

Universidad de las Ciencias Informáticas

FACULTAD 6



“Mercado de datos para el área de Planeación de la Empresa de Mantenimiento de Grupos Electrónicos Fuel-Oil.”

Trabajo de Diploma para optar por el Título de Ingeniero en Ciencias Informáticas

Autores: Yurislaine Pelegrin Cabrera.

Clara Elena Brizuela Figueredo.

Tutores: Ing. Arodys Eugenio Dominguez Vaillant.

Ing. Yoendris Lacoste Ricardo.

La Habana

Junio de 2012.

“Año 54 de la Revolución”.



"Seamos realistas y hagamos lo imposible."

DECLARACIÓN DE AUTORÍA

Declaramos ser autoras del presente trabajo “Mercado de datos para el área de Planeación de la Empresa de Mantenimiento de Grupos Electrógenos Fuel-Oil” y reconocemos a la Universidad de las Ciencias Informáticas (UCI) los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste se firma la presente a los ____ días del mes de _____ del año 2012.

Firma del Autor

Yurislaine Pelegrin Cabrera

Firma del Autora

Clara Elena Brizuela Figueredo

Ing. Arodys Eugenio Dominguez Vaillant

Firma del Tutor

Ing. Yoendris Lacoste Ricardo

Firma del Tutor

DATOS DE CONTACTO

Ing. Arodys Eugenio Dominguez Vaillant

Graduado en la Universidad de las Ciencias Informáticas en 2007.

E-mail: adominguez@uci.cu

Ing. Yoendris Lacoste Ricardo

Graduado en la Universidad de las Ciencias Informáticas en 2007.

E-mail: ylacoste@uci.cu

AGRADECIMIENTOS

De Clara Elena:

Muchas fueron las personas que de una forma u otra colaboraron con la realización de este Trabajo de Diploma, he recibido mucha confianza y apoyo de mi familia, amigos y compañeros de laboratorio.

A mis padres, mi abuela Georgina, mi tía Alina y muy especial a mi abuelito Cevero por ser el motor impulsor y motivarme a alcanzar mis metas.

A mis viejos y nuevos amigos por ser incondicionales y acompañarme en los momentos buenos y malos, en especial a Yadira, Soto, Lissette, Yudelkis, Alberto, Adonis y Diosmani.

A mis queridos tutores Arodys y Yoendris por haberme guiado durante la realización de este trabajo.

A todos los profes del centro DATEC, Nara, Leonel, Elio y Fabián por brindarme apoyo día a día a lo largo de estos años, por compartir tantas cosas buenas y su ayuda incondicional.

A todos los profesores que de una forma u otra han contribuido a mi formación para la vida.

A mis compañeros por compartir estos maravillosos años de mi vida.

En general a todos aquellos que verdaderamente se han preocupado por mí y me han ayudado a salir adelante.

Gracias por todo.

De Yurislaine

Primeramente quisiera agradecerle a Dios por permitirme realizar mi sueño y darme fuerzas para poder llegar hasta donde estoy.

A mi madre que aunque sé que las palabras no me van a alcanzar para decirle todas las cosas que ha hecho para que yo hoy logre este sueño, porque más que mío es de ella...Te tengo que agradecer por tus consejos, por tus mimos, por tu dedicación, por tu amor, por tus regaños, por siempre compartir todo conmigo, las alegrías, las tristezas, las lagrimas, por ser mi mejor amiga, mi paño de lágrimas, pero sobre todo por decirme cuando lo mas he necesitado: tu si puedes. A ti mamita...mil gracias.

A mi querida hermana Araimita, por ser mi motivación para ser como soy y poder servirle de ejemplo,

A mi abuela mami por siempre consentirme, complacerme y apoyarme.

A mi abuelito papi que siempre fue muy especial conmigo y hoy que no está físicamente entre nosotros se ha convertido en mi ángel de la guarda y sé que en estos momentos está muy orgulloso de mí.

A mis tíos Pipo, Tatá, Toti y Chuchi por la confianza que siempre tuvieron en mi.

A mi papá Juan, que ha sabido ser más que un padre para mí y siempre me ha brindado mucho amor, cariño y sobre todo mucho apoyo para seguir adelante.

