

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

FACULTAD DE CIENCIAS E INGENIERÍA



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DEL PERÚ

ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN DATAMART PARA EL ÁREA DE MANTENIMIENTO Y LOGÍSTICA DE UNA EMPRESA DE TRANSPORTE PÚBLICO DE PASAJEROS

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

Jaime Alexander Zambrano Alarcón

ASESOR: Carla Basurto Figueroa

Lima, noviembre del 2011

Resumen

El presente trabajo de tesis implementa un Datamart para el apoyo al proceso de toma de decisiones del área de mantenimiento y logística de una empresa de transporte público de pasajeros.

Las empresas de transporte público de pasajeros son un tipo de empresa que maneja una gran cantidad de información día a día. Sin embargo, muchas de ellas no saben cómo administrarlo adecuadamente, debido a que sus sistemas actuales no soportan el manejo adecuado de grandes volúmenes de información. Así, las empresas de transporte público tienen el problema de utilizar su información para emplearla en la toma de decisiones.

El objetivo principal es proveer una solución de Inteligencia de Negocios que dé soporte a las necesidades de información de los usuarios finales del área de mantenimiento y logística.

Para llevar adelante el desarrollo del Datamart se utilizó la metodología DWEP, la cual está basada en la metodología de implementación de desarrollo de software, Rational Unified Process. Además, para una adecuada gestión del proyecto se consideraron las actividades y entregables propuestos por el PMBOK.

La acertada selección de las actividades y las tareas de las metodologías nombradas han guiado y facilitado el desarrollo de la solución logrando un producto que cumple satisfactoriamente las necesidades de información de los usuarios. El proceso de gestión de proyecto siguiendo las recomendaciones del PMBOK, con sus tareas de planificación, estimación, seguimiento y control, ha permitido culminar el trabajo en un tiempo similar al estimado y con la calidad deseada.

Como futuras aplicaciones de la solución se propone implementar un componente de Inteligencia de Negocios basado en Balanced Scorecard, incorporar nuevas herramientas especializadas en Inteligencia de Negocios como también ampliar la funcionalidad incorporando otras áreas de una empresa de transporte público de pasajeros.

TEMA DE TESIS PARA OPTAR EL TÍTULO DE INGENIERO INFORMÁTICO

TÍTULO: ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN DATAMART PARA EL ÁREA DE MANTENIMIENTO Y LOGÍSTICA DE UNA EMPRESA DE TRANSPORTE PÚBLICO DE PASAJEROS
SISTEMAS DE INFORMACIÓN

PROPONENTE: Carla Basurto Figueroa

ASESOR: Carla Basurto Figueroa

ALUMNO: Jaime Alexander Zambrano Alarcón

CÓDIGO: 20030339

TEMA N°: _____

FECHA: San Miguel, 1 de octubre de 2011

DESCRIPCIÓN

En la actualidad, la información se ha convertido en un bien muypreciado. Las empresas buscan emplear dicha información para generar conocimiento útil dirigido a la mejora de sus procesos empresariales. De esta forma, la ventaja competitiva de las organizaciones radica en la forma de interpretar la información y convertirla en un elemento diferencial.

Las empresas de transporte público de pasajeros son un tipo de empresa que maneja una gran cantidad de información día a día. Este tipo de empresas realiza un alto número de transacciones generando una gran cantidad de datos. Sin embargo, muchas de ellas no saben cómo administrarla adecuadamente, debido a que sus sistemas actuales no soportan el manejo adecuado de grandes volúmenes de datos. Así, estas empresas tienen el problema de utilizar inadecuadamente su información para la toma de decisiones.

El Datamart propuesto será una herramienta que brindará información útil para la toma de decisiones en el área de mantenimiento y logística de una empresa de transporte público de pasajeros. Además, permitirá un fácil acceso a la información por parte de los usuarios de manera independiente y sin necesidad de conocimientos técnicos.

OBJETIVO

El objetivo del presente proyecto es realizar el análisis, diseño e implementación de un Datamart que responda a los requerimientos de análisis de información del área de mantenimiento y logística de una empresa de transporte público de pasajeros.

OBJETIVOS ESPECÍFICOS

- Identificar los requerimientos de análisis de información para el área de mantenimiento y logística del tipo de empresa mencionada.
- Elaborar un modelo de base de datos multidimensional que permita el análisis y explotación de la información identificada.

- Desarrollar los procesos de extracción, transformación y carga de los datos al Datamart.
- Implementar el Datamart utilizando una herramienta de explotación de información que permita una fácil interacción con el usuario final.
- Elaborar reportes y tableros de control que faciliten la toma de decisiones para los usuarios finales.

ALCANCE

El Datamart manejará variables propias del negocio como: autobús, tipo de mantenimiento, producto, proveedor. De esta manera, permitirá un análisis de: mantenimiento preventivo y correctivo de los autobuses, control de los productos del almacén, es decir, no sólo realizará un análisis del almacén sino también de mantenimiento.

El modelo multidimensional del Datamart contendrá las variables necesarias que respondan a los requerimientos citados anteriormente.

Se implementarán los procesos ETL a través de la herramienta libre Pentaho Data Integration, la cual permitirá cargar la información al Datamart.

Se empleará Pentaho como herramienta de explotación para que permita el acceso a la información del Datamart.

Los reportes y tableros de control elaborados contendrán gráficos e indicadores de gestión que ayudarán a los usuarios finales en la toma de decisiones.

Agradecimientos

Deseo expresar mis más sinceras muestras de agradecimientos a:

A mis padres y hermana por creer y confiar en siempre en mí, apoyándome en todas las decisiones que he tomado en mi vida.

A mis maestros, en especial a mi asesora Carla Basurto, por sus consejos y por compartir desinteresadamente sus amplios conocimientos y experiencia.



ÍNDICE

Introducción.....	10
1. Generalidades	11
1.1. Definición del problema.....	11
1.2. Marco conceptual del problema.....	15
1.2.1. Introducción al Data Warehouse y Datamart.....	15
1.2.2. Introducción a los sistemas ETL.....	17
1.2.3. Introducción al Modelo Multidimensional.....	17
1.2.4. Introducción a la Inteligencia de Negocios y a los Sistemas de Información Ejecutiva.....	20
1.2.5. Conceptos asociados al negocio de una empresa de transporte público de pasajeros.....	20
1.3. Plan del proyecto.....	22
1.3.1. Métodos y Procedimientos.....	22
1.3.2. Planificación.....	24
1.4. Estado del arte.....	28
1.4.1. Soluciones generales.....	28
1.4.2. Soluciones específicas.....	31
1.5. Descripción y sustentación de la solución.....	33
1.5.1. Descripción de la solución.....	33
1.5.2. Sustentación de la solución.....	35
2. Análisis.....	38
2.1. Definición de la metodología de la solución.....	38
2.2. Identificación de requerimientos.....	44
2.2.1. Requerimientos Funcionales.....	44
2.2.2. Requerimientos No Funcionales.....	48
2.3. Análisis dimensional.....	48
2.3.1. Diseño Conceptual del Datamart.....	49
2.3.2. Diseño Lógico del Datamart.....	55
3. Diseño.....	60
3.1. Arquitectura de la solución.....	60
3.1.1. Fuentes de Información.....	62
3.1.2. ETL.....	62
3.1.3. Datamart.....	64
3.2. Diseño de Extracción.....	65
3.3. Diseño de Explotación.....	67
3.3.1. Estándares de Reportes.....	69
4. Construcción.....	72
4.1. Construcción.....	72
4.1.1. Herramientas y soluciones en el mercado.....	72
4.1.2. Configuración de las herramientas.....	77
4.1.3. Proceso ETL.....	77
4.2. Pruebas.....	83
4.2.1. Planificación del plan de pruebas.....	84
4.2.2. Diseño del plan de pruebas.....	84
4.2.3. Determinación de los casos de prueba.....	85
4.2.4. Ejecución del plan de pruebas.....	86
4.2.5. Análisis y evaluación del plan de pruebas ejecutado.....	86
4.2.6. Caso de Prueba para el ETL de FT_DOCUMENTO_DETALLE.....	86
4.2.7. Caso de Prueba para el Cubo de Mantenimiento.....	89
5. Observaciones, conclusiones y recomendaciones.....	91
5.1. Observaciones.....	91
5.2. Conclusiones.....	92
5.3. Recomendaciones y trabajos futuros.....	94
5.3.1. Implementar un componente de Inteligencia de Negocios basado en Balanced Scorecard.....	94
5.3.2. Implementar herramientas de Inteligencia de Negocios.....	94
5.3.3. Ampliación de áreas y departamentos.....	95

5.3.4. Adaptación a empresas similares	95
Bibliografía	96



Índice de Figuras

Figura 1: Organigrama de empresa de transporte de pasajeros.....	13
Figura 2: Esquema Estrella.....	18
Figura 3: Esquema Constelación de Hechos.....	18
Figura 4: Estructura de un Cubo.....	19
Figura 5: WBS.....	25
Figura 6: Diagrama Gantt por fases.....	26
Figura 7: RBS.....	27
Figura 8: Estructura de Solución BI común.....	29
Figura 9: Metodología propuesta por Inmon.....	30
Figura 10: Metodología propuesta por Kimball.....	30
Figura 11: Metodología de Datamarts independientes.....	31
Figura 12: Fases y Disciplinas del DWEP.....	40
Figura 13: Relación entre los Esquemas del Datamart.....	42
Figura 14: Diagrama de Actividades de los principales pasos del DWEP.....	43
Figura 15: Jerarquía de Dimensiones.....	50
Figura 16: Esquema Conceptual del Modelo Estrella Documento Detalle.....	51
Figura 17: Esquema Conceptual del Modelo Estrella Documento.....	52
Figura 18: Esquema Conceptual del Modelo Estrella Ingresos y Salidas.....	53
Figura 19: Esquema Conceptual del Modelo Estrella Stock Valorizado.....	53
Figura 20: Esquema Conceptual del Modelo Estrella Mantenimiento.....	54
Figura 21: Esquema Lógico del Modelo Estrella Documento Detalle.....	56
Figura 22: Esquema Lógico del Modelo Estrella Documento.....	57
Figura 23: Esquema Lógico del Modelo Estrella Ingresos y Salidas.....	58
Figura 24: Esquema Lógico del Modelo Estrella Stock Valorizado.....	58
Figura 25: Esquema Lógico del Modelo Estrella Mantenimiento.....	59
Figura 26: Diagrama de Arquitectura.....	61
Figura 27: Diagrama de Clases de la base de datos fuente.....	62
Figura 28: Diagrama de Mapeo de Datos de Autobús.....	63
Figura 29: Diagrama de Proceso ETL para la Dimensión Autobús.....	64
Figura 30: Diagrama de Componente de la base de datos del Datamart.....	64
Figura 31: Diagrama de Componente del servidor de base de datos.....	65
Figura 32: Distribución de las partes de la Interfaz de Usuario.....	68
Figura 33: Ejemplo de Análisis de Cubo.....	68
Figura 34: Reporte tipo lista.....	69
Figura 35: Reporte tipo gráfico.....	70
Figura 36: Reporte híbrido.....	70
Figura 37: Proceso ETL principal de la carga periódica.....	79
Figura 38: Proceso ETL de las dimensiones para la carga periódica.....	80
Figura 39: Proceso ETL de los hechos de documento para la carga periódica.....	81
Figura 40: Reporte de Ranking de Proveedores.....	82
Figura 41: Tablero de control de Logística.....	83
Figura 42: Tabla fuente de DETALE_CTA_CTE (173 registros).....	87
Figura 43: ETL para cargar la FT_DOCUMENTO_DETALLE.....	88
Figura 44: Log del ETL para cargar la FT_DOCUMENTO_DETALLE.....	88
Figura 45: Tabla intermedia FT_DOCUMENTO_DETALLE (173 registros).....	88
Figura 46: ETL final para cargar la FT_DOCUMENTO_DETALLE.....	89
Figura 47: Log del ETL final para cargar la FT_DOCUMENTO_DETALLE.....	89
Figura 48: Tabla destino FT_DOCUMENTO_DETALLE (173 registros).....	89
Figura 49: Consulta en FT_MANTENIMIENTO.....	90
Figura 50: Consulta en Cubo Mantenimiento.....	90

Índice de Tablas

Tabla 1: Cuadro comparativo de metodologías	31
Tabla 2: Cuadro comparativo de sistemas para empresas de transporte público	33
Tabla 3: Cuadro comparativo entre Data Warehouse y Datamart.....	37
Tabla 4: Resumen de Entregables por Disciplinas del DWEP	40
Tabla 5: Hechos vs. Dimensiones	55
Tabla 6: Tabla de campos de la cabecera.....	69
Tabla 7: Reportes por Tema	71
Tabla 8: Cuadro comparativo de las herramientas preseleccionadas.....	76
Tabla 9: Objetos de prueba y sus criterios de aceptación.....	85
Tabla 10: Ejemplo de caso de prueba	86
Tabla 11: Caso de Prueba para el Cubo de Mantenimiento.....	90



Introducción

Las empresas actualmente caracterizan a la información como uno de los activos de la empresa [BIT 2002], debido a ello empiezan a tratarla más metódicamente, especialmente la información que da soporte al proceso de toma de decisiones.

Las empresas cuentan con un conjunto de aplicaciones de procesamiento transaccional que mecanizan las operaciones de su día a día. En este conjunto de aplicaciones se procesan grandes cantidades de datos referentes a las actividades rutinarias y se almacenan en bases de datos. De ellas se puede extraer información que básicamente sirve de soporte para apoyar en decisiones operativas que conducen actividades básicas, mas no sirve para realizar un análisis más profundo o estratégico, ya que no están diseñadas para este tipo de tareas.

Así muchas empresas si bien cuentan con una gran cantidad de información que podría generarle una ventaja competitiva, no cuentan con las herramientas necesarias para poder administrar los datos y se enfrentan al problema de procesar dichos datos y transformarla en información útil.

Como solución a los problemas de información de las empresas, es posible extraer un grupo de datos, a partir de una o varias bases de datos operacionales, que aporten un valor agregado a la gestión de la empresa, lo que constituirá un Data Warehouse o Datamart.

El presente proyecto tiene como objetivo principal implementar un Datamart para el área de mantenimiento y logística de una empresa de transporte público de pasajeros para brindarle una herramienta que facilitará a dicha área en la toma de decisiones.

1. Generalidades

En el presente capítulo se define claramente el problema que se desea resolver. Luego se presenta las definiciones necesarias para entender tanto el problema como la solución propuesta. Además, se realiza el plan de proyecto basado en entregables del PMBOK que permitirán gestionar el proyecto de fin de carrera. Finalmente, se realiza una especificación de la solución.

1.1. Definición del problema

En la actualidad, la información se ha convertido en un bien muy preciado. [BIT 2002] Las empresas buscan emplear dicha información para generar conocimiento útil dirigido a la mejora de sus procesos empresariales. De esta forma, la ventaja competitiva de las organizaciones radica en la forma de interpretar la información y convertirla en un elemento diferencial.

Las empresas de transporte público de pasajeros son un tipo de empresa que maneja una gran cantidad de información día a día. Este tipo de empresas realiza un alto número de transacciones, lo cual genera un gran repositorio de datos. Sin embargo, muchas de ellas no saben cómo administrarlo adecuadamente, debido a que sus sistemas actuales no soportan el manejo adecuado de grandes volúmenes

de información. Así, las empresas de transporte público tienen el problema de utilizar su información para emplearla en la toma de decisiones.

Uno de los más grandes problemas que enfrenta Lima es respecto al transporte público de pasajeros. Para marzo del 2008, Lima tenía un parque automotriz conformado por aproximadamente 800'000 vehículos para una población cercana a los 7.5 millones de habitantes. De ese gran grupo, cerca de 42'000 vehículos son destinados al transporte público y lo conforman los omnibuses, microbuses y camionetas rurales (combis), los cuales transitan por 418 rutas en toda la ciudad. Sin embargo, sólo es requerido 22'000 vehículos para atender a los 6.8 millones de pasajeros que emplean el transporte público, es decir, existe una sobre oferta de 20'000 vehículos. [GTU 2008]

Este exceso de flota en el transporte público incrementa la congestión vehicular, la pérdida de tiempo, la contaminación y la inseguridad ciudadana. Además la mayoría de las unidades de transporte público no cumplen con la ruta establecida, e infringen los requisitos exigidos por la autoridad generando incomodidad en los pasajeros. Adicionalmente, en los últimos años han aumentado considerablemente las combis, conformando casi el 50% de todos los vehículos dedicados al transporte público urbano. Estos últimos son los vehículos típicos de transporte público para las distancias cortas, pero ofrecen un servicio deficiente respecto a los estándares de seguridad y calidad. Por estas razones, los usuarios del sistema de transporte público reclaman cambios en la operación exigiendo un servicio de mejor calidad y eficiencia.

Desde el 2010, ha entrado en circulación un nuevo sistema de transporte público para Lima llamado "El Metropolitano". Se trata de un sistema basado en los autobuses de tránsito rápido que conecta el norte y sur de Lima atravesando cerca de 16 distritos y beneficiando a 700 mil usuarios al día. El Metropolitano surgió como alternativa para ofrecer un transporte público de mayor calidad, pensando en el medio ambiente y en las personas de la tercera edad o con discapacidad. Sin embargo, el nuevo sistema de transporte no ha reemplazado a los vehículos de transporte público que ya circulaban por la capital. Por lo tanto, sigue existiendo una sobre oferta respecto al transporte público.

Existen diversas formas de abordar los problemas del transporte público. Para ello es necesario conocer cómo está organizada una empresa de este rubro y sus

principales procesos de negocio. Una empresa de transporte público de pasajeros tiene un organigrama como se muestra en la Figura 1.

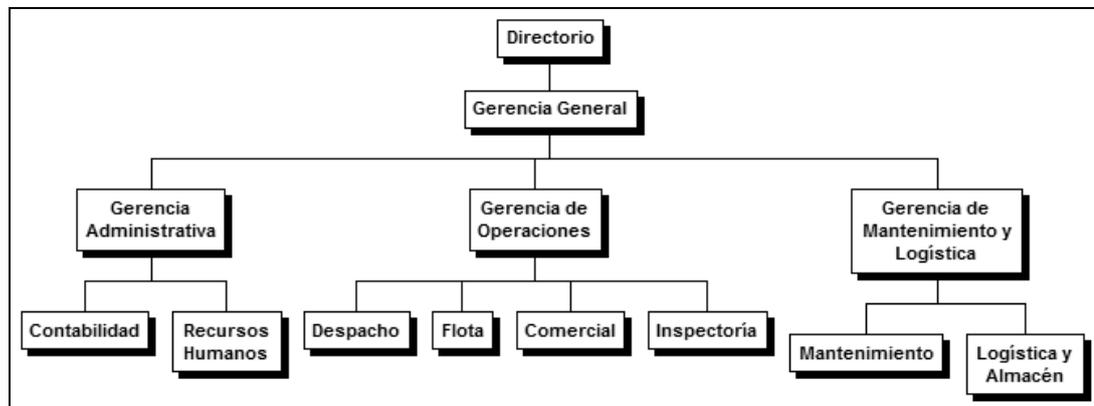


Figura 1: Organigrama de empresa de transporte de pasajeros

A grandes rasgos la empresa tiene tres áreas o gerencias principales: Gerencia Administrativa, Gerencia de Operaciones y Gerencia de Mantenimiento y Logística.

- **Gerencia Administrativa:** Conformada por el área de Contabilidad, que lleva la contabilidad de la empresa, y por el área de Recursos Humanos, encargada de la selección y reclutamiento del personal.
- **Gerencia de Operaciones:** Gerencia encargada de manejar los procesos principales de negocio. Está conformada por el área de Despacho, Flota, Comercial e Inspectoría. La primera es la encargada de realizar el despacho de los autobuses, es decir, de programar los horarios de salida de los autobuses. El área de Flota maneja todo lo relacionado a los autobuses y los choferes. El área Comercial es la encargada de la venta de los boletos. Por último, Inspectoría se ocupa de realizar inspecciones internas a los mismos choferes mientras realizan su trabajo.
- **Gerencia de Mantenimiento y Logística:** Conformada por el área de Mantenimiento, que se encarga de realizar el mantenimiento a los autobuses, y por el área de Logística y Almacén, encargada de manejar el ingreso y salida de productos y repuestos, necesarios para los mantenimientos de autobuses.

El proceso principal de negocio de una empresa de transporte de pasajeros es la venta de boletos. Este proceso se realiza de la siguiente manera: El pasajero sube al autobús desde uno de los paraderos autorizados y le informa al chofer su destino. De acuerdo al destino y tipo de pasajero (Adulto, universitario o escolar), el

chofer le indica el monto del pasaje al pasajero. El pasajero paga el monto y recibe su boleto de viaje.

Si bien el proceso principal es la venta de boletos, existen otros procesos internos que son muy importantes para las operaciones de la empresa. Entre ellos está la programación de los despachos de autobuses, el monitoreo del viaje de los autobuses y la liquidación de los boletos.

La programación de los despachos consiste en programar adecuadamente la flota de autobuses con sus respectivos choferes, teniendo en cuenta que cada autobús debe hacer 3 viajes completos al día. Además existen choferes que están en planilla y otros que trabajan por horas. De esta manera el Jefe de Flota realiza un cronograma de despachos cada mes que debe ser respetado.

Los autobuses siguen una ruta determinada hasta llegar al terminal destino, pero es necesario llevar un control para asegurarse que se está cumpliendo con el itinerario del viaje. El proceso de monitoreo del viaje consiste en registrar la hora en que cada bus pasa por un punto de control. Esto se realiza a través de GPS (Sistema de posicionamiento global) y el encargado del monitoreo puede hacer un seguimiento en tiempo real del autobús. De esta manera, es posible realizar un comparativo entre la hora programada y hora real para evitar posibles competencias entre los autobuses y asegurar que haya un intervalo equitativo entre cada autobús.

Cuando llega un autobús al terminal destino se inicia el proceso de liquidación de boletos. Este proceso consiste en hacer entrega del monto acumulado por la venta de los boletos durante el viaje, el cual debe coincidir con la cantidad de boletos entregados a los pasajeros.

Luego de describir algunos procesos del negocio, se puede afirmar que una forma de solucionar los problemas actuales de las empresas de transporte es mejorando sus procesos internos. Para el presente proyecto se busca mejorar el área de mantenimiento y logística. Esta área permite que los autobuses se encuentren en un óptimo estado y así ofrecer un servicio de calidad. Para ello el manejo de la información dentro de la empresa juega un papel determinante. Sin embargo, existen muchas empresas de transporte que operan de manera informal y ni

quiera cuentan con sistemas de información que le permitan un registro de sus datos.

Las soluciones basadas en la Inteligencia de Negocios permiten proveer información valiosa para la toma de decisiones. Entre los tipos de solución de Inteligencia de Negocios existentes en la actualidad, una de las más utilizadas es el Datamart. [KIM 2002] Un Datamart permite, por ejemplo, obtener cuál es el repuesto más solicitado, los productos que representan el 5% de los costos totales, el autobús que más falla, es decir, una diversidad de reportes que involucren una serie de variables e indicadores que permitirían encontrar oportunidades de mejora en el mantenimiento de los autobuses y almacén de los repuestos. Si bien las bases de datos relacionales también permiten obtener estos reportes, no permiten la flexibilidad de relacionar muchas variables e indicadores de manera intuitiva. De esta manera, estos reportes servirían para la toma de decisiones en el área de mantenimiento y logística del tipo de empresa que se hace mención.

El Datamart propuesto se convertirá en una herramienta que brindará información útil para la toma de decisiones en el área de mantenimiento y logística de una empresa de transporte público de pasajeros. Además, permitirá un fácil acceso a la información por parte de los usuarios de manera independiente y sin necesidad de conocimientos técnicos. Esto le permitirá a la empresa manejar adecuadamente su información para optimizar procesos internos, mejorar la calidad de los autobuses, prever posibles fallas o accidentes mediante un correcto mantenimiento preventivo, es decir, ofrecer un mejor servicio a los usuarios de Lima.

1.2. Marco conceptual del problema

A continuación se presentará algunas breves introducciones de conceptos que serán de utilidad para comprender mejor los objetivos del presente proyecto.

1.2.1. Introducción al Data Warehouse y Datamart

Los Data Warehouse nacen debido a la necesidad de contar con información útil de apoyo a la toma de decisiones, dado que los datos operacionales no cumplen con este objetivo. Un Data Warehouse es una colección de datos integrados, temáticos, no volátiles y variantes en el tiempo, organizados para soportar necesidades empresariales orientadas a la toma de decisiones. [INM 2002]

Se puede concluir, que un Data Warehouse, es el proceso de extraer datos de las operaciones diarias de la empresa, procedentes de diversos subsistemas, para integrarlos, sumarizarlos y almacenarlos en un depósito de datos, para poder acceder a ellos cada vez que el usuario lo requiera.

Basándose en la definición de Inmon, un Data Warehouse se caracteriza por ser:

- Integrado: Su característica más importante, pues posee la información integrada. Los datos almacenados en un Data Warehouse deben integrarse en una estructura consistente. Esta estructura permite tener la información en distintos niveles de detalle para adecuarse a las necesidades del usuario.
- Temático: Los datos se organizan por temas para facilitar su acceso y su entendimiento por parte de los usuarios finales.
- No volátil: La perspectiva estratégica que permite el análisis y la toma de decisiones requiere una base de datos estable, no cambiante.
- Variante en el tiempo: La información contenida en el Data Warehouse sirve para predecir tendencias. Por ello, esta se carga con los distintos valores que tiene una variable en el tiempo permitiendo comparaciones.

El acceso a toda la información por parte de los usuarios de la empresa no es conveniente, ya que muchas veces sólo necesitan un subconjunto de esta información. En estos casos utilizan los Datamarts. Según Inmon, el concepto de Datamart es una especialización de un Data Warehouse enfocado a un departamento o área específica dentro de una empresa, como por ejemplo, los departamentos de Finanzas o Recursos Humanos, permitiendo un mejor control de la información que se está abarcando. [INM 2002] Un Datamart permite acelerar las consultas reduciendo el volumen de datos a recorrer frente a un Data Warehouse. Inmon plantea modelar el Data Warehouse como primer paso para, a partir de este, crear uno o varios Datamarts según sea el caso.

Sin embargo, a diferencia de Inmon, Kimball plantea primero crear uno o varios Datamarts y el conjunto de ellos forma un Data Warehouse para la organización. Es decir, según Kimball un Datamart es una parte del Data Warehouse y no una especialización como plantea Inmon. Inmon y Kimball tienen planteamientos opuestos acerca de la definición de un Datamart, pero ambas definiciones son válidas.

1.2.2. Introducción a los sistemas ETL

Los sistemas ETL (de las siglas en inglés *Extraction, Transformation, Load*) se encargan de las funciones de extracción de distintas fuentes de datos, sean estas transaccionales o externas, transformación, realizando tareas de limpieza y consolidación de datos y la carga del Data Warehouse o Datamart.

Entre las principales funciones de los sistemas ETL tenemos [HER 2003]:

- La extracción de datos.
- El filtrado de datos.
- La carga inicial al Data Warehouse o Datamart.
- Refresco del Data Warehouse o Datamart: Operación periódica que actualiza los cambios de las fuentes externas al Data Warehouse o Datamart.

