ESP&PT=html&TACTICS=%26S_TACT%3D%26S_CMP%3D%26brand%3D&ibm-submit=Ver+precios+y+comprar)

Último acceso: marzo 2013

[INM 2003] W.H. Inmon, Wiley. Building the Datawarehouse. 2003.

[KET 2012] Data Integration - Kettle

<http://kettle.pentaho.com/>

Último acceso: enero 2012

[KIM 2002] Ralph Kimball, Wiley. The Datawarehouse toolkit: the complete guide to dimensional modeling. 2002.

[MIC 2010] Microstrategy Business Intelligence Solutions Web Page

<http://www.microstrategy.com/>

Último acceso: mayo 2010

[MOS 2003] Moss Larissa, Atre Shaku. Business Intelligence Roadmap. The Complete Project Lifecycle for Decisión-Support Applications. 2003 EE.UU.

[OCT 2010] Enhydra Octopus Web Page

<http://www.enhydra.org/tech/octopus/>

Último acceso: mayo 2010

[PAD 2010] Paulina Padilla Manzano, Yuli Alvarado Moncada, Jaime Lozada Loza. “Planeación operativa en el área de ventas basada en la metodología de Balanced Scorecard en una empresa dedicada a la fabricación y comercialización de repuestos de cauchos”. Instituto de Ciencias Matemáticas, Facultad de Ingeniería en Electricidad y Computación. Escuela Superior Politécnica del Litoral.

<http://www.dspace.espol.edu.ec/bitstream/123456789/6528/1/Planeaci%C3%B3n%20Operativa%20en%20el%20C3%81rea%20de%20ventas%20basada%20en%20la%20metodolog%C3%ADa.pdf>.

Último acceso: mayo 2010

[PEN 2010] Pentaho Web Page

<http://www.pentaho.com/>

Último acceso: mayo 2010

[RMR 2013] Rolando Moreno Reyes (Elaboración propia)

[SQL 2010] SQL Server Integration Services Web Page

<http://msdn.microsoft.com/es-es/library/ms141026.aspx>.

Último acceso: mayo 2010

[SQL 2011] SQL Server 2008 R2

<http://www.microsoft.com/downloads/details.aspx?familyid=265f08bc-1874-4c81-83d8-0d48dbce6297&displaylang=es>

Último acceso: enero 2011

[SUN 2010] Sunopsis Web Page

<http://www.oracle.com/sunopsis/index.html>

Último acceso: mayo 2010

[TCO 2012] TW Datos

[http://www.twdatos.com/sitio/\(S\(322e3w45kmcisuvvdhwxnv45\)A\(d27qdDPyzAEkAA AAMTI2Y2I2M2QtNDFmNy00ZWU2LWlzODktODc3OGE0YWUyMTIi3EUo71DzoiNTSa3_4nigxbVtf11\)\)/Publicaciones/entendiendoPrecioBI.aspx?AspxAutoDetectCookieSupport=1](http://www.twdatos.com/sitio/(S(322e3w45kmcisuvvdhwxnv45)A(d27qdDPyzAEkAA AAMTI2Y2I2M2QtNDFmNy00ZWU2LWlzODktODc3OGE0YWUyMTIi3EUo71DzoiNTSa3_4nigxbVtf11))/Publicaciones/entendiendoPrecioBI.aspx?AspxAutoDetectCookieSupport=1)

Último acceso: marzo 2013

[TEC 2013] Técnicas de navegación
http://en.wikipedia.org/wiki/OLAP_cube
Último acceso: mayo 2013