Al resto de mi familia que siempre de alguna manera me han apoyado. A mi tía Mili, Magela, Cherito, Julia, Adelaida, a mi querida prima Yaneiris porque es mi ejemplo a seguir. Te quiero mi primis, eres mi ídolo.

A mi dúo de tesis que apenas sin conocernos hicimos muy buen equipo de trabajo y gracias a todo nuestro esfuerzo es el fruto de este trabajo. Gracias.

A mis tutores Arodys y Yoendris por todo el apoyo que nos brindaron, por sabernos guiar en todo este tiempo y por soportar nuestras persistencias.

A Diosmani, Leonel, Fabián, Elio, por no cansarse y tener paciencia con nosotras cada vez que necesitamos de ellos, no cambien nunca, para mí son perfectos así, los quiero.

A todos amigos que conocí desde primer año y lo que he ido conociendo en el transcurso de toda la carrera, a Estela, Yelena, Tamara, Yoendy, Yaime, Orestes, Magyori, Evelin, Eugenio, Tahimy, Olivia, Yenei, Yadiris, Yisel, Gladis, mi amiguita Teresa, Katuska, Yasnelis, Popi, Flabio, entre muchos otros.

Muy especialmente a:

A mis amigas Yeni, Yamila y Marisleidis, que desde que nos conocemos siempre hemos sido más que amigas, por ser siempre el hombro en que me he apoyado, ese hombro fuerte que siempre está dispuesto para mí, por ser incondicionales conmigo. Por todas esas cosas lindas mil gracias.

A Marcos que ha sido una persona muy significativa en mi vida y ha sabido llegar muy hondo en mi corazón.

A Fifi que desde que la conozco ha sido otra Marisel aquí en la UCI. Gracias por todo, por ser mi confidente, mi paño de lágrimas, y sobre todo por el buen papel de madre que ha desempeñado conmigo. Te quiero mucho.

A Yordano que ha sido una persona a la que le tengo mucho cariño y le debo muchas cosas de las que soy hoy. Nunca cambies. Gracias a Ivis y a Lázaro por siempre estar pendientes de mis cosas y por preocuparse por mí como si yo fuera su hija también.

A todos muchas gracias.

DEDICATORIA

De Clara Elena

*A mi mamá y mis abuelos Georgina y Cevero,
por todo el sacrificio y esfuerzo realizado
en aras de hacer realidad mis sueños.*

De Yurislaine

*A mi adorada madre por ser un digno ejemplo a seguir, por
siempre estar ahí cuando más la necesito y por decirme siempre
tu si puedes. Te quiero mucho mima.*

RESUMEN

El Centro de Tecnología de Gestión de Datos (DATEC) de la Universidad de las Ciencias Informáticas (UCI) brinda apoyo a algunas empresas del país, entre las que se destaca la Empresa de Mantenimiento de Grupos Electrógenos Fuel-Oil (EMGEF). El departamento de Planeación es una de las áreas de la EMGEF, la cual se encarga de recopilar diariamente un gran volumen de información proveniente de cada una de las Centrales de todas las Unidades Empresariales de Base (UEB) con que cuenta; actualmente esta área dispone de un Sistema Estadístico para el Mantenimiento de Grupos Electrógenos Fuel-Oil (SEMGEF) que almacena y gestiona dicha información, pero esta no puede ser analizada desde diferentes perspectivas. Para darle solución a este problema se implementó el mercado de datos para el área de Planeación de la Empresa de Mantenimiento de Grupos Electrógenos Fuel-Oil, logrando un formato estándar, organizado y centralizado de los datos recogidos en todas las UEB de la empresa a lo largo de los años.

PALABRAS CLABES: DATEC, EMGEF, mercado de datos, Toma de decisiones, UCI, Unidades Empresariales de Base (UEB).