1.2.3. Introducción al Modelo Multidimensional

Un modelo de datos es un sistema formal y abstracto que permite describir los datos de acuerdo con reglas y convenios predefinidos. Es formal, pues los objetos del sistema se manipulan siguiendo reglas perfectamente definidas y utilizando exclusivamente los operadores definidos en el sistema, independientemente de lo que estos objetos y operadores puedan significar. [ULL 1999]

La arquitectura de Data Warehouse se basa en un modelo de datos llamado modelo multidimensional. Este modelo permite modelar base de datos simples y entendibles al usuario final, debido que presenta la información en un marco estándar e intuitivo que permite un acceso de alto rendimiento. Además, permite resolver con problemas planteados en sistemas transaccionales.

El modelo multidimensional está compuesto por dos componentes:

- El primer componente son las tablas que a su vez se dividen en dos tipos: Tablas de hechos y de dimensiones. Las tablas de hechos constituyen el objeto a analizar, poseen atributos de hechos que son del tipo cuantitativo cuyos valores se obtienen por aplicación de alguna función estadística que resumen un conjunto de valores en un único valor. Las tablas de dimensiones representan cada uno de los ejes en un espacio

multidimensional. Sus atributos son del tipo cualitativo que proporcionan el contexto en el que se obtienen las medidas en un esquema de hecho. Las dimensiones poseen jerarquías, que son varios atributos unidos mediante una relación del tipo jerárquico.

- El segundo componente son los esquemas. Existen dos esquemas también: Esquema estrella y de copo de nieve o constelación de hechos. El esquema estrella forma un diagrama en forma de estrella teniendo en el centro de la estrella una o más tablas de hechos y las puntas de las estrellas a las tablas de dimensiones. En el caso del esquema de copo de nieve, las tablas de dimensiones se encuentran normalizadas, es decir, cada tabla dimensional sólo contiene el nivel que es la clave primaria en la tabla y la llave foránea de su parentesco del nivel más cercano. [DIA 2002]

En la Figura 2 se muestra el esquema estrella y en la Figura 3 el esquema de constelación de hechos.

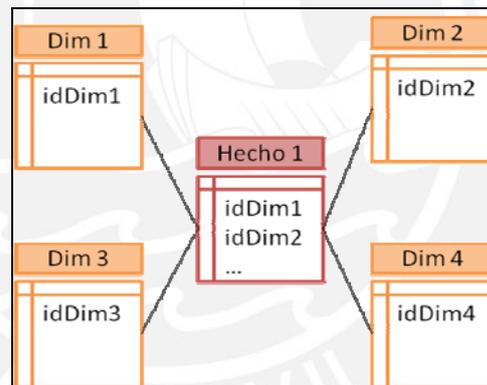


Figura 2: Esquema Estrella

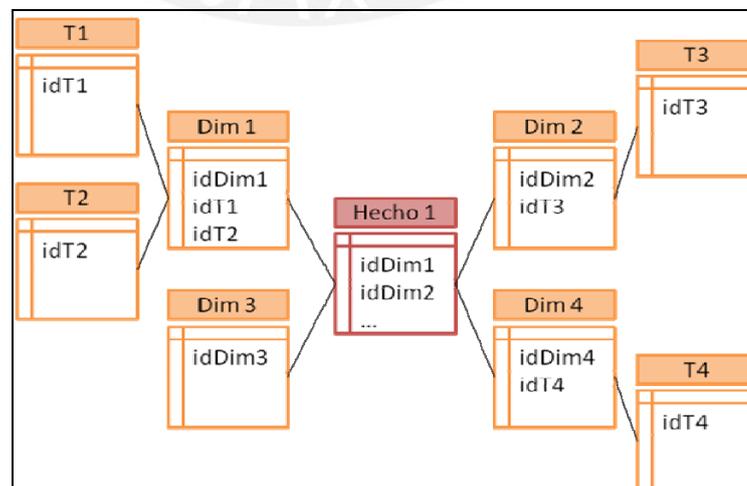


Figura 3: Esquema Constelación de Hechos

La estructura básica del modelo multidimensional se conoce como Cubo o Hipercubo, ya que la información se representa como una matriz multidimensional, en los ejes de esta matriz se encuentran los criterios de análisis y en los cruces están los valores a analizar.

Los Cubos están formados por:

- Dimensiones: Representan los criterios de análisis de los datos. Si una dimensión tiene más de un nivel entonces los miembros de la dimensión pueden ser organizados en una o más jerarquías.
- Medida: Dato numérico que representa una actividad específica de un negocio, mientras que una dimensión representa una perspectiva de los datos. Una medida contiene una propiedad numérica y fórmula.

En la Figura 4 se muestra la estructura de un Cubo. [YDI 2004]

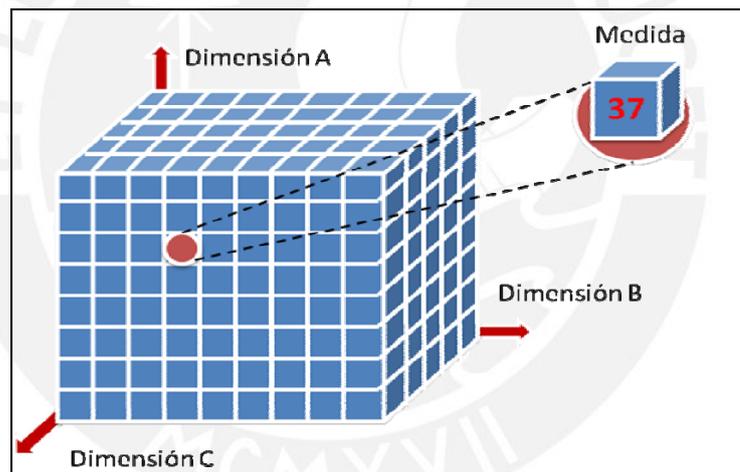


Figura 4: Estructura de un Cubo

Existen algunas operaciones que se realizan en el modelo multidimensional, a continuación se mencionarán las principales:

- Slice and Dice: Hacer una selección de valores de las dimensiones que queremos analizar.
- Rotación: Seleccionar el orden de visualizar las dimensiones.
- Movimientos en la jerarquía de una dimensión (*Drill Up* y *Drill Down*): Subir o bajar a los niveles más atómicos del esquema multidimensional.
- Consolidación: Realizar cálculos a las medidas en función de agrupamientos. Esta operación puede ser de suma, promedio, etc.
- Drill Across: Relacionar dos cubos.

- Drill Through: Acceder a los datos descriptivos del modelo.

1.2.4. Introducción a la Inteligencia de Negocios y a los Sistemas de Información Ejecutiva

La Inteligencia de Negocios (en inglés *Business Intelligence*) nació como un concepto que se asociaba totalmente con los niveles de los altos directivos ante la necesidad de contar con información para dirigir el rumbo de la empresa. Sin embargo, con el tiempo este alcance se ha ido ampliando hasta llegar a los niveles más bajos de la empresa. La Inteligencia de Negocios se define como un conjunto de tecnologías de colección de datos y manejo de información, que implementa soluciones orientadas al usuario final para apoyar en la toma de decisiones, aprovechando la información disponible en cualquier parte de la organización.

Para la colección de datos se construye almacenes de datos, los cuales no son más que los Data Warehouses o Datamarts. Entre las principales técnicas de análisis de la información están los Sistemas de Información Ejecutiva (EIS).

Un EIS es un software que muestra información ejecutiva de las diferentes áreas de la organización en un sólo sistema. Se caracteriza por proveer toda la información necesaria para la toma de decisiones de modo fácil y con la mínima interacción con el sistema. Las interfaces en este sistema deben ser más sofisticadas y deben incluir, en la menor cantidad de páginas, la mayor cantidad de información relevante que el usuario necesita para el monitoreo de la empresa. Generalmente, los EIS obtienen todos sus datos a través de los Cubos, estos a su vez del Data Warehouse de la organización.

El EIS tiene una serie de características, entre las principales están los Tableros de Control. Un Tablero de Control es una herramienta en la cual el usuario puede monitorear a la empresa a través de indicadores de cualquier tipo. Su especialización ha tomado camino hacia los Cuadros de Mando, pues son una poderosa herramienta para la dirección de la organización.

1.2.5. Conceptos asociados al negocio de una empresa de transporte público de pasajeros

Además de los conceptos asociados a la solución del problema es importante aclarar algunas definiciones asociadas al negocio para un mejor entendimiento

de la solución y del proyecto en sí. A continuación algunas definiciones asociadas al área de mantenimiento y logística de este tipo de empresas.

- Autobús: Medio de transporte público empleado para el transporte de personas.
- Almacén: Lugar donde se almacenan los ítems de la empresa. Para el caso de una empresa de transporte, se almacena los repuestos de las partes del autobús.
- Concepto: Categoría por la que es atendido un autobús como parte de su mantenimiento. Por ejemplo, el concepto puede ser: “Cambio de aceite”, “Reparación de frenos”, entre otros.
- Encargado del almacén: Persona responsable del almacén quien entrega los ítems a las personas que lo soliciten. Además se encarga de recibir correctamente los ítems que ingresan al almacén.
- Ítem: Generalmente está referido a los productos del almacén. Sin embargo, en algunos casos pueden ser servicios.
- Mecánico: Persona encargada de realizar el mantenimiento sobre los autobuses.
- Orden de Movimiento: Documento que contiene un listado de todos los ítems que deben ingresar o salir del almacén.
- Orden de Trabajo: Documento que indica una atención sobre un autobús, es decir, un mantenimiento. Este documento contiene uno o más Órdenes de Movimiento.
- Socio Estratégico: Persona jurídica que tiene un vínculo estratégico con la empresa. El socio será un proveedor si le vende ítems a la empresa o será cliente, si le compra ítems.
- Transportista: Persona que conduce el autobús.
- Transacción de Inventario: Transacción realizada en un almacén. Esta transacción puede ser: Ingreso, Salida, Cierre, entre otros.
- Vendedor: Persona encargada de realizar la venta de ítems a los clientes.

Adicionalmente, es importante conocer los procesos principales del área del negocio. Por ello, a continuación se presentarán los procesos de mantenimiento de autobuses y del movimiento en almacén.

Mantenimiento de autobuses: El supervisor del área de mantenimiento y logística revisa diariamente los mantenimientos preventivos o correctivos a realizar y genera una Orden de Trabajo para cada atención donde se asigna al

mecánico responsable del mantenimiento. El mecánico recibe la Orden de Trabajo y se dirige al almacén para solicitar los repuestos necesarios. Luego procede a realizar el mantenimiento sobre el autobús. Finalmente, el supervisor da el visto bueno del mantenimiento.

Movimiento en almacén: El mecánico llega al almacén y solicita los repuestos según la Orden de Movimiento que forma parte de la Orden de Trabajo. El encargado del almacén busca y entrega los repuestos al mecánico. Finalmente, el encargado actualiza el stock de los ítems en el sistema.

Finalmente, en el Anexo 5 están los Diagramas de Actividades del Negocio que muestran gráficamente los pasos a seguir en los procesos descritos anteriormente.

1.3. Plan del proyecto

En esta sección se presenta la planificación para el desarrollo de la solución. Primero se definirá los métodos y procedimientos a utilizar tanto para el desarrollo del proyecto como de la solución. Luego se explicará el plan de proyecto a realizar.

1.3.1. Métodos y Procedimientos

En esta sección se describirán brevemente las metodologías y procedimientos usados a lo largo del proyecto.

1.3.1.1. Métodos y Procedimientos en la gestión del proyecto

Para la gestión del proyecto se empleará una metodología basada en la guía del PMBOK (de las siglas en inglés *Project Management Body of Knowledge*). El PMBOK propone las mejores prácticas para la gestión de un proyecto, en este caso un proyecto informático. En el PMBOK se menciona que todo proyecto debe tener en cuenta nueve áreas de conocimiento para una buena gestión, las cuales son: Integración, Alcance, Tiempo, Costes, Calidad, Recursos Humanos, Comunicaciones, Riesgos y Adquisiciones. Estas áreas están a su vez agrupadas en áreas centrales y facilitadoras. Las áreas centrales corresponden a las 4 primeras áreas mencionadas excepto por la de Integración, el resto corresponden a las áreas facilitadoras. El presente proyecto sólo tomará las siguientes áreas de conocimiento: Alcance, Tiempo, Costes, Calidad y Riesgos, al ser consideradas las áreas más importantes. Las

demás áreas no aplican a un proyecto de tesis en donde sólo se tiene un recurso humano. A continuación se presenta una breve descripción de los métodos a utilizar por cada área. [PMI 2004]

En el área de Alcance, se definirá el alcance del proyecto basándose en la metodología de Estructura de Desglose de Tareas (en inglés conocido como *Work Breakdown Structure* - WBS). Ésta permite mostrar en un gráfico todos los entregables del proyecto, permitiendo una definición clara del alcance.

En el área de Tiempo, se definirá la secuencia de las actividades a realizar, así como la estimación de las mismas. En la estimación de tiempo se empleará la metodología PERT, el cual plantea para cada actividad un escenario de tiempo probable, optimista y pesimista, dando como tiempo estimado la combinación de estos tiempos. Finalmente, se expresará la secuencia y dependencia de actividades a través del diagrama Gantt.

En el área de Costes, se asignarán a las actividades definidas anteriormente un costo. Además, se le asignarán costos también a los recursos para poder obtener un costo final de todo el proyecto.

En el área de Calidad, se definirán los posibles requerimientos de calidad del proyecto teniendo como base lo que propone la guía del PMBOK.

En el área de Riesgos, se identificarán y clasificarán los posibles riesgos realizando un análisis cualitativo y cuantitativo del impacto que producen. Además, se elaborará un plan en respuesta a los riesgos como contingencia.

1.3.1.2. Métodos y Procedimientos en el desarrollo del Datamart

Para la gestión del desarrollo del Datamart se basará en la metodología DWEP (de sus siglas en inglés *Data Warehouse Engineering Process*), propuesto por Sergio Lujan-Mora y Juan Trujillo, la cual es una adaptación de la metodología RUP (*Rational Unified Process*) y de la herramienta UML (*Unified Modeling Language*) para el desarrollo de Data Warehouse o Datamart. En este caso se aplicará dicha metodología para el desarrollo de un Datamart. La metodología DWEP al igual que RUP divide el desarrollo en 4 fases: Concepción, Elaboración, Construcción y Transición. Sin embargo, el DWEP presenta algunas variantes que se adapta al desarrollo de un Datamart. [LUJ 2006]

En la fase de Concepción se abarca desde la captura de las principales necesidades de los usuarios finales, poniéndose énfasis en los requerimientos funcionales y termina con la elaboración del Documento de Análisis de Requerimientos.

En la fase de Elaboración se empieza desde el plan de proyecto, el cual contiene la secuencia de actividades a realizar. Además, se define la arquitectura del Datamart y acaba con un esquema lógico del Datamart.

En la fase de Construcción, se implementa el Datamart hasta obtener la primera versión operativa con datos reales. Además se desarrollan los procesos ETL necesarios para la carga de datos.

Por último, en la fase de Transición, se pone énfasis en la detección de errores y empieza cuando el Datamart entra en producción. Para el presente proyecto esta fase no se llevará a cabo pues el alcance del proyecto termina con la construcción del Datamart, es decir, en la fase de Construcción.

1.3.2. Planificación

En esta sección se detallará la planificación para llevar a cabo el proyecto empleando las metodologías y procedimientos descritos anteriormente.

1.3.2.1. Planificación del Alcance

Antes de definir el alcance del proyecto se debe definir el alcance del producto, es decir, del Datamart. El Datamart a desarrollar está orientado al uso del área de mantenimiento y logística de una empresa de transporte público de pasajeros. Con este modelo de Datamart se busca abarcar los posibles escenarios del área de una empresa de este tipo.

A grandes rasgos, el Datamart podrá realizar un análisis de:

- **Mantenimiento:** Permitirá el análisis del mantenimiento preventivo y correctivo de los autobuses a través de diversas variables como tiempo (fechas, estaciones, meses, años, etc.), autobús, personal, tipos de mantenimiento, etc.

- **Logística:** Permitirá analizar la eficiencia de la salida y entrada de repuestos del almacén para el mantenimiento de los autobuses. Además, de un control de las compras y ventas que se realizan en el almacén.

En resumen, el Datamart no sólo soportará un análisis de logística, sino también de mantenimiento.

En la definición del alcance del proyecto, este se ha dividido en 6 fases, las cuales tres corresponden a las fases propuestas por la metodología DWEP excepto la fase de Transición debido a que esta fase no aplica al presente proyecto. De las tres restantes, la primera fase está dedicada a la gestión misma del proyecto. La segunda fase corresponde la fase Preliminar, en la cual se definió el tema de tesis y se elaboró el plan de tesis. La tercera y última fase corresponde a una fase post cierre del proyecto en donde se establecen las conclusiones del proyecto. En la Figura 5 se muestra el diagrama de WBS con las fases mencionadas.

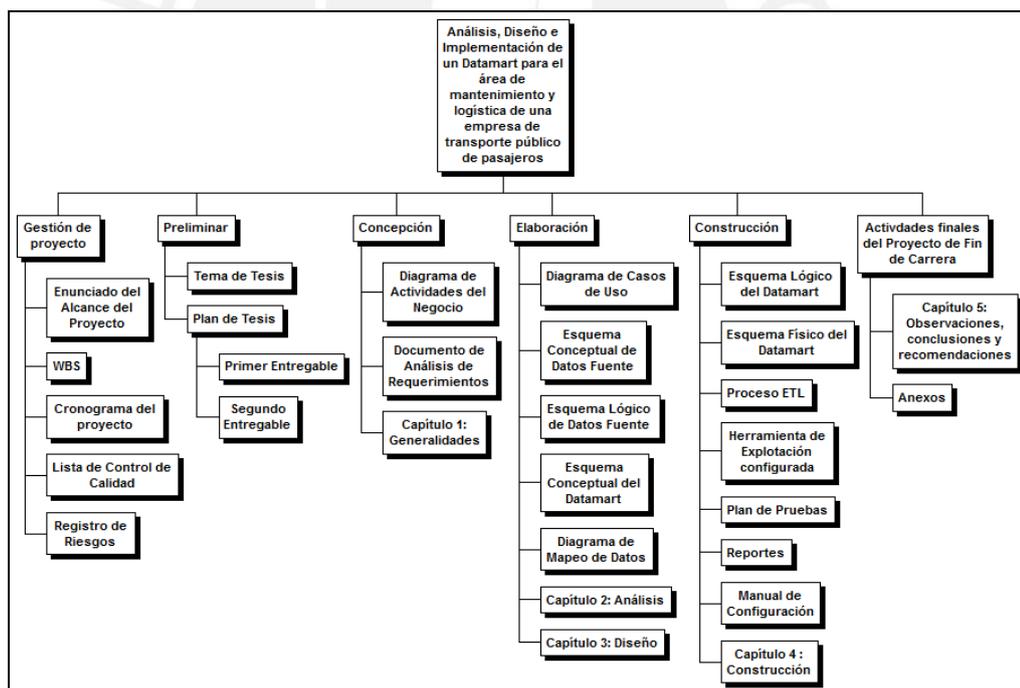


Figura 5: WBS

Para cada fase se ha determinado un entregable final que cierra dicha fase. Los entregables son:

- **Concepción:** Documento de Análisis de Requerimientos.
- **Elaboración:** Modelo multidimensional (lógico y físico) del Datamart.

- **Construcción:** Herramienta de explotación configurada para el acceso al Datamart.

Finalmente, en el Anexo 1 se presenta el Enunciado del Alcance del Proyecto el cual detalla el alcance a nivel de proyecto.

1.3.2.2. Planificación del Tiempo

Tomando como base el diagrama de WBS, se especificó la secuencia de actividades a realizar. Para la estimación de los tiempos por cada actividad se determinó con el método PERT. En la Figura 6 se muestra el diagrama Gantt mostrando sólo las fases que comprende el proyecto. El detalle las actividades agrupadas por fase, así como la duración de las mismas puede observarse en el Anexo 2. Como se observa, el proyecto se inició el 1 de febrero y está programado para que culmine el 30 de mayo del presente año. Se asumió que se trabaja 8 horas al día, 5 días a la semana, dando como resultado un total de 680 horas (85 días) de trabajo. Estas horas incluyen una fase Preliminar en donde se elaboró el Plan de Tesis.

EDT	Nombre de tarea	Duración	Comienzo	Fin
1	Análisis, Diseño e Implementación del Datamart	85 días	mar 01/02/11	lun 30/05/11
1.1	Gestión de proyecto	7 días	mié 23/02/11	jue 03/03/11
1.2	Preliminar	16 días	mar 01/02/11	mar 22/02/11
1.3	Concepción	10 días	vie 04/03/11	jue 17/03/11
1.4	Elaboración	19 días	vie 18/03/11	mié 13/04/11
1.5	Construcción	29 días	jue 14/04/11	mar 24/05/11
1.6	Actividades finales del Proyecto de Fin de Carrera	4 días	mié 25/05/11	lun 30/05/11

Figura 6: Diagrama Gantt por fases

1.3.2.3. Planificación de Costos

Sólo se cuenta con un recurso humano, el tesista. Este se encargará de realizar todas las actividades que figuran en el diagrama Gantt.

El costo del proyecto, teniendo en cuenta la cantidad de horas trabajadas, es:

$$1 \text{ hombre} * 680 \text{ horas/hombre} * 15 \text{ Nuevos Soles/hora} = \text{S/. } 10'200$$

El costo total del proyecto es de 10'200 Nuevos Soles. Este costo es un costo estimado y el costo real podrá obtenerse al desarrollar el presente proyecto.

1.3.2.4. Planificación de Calidad

Se debe planificar métricas de calidad que permitan el control tanto del proyecto como del producto. Para el caso del proyecto se utilizará el WBS y el diagrama de Gantt para saber si se están cumpliendo con los entregables propuestos y sobre todo en las fechas indicadas. Para el caso del producto, el Documento de Análisis de Requerimientos será el documento guía para verificar la calidad del producto.

Finalmente, se elaborará una Lista de Control de Calidad. En ella se listará los criterios de calidad que debe cumplir el producto para que sea considerado de calidad. Sólo cuando cumpla con todos los requisitos listados en la lista, se habrá verificado la calidad total del producto. La Lista de Control de Calidad se presenta en el Anexo 3.

1.3.2.5. Planificación de Riesgos

Se identifican los riesgos del proyecto para así tomar las acciones debidas frente a un riesgo ocurrido. Los riesgos pueden ser clasificados para poder identificarlos más rápidamente. Para ello se usa la Estructura de Desglose del Riesgo (RBS), éste clasifica las categorías en donde aparecen los riesgos típicos de un proyecto. En la Figura 7 se muestra el RBS para el presente proyecto.

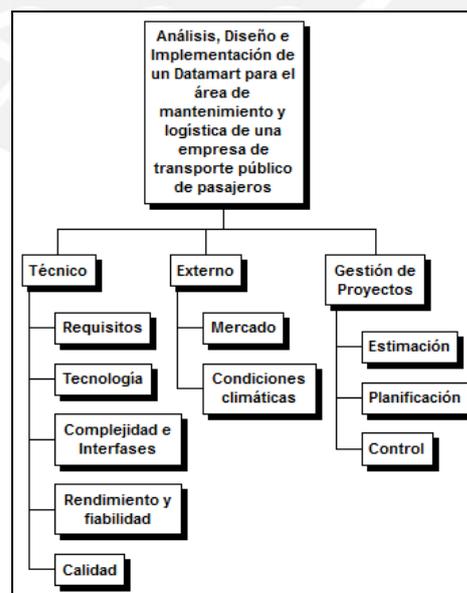


Figura 7: RBS

Como se puede observar en la Figura 7, los riesgos pueden ser divididos en tres categorías.

- **Técnico:** Estos riesgos son propios del producto. Los riesgos más críticos son los relacionados a los requisitos, ya que si no se definen adecuadamente los requisitos, el producto final no cumplirá con las expectativas y el resultado será un producto diferente al que se planteó en un primer momento. Existen otros riesgos como por ejemplo, el no contar con la tecnología para llevar a cabo el proyecto.
- **Externo:** Estos riesgos son inevitables, pues son causados por factores externos. Se deben plantear planes de contingencias ante algún posible riesgo de ese tipo.
- **Gestión de Proyectos:** Estos riesgos surgen ante la inadecuada gestión del proyecto. Los riesgos posibles son una mala estimación de tiempos en las actividades, mala planificación, un control inadecuado de las actividades, etc. Se debe tener mucho cuidado para evitarlos, ya que de ocurrir afectan directamente al ciclo de vida del proyecto.

En el Anexo 4 se presenta el Registro de Riesgos, el cual identifica a los riesgos basados en el RBS y los clasifica según su impacto y probabilidad de ocurrencia. Además, para cada riesgo se define un plan para mitigarlo.

1.4. Estado del arte

En esta sección primero se describirá las soluciones generales que propone la Inteligencia de Negocios para resolver los problemas en la toma de decisiones y luego se describirá qué soluciones específicas orientadas a las empresas de transporte público de pasajeros existen. Finalmente, se enumerará las herramientas que existen actualmente en el mercado.

1.4.1. Soluciones generales

La Inteligencia de Negocios (BI) plantea una serie de soluciones para proporcionar información útil. Una metodología usada por las grandes corporaciones es la propuesta por Inmon. Esta metodología consiste en implementar primero un Data Warehouse, ya que este constituye el repositorio para almacenar toda la información que posteriormente será usada por otras herramientas BI. La carga de los datos al Data Warehouse se realiza mediante los sistemas ETL. Para el análisis de los datos se usa la tecnología OLAP (de sus siglas en inglés *Online Analytical Processing*). Esta permite un uso eficaz de los Data Warehouses, pues proporciona respuestas rápidas a consultas analíticas complejas e

iterativas. Sin embargo, lo más resaltante es que presenta los datos al usuario a través de un modelo de datos intuitivo y natural que permite una fácil navegación. El OLAP plantea la creación de uno o varios Datamarts y/o Cubos.

Finalmente, una vez creado el Cubo o Datamart, se usa alguna herramienta de explotación que permita crear reportes o tableros de control. A continuación una breve descripción de las herramientas más usadas por las empresas.

- **Tableros de control:** Aplicaciones dirigidas a un perfil de usuario alto, no tecnológico. Muestra la información en forma de indicadores del negocio y conceptos de información de las áreas usuarias en función de las dimensiones de negocio.
- **Informes:** Permiten a los usuarios de más bajo nivel la posibilidad de crear reportes personalizados para su uso o para usuarios menos avanzados.
- **Data Mining:** Aplicación, que basado en el Data Warehouse, permite obtener patrones de comportamiento entre determinados conceptos. Es útil para hallar tendencias y realizar proyecciones. Estas aplicaciones usan diversas técnicas en su proceso entre ellas tenemos: la estadística clásica, modelos basados en árboles de decisiones, redes neuronales, etc.

En la Figura 8 se muestra el flujo que propone una solución de BI.

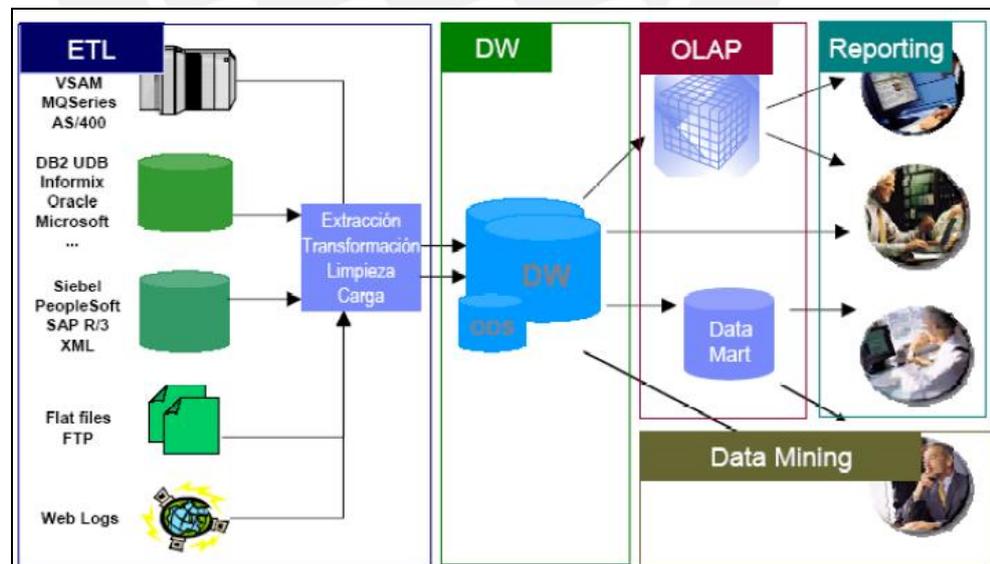


Figura 8: Estructura de Solución BI común

Existe otra metodología propuesta por Kimball. Esta consiste en implementar primero el Datamart sin necesidad de crear primero un Data Warehouse. Una vez creado el Datamart se elaboran los reportes y tableros de control necesarios. A través de la construcción de varios Datamart se va

implementando el Data Warehouse organizacional. Esta solución está siendo utilizada por muchas empresas, ya que es mucho más simple y menos costoso hacer un Datamart en vez de un Data Warehouse. Además, muchas empresas no requieren crear un repositorio con toda la información organizacional. Para el presente proyecto se empleará la metodología de Kimball, pues se creará el Datamart sin necesidad de haberse creado antes un Data Warehouse.

Por último, existe una metodología que plantea crear Datamarts independientes que no guarden relación alguna entre ellos. Esta metodología es la menos usada, ya que requiere un esfuerzo adicional en crear un ETL por cada Datamart. En los casos anteriores sólo se creaba un sistema ETL para la carga de datos.

En la Figura 9 se muestra la metodología que plantea Inmon, en la Figura 10 la propuesta por Kimball y en la Figura 11 la metodología de Datamarts independientes. [ACM 2006]

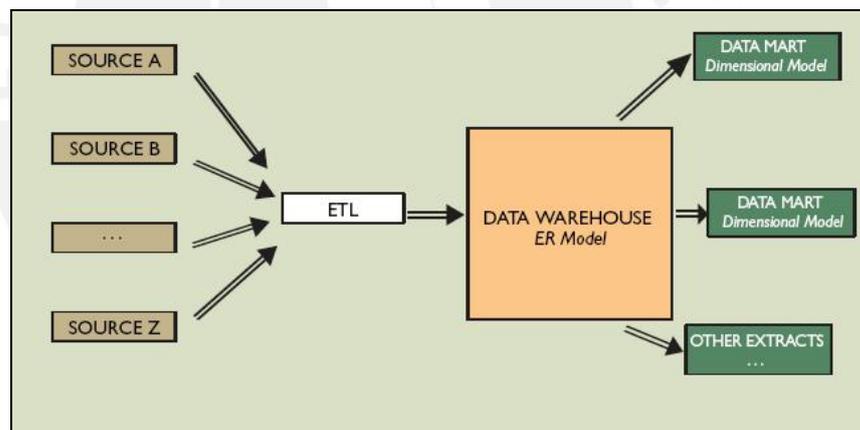


Figura 9: Metodología propuesta por Inmon

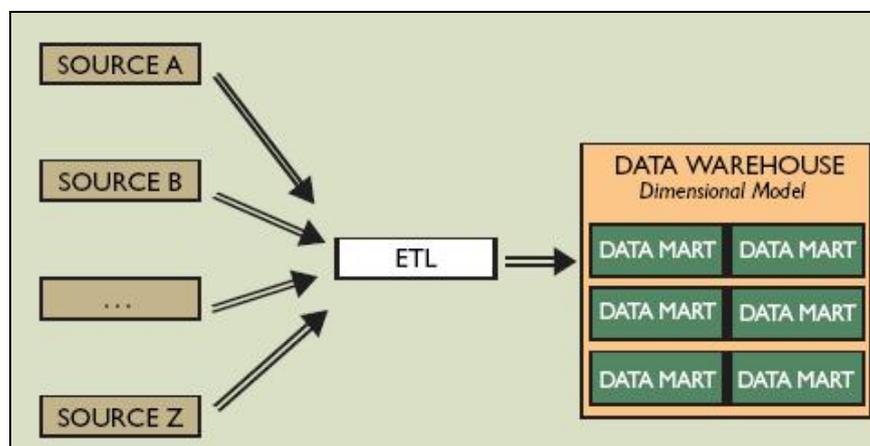


Figura 10: Metodología propuesta por Kimball

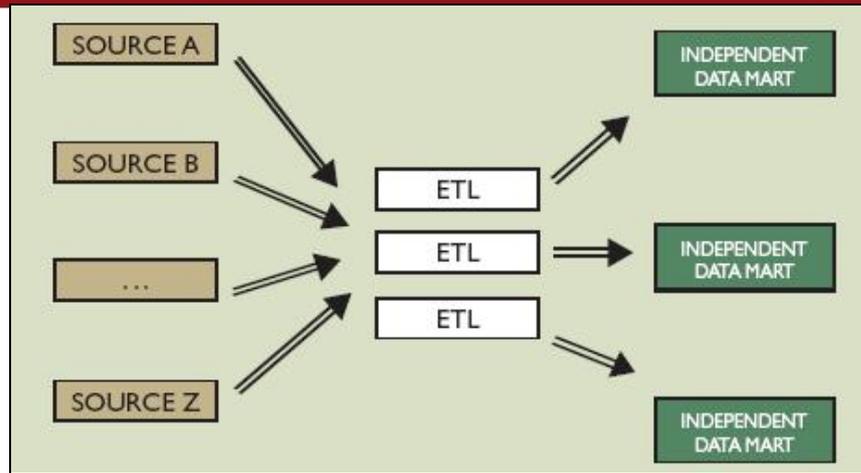


Figura 11: Metodología de Datamarts independientes

En la Tabla 1 se muestra una comparación de las características principales de las metodologías mencionadas y se observa que la metodología de Kimball cumple con la mayoría de las características presentadas.

Características	Metodología propuesta por Inmon	Metodología propuesta por Kimball	Metodología de Datamarts independientes
Rapidez en crear Datamart.	✗	✓	✓
Facilidad en el modelamiento multidimensional.	✗	✓	✓
Elaboración de un solo ETL.	✓	✓	✗
Facilidad en crear varios Datamarts.	✓	✗	✗
Dependencia entre los Datamarts.	✓	✓	✗

Tabla 1: Cuadro comparativo de metodologías

1.4.2. Soluciones específicas

Las soluciones planteadas para las empresas de transporte público de pasajeros son escasas, pues más se conocen sistemas transaccionales, a los cuales se le añade un módulo de reportes que deben ir modificando de acuerdo a las exigencias del usuario final. Esta tarea es laboriosa, pues para cada nuevo requerimiento se necesita modificar código fuente y este proceso está sujeto a errores. Además, el usuario depende del área desarrolladora del sistema, pues no es capaz de crear sus propios reportes. Por otro lado, se sabe que las empresas implementan sus propios Data Warehouses, es decir, crean la solución desde cero, pues no existen modelos estandarizados de Data Warehouse para empresas de este tipo. Se conoce muchos proyectos de empresas consultoras que desarrollan Data Warehouse para sus clientes, en ellas empresas de transporte público, demostrando que la solución de

Inteligencia de Negocios es la más utilizada actualmente. Sin embargo, las aplicaciones utilizadas o desarrolladas en estos proyectos generalmente no se conocen, al ser consideradas propiedad intelectual de las empresas desarrolladoras.

Se encontraron tres aplicaciones que solucionan, en parte, el problema planteado:

- **JR SOFTWARE Transporte de Pasajeros:** Es un software dedicado a las empresas que cuentan con una flota ya sea de colectivos, aviones, etc. Está formado por varios módulos: Recorridos y Ventas, Proveedores, Compras, Cuenta Corriente, Caja Diaria, entre otros. No se menciona un módulo dedicado a los reportes, pero se asume que cada módulo cuenta con algunos reportes básicos. Sin embargo, no permite la creación de nuevos reportes. Es un claro ejemplo de un sistema transaccional con reportes, considerada como la solución más común y rápida si ya se cuenta con un sistema transaccional. [JRS 2011]
- **MGX ERP:** ERP que abarca los procesos de una empresa de transporte de pasajeros. Contiene varios módulos llamados “Soluciones” y cada módulo presenta varios reportes. Sin embargo, los reportes están más orientados a los procesos del día a día como por ejemplo: Planilla de viajes por día, Consumo totalizado diario de combustible por vehículo, Listado de vehículos por ruta, entre otros. Si bien el ERP si abarca el área mantenimiento de una empresa de transporte público de pasajeros, no permite analizar información para tomar una decisión estratégica. [BIE 2010]
- **TransMTO:** Sistema de gestión de mantenimiento vehicular para el transporte urbano desarrollado por la empresa peruana MRVisual Corp. SAC. Adicionalmente, este sistema se integra con otros que han sido desarrollados por la misma empresa como es TransRECAUDO, TransGPS, TransOPERACION, entre otros. Cuenta con diversas funcionalidades como la gestión de inspecciones, registro de mantenimiento correctivo y preventivo, teniendo la posibilidad de generar reportes que pueden ser exportador a Microsoft Office Excel. Sin embargo, estos reportes no son configurables y si se requieren nuevos parámetros o alguna modificación en

el sistema es necesario contactarse con el área de soporte de MRVisual.
[MRV 2010]

En la Tabla 2 se muestra una comparación de las características de los sistemas encontrados con la solución de BI.

Características	JR SOFTWARE	MGX ERP	TransMTO	Datamart
El sistema genera reportes.	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
El usuario puede seleccionar criterios a los reportes.	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
El usuario puede crear nuevos reportes.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
El sistema contiene tableros de control.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
El sistema contiene gráficos.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
El sistema está orientado al área de mantenimiento y logística de una empresa de transporte público de pasajeros.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Tabla 2: Cuadro comparativo de sistemas para empresas de transporte público

Se concluye que no existe actualmente, o no es conocida, una herramienta dedicada exclusivamente para la toma de decisiones en el área de mantenimiento y logística de una empresa de transporte público de pasajeros.

1.5. Descripción y sustentación de la solución

En esta sección se detallará la solución a desarrollar y posteriormente se justificará la elección de la solución planteada.

1.5.1. Descripción de la solución

La solución a desarrollar en el presente proyecto consiste en la creación de un Datamart para el área de mantenimiento y logística de una empresa de transporte público de pasajeros a fin de que se convierta en una herramienta útil que ayude a los usuarios de esta área en la toma de decisiones.

Sin embargo, la solución no sólo consiste en el modelo multidimensional del Datamart. Esta solución abarca varios puntos que a continuación se describirán detalladamente. La integración de todos estos puntos forma la solución integral al problema.

1.5.1.1. Funcionalidad

El Datamart debe permitir cumplir con las necesidades de información requeridas, es decir, debe contener la funcionalidad adecuada. Sin ella, el Datamart no cumple con su objetivo principal y el proyecto no logra solucionar el problema planteado. Por ello, el Datamart debe estar orientado a satisfacer las necesidades del área de mantenimiento y logística de una empresa de transporte público de pasajeros. Sin embargo, existen varios tipos de empresa de este rubro y el Datamart debe ser flexible a estos escenarios de las empresas reales.

1.5.1.2. Modelo multidimensional

Constituye el punto fundamental de la solución debido a que el modelo multidimensional es la base del Datamart. El modelo multidimensional a desarrollar será del tipo estrella y tendrá las dimensiones y hechos necesarios para abarcar los posibles escenarios y supuestos de una empresa de transporte público. Para llevar a cabo esta labor se habrá tenido que levantar información y definido claramente los requerimientos de los usuarios del área de mantenimiento y logística de una empresa de transporte público. Sólo así se puede pasar a la construcción del Datamart. En la construcción del Datamart, primero se elaborará un modelo lógico para ver las relaciones e interacciones entre las tablas. Luego se elaborará el modelo físico en el cual se definirá los tipos de datos y tamaños más adecuados para los campos de las tablas.

1.5.1.3. Proceso ETL

Un Datamart es una base de datos departamental y como tal requiere de información. Por ello es necesario algún procedimiento que permita cargarle de datos válidos. La solución incluye un proceso ETL el cual extraerá la información de una o más fuentes de datos, transformará la estructura de datos a fin de que coincida con la estructura del modelo del Datamart y realizará la carga de datos a éste. La fuente de datos puede ser una base de datos o archivos. De esta manera, por ejemplo, habrá una tabla con los datos del autobús, otra tabla con los datos del producto, etc. El proceso ETL extraerá la información de estas tablas y se realizará las transformaciones necesarias para cargar los datos al motor de base de datos escogido.

1.5.1.4. Herramienta de Explotación

Si bien el modelo multidimensional constituía en gran parte la solución, es la herramienta de explotación finalmente con la que el usuario puede interactuar y ver el Datamart. La herramienta de explotación es un sistema que recibe una base de datos de modelo multidimensional y permite visualizarla de una manera fácil e intuitiva. La solución no incluye la implementación de una herramienta de explotación. Existen diversas herramientas en el mercado y se seleccionará una de ellas. La herramienta seleccionada se configurará para que acepte al modelo multidimensional construido. Además, se personalizará la herramienta para una mejor interacción con el usuario final.

1.5.1.5. Reportes y tableros de control

Con la herramienta de explotación configurada y con la información cargada el usuario ya podría empezar a cruzar las diversas dimensiones y crear sus propios reportes. Sin embargo, como parte de la solución propuesta se crearán algunos reportes que se consideran útiles para la toma de decisiones. Entre los reportes a elaborar están: reporte de ingreso y salida de repuestos, reporte del mantenimiento preventivo y correctivo por autobuses. Además, se crearán tableros de control orientados a usuarios de más alto nivel. Estos tableros de control mostrarán una serie de indicadores y gráficos que permitirán llevar a cabo una mejor gestión del área.

1.5.2. Sustentación de la solución

La solución planteada es la más adecuada debido a dos razones principales: es una solución de Inteligencia de Negocios y está basado en la creación de un Datamart. A continuación se explicará por qué estas dos razones hacen que la solución propuesta sea considerada como la más adecuada y ventajosa. Adicionalmente se realizará un pequeño análisis de costo y beneficio.

1.5.2.1. Solución de Inteligencia de Negocios

Al ser una solución de Inteligencia de Negocios, el Datamart propuesto permite generar conocimiento útil a partir de una gran cantidad de información organizacional. Las soluciones de este tipo permiten administrar adecuadamente la información para utilizar sólo la información que los usuarios requieran al momento deseado. Los tiempos de respuesta a la información deseada son considerablemente superiores a los sistemas transaccionales.

Esta información presentada puede ser resumida o detallada según la necesidad del usuario.

Además, ofrece la capacidad de crecimiento de la información a medida que la organización realiza sus operaciones del día a día y permite actualizar la información a través de los procesos ETL. De esta manera se asegura que siempre se cuente con la información actualizada y real de la empresa. Por último, es relativamente fácil agregar alguna dimensión al Data Warehouse o Datamart. En conclusión, el Datamart, con el uso correcto, podrá ser una herramienta de soporte en la toma de decisiones.

1.5.2.2. Creación de Datamart

Se eligió crear un Datamart frente a un Data Warehouse debido a que el primero ofrece una serie de ventajas. Un Datamart es más fácil de implementar e instalar que un Data Warehouse corporativo. De esto se concluye que el costo de construcción de un Datamart es considerablemente menor al del Data Warehouse. Así muchas empresas pueden tener acceso al Datamart debido a que está dentro de sus posibilidades económicas.

Por otro lado, los Datamarts, al ser más pequeños, están enfocados a satisfacer las necesidades de un grupo de usuarios en concreto y permite que el tiempo de respuesta a sus consultas sea más rápido. Además, a diferencia del Data Warehouse, la captura de requerimientos se realiza mucho más rápido y concentra su atención en el negocio del área de la organización deseada. En cambio, los Data Warehouse al tener la información de toda la organización dificulta al usuario al momento de acceder a la información requerida.

En la Tabla 3 se muestra la comparación entre las características principales de un Data Warehouse y Datamart. Se observa que las características del Datamart están más orientadas a solucionar el problema en la toma de decisiones del área comercial de una empresa.

Data Warehouse	Datamart
Construido para satisfacer las necesidades de información de una empresa.	Construido para satisfacer las necesidades de una función o unidad comercial específica.
Diseñado para optimizar la integración y la administración de los datos fuente.	Diseñado para optimizar la entrega de información de soporte a decisiones.
Administra grandes cantidades de información histórica.	Primordialmente se concentra en administrar resúmenes y datos de

	muestreo.
Pertenece a, y se administra por, las organizaciones de Sistemas de Información de la empresa.	Puede ser propiedad, y administrado por, el grupo de Sistemas de Información en la Línea de Negocio.

Tabla 3: Cuadro comparativo entre Data Warehouse y Datamart

1.5.2.3. Análisis Costo-Beneficio

Se realizará un pequeño análisis costo-beneficio para una empresa de transporte público al implementar una solución de Datamart a su área comercial.

Análisis de costos:

- Inversión inicial en compra de servidor.
- Capacitación de los usuarios finales del área en el uso de la herramienta de explotación.
- Costo de soporte y mantenimiento del Datamart.

Análisis de beneficios:

- No requiere altos niveles de conocimiento para implantar la solución.
- Se empleará la herramienta libre Pentaho por lo que no requiere gasto en licencias.
- El tiempo de dedicación en la implantación de la solución es de poca duración y no interfiere con las actividades del día a día del personal del área de mantenimiento y logística.
- Ahorro en el área de sistemas que apoyaba a las demás áreas para realizar estas tareas.
- Menor tiempo de dedicación en elaborar los reportes para la toma de decisiones.

Se puede observar claramente que la implementación de la presente solución ofrece una serie de beneficios frente a un costo aceptable, debido a que no hay costo de licencia por emplear software libre.

Finalmente, se concluye, por las razones presentadas, que la solución propuesta es la más adecuada para resolver el problema de los usuarios del área de mantenimiento y logística de una empresa de transporte público en la toma de decisiones.

2. Análisis

En el presente capítulo se definirá la metodología de la solución a emplear. Luego se identificarán los requerimientos funcionales y no funcionales que debe satisfacer la solución. Finalmente, se realizará el análisis dimensional que satisface a los requerimientos.

2.1. Definición de la metodología de la solución

Se eligió como metodología de la solución a una metodología basada en RUP y en la herramienta UML para desarrollar un Data Warehouse o Datamart llamada Data Warehouse Engineering Process (DWEP). Esta metodología fue propuesta por Sergio Luján-Mora y Juan Trujillo en el año 2006. Para el presente proyecto dicha metodología será aplicada al desarrollo de un Datamart.

El DWEP divide el desarrollo del almacén de datos, al igual que en RUP, en 4 fases: Concepción, Elaboración, Construcción y Transición. Los objetivos generales por cada fase son los mismos que en RUP. Sin embargo, existe una gran diferencia en los entregables que propone el DWEP, pues estos entregables son propios de un proyecto de desarrollo de Data Warehouse o Datamart.

Además de las mismas fases de RUP, el DWEP también propone disciplinas, pero le añade dos más que son de importancia en un proyecto de Datamart. Sin embargo, estas disciplinas adicionales se llevan a cabo cuando el usuario ya tiene tiempo usando el Datamart. Para el caso de este proyecto no se llevarían a cabo porque este culmina cuando se tiene el Datamart listo para usarse. A continuación una breve explicación de cada disciplina.

- **Requerimientos:** Se enfoca en las necesidades del usuario final porque los Datamarts suelen ser únicos para cada tipo de empresa. Durante esta disciplina el usuario debe especificar qué tipo de análisis y agregaciones, le serán de utilidad para generar reportes y tableros de control que le ayuden en la toma de decisiones. La captura de requerimientos se hará mediante los diagramas de Caso de Uso, los cuales permiten representar lo que el usuario quiere que haga el Datamart.
- **Análisis:** Empieza por definir y analizar claramente la especificación de los Casos de Uso encontrados en la etapa anterior. Además, se analiza los datos que servirán como fuente para el Datamart. Se emplearán Diagramas de Esquemas para modelar los datos de carga y se realizan los esbozos para los procesos de ETL.
- **Diseño:** Se define la estructura del Datamart tanto al nivel lógico como físico. Además, se realiza un Diagrama de Mapeo de Datos, el cual muestra la relación de cómo los datos fuentes están relacionados con los datos del modelo multidimensional.
- **Implementación:** Durante esta etapa se construye el Datamart. La estructura física del Datamart es construida. También, se desarrollan los procesos de extracción, transformación y carga de los datos al Datamart.
- **Pruebas:** Se prueba que lo implementado cumpla con las especificaciones del usuario. Se debe elaborar un plan de pruebas que permite analizar los resultados de las pruebas.
- **Mantenimiento:** Esta etapa comienza cuando los usuarios ya están usando el Datamart y acaba cuando el ciclo de vida del Datamart concluya, pues durante toda su vida útil se le debe de actualizar con la información de la empresa.
- **Revisión post-desarrollo:** Se revisa la documentación existente y se busca oportunidades de mejora al Datamart que puede terminar generando futuros proyectos.

El DWEP plantea que el desarrollo del Datamart está dividido en pequeños pasos o iteraciones, los cuales son más fáciles de controlar y llevar a cabo. En cada

disciplina se elaborarán entregables basados en diagramas o técnicas UML. La Figura 12 muestra cómo las siete disciplinas propuestas se relacionan con las fases de DWEP. En la Tabla 4 se muestra un resumen de los diagramas o entregables por cada disciplina del DWEP que se usarán para el presente proyecto, es decir, no se elaborarán todos los diagramas que propone el DWEP, ya que algunos de ellos no son relevantes para el desarrollo del Datamart. Sin embargo, adicionalmente a los entregables propuestos por el DWEP, se elaborarán otros entregables como Plan de Pruebas y Manual de Configuración.

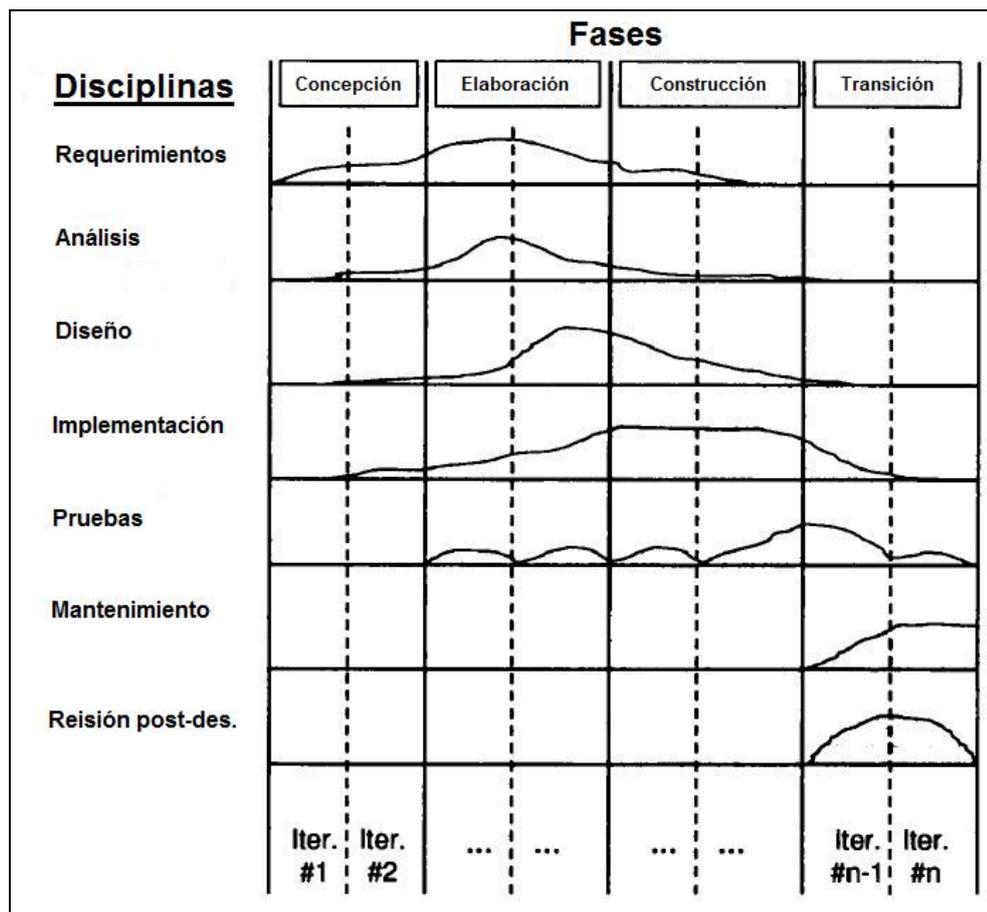


Figura 12: Fases y Disciplinas del DWEP

Disciplinas del DWEP	Entregables
Requerimientos	Diagrama de Actividades Diagrama de Casos de Uso
Análisis	Esquema Conceptual de los Datos Fuente Esquema Lógico de los Datos Fuente
Diseño	Esquema Conceptual del Datamart Diagrama de Mapeo de Datos
Implementación	Esquema Lógico del Datamart Esquema Físico del Datamart Proceso ETL
Pruebas	No hay
Mantenimiento	No hay
Revisión post-desarrollo	No hay

Tabla 4: Resumen de Entregables por Disciplinas del DWEP

A continuación se explicará cada entregable presentado en la Tabla 4:

- **Diagrama de Actividades:** Muestra en orden la secuencia de actividades y las decisiones a tomar para llevar a cabo un proceso, que puede ser una actividad de negocio.
- **Diagrama de Casos de Uso:** Muestra los requerimientos funcionales o necesidades de los usuarios.
- **Esquema Conceptual de Datos Fuente:** Muestra a nivel conceptual cómo está organizada la estructura que forma parte de la información fuente para el Datamart.
- **Esquema Lógico de Datos Fuente:** Muestra a nivel lógico la estructura de la información fuente para el Datamart.
- **Esquema Conceptual del Datamart:** Muestra a nivel conceptual el modelo multidimensional del Datamart.
- **Diagrama de Mapeo de Datos:** Muestra la relación entre la estructura de la información fuente del Datamart con el mismo Datamart.
- **Esquema Lógico del Datamart:** Muestra a nivel lógico el modelo multidimensional del Datamart.
- **Esquema Físico del Datamart:** Muestra a nivel físico el modelo multidimensional del Datamart como la configuración de los servidores y cómo está almacenado en los discos.
- **Proceso ETL:** Muestra el proceso de extracción de la información fuente, transformación de la misma y su carga en el Datamart.

En la Figura 13 se muestra la relación que existe entre los diferentes esquemas del Datamart que propone el DWEP: Esquema Conceptual, Lógico y Físico.