TABLA DE CONTENIDOS	
AGRADECIMIENTOS	I
DEDICATORIA	IV
RESUMEN	V
INTRODUCCIÓN	1
CAPÍTULO 1: FUNDAMENTOS TEÓRICOS	4
1.1. SISTEMAS PARA EL APOYO A LA TOMA DE DECISIONES	4
1.2. ALMACENES DE DATOS	8
1.3. INTELIGENCIA DE NEGOCIO	17
1.4. METODOLOGÍA DE DESARROLLO DE ALMACENES DE DATOS	17
1.5. HERRAMIENTAS PARA EL DESARROLLO DEL MERCADO DE DATOS	20
1.6. CONCLUSIONES DEL CAPÍTULO	24
CAPÍTULO 2 ANÁLISIS Y DISEÑO	25
2.1. ANÁLISIS	25
2.2. DISEÑO	35
2.3. CONCLUSIONES DEL CAPÍTULO	38
CAPÍTULO 3 IMPLEMENTACIÓN	39
3.1. IMPLEMENTACIÓN DEL MODELO DE DATOS	39
3.2. IMPLEMENTACIÓN DE LOS PROCESOS DE EXTRACCIÓN, TRANSFORMACIÓN Y CARGA	39
3.3. IMPLEMENTACIÓN DE LOS PROCESOS DE INTELIGENCIA DE NEGOCIO	41
3.4. CONCLUSIONES DEL CAPÍTULO	46
CAPÍTULO 4 PRUEBAS DE SOFTWARE	47
4.1. PRUEBAS DE LA SOLUCIÓN.....	47
4.2. DISEÑO DE CASOS DE PRUEBA	48
4.3. LISTA DE CHEQUEO	49
4.4. APLICACIÓN DE LA LISTA DE CHEQUEO	51
4.5. RESULTADOS DE LAS PRUEBAS	52
4.6. CONCLUSIONES DEL CAPÍTULO	53
CONCLUSIONES	54
RECOMENDACIONES	55

REFERENCIAS BIBLIOGRÁFICAS	56
BIBLIOGRAFÍA	58
ANEXOS	61
GLOSARIO DE TÉRMINOS	63

ÍNDICE DE FIGURAS

Figura 1: Relación entre los componentes de un Almacén de Datos.	10
Figura 2: Estructura de un Cubo OLAP.....	12
Figura 3: Esquema Estrella.	13
Figura 4: Esquema Copo de Nieve.	14
Figura 5: Esquema Constelación de Hechos.....	15
Figura 6: Diagrama de casos de uso del sistema.....	30
Figura 7: Modelo de datos.	37
Figura 8: Diseño de las transformaciones.....	37
Figura 9: Transformación del hecho ocurrencia.....	40
Figura 10: Implementación del trabajo general.	40
Figura 11: Implementación del cubo OLAP.	41
Figura 12: Mapa de navegación.	42
Figura 13: Vista de análisis.....	43
Figura 14: Consulta mdx.....	43
Figura 15: Tablero de mando: Ocurrencias por causas.....	44
Figura 16: Tablero de mando: Ocurrencias por causas y UEB.....	45
Figura 17: Caso de prueba Presentar información sobre ocurrencias.....	49
Figura 18: Resultado de la lista de chequeo.	53

INDICE DE TABLAS

Tabla 1: Descripción de los actores del sistema.....	26
Tabla 2: CU Realizar extracción de datos.....	30
Tabla 3: CU Realizar Transformación y Carga de los datos.....	32
Tabla 4: CU Presentar información sobre ocurrencias.....	33
Tabla 5: Tabla de hechos.....	36
Tabla 6: Matriz BUS.....	36
Tabla 7: Roles y permisos.....	38
Tabla 8: Esquemas y tablas del mercado.....	39
Tabla 9: Aplicación de la Lista de chequeo para el artefacto “Mapa Lógico”.....	51

INTRODUCCIÓN

La información siempre ha sido el recurso más valioso y la materia prima indispensable para el desarrollo, equilibrio y adaptabilidad de los sistemas empresariales, a través de ella se obtienen conocimientos útiles para las empresas que representan factores críticos en la determinación del éxito o fracaso de las mismas.

Con el tiempo el cúmulo de información obtenida es rico y abundante, haciéndose engorroso el proceso de organización, recuperación, análisis y manejo de esta. Una solución a estos problemas han sido los sistemas de información enfocados al apoyo de toma de decisiones, sistemas basados en inteligencia de negocio (SINNEXUS, 2011) para extraer una cierta inteligencia o conocimientos a partir de grandes volúmenes de información, algunos ejemplos de estos sistemas son los Sistemas de Soporte a las Decisiones en Grupo (GDSS), Sistemas de Información para Ejecutivos (EIS), entre otros sistemas.