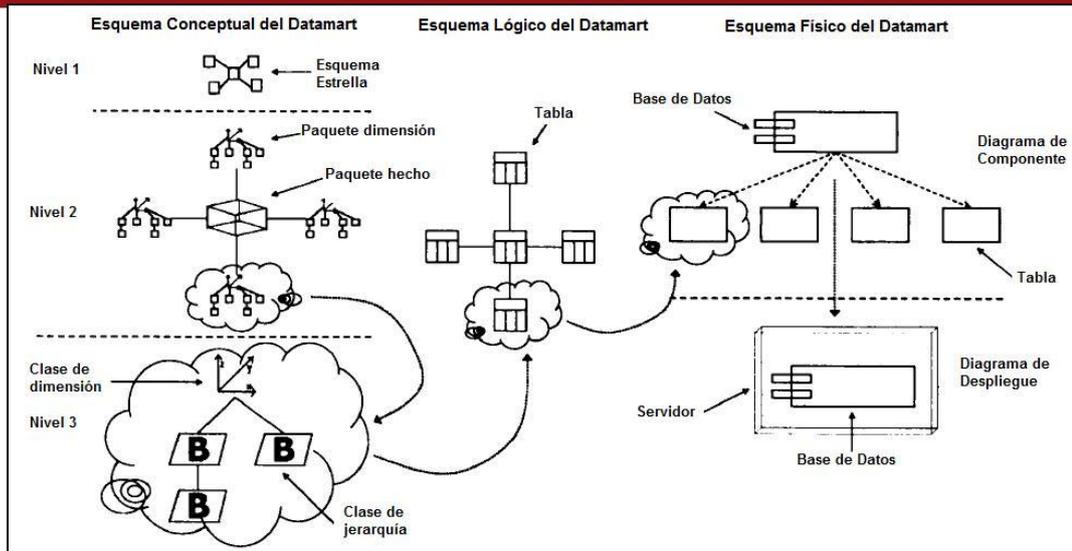


Figura 13: Relación entre los Esquemas del Datamart

Se observa en la Figura 13 que el Esquema Conceptual del Datamart consta de 3 niveles. En el nivel 1 define el modelo multidimensional y cada paquete consta de un esquema estrella. En el nivel 2 se especifica cada esquema estrella del nivel anterior y muestra con qué dimensiones se relaciona cada hecho. Por último, en el nivel 3 se especifica cada dimensión u hecho presentando los niveles jerárquicos que existen en cada dimensión. El detalle de la explicación de cada entregable está en la parte de los Anexos.

En el desarrollo de un proyecto de almacén de datos es importante la participación de los usuarios finales debido a que algunos entregables del DWEP deben contar con su participación para su elaboración. Así en la Figura 14 se muestra los principales pasos que se deben realizar en las disciplinas de Análisis, Diseño e Implementación agrupados por las funciones de los actores que están involucrados en el proyecto. Para el presente proyecto los usuarios finales del Datamart es el personal del área de mantenimiento y logística de la empresa de transporte público de pasajeros y el administrador del Datamart es el tesista.

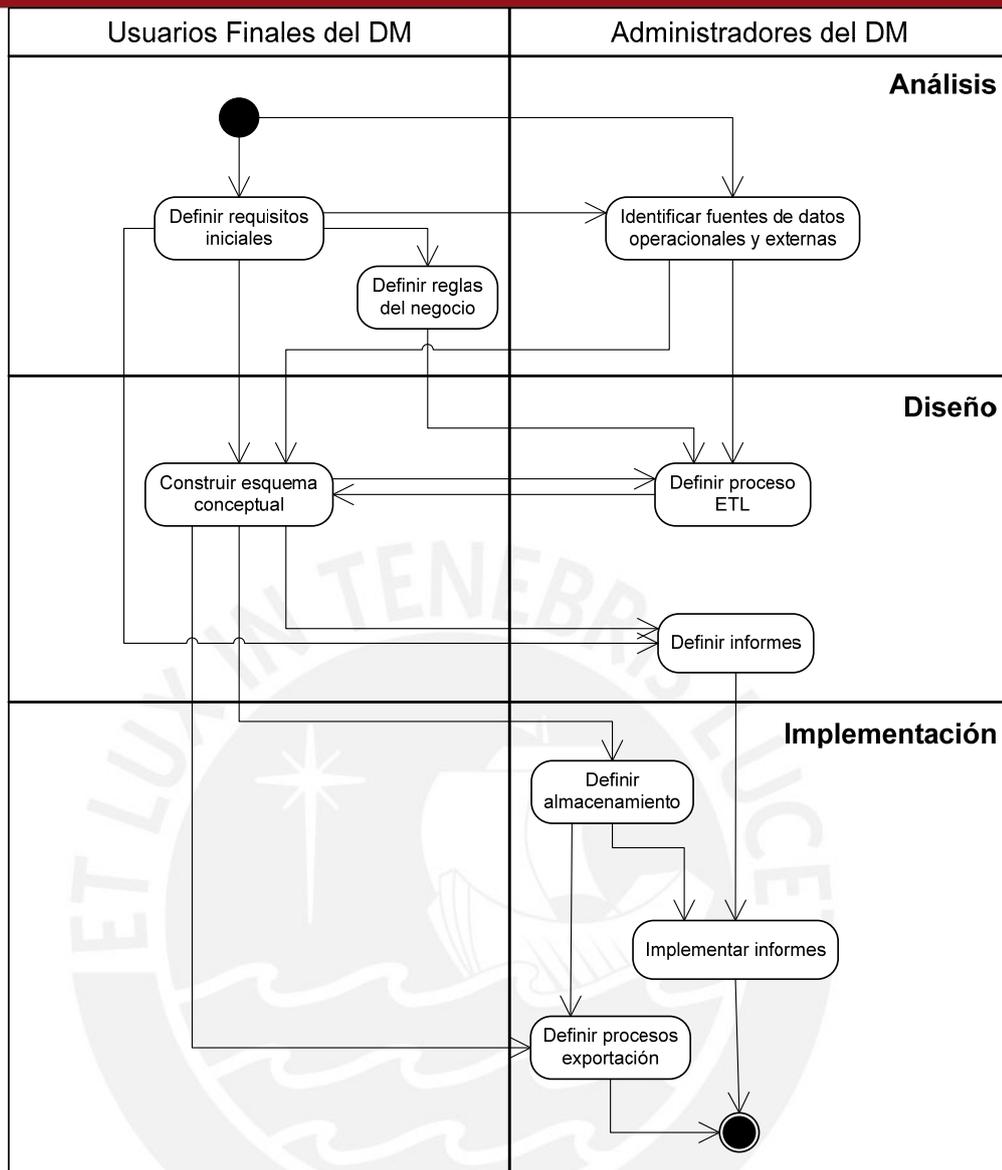


Figura 14: Diagrama de Actividades de los principales pasos del DWEP

La elección de la metodología DWEP como la metodología seleccionada para desarrollar la solución planteada está basada en las siguientes razones:

Primero, es una metodología basada en RUP (*Rational Unified Process*), es decir, está basada en una metodología ampliamente conocida y sobre todo propone las mejores prácticas en el desarrollo de software. Por ello, es fácil comprender las etapas que comprenden así como sus entregables. De esta manera DWEP adapta la metodología RUP a los proyectos de desarrollo de almacenes de datos. Además, también se basa en notación UML para elaborar sus propios entregables. El emplear UML tiene las siguientes ventajas: es un lenguaje visual que permite complejidad en el modelamiento, provee flexibilidad y especialización en los

diagramas, permite crear estereotipos personalizados, permite modelar dimensiones que serían difíciles con un diagrama de Entidad - Relación.

Segundo, si bien existen una gran variedad de modelos en el diseño de almacenes de datos, no existe una metodología que esté presente durante todo el desarrollo de dichos almacenes. La metodología DWEP brinda una serie de entregables y herramientas desde la etapa de levantamiento de información hasta la construcción del Datamart cubriendo etapas importantes como el proceso ETL. Así se emplea una metodología que irá guiando al desarrollador durante todas las etapas del proyecto.

2.2. Identificación de requerimientos

La solución planteada tiene como objetivo principal ofrecer una herramienta que mediante su uso adecuado mejore la toma de decisiones en un área de mantenimiento y logística de una empresa de transporte público de pasajeros. Es decir, el requerimiento principal es que *el Datamart propuesto brinde un conjunto de facilidades que permitan utilizar la información disponible del área de mantenimiento y logística para hacer un mejor análisis, descubrir nuevas oportunidades y sobre todo mejorar la toma de decisiones*. De esta premisa general se puede identificar los requerimientos funcionales y no funcionales que debe cumplir la solución.

2.2.1. Requerimientos Funcionales

Los requerimientos funcionales en el desarrollo de un Datamart constituyen las necesidades de información de los usuarios finales que en este caso es el personal del área de mantenimiento y logística. Este personal está conformado por personas que toman decisiones a diferentes niveles como: asistentes, jefes y gerentes del área.

El desarrollo del presente proyecto se basó en las necesidades de información del personal de una empresa de transporte urbano. Esta empresa está dedicada a ofrecer el servicio de transporte público en Lima desde hace 15 años. En la actualidad cuenta con un flota superior a los 70 autobuses y con más de 150 trabajadores. Se estima que anualmente transporta más de 10 millones de pasajeros.

Luego de algunas reuniones con el personal de la gerencia de Mantenimiento y Logística se ha podido obtener sus necesidades de información:

- **Contar con una herramienta para facilitar la toma de decisiones:** Es la necesidad principal dentro de toda el área y a su vez el objetivo principal de la solución planteada. Actualmente emplean el Microsoft Excel para la elaboración de reportes. Sin embargo, al incorporar el Datamart dentro del área habrá un cambio en el proceso de toma de decisiones, es decir, cambia el circuito de solicitud, búsqueda, preparación, entrega de información para finalmente tomar la decisión. El tomador de decisiones debe poder acceder a la información requerida lo más rápido posible sin necesidad de solicitarla al área de sistemas o tener que estar elaborando reportes manualmente.
- **Tener mejor organizada la información del área:** El personal de la gerencia de Mantenimiento y Logística maneja una gran cantidad de información en el día a día debido al ingreso y salida de repuestos y servicios. Por ello, requieren contar con la información actualizada y organizada ya sea para tomar decisiones u otras actividades. Además desean que el acceso a la información sea a través de roles y usuarios. Si bien la información se almacena en sus sistemas transaccionales, la única forma para poder ver y analizar la información sumariada o agrupada es mediante los reportes. Estos reportes muchas veces no son flexibles a las necesidades del usuario teniéndose que crear un reporte por cada necesidad. El Datamart permitirá contar con la información organizada de manera intuitiva de tal forma que el personal pueda acceder a ella a través de usuarios. Estos usuarios tendrán roles de acuerdo a su jerarquía dentro de la empresa.
- **Poder acceder a la información deseada rápidamente:** El tener acceso a la información requerida es una necesidad de suma importancia para las empresas de transporte, mucho más para el área de logística que maneja información de clientes y proveedores. Los sistemas actuales de la gerencia están diseñados para el almacenamiento de la información de manera rápida, pero no para la consulta de ella. Los reportes muchas veces no satisfacen las necesidades de información del personal y consumen trabajo y tiempo en la elaboración de reportes que permitan contar con la

información rápidamente. El Datamart permitirá acceder a la información de logística de manera rápida y además permitirá la creación de reportes personalizados por el propio personal ahorrando tiempo y carga al área de sistemas.

- **Contar con una herramienta fácil de usar:** La solución planteada no sirve si es que los usuarios no la utilizan. El personal actualmente está acostumbrado a emplear el Microsoft Excel como herramienta de análisis de datos. La nueva herramienta debe ser fácil de usar de tal forma que todo el personal pueda emplearlo en la toma de decisiones. El Datamart debe contar un una interfaz gráfica que permita una fácil interacción con la misma. De esta manera, para el personal no será difícil aprender a utilizarla y podrá verla como una herramienta útil en su labor diario.

Los requerimientos funcionales que se presentarán a continuación han sido agrupados por módulos lógicos para un mejor entendimiento en: Mantenimiento y Logística. Estos requerimientos pueden ser representados mediante diagramas de Casos de Uso que se encuentran en el Anexo 7.

2.2.1.1. Mantenimiento

La calidad en el servicio de una empresa de transporte público depende directamente del desempeño de sus autobuses. Por ello, es de suma importancia poder analizar la información del estado de los autobuses y sus respectivos mantenimientos ya sean preventivos o correctivos. Con esta información se podrá responder a preguntas como: ¿Cuáles son las fallas más comunes en los autobuses?, ¿Cuáles son los mecánicos más solicitados?, ¿Cuáles son los autobuses con más mantenimiento? De esta manera se podrán tomar las acciones correctivas necesarias.

Los reportes o tableros de control requeridos de este módulo son:

- Evolución de Mantenimientos agrupados por Tipo de Mantenimiento (Correctivo y Preventivo) y Periodo, mostrando información de las atenciones por mantenimiento y teniendo como parámetro un año.
- Atención de Autobuses agrupados por Tipo de Autobús (Normal y Eléctrico) y Propietario, mostrando información de las atenciones con su respectiva duración en minutos y teniendo como parámetro un rango de fechas.

- Ranking de Atenciones por Concepto agrupados por Tipo de Falla, mostrando información de las atenciones teniendo como parámetros un rango de fechas y Tipo de Sistema (Mecánico y Eléctrico).
- Tablero de control de mantenimiento que muestre 3 gráficos sobre información consolidada de mantenimiento: Atención de Autobuses por Periodo, Ranking de Autobuses más atendidos en el Año y Ranking de Atenciones por Concepto.

2.2.1.2. Logística

El área de mantenimiento trabaja conjuntamente con el área logística para proveerle adecuadamente los repuestos que sean necesarios. Por ello es importante contar con la información de los productos que ingresan y salen del almacén. Así se podrá tener un control adecuado del stock ante cualquier requerimiento del área de mantenimiento. Con esta información se podrá responder a preguntas como: ¿Cuál es el producto más caro de mi almacén?, ¿Cuál es el repuesto que suele acabarse más rápido?, ¿Cuáles son las 5 empresas proveedoras más importantes de repuestos?, ¿Cuáles son los servicios de reparación más comunes que se realizan sobre los autobuses?

Los reportes o tableros de control requeridos de este módulo son:

- Productos por Catálogo agrupados por Familia y Subfamilia, mostrando información de los productos.
- Movimiento de Productos del Almacén agrupados por Tipo de Transacción (Ingresos y Salidas), mostrando información de los productos y teniendo como parámetro un rango de fechas.
- Compra de Productos agrupados por Periodo, mostrando información de la cantidad y monto de los productos comprados y teniendo como parámetro un año.
- Ranking de Proveedores mostrando información de los Socios Estratégicos, cantidad y montos de compra y teniendo como parámetro un rango de fechas.
- Tablero de control de logística que muestre 3 gráficos sobre información consolidada de logística: Compra de productos por Periodo, Principales productos agrupados por Familia y Ranking de Proveedores.

Los requerimientos presentados podrán satisfacer las necesidades de información del área de mantenimiento y logística. De esta manera el personal involucrado podrá contar con la información adecuada, organizada y en el tiempo adecuado de tal forma que le ayude en la toma de decisiones.

2.2.2. Requerimientos No Funcionales

Los requerimientos no funcionales no pueden asociarse a un Caso de Uso en particular, pero son de mucha importancia en el desarrollo de un Datamart. A continuación se presenta los requerimientos no funcionales identificados:

- El modelo multidimensional de la solución debe permitir la creación de nuevas dimensiones en el futuro.
- La herramienta de explotación seleccionada debe permitir realizar las técnicas de consulta multidimensional como Slicing, Dicing, Drilling, etc. Además, debe permitir la creación de reportes personalizados por el usuario final.
- La solución debe contener reportes y tableros de control elaborados con la herramienta de explotación seleccionada. Además de la posibilidad de exportarlos a archivos de formatos estándar como archivo de texto (.txt) o Excel (.xls).
- La herramienta de explotación debe permitir mostrar la información del Datamart en base a los siguientes 3 tipos: Reportes tradicionales (basados en columnas), Tablas pivote (matriz de doble entrada) y Gráficos.
- La herramienta de explotación debe poder ser accedida vía Web y tener una disponibilidad de 24 x 7.

El análisis de los requerimientos tanto funcionales como no funcionales se muestra en el entregable Documento de Análisis de Requerimientos que se encuentra en el Anexo 6.

2.3. Análisis dimensional

En esta sección se analiza el modelo dimensional que responde a los requerimientos expuestos en la sección anterior.

La metodología DWEP divide el diseño de un Datamart en fases bien definidas: Diseño conceptual, lógico y físico. En cada fase se genera un entregable y estos

son: Esquema Conceptual del Datamart, Esquema Lógico del Datamart y Esquema Físico del Datamart.

Para el diseño de una base de datos generalmente sólo se realiza un diseño lógico y físico. El primero muestra las tablas y la relación entre ellas. El segundo, diseño físico, muestra las tablas con sus campos y sus respectivos tipos de datos. Sin embargo, en la metodología DWEP el diseño conceptual sería lo equivalente al diseño lógico tradicional, pues muestra los esquemas y tablas así como la relación entre ellas. De la misma manera, el diseño lógico según DWEP sería el equivalente al diseño físico tradicional que muestra los campos de las tablas así como sus tipos de datos. Finalmente, el diseño físico según DWEP muestra cómo está estructurada a nivel de archivos la base de datos y cómo se almacena en el servidor. Este diseño no existe en el diseño tradicional de bases de datos por lo que no existe ninguna equivalencia con el diseño tradicional, lo cual se explicó en la Arquitectura de la solución.

A continuación se detallará los aspectos que se tomaron en cuenta en el diseño del modelo de datos del Datamart.

2.3.1. Diseño Conceptual del Datamart

Para satisfacer las necesidades de información de los usuarios del área de mantenimiento y logística se definió cinco esquemas de modelo estrella: Documentos Detalle, Documentos, Ingresos y Salidas, Stock Valorizado y Mantenimiento. Cada esquema abarca un tema de análisis de información en particular organizando las variables y dimensiones necesarias para satisfacer la demanda de información. Antes de explicar acerca de cada esquema se deberá abarcar algunos conceptos relacionados al diseño conceptual. Por otro lado, el detalle del Esquema Conceptual del Datamart se encuentra en el Anexo 9.

- Nivel de granularidad: Indica hasta qué nivel de detalle puede ir el análisis del modelo estrella. Está relacionado directamente a la tabla de hechos, pues un registro de esta tabla indica el nivel de granularidad del modelo estrella.

- Nivel de agregación: Indica cómo se está agrupada o sumariada la información del esquema. También está relacionado a la tabla de hechos pues contiene cantidades o montos sumariados.
- Jerarquía de las dimensiones: Algunas dimensiones tienen campos ordenados de una manera jerárquica que permite un análisis de datos más generales a específicos y viceversa. Las jerarquías se encuentran en las tablas de dimensiones y cada una puede tener distintos niveles de jerarquías.

Las jerarquías definidas para el Datamart se muestran en la Figura 15.

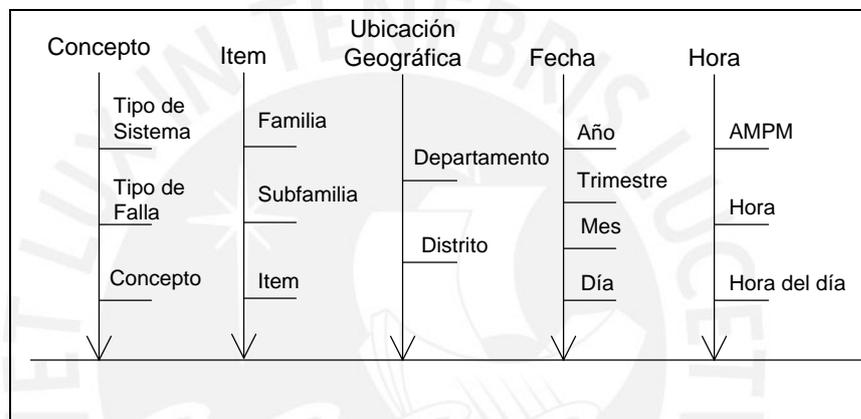


Figura 15: Jerarquía de Dimensiones

Por ejemplo, la jerarquía definida para la ubicación geográfica permite el análisis de los documentos de pago no sólo por los distritos de la oficina principal del proveedor sino también por los departamentos. De la misma manera permite se puede analizar con las demás jerarquías.

2.3.1.1. Esquema de Documento Detalle

Este esquema tiene como objetivo permitir el análisis de los documentos de compra y venta a un nivel detallado pudiéndose distinguir cada ítem en particular. Satisface los requerimientos del análisis de los documentos. Su nivel de granularidad está dado por los ítems comprados o vendidos en cada documento. Su nivel de agregación contiene el monto de la compra y venta tanto en moneda nacional como extranjera. Las dimensiones de Ítems, Socios Estratégicos, Fechas y Horas presentan niveles de jerarquía.

En la Figura 16 se muestra las dimensiones con las que se relaciona.

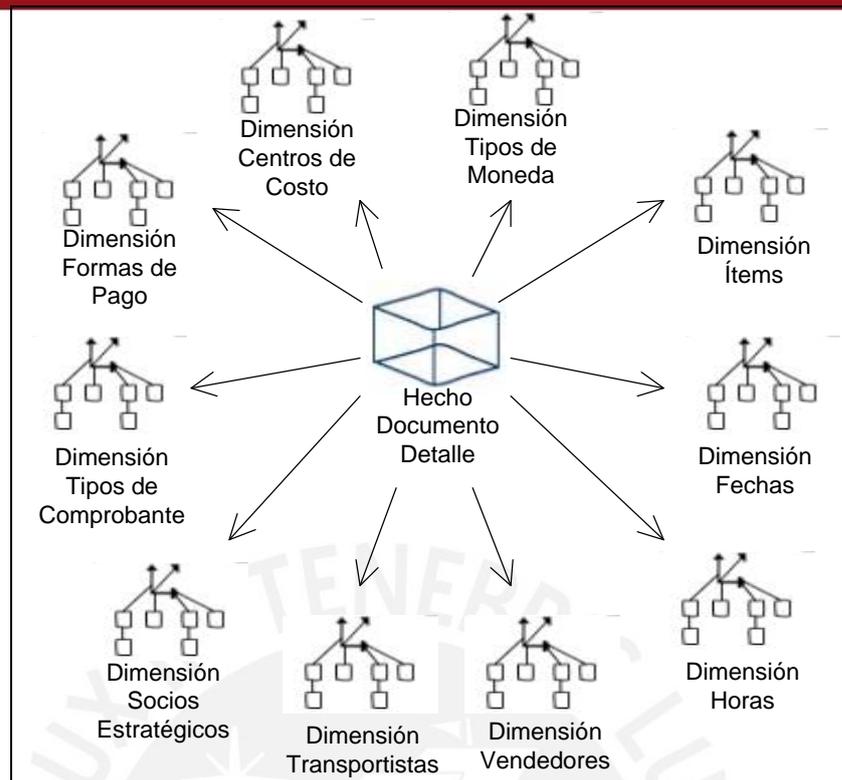


Figura 16: Esquema Conceptual del Modelo Estrella Documento Detalle

2.3.1.2. Esquema de Documento

Este esquema tiene como objetivo permitir el análisis de los documentos a un nivel más macro. Satisface los requerimientos del análisis de los documentos. Su nivel de granularidad está dado por cada documento de compra y venta. Su nivel de agregación contiene la suma de la venta o compra tanto en moneda nacional como extranjera. Las dimensiones de Socios Estratégicos, Fechas y Horas presentan niveles de jerarquía.

En la Figura 17 se muestra las dimensiones con las que se relaciona.

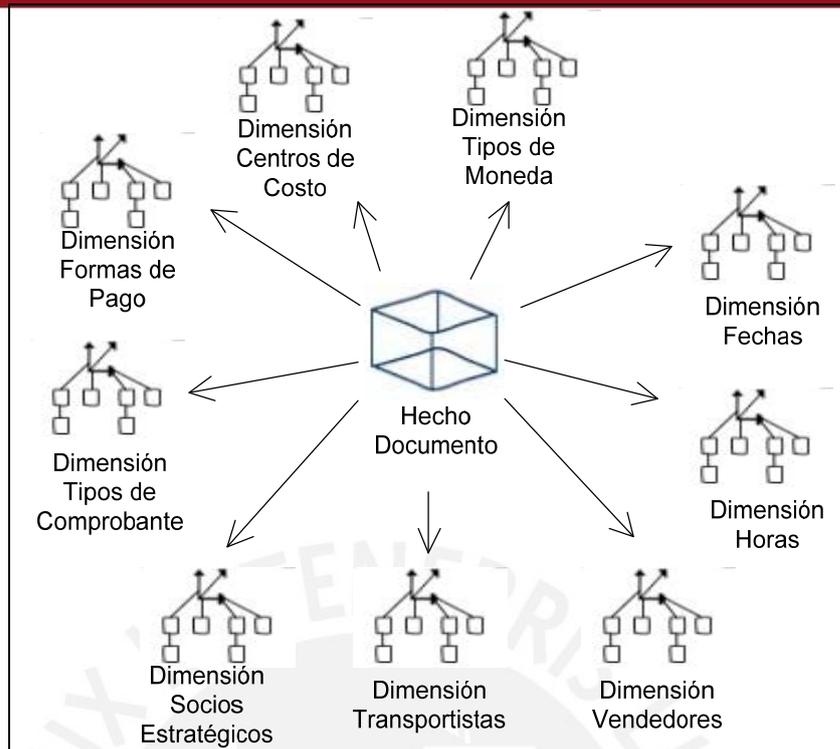


Figura 17: Esquema Conceptual del Modelo Estrella Documento

2.3.1.3. Esquema de Ingresos y Salidas

Este esquema tiene como objetivo permitir el análisis del movimiento del almacén. Satisface los requerimientos del análisis de los ingresos y salidas de los ítems. Su nivel de granularidad está dado por cada ítem que ingresa o sale dado una orden de movimiento. Su nivel de agregación contiene la cantidad y precio o costo unitario del ítem que ingresó o salió del almacén. Las dimensiones de Almacenes, Ítems, Fechas, Horas y Transacciones de Inventario del día presentan niveles de jerarquía.

En la Figura 18 se muestra las dimensiones con las que se relaciona.

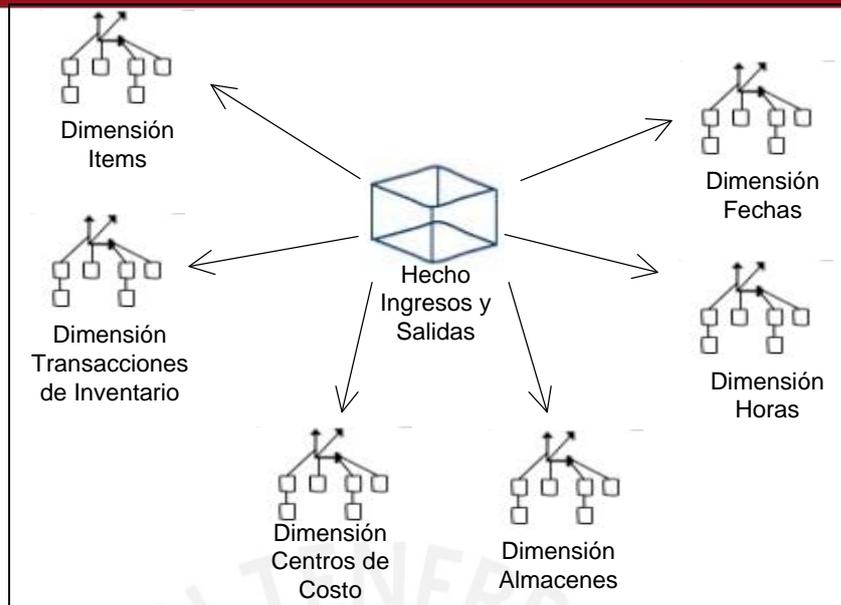


Figura 18: Esquema Conceptual del Modelo Estrella Ingresos y Salidas

2.3.1.4. Esquema de Stock Valorizado

Este esquema tiene como objetivo permitir el análisis del stock del almacén. Satisface los requerimientos del análisis del stock valorizado por ítem. Su nivel de granularidad está dado por cada ítem. Su nivel de agregación contiene el stock de cada ítem al final del día. Las dimensiones de Almacenes, Ítems, Fechas y Transacciones de Inventario presentan niveles de jerarquía.

En la Figura 19 se muestra las dimensiones con las que se relaciona.

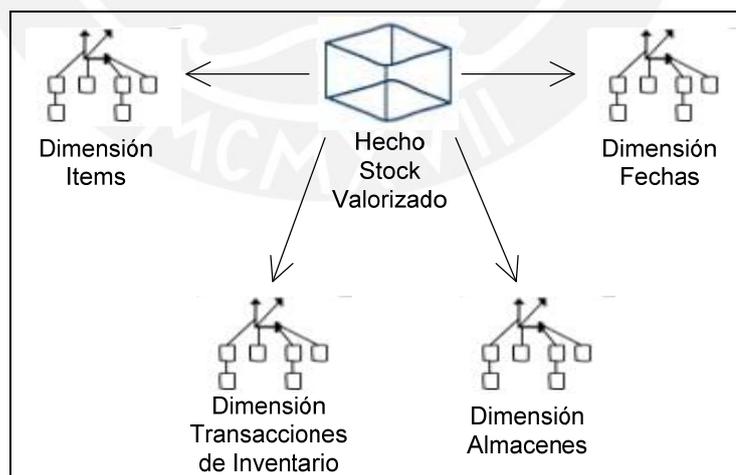


Figura 19: Esquema Conceptual del Modelo Estrella Stock Valorizado

2.3.1.5. Esquema de Mantenimiento

Este esquema tiene como objetivo permitir el análisis los mantenimientos tanto preventivos como correctivos que se realizan a los autobuses. Satisface los requerimientos del análisis de mantenimiento. Su nivel de granularidad está

dado por cada concepto que se realiza sobre un autobús como parte de su mantenimiento. Su nivel de agregación contiene el indicador de autobús atendido. Las dimensiones de Autobuses, Concepto, Fechas y Horas del día presentan niveles de jerarquía.

En la Figura 20 se muestra las dimensiones con las que se relaciona.

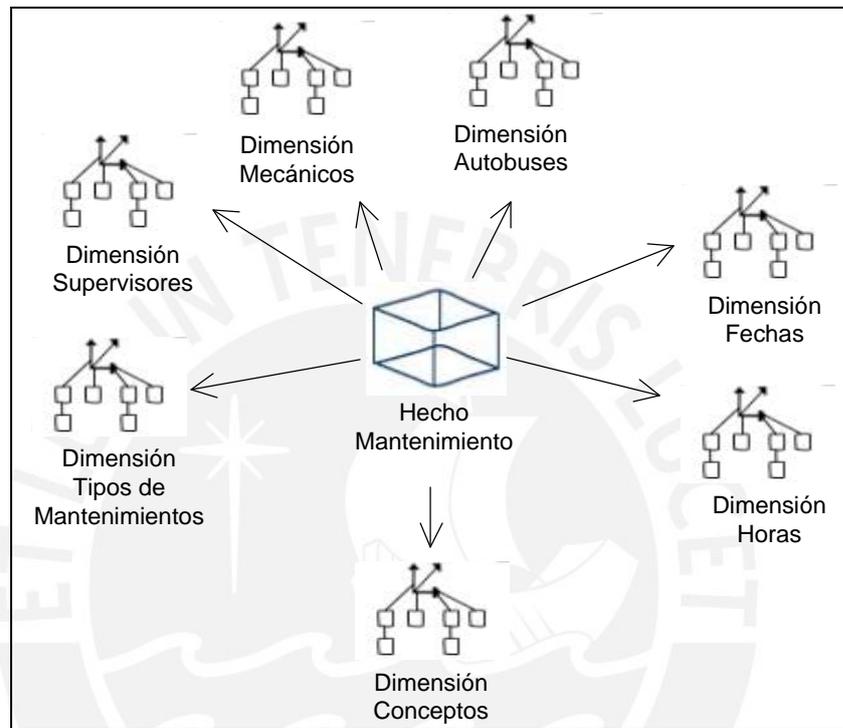


Figura 20: Esquema Conceptual del Modelo Estrella Mantenimiento

De manera resumen se puede observar en la Tabla 5 un cuadro de doble entrada en donde tenemos a los cinco modelos tipo estrella en las columnas y a las dimensiones en las filas. Se aprecia en la tabla que existen dimensiones que son empleadas en más de una tabla de hecho.

Dimensiones	Facts				
	Documento Detalle	Documento	Ingresos y Salidas	Stock Valorizado	Mantenimiento
Almacenes			✓	✓	
Autobuses					✓
Centros de Costo	✓	✓	✓		
Conceptos					✓
Fechas	✓	✓	✓	✓	✓
Formas de Pago	✓	✓			

Dimensiones	Facts				
	Documento Detalle	Documento	Ingresos y Salidas	Stock Valorizado	Mantenimiento
Horas del día	✓	✓	✓		✓
Ítems	✓		✓	✓	
Mecánicos					✓
Socios Estratégicos	✓	✓			
Supervisores					✓
Tipos de Moneda	✓	✓			
Tipos de Comprobante	✓	✓			
Tipos de Mantenimiento					✓
Transportistas	✓	✓			
Transacciones de Inventario			✓	✓	
Vendedores	✓	✓			

Tabla 5: Hechos vs. Dimensiones

2.3.2. Diseño Lógico del Datamart

Para diseñar un modelo lógico primero se deben definir los campos de las tablas con sus tipos de datos más adecuados. Todo campo a emplear en el modelo del Datamart está entre alguno de los siguientes grupos de tipos de datos: Numéricos exactos, numéricos aproximados, fecha y hora y cadena de caracteres. [KIM 2002]

- Numérico exacto: Se empleó el tipo de dato INTEGER para los identificadores de todas las tablas tanto de dimensiones como hechos. Esto debido a que las búsquedas y consultas son más rápidas cuando llave primaria es un número entero.
- Numérico aproximado: Se empleó el tipo de dato FLOAT para los valores de los montos de las ventas y compras debido a que debe manejarse decimales en sus valores.
- Fecha y hora: Se empleó el tipo de dato VARCHAR para los valores de las fechas. No se empleó el tipo de dato TIMESTAMP porque no era necesario almacenar la fecha y hora conjuntamente sino por separado.
- Cadena de caracteres: Se empleó el tipo de dato VARCHAR para los valores de texto debido a que estos pueden ser de longitud muy variable.

Sólo se empleó CHAR para los datos que son códigos de la base de datos fuente.

A continuación el diseño lógico de los esquemas estrellas. El detalle de sus atributos y tipos de datos se encuentra en el Anexo 11.

2.3.2.1. Esquema Documento Detalle

Este esquema brinda información detallada de los de documentos. Por ello, tiene medidas como los montos de las ventas y compras de ítems en moneda nacional y extranjera.

En la Figura 21 se muestra el modelo lógico del esquema Documento Detalle.

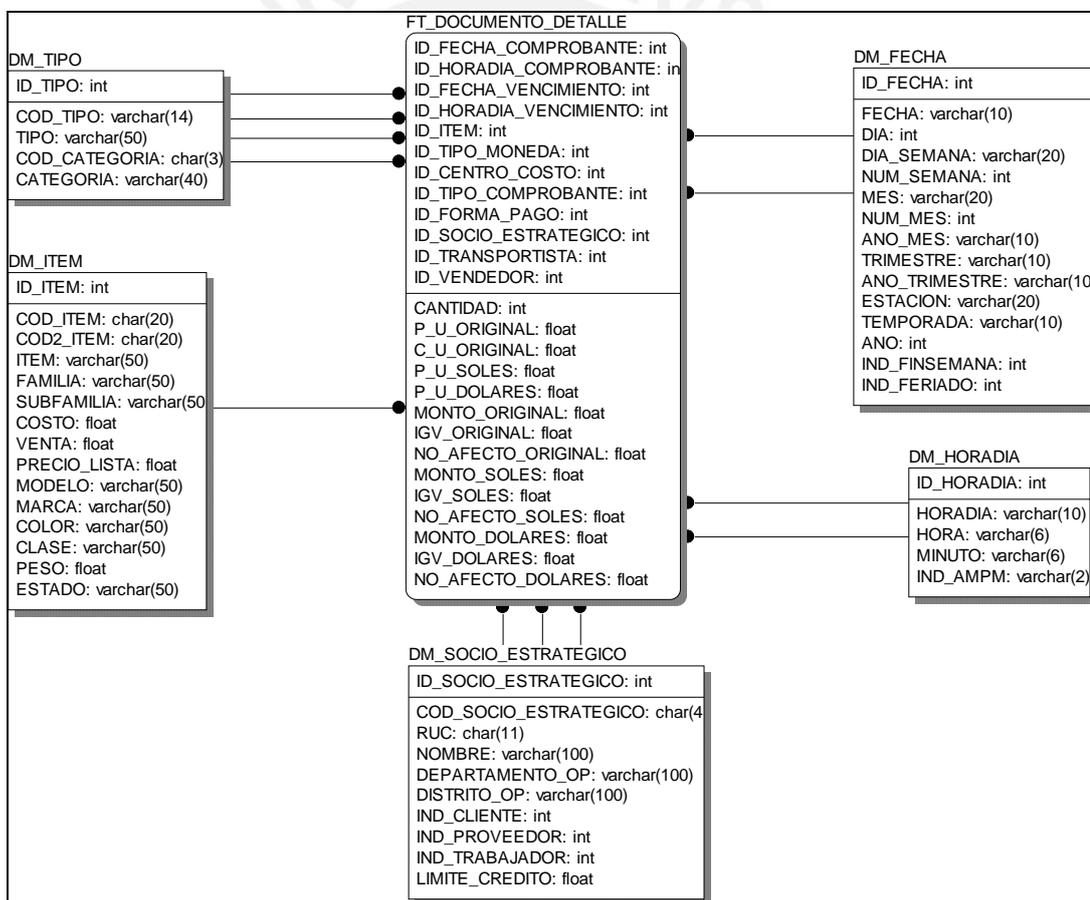


Figura 21: Esquema Lógico del Modelo Estrella Documento Detalle

2.3.2.2. Esquema Documento

Este esquema brinda información de los documentos. Por ello, tiene medidas como los montos de las ventas o compras en moneda nacional y extranjera. A diferencia del esquema anterior, este esquema no permite identificar el ítem en particular.

En la Figura 22 se muestra el modelo lógico del esquema Documento.

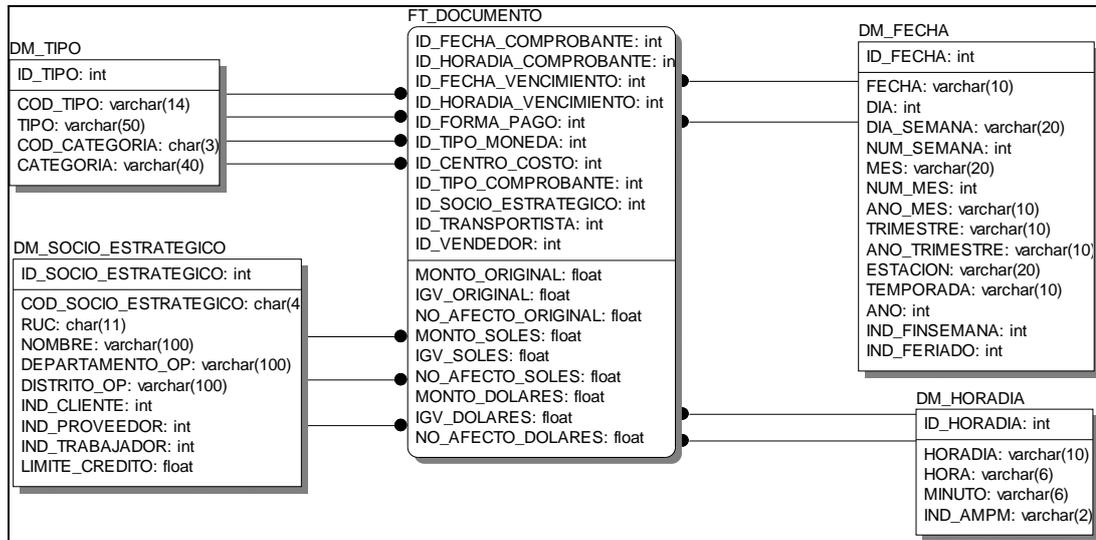


Figura 22: Esquema Lógico del Modelo Estrella Documento

2.3.2.3. Esquema Ingresos y Salidas

Este esquema brinda información de los ingresos y salidas de los ítems del almacén. Por ello, tiene medidas como la cantidad y el precio o costo unitario del ítem. Las medidas de este esquema permiten analizar cómo es el movimiento del almacén.

En la Figura 23 se muestra el modelo lógico del esquema Ingresos y Salidas.

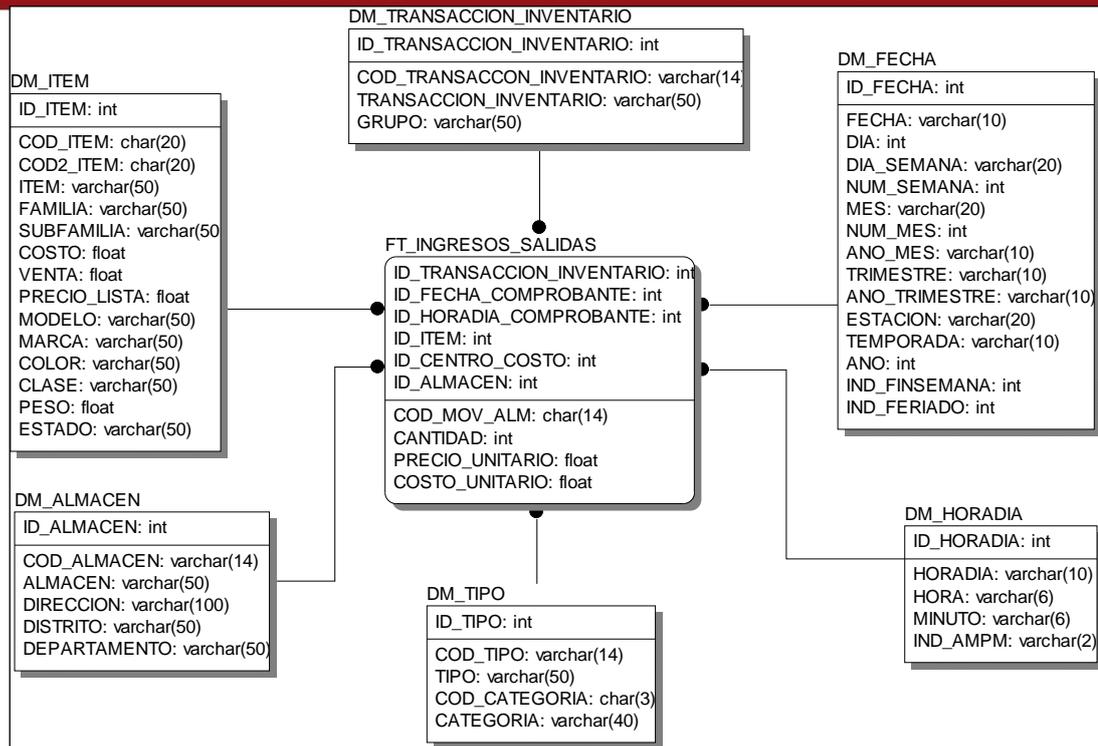


Figura 23: Esquema Lógico del Modelo Estrella Ingresos y Salidas

2.3.2.4. Esquema Stock Valorizado

Este esquema brinda información del stock valorizado. Las medidas de este esquema permiten conocer el stock a nivel diario por cada ítem.

En la Figura 24 se muestra el modelo lógico del esquema Stock Valorizado.

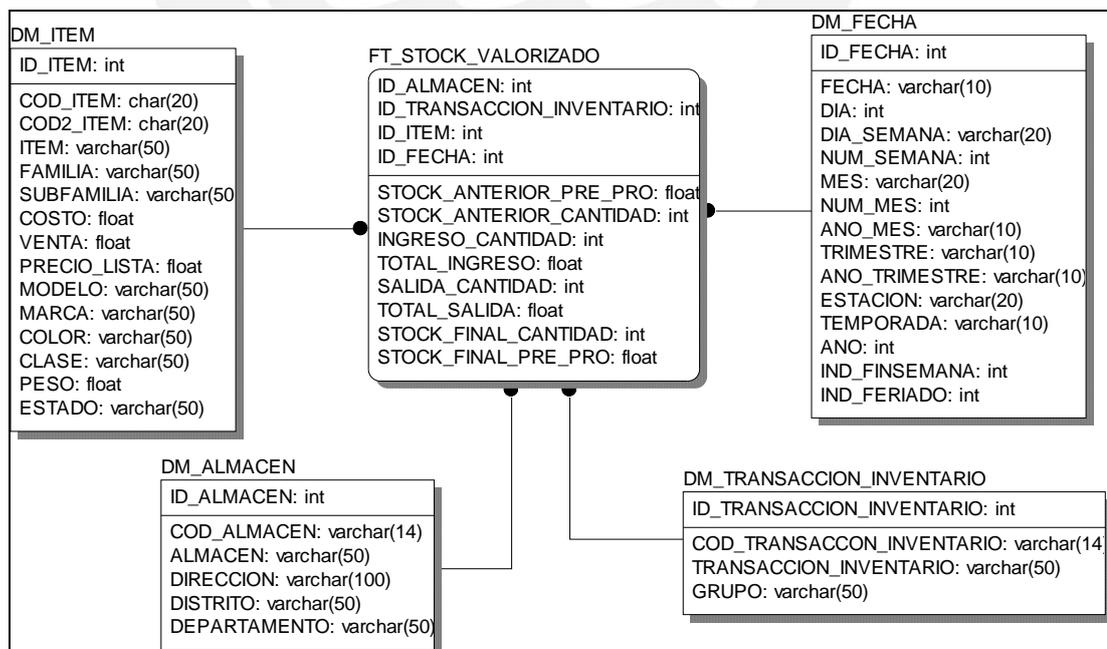


Figura 24: Esquema Lógico del Modelo Estrella Stock Valorizado

2.3.2.5. Esquema Mantenimiento

Este esquema brinda información de los mantenimientos. Las medidas de este esquema permiten conocer la cantidad de los autobuses que han pasado por un mantenimiento ya sea correctivo o preventivo.

En la Figura 25 se muestra el modelo lógico del esquema Mantenimiento.

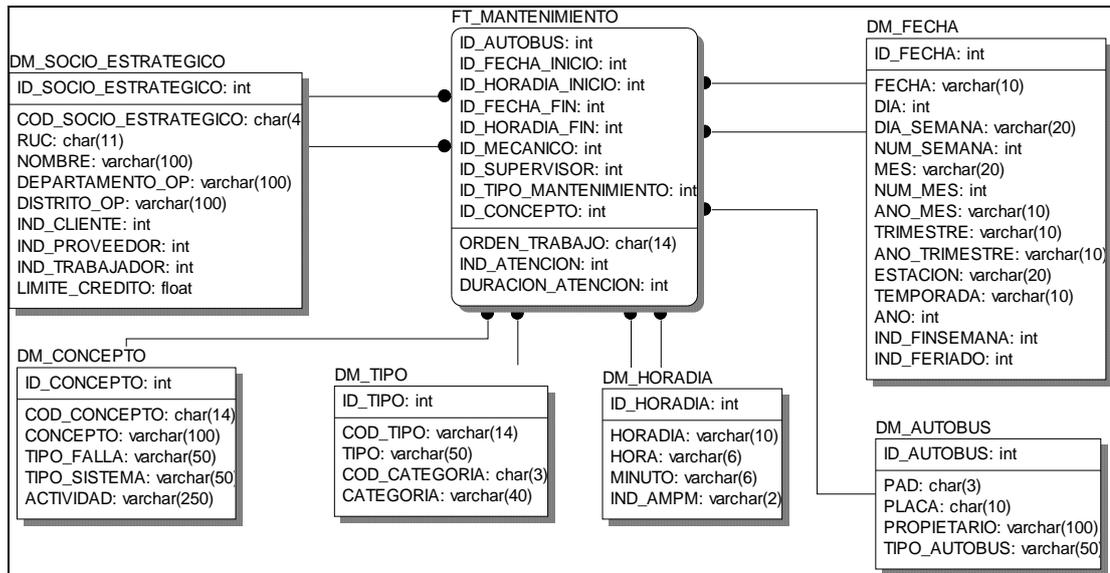


Figura 25: Esquema Lógico del Modelo Estrella Mantenimiento

Se observa que si bien a nivel conceptual existirían muchas dimensiones, a nivel lógico la cantidad de dimensiones disminuye. Esto debido a que varias dimensiones son idénticas a nivel lógico y por ello pueden agruparse.

3. Diseño

En el presente capítulo se mostrará la arquitectura que conformará la solución. Luego se detallará los criterios que se consideraron en el diseño de extracción y explotación.

3.1. Arquitectura de la solución

La arquitectura que permite la implantación del Datamart está conformada por diversos componentes. A continuación se mencionará cada componente o interfaz:

- **Componente Fuentes de Información:** Representa a las fuentes de información para el presente proyecto y por ello será la fuente de datos para el Datamart. Está formado por la base de datos que contiene la información transaccional del área de mantenimiento y logística de la empresa de transporte público.
- **Componente ETL:** Implementa los procesos necesarios para extraer la información de los datos fuentes, transformarlos y cargarlos en la base de datos de Datamart. Se empleará la herramienta Pentaho Data Integration para la elaboración de los procesos ETL.

- **Componente Datamart:** Representa a la base de datos multidimensional del Datamart. Tiene la estructura adecuada para cargarse en una herramienta de explotación y permitir el análisis de la información. Este componente como el de Fuentes de información estarán en la base de datos PostgreSQL.
- **Componente OLAP:** Contiene los servicios OLAP que sirviéndose de la base de datos de Datamart permite mostrar los datos en Cubos multidimensionales para su mejor análisis. Se empleará Pentaho Schema Workbench para crear los cubos. Estos servicios son transparentes al usuario. Los componentes mencionados incluyendo éste corresponden al Back-End de la solución, es decir, son componentes transparentes para el usuario final.
- **Componente Interfaz de Usuario:** Representa a la interfaz con la que interactúa el usuario. A través de ella les es posible tomar las decisiones correspondientes en base a la información que se le presenta. Se empleará la Consola de Usuario de Pentaho como herramienta de explotación e interfaz de usuario.

En la Figura 26 se muestra la interacción entre los componentes que conforman la arquitectura de la solución.

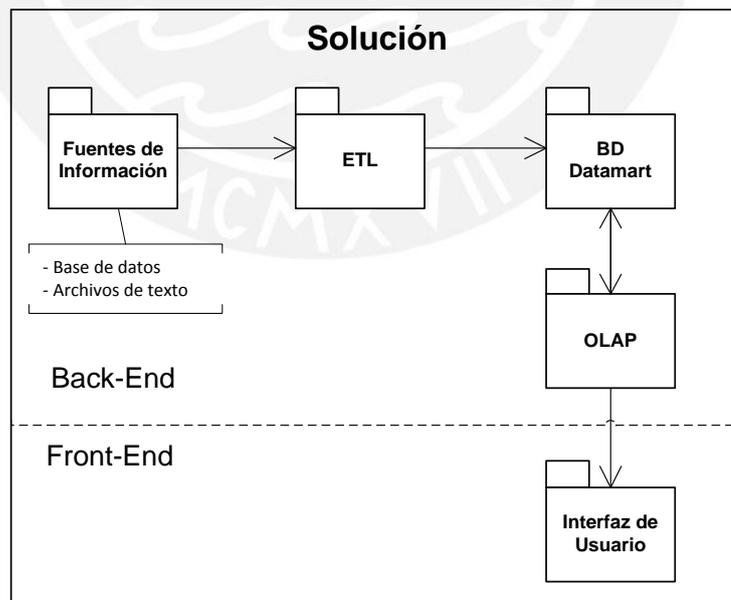


Figura 26: Diagrama de Arquitectura

A continuación se desarrollará a mayor detalle los principales componentes que forman parte del Back-End, el componente de Interfaz de Usuario se desarrollará en el subcapítulo Diseño de Explotación.

3.1.1. Fuentes de Información

Esta base de datos está formada por la base de datos y otras fuentes de la Gerencia de Mantenimiento y Logística. Será considerada como la base de datos transaccional que contiene la información de una empresa de transporte público de pasajeros producto del ejercicio de su negocio.

El diagrama de clases de esta base de datos se muestra en la Figura 27. El diccionario de datos se encuentra en el Anexo 8. Esta base de datos es la que proporcionará de información al Datamart. Sin embargo, dado que sólo se ha tenido acceso a la estructura del modelo de base de datos, ésta no cuenta realmente con información. Por ello, se ha establecido cargar dicho modelo a través de archivos planos con información ficticia. La razón de ser de estos archivos es agilizar el proceso de carga de la información a la base de datos transaccional.

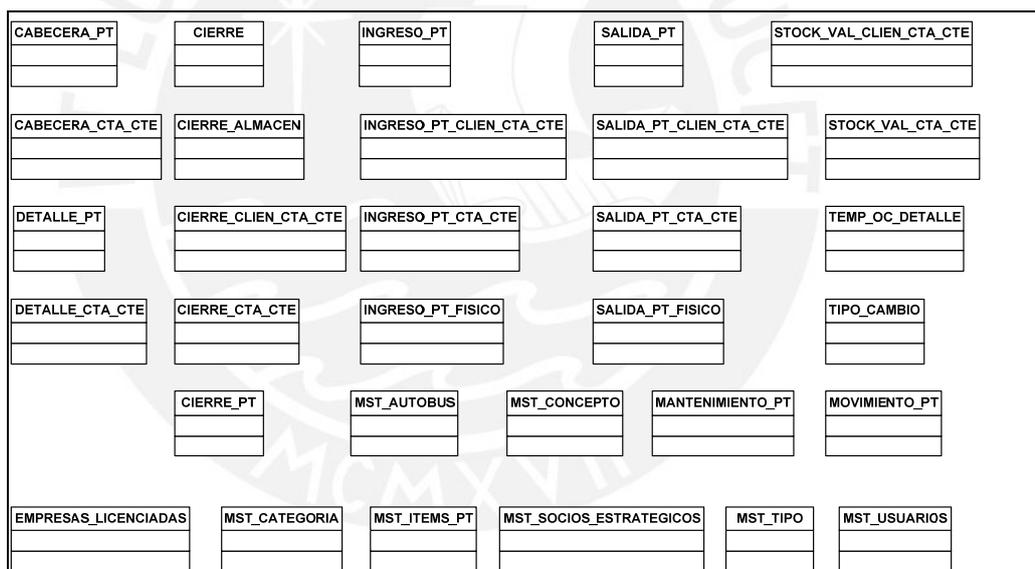


Figura 27: Diagrama de Clases de la base de datos fuente

3.1.2. ETL

El proceso ETL se encarga básicamente del paso de la información de un modelo de base de datos relacional a un modelo multidimensional. El proceso extrae la información que está almacenada en el modelo relacional, realiza la transformación de los tipos de datos de los campos correspondientes y carga de datos las tablas de dimensiones y hechos de la base de datos multidimensional.