Cuba no ha quedado al margen de estos desarrollos informáticos, y ha ido madurando en estos temas con gran aceptación. Uno de los resultados de este desarrollo en la rama de la informática fue la creación de la Universidad de las Ciencias Informáticas (UCI) en el año 2002, con el objetivo de formar profesionales capaces de informatizar los procesos fundamentales de la sociedad cubana. Actualmente la UCI cuenta con varios centros de desarrollo, distribuidos entre sus diferentes facultades, entre los que se encuentra el Centro de Tecnología de Gestión de Datos (DATEC), perteneciente a la Facultad 6. Este centro en estos momentos brinda apoyo a algunas empresas del país, entre las que se encuentra la Empresa de Mantenimiento de Grupos Electrógenos Fuel-Oil (EMGEF), la cual ha estado presentando problemas en el proceso de toma de decisiones.

El área de Planeación de la EMGEF recopila diariamente por vía telefónica un gran volumen de información proveniente de cada una de las centrales de todas las provincias del país; actualmente dicha área dispone del Sistema Estadístico para el Mantenimiento de Grupos Electrógenos Fuel-Oil (SEMGEF) que almacena y gestiona dicha información. Este sistema informático proporciona a los directivos de la empresa un conjunto de reportes que de alguna manera contribuyen a la toma de decisiones; pero dichos reportes no permiten ampliar un dominio de cómo marcha la empresa con respecto a otro período de tiempo determinado, ni que la información sea analizada desde distintas perspectivas.

Por todo lo anteriormente descrito se plantea como **Problema de la Investigación:** ¿Cómo contribuir al proceso de toma de decisiones en el área de Planeación de la Empresa de Mantenimiento de Grupos Electrónos Fuel-Oil?

Definiendo como **Objeto de estudio:** Los almacenes de datos; centrando su **Campo de acción:** Mercado de datos para el área de Planeación de la Empresa de Mantenimiento de Grupos Electrónos Fuel-Oil. Para dar respuesta al problema planteado se plantea como **Objetivo general:** Desarrollar un mercado de datos para el área de Planeación de la Empresa de Mantenimiento de Grupos Electrónos Fuel-Oil que contribuya a la toma de decisiones.

A partir del objetivo general se desglosan los siguientes **objetivos específicos:**

1. Fundamentar la selección de las herramientas, metodologías y tecnologías para el desarrollo de almacenes de datos.
2. Realizar el análisis y diseño del mercado de datos para el área de Planeación de la EMGEF.
3. Implementar el mercado de datos para el área de Planeación de la EMGEF.
4. Realizar pruebas al mercado de datos para el área de Planeación de la EMGEF.

Para el cumplimiento de los mismos se proponen las siguientes **tareas de investigación:**

1. Analizar los fundamentos teóricos que sustentan los sistemas informáticos de apoyo a la toma de decisiones.
2. Caracterización de las herramientas, metodologías y tecnologías para el desarrollo de almacenes de datos.
3. Levantamiento de requisitos del mercado de datos.
4. Descripción de los casos de uso del mercado de datos.
5. Definición de los hechos, las medidas y las dimensiones del mercado de datos.
6. Diseño del modelo de datos.
7. Definición de la arquitectura del mercado de datos.
8. Diseño del subsistema de integración.
9. Diseño del subsistema de visualización.
10. Diseño de los casos de prueba.
11. Implementación del subsistema de integración.
12. Implementación del subsistema de visualización.
13. Aplicación de las listas de chequeo.
14. Aplicación de los casos de prueba.

Posibles resultados

Mercado de datos para el área de Planeación de la Empresa de Mantenimiento de Grupos Electrógenos Fuel-Oil poblado y funcional.

Estructura del documento

El documento está estructurado en 4 capítulos.

El Capítulo 1: Está referido al fundamento teórico de los sistemas de apoyo a la toma de decisiones, sus características, así como sus ventajas y desventajas. Se realizará una investigación acerca de los diferentes sistemas de apoyo a la toma de decisiones, y con esta investigación se definirá una propuesta de solución al problema anteriormente planteado. A partir de esta propuesta se realizará la selección de las metodologías, herramientas y tecnologías a utilizar en el desarrollo del sistema de apoyo a la toma de decisiones seleccionada.