El primer paso del diseño del proceso ETL es realizar un mapeo de los datos de la base de datos fuente con los de la base de datos del Datamart. De esta manera se puede establecer la relación de los campos en ambas bases de datos. La metodología DWEP propone el Diagrama de Mapeo de Datos que se especifica en el Anexo 10. En la Figura 28 se muestra un ejemplo de este diagrama para los datos de los autobuses. En la figura se observa que las clases del lado izquierdo corresponden a las clases de la base de datos fuente y las del lado derecho a la dimensión Autobús del Datamart.

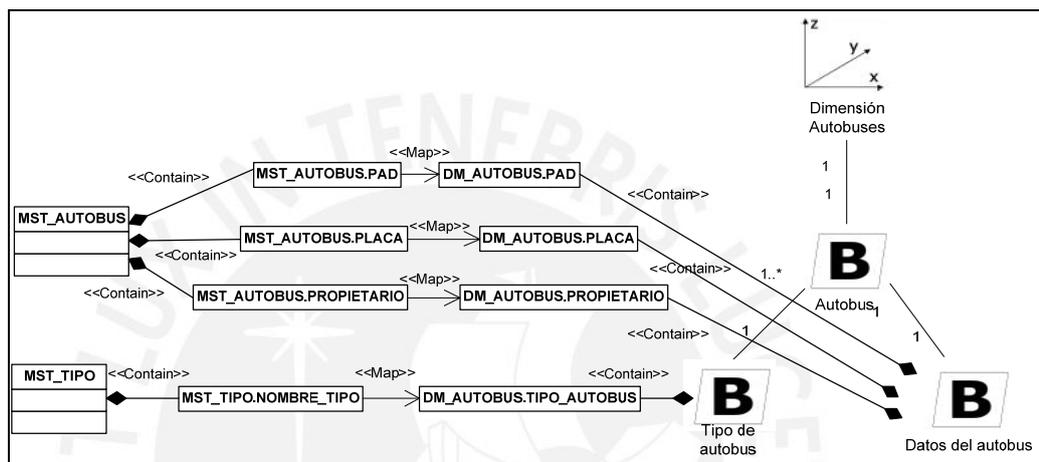


Figura 28: Diagrama de Mapeo de Datos de Autobús

Luego de realizar el mapeo de datos se debe diseñar el flujo del proceso de ETL para cargar cada dimensión o hecho para tener una idea clara sobre las transformaciones que sufrirán los valores al pasar de una base de datos a otra. Esto se debe realizar para facilitar la implementación de los procesos ETL en la fase de Construcción. El DWEP propone un diagrama que define el diseño de los procesos ETL. Todos los diagramas del proceso ETL se encuentran especificados en el Anexo 13. En la Figura 29 se muestra un ejemplo del diagrama de Proceso ETL para la carga inicial de la dimensión de Autobús. En la figura se observa que se requiere la información de una tabla (MST_AUTOBUS) y se carga en una nueva tabla la cual tiene una nueva llave primaria.

El diseño de la extracción y el detalle del mapeo se aprecia en el siguiente punto llamado Diseño de Extracción.

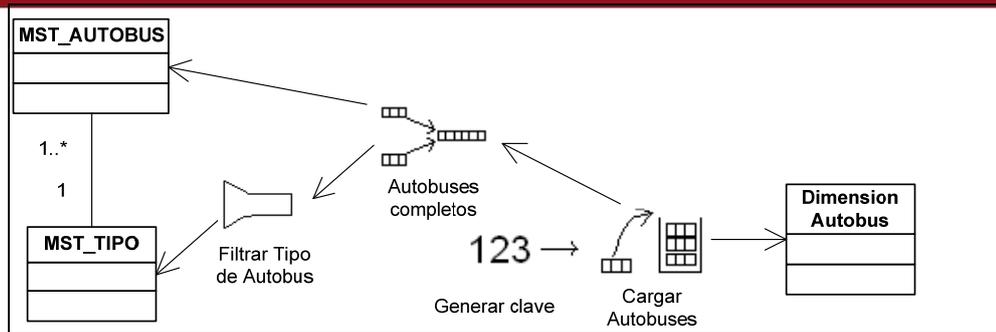


Figura 29: Diagrama de Proceso ETL para la Dimensión Autobús

3.1.3. Datamart

Este componente está formado por la entidad de base de datos. A nivel físico, las tablas están almacenadas en diversos archivos planos, pero podemos dividirlos en dos grupos. Los dos grupos son: Dimensiones y Hechos. Además de permitir una mejor organización, permite agregar de manera más simple futuras tablas a medida que crezca la base de datos. En la Figura 30 se observa el diagrama de componentes que conforma la base de datos del Datamart. El detalle de este diagrama se muestra en el Anexo 12.

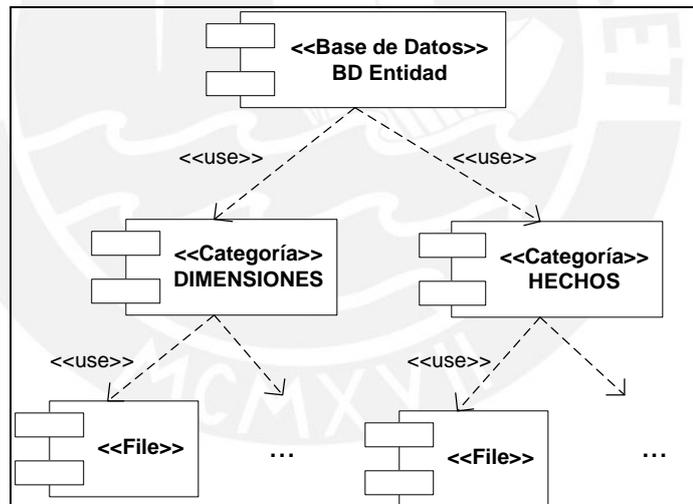


Figura 30: Diagrama de Componente de la base de datos del Datamart

Es importante mencionar que los componentes que conforman el Back-End se encontrarán en el servidor de la empresa de transporte y sólo la interfaz del usuario estará en cada computadora cliente de los usuarios finales. En la Figura 31 se muestra el diagrama de componente de la computadora que será el servidor que contenga la base de datos del Datamart.

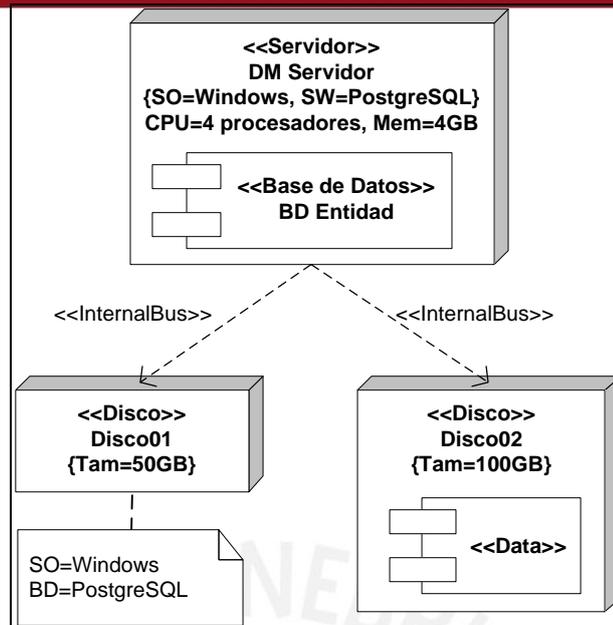


Figura 31: Diagrama de Componente del servidor de base de datos

La computadora que almacenará el Datamart debe tener como mínimo las siguientes características de hardware:

Procesador	Procesador de 4 núcleos
Memoria RAM	4 GB
Disco Duro	150 GB

Además, debe tener el sistema operativo Windows Xp y tener instalado PostgreSQL y las herramientas de Pentaho: Servidor de BI, Data Integration, Schema Workbench, Report Designer. Por otro lado, las computadoras con las que interactuará el usuario deben tener como mínimo las siguientes características de hardware:

Procesador	Procesador de 2 núcleos
Memoria RAM	2 GB
Disco Duro	50 GB

Además, debe tener el sistema operativo Windows Xp y la herramienta para la visualización de datos: Pentaho.

3.2. Diseño de Extracción

A continuación, se presentará el mapeo de cómo se cargarán las tablas destino y de sus tipos de datos, indicando las diversas fuentes.

DM_AUTOBUS

Tabla: DM_AUTOBUS		
Campo	Tipo	Mapeo
ID_AUTOBUS	Integer	Autogenerado
PAD	Char(3)	MST_AUTOBUS.PAD
PLACA	Char(10)	MST_AUTOBUS.PLACA
PROPIETARIO	Varchar(100)	Mayuscula(MST_SOCIOS_ESTRATEGICOS.NOMBRE_ANEXO)
TIPO_AUTOBUS	Varchar(50)	Mayuscula(MST_TIPO.NOMBRE_TIPO)

DM_CONCEPTO

Tabla: DM_CONCEPTO		
Campo	Tipo	Mapeo
ID_CONCEPTO	Integer	Autogenerado
COD_CONCEPTO	Char(14)	MST_CONCEPTO.COD_CONCEPTO
CONCEPTO	Varchar(100)	Mayuscula(MST_CONCEPTO.NOMBRE_CONCEPTO)
TIPO_FALLA	Varchar(50)	Mayuscula(MST_TIPO.NOMBRE_TIPO)
TIPO_SISTEMA	Varchar(50)	Mayuscula(MST_TIPO.NOMBRE_TIPO)
ACTIVIDAD	Varchar(250)	Mayuscula(MST_CONCEPTO.ACTIVIDAD)

DM_ITEM

Tabla: DM_ITEM		
Campo	Tipo	Mapeo
ID_ITEM	Integer	Autogenerado
COD_ITEM	Char(20)	MST_ITEM_PT.ITEM_COD
COD2_ITEM	Char(20)	MST_ITEM_PT.ITEM_SEGUNDO_COD
ITEM	Varchar(50)	Mayuscula(MST_ITEM_PT.ITEM_NOMBRE)
FAMILIA	Varchar(50)	Mayuscula(MST_TIPO.NOMBRE_TIPO)
SUBFAMILIA	Varchar(50)	Mayuscula(MST_TIPO.NOMBRE_TIPO)
COSTO	Float	MST_ITEM_PT.ITEM_COSTO
VENTA	Float	MST_ITEM_PT.ITEM_VENTA
PRECIO_LISTA	Float	MST_ITEM_PT.ITEM_PRECIO_LISTA
MODELO	Varchar(50)	Mayuscula(MST_TIPO.NOMBRE_TIPO)
MARCA	Varchar(50)	Mayuscula(MST_TIPO.NOMBRE_TIPO)
COLOR	Varchar(50)	Mayuscula(MST_TIPO.NOMBRE_TIPO)
CLASE	Varchar(50)	Mayuscula(MST_TIPO.NOMBRE_TIPO)
PESO	Float	MST_ITEM_PT.ITEM_PESO
ESTADO	Varchar(50)	Mayuscula(MST_TIPO.NOMBRE_TIPO)

FT_DOCUMENTO_DETALLE

Tabla: FT_DOCUMENTO_DETALLE		
Campo	Tipo	Mapeo
ID_FECHA_COMPROBANTE	Integer	DM_FECHA.ID_FECHA
ID_HORADIA_COMPROBANTE	Integer	DM_HORADIA.ID_HORADIA
ID_FECHA_VENCIMIENTO	Integer	DM_FECHA.ID_FECHA
ID_HORADIA_VENCIMIENTO	Integer	DM_HORADIA.ID_HORADIA
ID_ITEM	Integer	DM_ITEM.ID_ITEM
ID_FORMA_PAGO	Integer	DM_TIPO.ID_TIPO
ID_TIPO_MONEDA	Integer	DM_TIPO.ID_TIPO
ID_CENTRO_COSTO	Integer	DM_TIPO.ID_TIPO
ID_TIPO_COMPROBANTE	Integer	DM_TIPO.ID_TIPO
ID_SOCIO_ESTRATEGICO	Integer	DM_SOCIO_ESTRATEGICO.ID_SOCIO_ESTRATEGICO
ID_TRANSPORTISTA	Integer	DM_SOCIO_ESTRATEGICO.ID_SOCIO_ESTRATEGICO
ID_VENDEDOR	Integer	DM_SOCIO_ESTRATEGICO.ID_SOCIO_ESTRATEGICO
CANTIDAD	Integer	DETALLE_CTA_CTE.ITEM_CANTIDAD
P_U_ORIGINAL	Float	DETALLE_CTA_CTE.ITEM_P_U

Tabla: FT_DOCUMENTO_DETALLE		
Campo	Tipo	Mapeo
C_U_ORIGINAL	Float	MST_ITEM_PT.ITEM_COSTO
P_U_SOLES	Float	DETALLE_CTA_CTE.ITEM_P_U_SOLES
P_U_DOLARES	Float	DETALLE_CTA_CTE.ITEM_P_U_DOLARES
MONTO_ORIGINAL	Float	DETALLE_CTA_CTE.MONTO_ORIGINAL
IGV_ORIGINAL	Float	DETALLE_CTA_CTE.IGV_ORIGINAL
NO_AFECTO_ORIGINAL	Float	DETALLE_CTA_CTE.NO_AFECTO_ORIGINAL
MONTO_SOLES	Float	DETALLE_CTA_CTE.MONTO_SOLES
IGV_SOLES	Float	DETALLE_CTA_CTE.IGV_SOLES
NO_AFECTO_SOLES	Float	DETALLE_CTA_CTE.NO_AFECTO_SOLES
MONTO_DOLARES	Float	DETALLE_CTA_CTE.MONTO_DOLARES
IGV_DOLARES	Float	DETALLE_CTA_CTE.IGV_DOLARES
NO_AFECTO_DOLARES	Float	DETALLE_CTA_CTE.NO_AFECTO_DOLARES

FT_MANTENIMIENTO

Tabla: FT_MANTENIMIENTO		
Campo	Tipo	Mapeo
ID_AUTOBUS	Integer	DM_AUTOBUS.ID_AUTOBUS
ID_FECHA_INICIO	Integer	ID_FECHA.ID_FECHA
ID_HORADIA_INICIO	Integer	DM_HORADIA.ID_HORADIA
ID_FECHA_FIN	Integer	ID_FECHA.ID_FECHA
ID_HORADIA_FIN	Integer	DM_HORADIA.ID_HORADIA
ID_MECANICO	Integer	DM_SOCIO_ESTRATEGICO.ID_SOCIO_ESTRATEGICO
ID_SUPERVISOR	Integer	DM_SOCIO_ESTRATEGICO.ID_SOCIO_ESTRATEGICO
ID_TIPO_MANTENIMIENTO	Integer	DM_TIPO.ID_TIPO
ID_CONCEPTO	Integer	DM_CONCEPTO.ID_CONCEPTO
ORDEN_TRABAJO	Char(14)	MANTENIMIENTO_PT.ORDEN_TRABAJO
IND_ATENCION	Integer	1
DURACION_ATENCION	Integer	(MANTENIMIENTO_PT.FECHA_FIN - MANTENIMIENTO_PT.FECHA_INICIO)

El detalle del mapeo de las tablas restantes se encuentra en el Anexo 13.

3.3. Diseño de Explotación

El usuario final debe interactuar a través de una interfaz que le permita el análisis y explotación adecuada de la información del Datamart. Se empleó como herramienta de explotación a la Consola de Usuario de Pentaho. Esta herramienta cuenta con una serie de ventajas:

- Interfaz familiar para los usuarios acostumbrados a programas de Microsoft.
- Permite organizar la información en temas que tienen relación a través de una lista de tabla pivote. Para el caso de un cubo es importante, pues facilita la búsqueda de las dimensiones y hechos requeridos.
- Se puede acceder desde la Web facilitando las consultas desde cualquier parte del mundo con una disponibilidad de 24 x 7.

En la Figura 32 se muestra un ejemplo de la distribución de los componentes que están presentes en la interfaz de usuario.

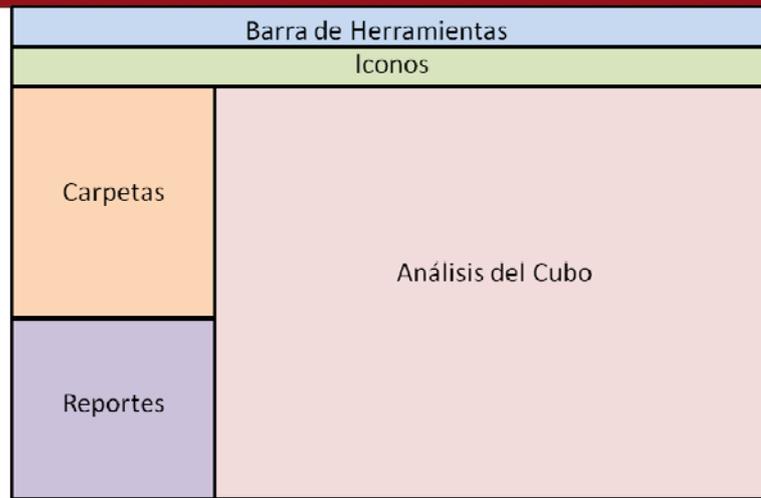


Figura 32: Distribución de las partes de la Interfaz de Usuario

El objetivo principal de dicha distribución es tener un fácil y rápido acceso a las carpetas que contienen los reportes. En la sección del análisis del cubo, el usuario puede seleccionar las dimensiones, medidas y realizar sus propios análisis de información.

El área de análisis del cubo tiene una forma similar a la que se muestra en la Figura 33. En este ejemplo, se puede observar dos dimensiones: Fecha y Tipo de Mantenimiento. En el gráfico de la figura se observa las atenciones de mantenimiento por trimestre del 2010 clasificados por Tipo de Mantenimiento.

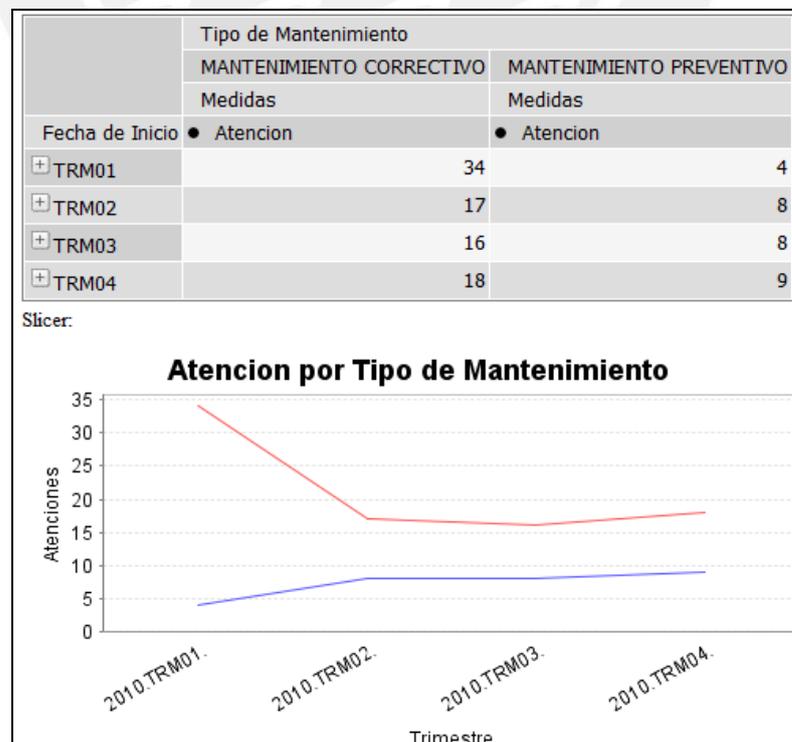


Figura 33: Ejemplo de Análisis de Cubo

3.3.1. Estándares de Reportes

Los reportes elaborados en Pentaho Report Designer forman parte de la interfaz gráfica debido a que los usuarios también interactúan con ellos. Por este motivo, es necesario establecer algunos estándares.

3.3.1.1. Configuración del Reporte

- Tamaño de Hoja: Web
- Orientación: Vertical u Horizontal (Depende del tipo de reporte)
- Márgenes: Se ocupará toda la ventana, dejando un margen del 10% a cada extremo.
- Imágenes: Formato PNG

3.3.1.2. Cabecera

Dato	Posición	Tamaño/ Tipo Letra	Color	Formato	Observación
Fecha del reporte	Izquierda superior	Arial 7	Negro	dd/mm/yyyy hh:mm	Fecha y hora de impresión del reporte.
Logo	Izquierda superior				Logo de la empresa.

Tabla 6: Tabla de campos de la cabecera

3.3.1.3. Pie de Página

- Número de página.
- Nombre físico del reporte.
- Nombre de la empresa.

3.3.1.4. Cuerpo del Reporte

- Reporte tipo lista

<Logo>		<Fecha del reporte>			
Titulo					
Filtro: <Dimensión X>					
XXXXX	0.00	0.00	0.00	0.00	0.00
XXXXX	0.00	0.00	0.00	0.00	0.00
XXXXX	0.00	0.00	0.00	0.00	0.00
XXXXX	0.00	0.00	0.00	0.00	0.00
XXXXX	0.00	0.00	0.00	0.00	0.00
TOTAL	0.00	0.00	0.00	0.00	0.00
Nombre: <Reporte>		Nombre de la empresa			Pág. ##

Figura 34: Reporte tipo lista

- Reporte tipo gráfico

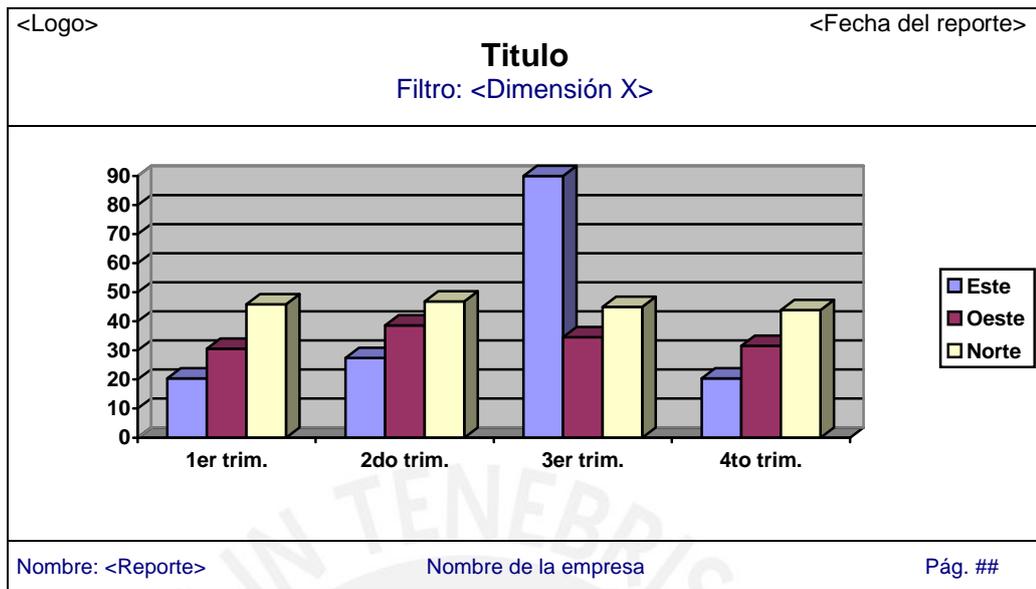


Figura 35: Reporte tipo gráfico

- Reporte híbrido

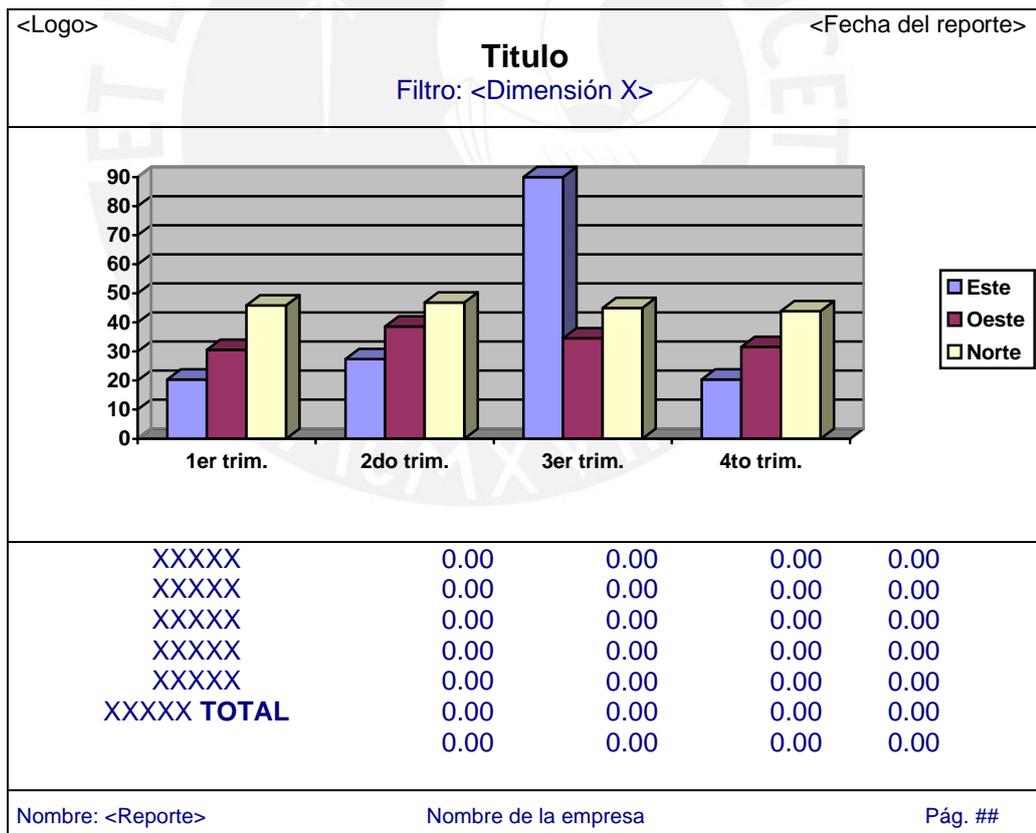


Figura 36: Reporte híbrido

En total se elaborarán siete reportes y 2 tableros de control que se muestran en la Tabla 7. El detalle de estos reportes y tableros de control se encuentra en el Anexo 15, Especificación de Reportes.

Tema	Reporte
Análisis de Logística	Productos por Catálogo
	Movimiento de Productos
	Compra de Productos
	Ranking de Proveedores
	Dashboard de Logística
Análisis de Mantenimiento	Evolución de Mantenimientos
	Atenciones por Autobuses
	Ranking de Atenciones por Conceptos
	Dashboard de Mantenimiento

Tabla 7: Reportes por Tema



4. Construcción

En el presente capítulo se mostrará los criterios y aspectos tomados en cuenta en la etapa de construcción. Además, se presentarán la estrategia, el plan y los tipos de prueba a utilizar.

4.1. Construcción

La fase de Construcción del proyecto está definida principalmente por la implementación de los procesos ETL debido a que los modelos de bases de datos fuentes y del Datamart ya están definidos en los capítulos anteriores. Adicionalmente, para la interfaz gráfica de usuario se empleará como herramienta de explotación a Pentaho. Sin embargo, la elección de dicha herramienta es el resultado de haber revisado y evaluado las herramientas del mercado.

4.1.1. Herramientas y soluciones en el mercado

A continuación se muestran las herramientas y soluciones existentes en el mercado que aportan el soporte tecnológico de Inteligencia de Negocios y son utilizadas por empresas obteniéndose resultados satisfactorios:

Base de Datos y Construcción de Data Warehouse

- Microsoft SQL Server (Microsoft Corporation)
- Oracle Express (Oracle Corporation)

- PostgreSQL (PostgreSQL Global Development Group)

Herramientas OLAP

- Microsoft SQL Server (Microsoft Corporation)
- Oracle Express (Oracle Corporation)
- Business Object (Business Object Corporation)
- MicroStrategy (MicroStrategy Corporation)
- Power Play (Cognos Corporation)
- Pentaho (Pentaho Corporation)

Consultas y Reportes

- Microsoft SQL Server (Microsoft Corporation)
- Oracle Discoverer (Oracle Corporation)
- Business Query (Business Object Corporation)
- MicroStrategy (MicroStrategy Corporation)
- Impromptu (Cognos Corporation)
- Microsoft Office Excel (Microsoft Corporation)
- Pentaho (Pentaho Corporation)

Dentro de este gran grupo de productos, teniendo en cuenta las posibilidades económicas como es el costo de las licencias, se realiza una preselección de las herramientas presentadas que son factibles de emplear para la implementación de la solución planteada. A continuación se presenta las herramientas preseleccionadas y sus principales ventajas.

- Microsoft SQL Server
- Oracle
- PostgreSQL
- Pentaho
- Cognos
- Microsoft Office Excel

4.1.1.1. Microsoft SQL Server

La compañía Microsoft ofrece la herramienta SQL Server 2000/2005/2008 como base de datos y como herramienta OLAP para dar soporte a las soluciones de Inteligencia de Negocios. Las principales ventajas de esta herramienta son:

- Facilidad de uso: Microsoft SQL facilita las tareas de administración de la base de datos debido a que automatizan la administración de la misma mediante una interfaz gráfica fácil de usar.
- Autoadministración dinámica: Automatiza muchas tareas de rutina como: los recursos de memoria, el tamaño de los archivos, etc.
- Integración: Permite la integración con los demás productos de la familia de Microsoft ampliamente conocidos como Windows y Microsoft Office.
- Desempeño, confiabilidad y escalabilidad.

4.1.1.2. Oracle

La compañía Oracle ofrece Oracle 9i/10g/11g que es una suite de herramientas que contiene desde base de datos hasta aplicaciones para soluciones de Inteligencia de Negocios. Las principales ventajas de esta suite son:

- Alta escalabilidad: Permite aumentar la capacidad de proceso de la información mediante los clusters.
- Menor tiempo en la recuperación de la información: Ya sea por errores humanos, fallos del sistema o de mantenimiento, Oracle tiene un componente que permite la recuperación de la información.
- Entorno integrado con Java y XML que permite el desarrollo de aplicaciones en estos lenguajes.
- Seguridad en la información.

4.1.1.3. PostgreSQL

PostgreSQL es un sistema de gestión de base de datos relacional orientada a objetos y libre. A pesar de ser un proyecto libre y manejado por una comunidad de desarrolladores, ha tenido un gran crecimiento y en los últimos años convirtiéndose en una base de datos confiable y robusta. Sus principales características son:

- Alta concurrencia: Permite acceder a una tabla que está siendo accedida por otro proceso sin necesidad de bloqueos.
- Amplia variedad de tipos nativos como texto largo ilimitado, figuras geométricas, direcciones IP, entre otras.

4.1.1.4. Pentaho

Pentaho es una plataforma libre de Inteligencia de Negocios orientada a la solución a través de sus distintos módulos: Reporting, Analysis, Dashboards, Data Mining e Integración de Datos. Sus principales ventajas son:

- Plataforma basada en Java.
- Acepta una amplia gama de gestores de base de datos como SQLServer, Oracle, MySQL, PostgreSQL, entre otros.
- Costo en licencias es cero.

4.1.1.5. Cognos

Cognos ha desarrollado herramientas para proveer la información requerida en la toma de decisiones. Este permite la creación de cubos, de reportes, realizar minería de datos y Balanced Scorecard. Cognos suite ofrece un conjunto de herramientas que se dividen en herramientas de análisis y de reportes. Las principales características de esta suite son:

- Soporta la información desde empresas grandes hasta pequeñas: Posee una arquitectura que permite integrarse al entorno de cualquier empresa.
- Entorno de usuario integrado, completo y sencillo: A través de un portal web se ofrece a los usuarios diferentes opciones dependiendo de sus necesidades como: informes predefinidos, realizar consultas, visualización de información a través de cuadros de mando o indicadores.

4.1.1.6. Microsoft Office Excel

Si bien es una herramienta de oficina de Microsoft, las últimas versiones de Excel como Xp, 2003 y 2007 han incorporado facilidades de visualización de datos que tienen como fuente de datos a los cubos o modelos multidimensionales. Además, se integra con el Microsoft SQL Server. Las principales características de esta herramienta son:

- Facilidad de uso para los usuarios: Muchos usuarios ya están familiarizados con los programas de Microsoft. Además cuenta con una interfaz muy intuitiva.
- Permite adecuar los reportes y gráficos como el usuario desee.
- Permite realizar operaciones de consulta multidimensional como Slicing, Dicing, Drilling, etc.

- Se integra fácilmente a otras herramientas de Inteligencia de Negocios.

De las herramientas preseleccionadas, las tres primeras, SQL Server, Oracle y PostgreSQL, tienen como objetivo soportar el modelo multidimensional del Datamart. Además, poseen facilidades para los procesos ETL. La cuarta es una herramienta para los procesos ETL, soporte OLAP y visualización de los datos. Las dos últimas herramientas, Cognos y Excel, tienen como objetivo dar soporte al usuario final en las tareas de Inteligencia de Negocios. Ambas se conectan a una base de datos y aportan facilidades en la visualización y análisis de los datos.

En la Tabla 8 se muestra un cuadro comparativo de las características herramientas preseleccionadas anteriormente.

Criterios	SQL Server	Oracle	Postgre SQL	Pentaho	Cognos	Excel
Soporte para Base de Datos o Data Warehouse	Alto	Alto	Alto	Nulo	Nulo	Nulo
Soporte para proceso ETL	Alto	Alto	Nulo	Alto	Alto	Nulo
Soporte de herramientas OLAP	Alto	Alto	Nulo	Alto	Alto	Bajo
Soporte al acceso de datos	Bajo	Alto	Medio	Alto	Alto	Alto
Soporte a la seguridad	Alto	Alto	Medio	Medio	Alto	Bajo
Facilidad de uso	Alto	Medio	Medio	Medio	Bajo	Alto
Facilidad de integración con otras herramientas	Alto	Alto	Medio	Alto	Alto	Alto
Facilidad de administración	Alto	Medio	Medio	Medio	Medio	Alto
Costo de la herramienta	Alto	Alto	Nulo	Nulo	Alto	Nulo
Adaptabilidad al proyecto	Alto	Bajo	Alto	Alto	Bajo	Alto

Tabla 8: Cuadro comparativo de las herramientas preseleccionadas

Se observa en la Tabla 8 que los costos del PostgreSQL y Pentaho son nulos, esto se debe a que son herramientas libres. Además, la empresa de transporte público ya tiene la licencia de Excel dado que trabaja con ese programa en la actualidad, por esta razón también su costo es nulo. Esto no sólo permite un ahorro sino también un menor tiempo en la enseñanza de la utilización de la herramienta.

Teniendo en cuenta los criterios de comparación presentados y el costo de las herramientas, las herramientas seleccionadas son: PostgreSQL para soportar

el modelo multidimensional del Datamart y Pentaho para el soporte OLAP, procesos ETL y la explotación de datos.

4.1.2. Configuración de las herramientas

Para la puesta en producción de la solución es necesario instalar y configurar adecuadamente las herramientas a utilizar. A continuación los programas que deben ser instalados en el servidor:

- Motor de base de datos PostgreSQL (para el Datamart)
- Pentaho Data Integration (para el proceso ETL)
- Pentaho Schema Workbench (para la elaboración de los cubos)
- Pentaho Business Intelligence Server (para la explotación de la información)
- Pentaho Report Designer (para la elaboración de los reportes)
- Community Dashboard Editor (para la elaboración de los tableros de control)

El detalle de cómo configurar estas herramientas se encuentra en el Anexo 16, Manual de Configuración.

4.1.3. Proceso ETL

Se implementará los procesos ETL a través de transformaciones y trabajos del Pentaho Data Integration.

A grandes rasgos deben implementarse dos tipos de procesos ETL. Un proceso que permita cargar toda la información histórica de la base de datos fuente a la base de datos del Datamart, el cual deberá ser ejecutado primero y por única vez. Luego una vez que se tiene el Datamart con información histórica ésta debe ser actualizada. Este proceso se conoce como carga periódica y debe ser un proceso reiterativo con una frecuencia que debe definir el área de mantenimiento y logística.

4.1.3.1. Proceso ETL para carga histórica

Este proceso debe poder cargar todas las dimensiones y hechos de la base de datos del Datamart desde que la empresa de transporte inició su negocio o desde que empezó a registrar sus datos.

Aspectos a tomar en cuenta:

- El proceso será implementado usando programación descendente, es decir, habrá un trabajo (*Job* en inglés) principal el cual deberá ser invocado para realizar la carga. Este trabajo llamará a dos trabajos, uno dedicado a cargar la información de las dimensiones y el otro la de los hechos. Cada uno de estos trabajos llamará a su vez a otros y así sucesivamente hasta implementar las transformaciones de las tablas en particular.
- El proceso tiene una base de datos intermedia que tiene una estructura similar a la base de datos del Datamart. Esto permite no saturar a la base de datos transaccional durante la ejecución del proceso ETL.
- El proceso de carga de la dimensión del periodo de tiempo es diferente al resto de las dimensiones debido a que no requieren de una tabla de la base de datos fuente. A este procedimiento se le debe indicar un año para que desde esa fecha hasta la actualidad genere los datos necesarios y cargue la dimensión de fechas. Este año debe ser el año en que empezó a registrar la información la empresa.
- Dado que se va a cargar información histórica, el proceso debe ser rápido, pero sobre todo debe permitir monitorear la carga de información. Ante alguna eventualidad se debe poder identificar hasta qué parte de la información pudo ser cargada a la base de datos del Datamart. El Pentaho Data Integration ya cuenta con herramientas que permiten hacer el seguimiento de sus procesos.

4.1.3.2. Proceso ETL para carga periódica

Este proceso debe cargar la nueva información de los hechos y actualizar las dimensiones en caso se haya agregado información nueva. Este proceso debe ser repetitivo teniendo una frecuencia variable. Se asumirá que la frecuencia mínima es el día, es decir, que al finalizar el día en la empresa se ejecutará dicho proceso para actualizar el Datamart.

Aspectos a tomar en cuenta:

- Su implementación será muy parecida al proceso de carga histórica debido a que se deben cargar dimensiones y hechos. Sin embargo, este proceso

tendrá parámetros de entrada que permitan cargar sólo la información del día. Estos parámetros están conformados por la fecha inicial (Año, mes y día) y fecha final, es decir, el rango de fechas del cual se desea cargar la información, es decir, al ejecutarse el proceso deberá indicarse estos valores.

- El proceso tiene una base de datos intermedia que tiene una estructura similar a la base de datos del Datamart. Esto permite realizar un comparativo con la información del Datamart e insertar o actualizar según sea conveniente.
- Para el caso de las dimensiones, el proceso deberá validar si ya existe el registro y sólo agregar los datos en caso no existan. Si bien los datos de las dimensiones podrían borrarse y volverse a cargar en su totalidad. Esto genera un error de llave foránea con los datos ya registrados en los hechos.
- Para el caso de los hechos, el proceso deberá borrar los registros del rango de fechas ingresado como parámetro y volverlos a cargar. Esto se hace para evitar registros duplicados y en caso sea necesario un reproceso.

A continuación se presentan algunos procesos ETL para la carga periódica:

- El proceso de la Figura 37 corresponde al ETL principal que realiza la carga de la información periódica. Está conformado por dos procesos. El primero se encarga de extraer los datos de las fuentes, realizar las transformaciones necesarias y cargarlos a una base de datos temporal. El segundo extrae la información de la base de datos temporal y carga la información, según el periodo y frecuencia, a la base de datos de Datamart.

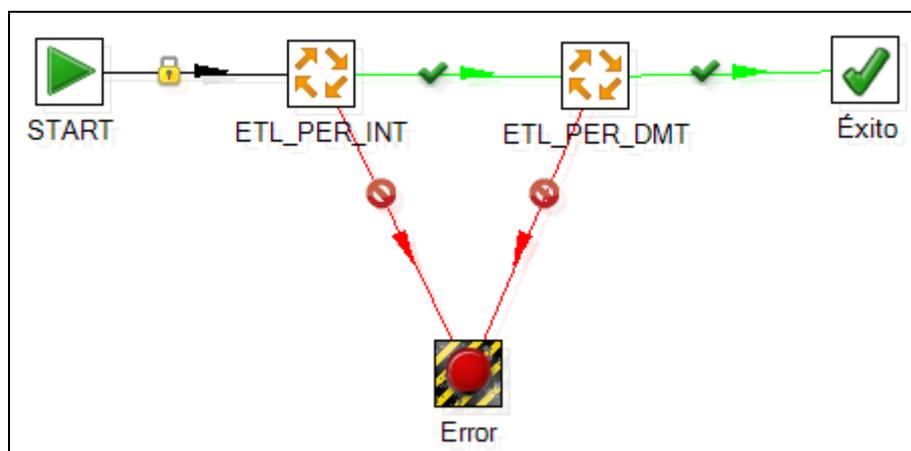


Figura 37: Proceso ETL principal de la carga periódica

- El proceso de la Figura 38 corresponde al ETL que realiza la extracción, transformación y carga de todas las dimensiones del Datamart. En algunos casos la fuente de información es una tabla de base de datos y en otros, un archivo de texto. La información de las dimensiones es cargada a una base de datos temporal.

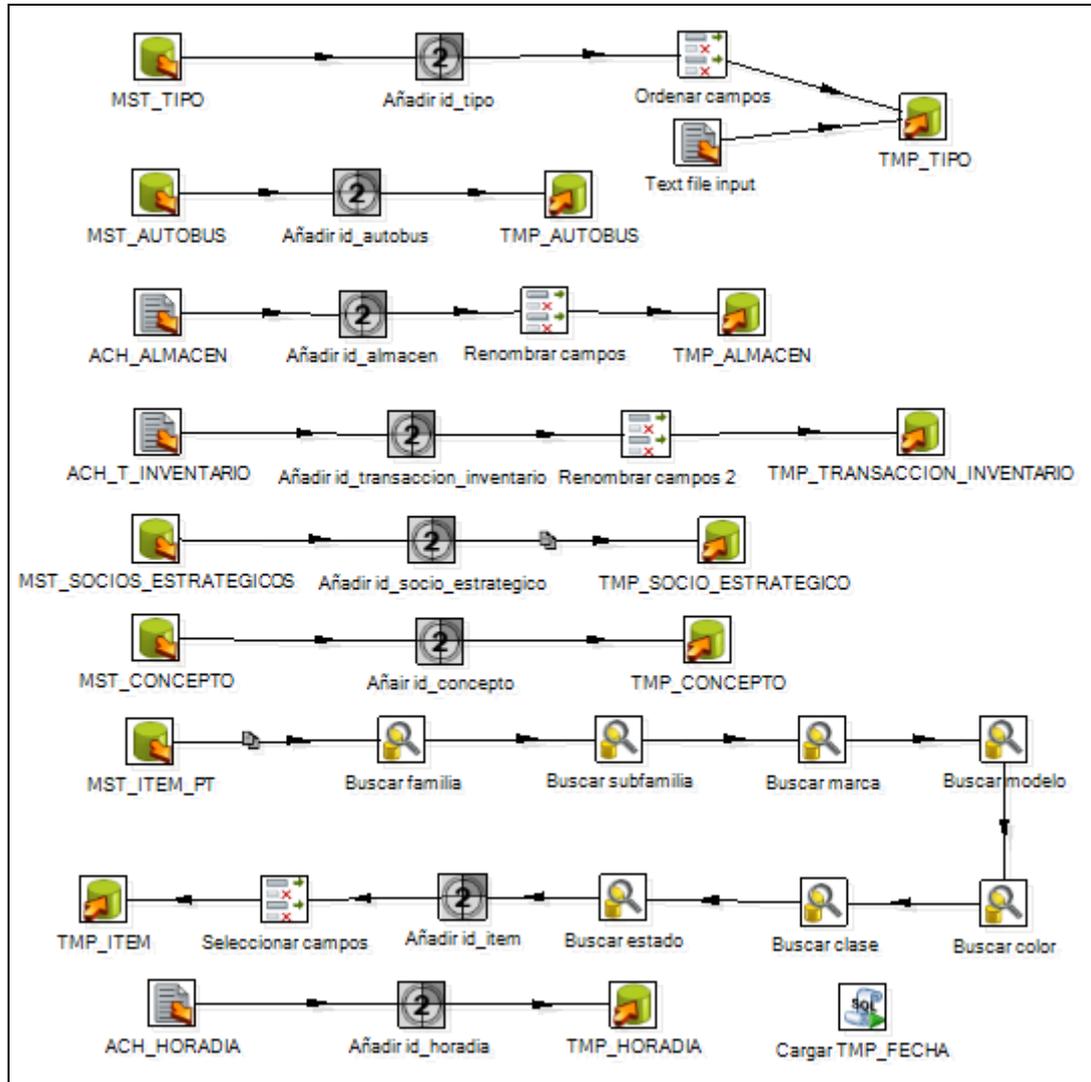


Figura 38: Proceso ETL de las dimensiones para la carga periódica

- El proceso de la Figura 39 corresponde al ETL que realiza la extracción, transformación y carga de la información de los documentos de pago. Extrae la información de la tabla del detalle de documentos de pagos y realiza las búsquedas y transformaciones necesarias para cargarla a las tablas de hechos de documentos del Datamart.

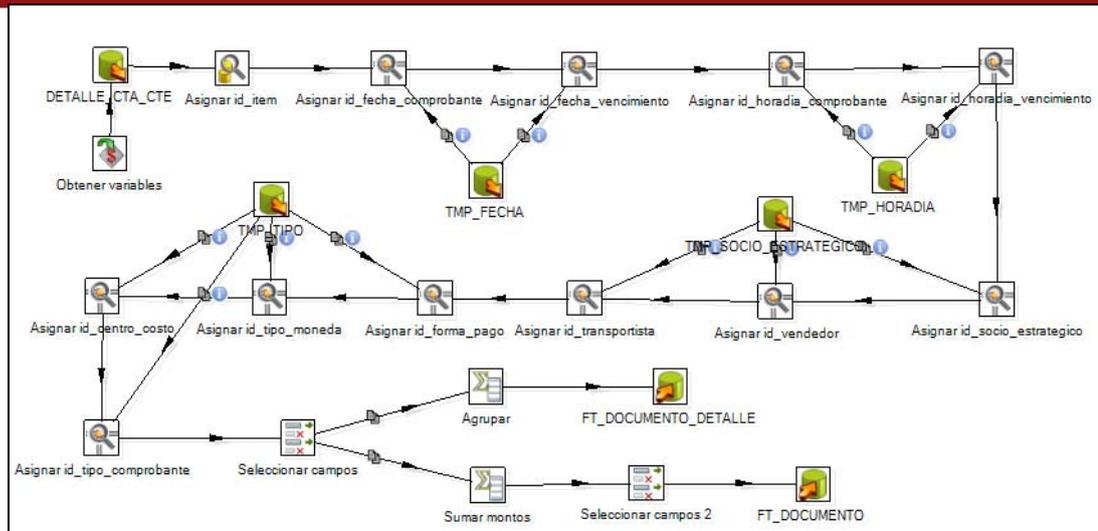


Figura 39: Proceso ETL de los hechos de documento para la carga periódica

El detalle y especificación de los procesos ETL se encuentran en el Anexo 13.

Además se presenta un reporte y un tablero de control de ejemplo:

- El reporte de la Figura 40 corresponde a un ranking de proveedores de ítems para el almacén. Está formado por un gráfico de barras horizontales que muestra de mayor a menor, el ranking de las empresas que son las principales proveedoras según el monto comprado. Debajo del gráfico se lista los datos principales de las empresas proveedoras. Además se ha empleado el uso de semáforos para los montos, de tal manera que a simple vista se pueda saber a qué empresas son las que se compra más.

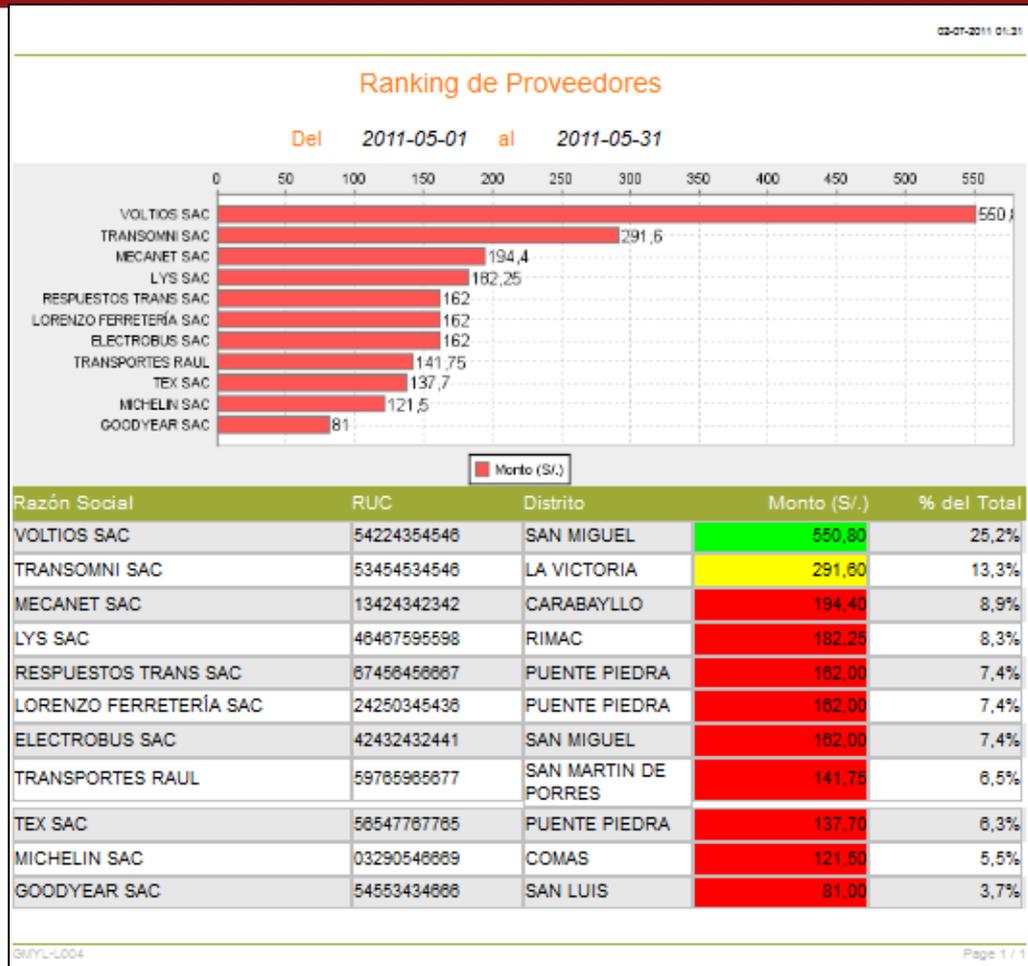


Figura 40: Reporte de Ranking de Proveedores

- El tablero de control de la Figura 41 muestra indicadores para el área de logística de la gerencia de Mantenimiento y Logística. El tablero está formado por 3 gráficos. El primero es un gráfico de barras verticales que muestra el total de las compras de ítems por mes para el año actual. El segundo es un gráfico del tipo pie que muestra la distribución de los productos que más se mueven dentro del almacén. Por último, el tercer gráfico es un resumen del reporte de la Figura 40 que mostraba el ranking de los proveedores, pero sin listar los datos principales de las empresas.

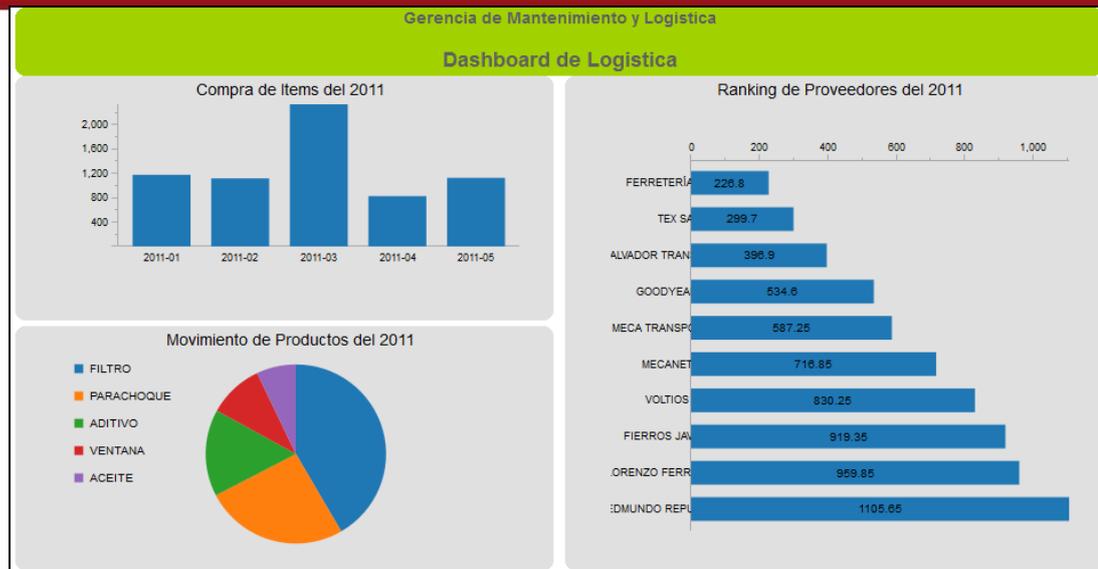


Figura 41: Tablero de control de Logística

Los datos y valores que se muestran en las Figuras 40 y 41 son datos ficticios que fueron creados para mostrar información en los reportes y tableros de control. Finalmente, el detalle y especificación de los reportes y tableros de control se encuentran en el Anexo 15.

4.2. Pruebas

Dentro del proceso de la construcción de la solución, las pruebas constituyen una de las actividades indispensables debido a que a través de ellas se asegura la calidad del producto. Por ello es necesario definir claramente una metodología de plan de pruebas que permita verificar los componentes de la solución y así asegurar que se ha construido un producto que cumpla con los requerimientos planteados.

La metodología del plan de pruebas está compuesta por:

- Planificación del plan de pruebas
- Diseño del plan de pruebas
- Determinación de los casos de prueba
- Ejecución del plan de pruebas
- Análisis y evaluación del plan de pruebas ejecutado

A continuación se detallará cada paso de la metodología definida.