El Capítulo 2: Análisis y diseño del Mercado de datos para el área de Planeación de la Empresa de Mantenimiento de Grupos Electrógenos Fuel-Oil (EMGEF). En el capítulo se efectuará un estudio preliminar del negocio, se expondrán las necesidades de información, las reglas del negocio, la descripción de los casos de uso, se desarrollará la matriz BUS y se definirán los reportes candidatos.

El Capítulo 3: Implementación del Mercado de datos para el área de Planeación de la Empresa de Mantenimiento de Grupos Electrógenos Fuel-Oil (EMGEF). En este capítulo se lleva a cabo la implementación del modelo de datos, así como de los reportes candidatos.

El Capítulo 4: Realizar pruebas al Mercado de datos para el área de Planeación de la Empresa de Mantenimiento de Grupos Electrógenos Fuel-Oil (EMGEF). En el capítulo se realizará la configuración de la seguridad de los usuarios y las pruebas necesarias.

CAPÍTULO 1: FUNDAMENTOS TEÓRICOS

Durante el desarrollo de este capítulo se realizará un estudio sobre los almacenes de datos y los mercados de datos, así como sus características y los elementos que lo integran; se realizará un análisis de los componentes así como las metodologías y herramientas a utilizar en su implementación.

1.1. Sistemas para el apoyo a la toma de decisiones

Existen diferentes tipos de sistemas de apoyo a la toma de decisiones, entre los cuales se encuentran los Sistemas de Soporte a las Decisiones en Grupo (GDSS por sus siglas en inglés), Sistemas de Información para Ejecutivos (EIS por sus siglas en inglés), Sistemas Expertos de Soporte a la toma de decisiones (EDSS por sus siglas en inglés) y los Sistemas de Soporte a la Toma de Decisiones (DSS por sus siglas en inglés).

1.1.1. Sistemas de Soporte a las Decisiones en Grupo

Los GDSS se pueden definir como un Sistema de Información interactivo que facilita la solución de problemas pocos estructurados por un conjunto de tomadores de decisiones que trabajan conjuntamente (SOTO). Un GDSS está compuesto por:

- Hardware.
- Software.
- Personal.
- Procedimientos.

Entre las principales características de un GDSS se encuentran:

- Es un sistema fácil de aprender y de utilizar, incluso para personas de diferentes niveles de preparación o con pocos conocimientos en el área de Informática.
- Es un sistema diseñado y desarrollado específicamente para el apoyo a las decisiones en grupo.
- El sistema puede ser específico para un problema en particular o general, capaz de apoyar la solución de diversos problemas.

Principales ventajas de los GDSS:

- Ayuda al tomador de decisiones a formular su escenario de decisión, incluyendo las diferentes alternativas de solución que satisfacen las necesidades y objetivos requeridos.
- Provee la facilidad de que el tomador de decisiones evalúe el impacto de cada una de las alternativas de solución.
- Permite al tomador de decisiones interpretar el impacto de las decisiones y analizar los objetivos planteados.

Principales desventajas de los GDSS:

- Falta de costumbre al utilizar un sistema para soportar el proceso de toma de decisiones.
- Resistencia al cambio por parte de los administradores.
- La responsabilidad al tomar una decisión puede diluirse (SOTO).

1.1.2. Sistemas de Información para Ejecutivos

Los EIS se pueden definir como un sistema que presenta información interna y externa en forma gráfica, que permite profundizar en datos detallados y capacita al usuario para manipular fácilmente la información. Su propósito es apoyar el proceso de toma de decisiones de los altos ejecutivos de una organización y presentar la información relevante usando recursos visuales y de fácil interpretación con el objetivo de mantenerlos informados (VALENCIA, 1969).

La mayoría de las veces tiende a confundirse el concepto de EIS con el Sistema de Soporte a la Toma de Decisiones (DSS), sin embargo, la diferencia estriba en que los EIS son presentadores de información y van dirigidos a la alta gerencia de la organización.