4.2.1. Planificación del plan de pruebas

El primer paso es definir el objetivo del plan de pruebas. Este objetivo difiere de los planes de pruebas tradicionales elaborados para los proyectos de desarrollo de software debido a que el objetivo del presente plan está relacionado con la calidad de información que se procesa y entrega al usuario. Por ello, el objetivo está dado por asegurar la calidad de los datos del Datamart cubriendo específicamente los procesos ETL y asegurar la veracidad de la información que se muestra a través de la interfaz de usuario. Significa poder verificar que la información que se encuentra en los sistemas fuentes es igual a la información mostrada a través de la herramienta de explotación.

Para llevar a cabo el objetivo del plan se deben identificar los objetos sobre los que se ejecutará el plan de pruebas. Los objetos identificados son:

- Los procesos ETL, desde la base de datos fuente hasta la base de datos del Datamart.
- La base de datos del Datamart.
- Servicios OLAP.

Sobre estos objetos se realizarán las siguientes pruebas empleado el método de caja negra:

- Prueba de Extracción, Transformación y Carga desde la base de datos fuente a la base de datos del Datamart.
- Prueba de Funcionalidad en Servicios OLAP. Se prueba que la información en cada uno de los cubos corresponda a la información de la base de datos fuente.

Este plan de pruebas deberá ser ejecutado según el Cronograma del Proyecto. En la fase de Construcción dentro del entregable Plan de Pruebas se puede observar la actividad de la realización de pruebas.

Una vez ejecutado el plan de pruebas se registrará los resultados en un informe el cual es definido en el Diseño del plan de prueba.

4.2.2. Diseño del plan de pruebas

El plan tendrá la siguiente estructura: se definirán los casos de prueba los cuales se ejecutarán sobre los objetos de prueba. Se registrará la salida

obtenida y se comparará dicho resultado con la salida esperada para dicha entrada. De esta manera se determinará si cada objeto de prueba cumple con los criterios de calidad.

La Tabla 9 muestra los criterios de aceptación de manera general por cada objeto de prueba.

Objeto de Prueba	Proceso involucrado	Criterio
Proceso ETL y Base de Datos del Datamart	Extracción	Los datos extraídos de la base de datos fuente deben ser los mismos que los datos cargados en la base de datos del Datamart.
Servicios OLAP	Explotación	La información de los cubos debe corresponder a la obtenida de la base de datos del Datamart.

Tabla 9: Objetos de prueba y sus criterios de aceptación

4.2.3. Determinación de los casos de prueba

Se ha elaborado una plantilla que servirá para especificar los casos de prueba que conformarán el plan de pruebas del proyecto. Esta plantilla tendrá, a grandes rasgos, la siguiente información:

- Nombre del caso de prueba
- Objeto a analizar en el caso de prueba
- Objetivo del caso de prueba
- Datos de entrada del caso de prueba
- Resultado y salida del caso de prueba

La Tabla 10 muestra un ejemplo de un caso de prueba.

Caso de Prueba N° 1
Objeto a analizar del caso de prueba:
Proceso ETL y Base de Datos del Datamart.
Objetivo del caso de prueba:
Verificar que la tabla FT_DOCUMENTO_DETALLE sea correctamente cargada por medio del ETL teniendo como fuente la tabla DETALLE_CTA_CTE.
Pre-requisitos del caso de prueba:
<ul style="list-style-type: none"> - Disponibilidad de la BD Fuente. - BD Intermedia creada. - BD Datamart creada. - El job ETL_PER debe estar creado y listo para usarse.
Parámetros requeridos:
<ul style="list-style-type: none"> - Fecha inicio - Fecha fin
Datos de Entrada:
<ul style="list-style-type: none"> - Tabla MST_TIPO - Tabla MST_CATEGORIA - Tabla MST_SOCIOS ESTRATEGICOS - Tabla DETALLE_CTA_CTE

- Archivo ACH_HORADIA				
N°	Función a probar	Acción	Resultado esperado	Resultado obtenido
1	Cantidad de registros por tabla en la BD Fuente frente a la cantidad de registros por tabla en BD Datamart.	Ejecutar el job ETL_PER.	La tabla FT_DOCUMENTO_DETALLE muestre la misma cantidad de documentos que la tabla DETALLE_CTA_CTE. Para mayor detalle ver la sección de Resultados intermedios y finales del Anexo 14.	Se cargó los datos correctamente.

Tabla 10: Ejemplo de caso de prueba

Tanto los casos de prueba definidos como el plan de pruebas se muestran en el Anexo 14.

4.2.4. Ejecución del plan de pruebas

Una vez definidos todos los casos de prueba, se debe empezar con la ejecución de las mismas. El probador analiza el caso de pruebas siguiendo las instrucciones paso a paso indicadas en la especificación del caso de prueba. En caso se produzca un resultado incorrecto o inesperado se deben registrar en la parte de Resultado obtenido.

4.2.5. Análisis y evaluación del plan de pruebas ejecutado

Una vez culminada la ejecución de todos los casos de prueba es posible determinar si los objetos de prueba cumplen con los criterios de aceptación y por ende si el producto final cumple con los requerimientos. En caso hayan surgido casos de prueba con resultados incorrectos estos deben ser corregidos y atendidos hasta obtener los resultados esperados en todos los casos.

De manera de ejemplo, a continuación se presentan dos casos de prueba: el primero sobre el ETL del FT_DOCUMENTO_DETALLE (Proceso de Extracción) y el segundo sobre el Cubo de Mantenimiento (Proceso de Explotación):

4.2.6. Caso de Prueba para el ETL de FT_DOCUMENTO_DETALLE

El caso de prueba se presentó en la tabla 10.

Datos de Entrada:

- Tabla MST_TIPO: Se listan las columnas significativas para el proceso y una porción del número de registros.

cod_categoria	cod_tipo	nombre_tipo
003	00000000000001	NUEVOS SOLES
003	00000000000002	DÓLARES AMERICANOS
004	00000000000003	EFFECTIVO
004	00000000000004	A 30 DÍAS
004	00000000000005	A 60 DÍAS

- Tabla MST_SOCIOS ESTRATEGICOS: Se listan las columnas significativas para el proceso y una porción del número de registros.

cod_anexo	ruc_anexo	nombre_anexo
0001	84767393755	CONSORCIO VIA SAC
0002	46754965464	TRANS SANCHEZ
0003	46576978797	EL RAPIDO
0004	78768786787	ORION TRANSPORTES
0005	46565467	CÁCERES ARMANDO, JOSÉ LUIS

- Tabla DETALLE_CTA_CTE: Se listan las columnas significativas para el proceso y una porción del número de registros.

cod_mov_alm	nro_linea	fecha_comprobante	monto_original	item_cantidad
000000000000001	1	04/01/2010 08:34	113.4	4
000000000000001	2	04/01/2010 08:34	121.5	3
000000000000001	3	04/01/2010 08:34	413.1	3
000000000000002	1	05/01/2010 09:34	550.8	4
000000000000002	2	05/01/2010 09:34	413.1	3

Resultado intermedio:

Tabla fuente:

	cod_mov_alm character(14)	nro_linea integer	ruc_anexo character varying(11)	vendedor character varying(11)	fecha_comprobante timestamp without time zone	fecha_vencimiento timestamp without time zone	forma_pago character varying(14)
1	000000000000001	1	59765965677	05465576	2010-01-04 08:34:00	2010-01-11 08:34:00	00000000000003
2	000000000000001	2	59765965677	05465576	2010-01-04 08:34:00	2010-01-11 08:34:00	00000000000003
3	000000000000001	3	59765965677	05465576	2010-01-04 08:34:00	2010-01-11 08:34:00	00000000000003
4	000000000000002	1	46467595598	04654566	2010-01-05 09:34:00	2010-01-12 09:34:00	00000000000004
5	000000000000002	2	46467595598	04654566	2010-01-05 09:34:00	2010-01-12 09:34:00	00000000000004
6	000000000000003	1	56547767765	05465576	2010-01-05 10:00:00	2010-01-12 10:00:00	00000000000005
7	000000000000004	1	13424342342	04654566	2010-01-07 11:45:00	2010-01-14 11:45:00	00000000000006
8	000000000000005	1	42432432441	05465576	2010-01-09 11:34:00	2010-01-16 11:34:00	00000000000003
9	000000000000006	1	53454534546	04654566	2010-01-10 13:34:00	2010-01-17 13:34:00	00000000000004
10	000000000000007	1	54224354546	05465576	2010-01-10 17:34:00	2010-01-17 17:34:00	00000000000005

Figura 42: Tabla fuente de DETALE_CTA_CTE (173 registros)

ETL intermedio:

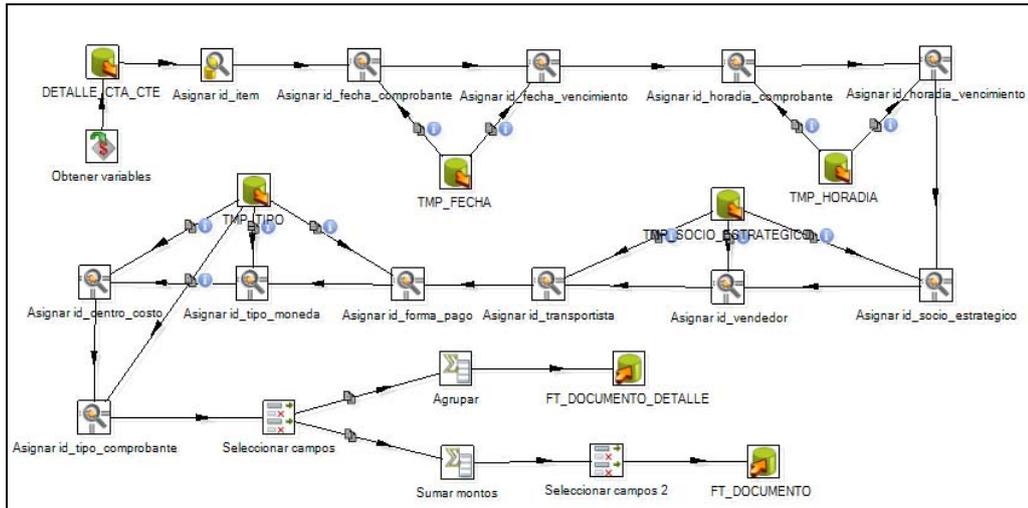


Figura 43: ETL para cargar la FT_DOCUMENTO_DETALLE

#	Stepname	Copynr	Read	Written	Input	Output
1	DETALLE_CTA_CTE	0	0	173	173	0
2	Asignar id_item	0	173	173	173	0
3	Asignar id_fecha_comprobante	0	904	173	0	0
4	TMP_FECHA	0	0	1462	731	0
5	Asignar id_fecha_vencimiento	0	904	173	0	0
6	Asignar id_horadia_comprobante	0	1613	173	0	0
7	TMP_HORADIA	0	0	2880	1440	0
8	Asignar id_horadia_vencimiento	0	1613	173	0	0
9	Asignar id_socio_estrategico	0	223	173	0	0
10	TMP_SOCIO_ESTRATEGICO	0	0	150	50	0
11	Asignar id_vendedor	0	223	173	0	0
12	Asignar id_transportista	0	223	173	0	0
13	Asignar id_forma_pago	0	247	173	0	0
14	TMP_TIPO	0	0	296	74	0
15	Asignar id_tipo_moneda	0	247	173	0	0
16	Asignar id_centro_costo	0	247	173	0	0
17	Asignar id_tipo_comprobante	0	247	173	0	0
18	Seleccionar campos	0	173	346	0	0
19	Sumar montos	0	173	170	0	0
20	Seleccionar campos 2	0	170	170	0	0
21	FT_DOCUMENTO	0	170	170	0	170
22	Agrupar	0	173	173	0	0
23	FT_DOCUMENTO_DETALLE	0	173	173	0	173
24	Obtener variables	0	1	1	0	0

Figura 44: Log del ETL para cargar la FT_DOCUMENTO_DETALLE

Tabla intermedia:

	id_fecha_comprobante integer	id_horadia_comprobante integer	id_fecha_vencimiento integer	id_horadia_vencimiento integer	id_item integer	id_tipo_moneda integer	id_centro_costo integer
1	4	515	11	515	1	1	7
2	4	515	11	515	2	1	7
3	4	515	11	515	24	1	7
4	5	575	12	575	23	2	8
5	5	575	12	575	32	2	8
6	5	601	12	601	6	1	9
7	7	706	14	706	7	2	7
8	9	695	16	695	8	1	8
9	10	815	17	815	9	2	9
10	10	1055	17	1055	10	1	7

Figura 45: Tabla intermedia FT_DOCUMENTO_DETALLE (173 registros)

Resultado final:

ETL final:



Figura 46: ETL final para cargar la FT_DOCUMENTO_DETALLE

#	Stepname	Copynr	Read	Written	Input	Output
1	Borrar campos 2	0	72	72	0	0
2	FT_DOCUMENTO_DETALLE	0	0	173	173	0
3	FT_DOCUMENTO_DETALLE_DMT	0	173	173	0	173
4	FT_DOCUMENTO_DETALLE_BOR	0	0	72	72	0
5	Obtener variables	0	1	1	0	0

Figura 47: Log del ETL final para cargar la FT_DOCUMENTO_DETALLE

Tabla final:

	id_fecha_comprobante integer	id_horadia_comprobante integer	id_fecha_vencimiento integer	id_horadia_vencimiento integer	id_item integer	id_tipo_moneda integer	id_centro_costo integer
1	4	515	11	515	1	1	7
2	4	515	11	515	2	1	7
3	4	515	11	515	24	1	7
4	5	575	12	575	23	2	8
5	5	575	12	575	32	2	8
6	5	601	12	601	6	1	9
7	7	706	14	706	7	2	7
8	9	695	16	695	8	1	8
9	10	815	17	815	9	2	9
10	10	1055	17	1055	10	1	7

Figura 48: Tabla destino FT_DOCUMENTO_DETALLE (173 registros)

Finalmente, al revisar los contenidos de la Fact se comprueba que los datos son los esperados.

4.2.7. Caso de Prueba para el Cubo de Mantenimiento

Caso de Prueba N° 2				
Objeto a analizar del caso de prueba:				
Servicios OLAP.				
Objetivo del caso de prueba:				
Verificar que el cubo de Mantenimiento muestre la misma información de la tabla FT_MANTENIMIENTO.				
Pre-requisitos del caso de prueba:				
<ul style="list-style-type: none"> - Disponibilidad de la BD Datamart. - Disponibilidad del cubo de Mantenimiento. 				
Parámetros requeridos:				
Ninguno				
Datos de Entrada:				
- Realizar la misma consulta:				
N°	Función a probar	Acción	Resultado esperado	Resultado obtenido
1	Valores en la BD Datamart frente a los valores del cubo.	Realizar la consulta en la base de	El cubo de Mantenimiento muestre la misma cantidad de atenciones para el año 2010 que FT_MANTENIMIENTO. Para mayor detalle ver la sección de Resultados del Datamart y del Cubo	Muestra la misma información.

		datos del Datamart y en el cubo.	del Anexo 14.	
--	--	----------------------------------	---------------	--

Tabla 11: Caso de Prueba para el Cubo de Mantenimiento

Resultado del Datamart:

	ano integer	mes character varying(20)	num_mes integer	atencion bigint
1	2010	ENERO	1	16
2	2010	FEBRERO	2	15
3	2010	MARZO	3	7
4	2010	ABRIL	4	5
5	2010	MAYO	5	10
6	2010	JUNIO	6	10
7	2010	JULIO	7	8
8	2010	AGOSTO	8	6
9	2010	SETIEMBRE	9	10
10	2010	OCTUBRE	10	11
11	2010	NOVIEMBRE	11	9
12	2010	DICIEMBRE	12	7
13	2011	ENERO	1	13
14	2011	FEBRERO	2	9
15	2011	MARZO	3	17
16	2011	ABRIL	4	6
17	2011	MAYO	5	11

Figura 49: Consulta en FT_MANTENIMIENTO

Resultado del Cubo:

Fecha Fin	Medidas
	● Atencion
+ 2010	114
+ ENERO	16
+ FEBRERO	15
+ MARZO	7
+ ABRIL	5
+ MAYO	10
+ JUNIO	10
+ JULIO	8
+ AGOSTO	6
+ SETIEMBRE	10
+ OCTUBRE	11
+ NOVIEMBRE	9
+ DICIEMBRE	7
+ 2011	56
+ ENERO	13
+ FEBRERO	9
+ MARZO	17
+ ABRIL	6
+ MAYO	11

Figura 50: Consulta en Cubo Mantenimiento

Finalmente, al revisar ambos resultados se comprueba que los datos son los esperados.

5. Observaciones, conclusiones y recomendaciones

En el presente capítulo se describirán algunas observaciones del proyecto, luego se describirán las conclusiones y finalmente algunas consideraciones y ampliaciones para trabajos futuros.

5.1. Observaciones

Luego de haber culminado el proyecto se plantea las siguientes observaciones:

- La guía del PMBOK propone procesos para planificar, ejecutar y controlar las actividades del proyecto y divide estos procesos en nueve áreas de conocimiento. Se sugiere abarcar las nueve áreas para una mejor administración del proyecto. Sin embargo, existen áreas que no se justifican utilizar para un proyecto de tesis donde el único ejecutante del proyecto es el tesista. Estas áreas son de Recursos Humanos y Comunicación. Adicionalmente, no es necesario elaborar todos los entregables propuestos en el PMBOK. Los entregables de Gestión de Proyecto que sí fueron elaborados en el presente proyecto están basados en los entregables descritos en la guía del PMBOK.
- La metodología DWEP es una metodología dedicada originalmente al desarrollo de un Data Warehouse y propone una gran cantidad de entregables. Dada la similitud entre un Data Warehouse y Datamart se pudo adaptar la metodología al

presente proyecto para construir un Datamart. Para ello, se le cambió los nombres a algunos entregables y dado al alcance que tenía el proyecto, algunos entregables no fueron elaborados debido a que no fue necesario. Sin embargo, los entregables y diagramas elaborados en el presente proyecto siguen fielmente las indicaciones que propone la metodología DWEP.

- Los procesos de ETL fueron implementados mediante procesos de Pentaho Data Integration y para ejecutarlos se debe ingresar a la interfaz de ejecución de esta herramienta. De manera similar debe ejecutarse el código de la creación de las base de datos del Datamart a través de la interfaz de PostgreSQL. No se implementó una interfaz adicional que permita al usuario crear la base de datos del Datamart ni ejecutar los procedimientos ETL.
- Las herramientas de Pentaho que se hacen mención en todo el documento corresponden a la versión de Pentaho Community, es decir, son herramientas elaboradas por el grupo desarrollador de la comunidad de Pentaho y por lo tanto no requieren licencia. Por otro lado, Pentaho si ofrece una versión con licencia llamada Enterprise Edition, pero ésta no ha sido utilizada para el presente proyecto.
- En caso se desee implementar la solución en una empresa, es recomendable realizar una etapa previa al desarrollo e implementación del Datamart. La etapa que se hace mención es la Limpieza de Datos, la cual abarca el descubrimiento y correjimiento o eliminación los datos erróneos de una base de datos. Prescindir de dicha etapa en una solución de Inteligencia de Negocios puede ocasionar que el Datamart contenga datos erróneos o duplicados.

5.2. Conclusiones

Luego de haber culminado el proyecto se puede formular las siguientes conclusiones:

- Los procesos de Gestión del Proyecto que incluyeron las tareas de planificación, ejecución, seguimiento y control junto con los entregables propuestos por el PMBOK permitieron desarrollar el proyecto satisfactoriamente en el tiempo similar al estimado y con la calidad deseada. Por ello, es importante contar con procesos y entregables de Gestión de Proyectos durante el desarrollo de todo proyecto.

- La metodología DWEP logró guiar exitosamente toda la construcción de la solución. Su integración con la metodología RUP y los diagramas UML permitió documentar los requerimientos, análisis, diseño e implementación de la solución de manera clara y sin ambigüedades. Si bien UML está orientado a la construcción de sistemas orientados a objetos, se ha demostrado en el presente proyecto que, empleando estereotipos basados en UML, puede modelarse la construcción de una solución basada en Inteligencia de Negocios.
- Las necesidades de información del área de mantenimiento y logística de una empresa de transporte público fueron identificadas satisfactoriamente debido a que se consideró los posibles escenarios, actores y supuestos en toda empresa del mismo rubro. Esto contribuyó a identificar requerimientos claros y precisos que fueron documentados y utilizados para la construcción del modelo multidimensional.
- El modelo multidimensional de la solución logró abarcar las necesidades de información identificadas y fue representado utilizando diagramas de fácil comprensión que permitieron una correcta validación del mismo.
- Los procesos de extracción, transformación y carga de los datos lograron contar con un Datamart con datos correctos y coherentes provenientes de la base de datos fuente. Los procesos fueron implementados empleando estándares de programación y luego se verificó su funcionamiento a través de un plan de casos de prueba.
- La elección de la herramienta de explotación fue la adecuada debido a que permitió una fácil interacción con usuarios que estuviesen familiarizados con el uso de hojas de cálculo y sin conocimientos avanzados de computación.
- Los reportes y tableros de control elaborados permitieron mostrar la importancia de la explotación de la información y de los indicadores que generan una ventaja competitiva en las empresas de transporte público.
- La selección de la suite Pentaho como plataforma de solución fue la adecuada porque a través de sus herramientas se pudo abarcar las distintas etapas de una solución de Inteligencia de Negocios. Si bien la edición Community tiene algunas

carencias respecto a la versión Enterprise, se logró obtener un producto que cumpla como los requerimientos y la calidad deseada.

5.3. Recomendaciones y trabajos futuros

El presente proyecto puede tomarse como base para futuros proyectos o soluciones similares. Por ello, a continuación se describen cuatro ampliaciones posibles de la solución que pueden implementarse o tomarse en cuenta, ya sea incorporando nuevas herramientas, nuevos usuarios o nueva funcionalidad.

5.3.1. Implementar un componente de Inteligencia de Negocios basado en Balanced Scorecard

Un Balanced Scorecard, es una herramienta que permite traducir la visión de una organización, expresada a través de su estrategia, en términos y objetivos específicos, estableciendo un sistema de que permita medir los logros de dichos objetivos a través de indicadores.

Implementar un Balanced Scorecard para el área de mantenimiento y logística es relativamente sencillo debido que se cuenta con un Datamart que cuenta con información actualizándose periódicamente y con dimensiones y medidas que pueden ser reutilizadas como indicadores.

5.3.2. Implementar herramientas de Inteligencia de Negocios

Existen diversas herramientas de Inteligencia de Negocios que pueden incorporarse al Datamart construido. El objetivo principal de estas herramientas es brindar medios más efectivos y fáciles de utilizar para la generación de información útil de manera rápida, integrada y con la seguridad de contar con información consistente. A continuación se listan algunas herramientas a integrarse con el Datamart.

- Minería de Datos (*Data Mining*): Herramienta que explora las bases de datos en busca de patrones ocultos, tendencias y comportamientos encontrando información predecible que ni un experto puede llegar a encontrar fácilmente. Para el área de mantenimiento y logística, por ejemplo, podría encontrar tendencias del almacén, predecir comportamiento de fallas en los autobuses, entre otros.

- Mantenimiento de metadatos: Herramientas que ayudan a mantener la información y documentación acerca de las estructuras de datos de los sistemas transaccionales, procesos ETL y Datamarts. Permite tener un repositorio de los datos acerca de los propios datos y ante algún cambio en las bases de datos puede mostrar los impactos que implica dicho cambio.

5.3.3. Ampliación de áreas y departamentos

Entre las posibles ampliaciones de la funcionalidad se puede desarrollar Datamarts para las otras áreas de la empresa e ir construyendo un Data Warehouse corporativo. De esta manera, el alcance no sólo estará limitado por las necesidades de información del área de mantenimiento y logística sino que podría abarcarse áreas como Recursos Humanos, Contabilidad y tener una visión más completa de la organización.

A efectos de incorporar esta ampliación se debe tener en cuenta la integración de todos los Datamarts a construir para no tener Datamarts independientes y evitar la redundancia de información. Se recomienda elaborar un solo diseño integral del Data Warehouse e ir construyendo Datamarts según las necesidades de los usuarios, o en su defecto buscar integración como criterio primordial entre los Datamarts que se vayan desarrollando.

5.3.4. Adaptación a empresas similares

Si bien la solución satisface las necesidades de información de un área de mantenimiento y logística de una empresa transporte público, ésta puede adaptarse a otras empresas de transporte. Existen similitudes en cuanto a necesidades de información, dimensiones, medidas en toda empresa de transporte, ya sea de personas o de carga. Este proyecto puede tomarse como base para satisfacer los requerimientos del área de mantenimiento y logística de una empresa de transporte aéreo, marítimo o terrestre. Asimismo se podría aplicar para una empresa de trenes, teniendo en cuenta que desde el 2011 empieza a funcionar el tren eléctrico de Lima como transporte público.

A efectos de desarrollar una solución para una empresa de transporte se debe tener en cuenta los requerimientos específicos de la empresa e identificar claramente las dimensiones y medidas que apliquen al contexto de la empresa de transporte escogida.

Bibliografía

- [ACM 2006] ACM
Communications of the ACM: Modeling Strategies and
Alternatives for Data Warehousing Projects
- [BIE 2010] BIEXPERTS
<http://www.biexpertsia.com/transporteurbano.php>
- [BIT 2002] BITAM,
“Business Intelligence”.
<http://www.bitam.com/spanish/AcercaDeBI.htm>
- [DIA 2002] DÍAZ, WLADIMIRO,
<http://www.uv.es/~diaz/>
Departament d'Informàtica,
Universitat de València, València, España.
- [GTU 2008] GERENCIA DE TRANSPORTE URBANO
<http://www.gtu.munlima.gob.pe/proyectos/concesionrutas.htm>
Gerencia de Transporte Urbano de la Municipalidad
Metropolitana de Lima
Lima, Perú
- [HER 2003] HERNANDEZ, JOSE,
<http://www.dsic.upv.es/~jorallo/cursoDWDWM/dwdm-I.pdf>
Departamento de Sistemas Informáticos y Computación
Universidad
Politécnica de Valencia, España.
- [INM 2002] INMON, W.H,
Building the Data Warehouse,
Tercera Edición. John Wiley & Sons, 2002.
- [JRS 2011] JR SOFTWARE
<http://www.softwarejr.com.ar/software-transporte.htm>

Santa Fe, Argentina

- [KIM 2002] KIMBALL, RALPH; ROSS, MARGY
The Data Warehouse Toolkit
- [LUJ 2006] LUJAN-MORAN, SERGIO; TRUJILLO, JUAN,
“Applying the UML and the Unified Process to Design of Data
Warehouses”
- [MRV 2010] MRVISUAL CORPORATION SAC
<http://www.mrvisual.com/>
Lima, Perú
- [PMI 2004] PROJECT MANAGEMENT INSTITUTE
Guía de los Fundamentos de la Dirección de Proyectos
Tercera Edición, 2004
- [ULL 1999] ULLMAN, J.D.; Widom, J.
“Introducción a los Sistemas de Bases de Datos”. Prentice
Hall, 1999.
- [YDI 2004] YDIRÍN, PRÉSTAMO,
http://140.148.3.250/u_dl_a/servlet/mx.udlap.ict.tales.html,
Escuela de Ingeniería, Universidad de las Américas, Puebla,
México.