Las características básicas que debe reunir un EIS son:

- Consolidación de la información interna y externa.
- Diseño altamente gráfico y visual.
- Permite al alto ejecutivo realizar inferencias y extrapolaciones.
- Disponibilidad (VALENCIA, 1969).

1.1.3. Sistemas Expertos de Soporte a la toma de Decisiones

Los sistemas expertos, también llamados sistemas basados en conocimiento, utilizan redes neuronales para simular el conocimiento de un experto y utilizarlo de forma efectiva para resolver un problema concreto (SINNEXUS, 2011).

Entre sus características se encuentran:

- Utilización de varios expertos dentro del trabajo rutinario.
- Las decisiones que se toman son complejas y siguen una secuencia lógica.
- Las decisiones lógicas, así como las soluciones del problema, pueden expresarse o traducirse a reglas heurísticas.
- El conocimiento que se está modelando se encuentra bien delimitado y es profundo, no amplio y superficial.
- El problema no tiene solución analítica, al contrario, se sugiere la solución a través de técnicas analíticas.

Principales desventajas de los EDSS:

- Requieren consultoría especializada.
- Contratación de ingenieros del conocimiento.
- Elevados costos de implantación.
- Elevados costos involucrados con el mantenimiento y seguimiento del sistema.

1.1.4. Sistema de Soporte a la Toma de Decisiones

Son los más utilizados en todo el mundo, y se definen como el conjunto de programas y herramientas que permiten obtener de manera oportuna la información que se requiere durante el proceso de toma de decisiones que se desarrolla en un ambiente de incertidumbre.

Los DSS nacen como una necesidad de las empresas para tener información exacta cuando es requerida, reducir costos, realizar cálculos rápidos, eliminar límites cognitivos y de almacenamiento de información y para un soporte rápido y eficiente. Como sistemas informáticos, consisten habitualmente en varios componentes: bases de datos fuentes, proceso de Extracción-Transformación-Carga de datos, Data Warehouse (DW) o Almacenes de Datos, herramientas de Procesamiento Analítico en Línea (OLAP), bases de datos multidimensionales y otras herramientas de análisis de información.

Entre sus características se encuentran:

- No requiere conocimientos técnicos para su utilización.
- Rapidez en el tiempo de respuesta, ya que la base de datos subyacente suele ser un almacén de datos o un mercado de datos.
- Integración entre todos los sistemas/departamentos de la compañía.
- Cada usuario dispone de información adecuada a su perfil.
- Disponibilidad de información histórica.

Descripción de algunos componentes de los DSS:

Almacén de datos: es un almacén centralizado de información, que contiene datos que se integran de las diferentes fuentes de información de una organización, y que surge de la necesidad de un soporte de almacenamiento a las aplicaciones de análisis de datos, por lo tanto, se incluyen herramientas para la inteligencia empresarial, herramientas para extraer, transformar y cargar datos en el almacén de datos, y herramientas para gestionar y recuperar los metadatos (LÓPEZ, 2008).

Mercado de datos: son pequeños almacenes de datos centrados en una determinada línea de negocio.

Proceso de Extracción-Transformación-Carga de datos: permite el desarrollo rápido de nuevas aplicaciones que requieren información de múltiples fuentes.

Procesamiento Analítico en Línea (OLAP): permite a los usuarios analizar diferentes dimensiones de datos multidimensionales, con el objetivo de agilizar la consulta de grandes cantidades de datos.

Sobre algunos de estos componentes de los DSS se profundizará más adelante en este capítulo.

Principales ventajas de los DSS:

- No son costosos.
- Disponen de una gran variedad de herramientas en el mercado que operan en el ambiente de microcomputadoras.
- Presentan muy baja dependencia de personas que se encuentran fuera del control del tomador de decisiones.

Después de haber estudiado algunas de las principales características, componentes, ventajas y desventajas de algunos sistemas de apoyo a la toma de decisiones, se decide que con el desarrollo de uno de los componentes de los Sistemas de Soporte a la toma de Decisiones se erradica el problema existente en la Empresa de Mantenimiento de Grupos Electrónos Fuel- Oil (EMGEF). La construcción de un mercado de datos, incluyendo metodologías para el desarrollo de almacenes de datos y herramientas para la inteligencia empresarial y la gestión de la información, permitirá el análisis y obtención de resúmenes e informes complejos, y la extracción de conocimiento.

1.2. Almacenes de Datos

1.2.1. Definición de Almacén de datos

Definición de Bill Inmon

Bill Inmon a principios de los noventa definió el tema de los Almacenes de Datos (Data Warehouse, DW): “Un almacén de datos es una colección de datos orientados por temas, integrados, variables en el tiempo y no volátiles para el apoyo de la toma de decisiones”. Orientados por temas porque los datos son estructurados por temas de interés para facilitar su análisis por parte de los usuarios; integrado debido a que los datos que se introducen en el almacén de datos se obtienen de una variedad de fuentes de datos y el proceso de Extracción, Transformación y Carga (ETL) es el encargado de su integración; variables en el tiempo porque es importante conocer la variación de los datos en el tiempo, por esto los datos siempre tienen que estar ligados a un instante de tiempo y no volátil porque en el almacén los datos pueden ser consultados pero no modificados, por tanto la información es permanente y la actualización consiste exclusivamente en la incorporación de nuevos datos.

Definición de Ralph Kimball

Ralph Kimball es reconocido a nivel mundial en el diseño de los almacenes de datos como creador del enfoque multidimensional. Se ha dedicado al desarrollo de su metodología para que este concepto sea bien aplicado en las organizaciones y se asegure la calidad en el desarrollo de estos proyectos. Define almacén de datos como "una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis". También fue Kimball quien determinó que un almacén de datos no era más que: "la unión de todos los Data Marts de una entidad" (KIMBALL y ROSS, 2002).

Arquitectura de un almacén de datos

La arquitectura de un almacén de datos se suele representar como varias capas a través de las cuales circulan datos, de modo que los datos de una capa se obtienen a partir de los datos de la capa previa. A partir de esta arquitectura se considera que el desarrollo de un almacén de datos se puede estructurar en un marco integrado por tres niveles que definen los diferentes diagramas empleados para modelar un almacén de datos:

Conceptual: define el almacén de datos desde un punto de vista conceptual, o sea, desde el mayor nivel de abstracción y contiene únicamente los objetos y relaciones más importantes.

Lógico: abarca aspectos lógicos del diseño del almacén de datos, como la definición de las tablas y claves, la definición de los procesos de Extracción, Transformación y Carga (ETL) y otros.

Físico: define los aspectos físicos del almacén de datos, como el almacenamiento de las estructuras lógicas o la configuración de los servidores que mantienen el almacén de datos.

1.2.2. Componentes de un almacén de datos

Los almacenes de datos están compuestos por una serie de procesos que definen, en su conjunto, el ambiente que estos poseen. Aunque cada desarrollo de un almacén de datos es diferente debido a las características de las organizaciones, generalmente cumplen con los siguientes componentes descritos en la figura que a continuación se muestra (KIMBALL y ROSS, 2002).



Figura 1: Relación entre los componentes de un Almacén de Datos.

El primer componente hace referencia al **Área de procesamiento**, encargada de los datos temporalmente y permite la realización de un conjunto de procesos comúnmente llamados de Extracción, Transformación y Carga (ETL). Además, constituye la interfaz entre las fuentes operacionales y el área de presentación.

El segundo componente es el **Área de almacenamiento**. En este componente los datos se encuentran organizados, almacenados y disponibles para ser consultados, reportados o analizados por parte de los usuarios finales. Generalmente esta área es referenciada como una serie de mercados de datos integrados donde cada uno se encuentra representando a un proceso específico del negocio.

El tercer componente lo constituyen las **Herramientas de acceso a datos**. En este componente se usa la palabra herramientas para referirse a la variedad de capacidades que pueden ser provistos a los usuarios del negocio para el soporte a la toma de decisiones. Su actividad principal es la de consultar el área de presentación del almacén de datos. El mismo puede abarcar desde una simple o personalizada herramienta de consulta hasta una compleja y sofisticada aplicación de modelado o de minería de datos¹.

¹ Proceso conocido como *Data Mining*, consiste en la extracción no trivial de información que reside de manera implícita en los datos. En otras palabras, la minería de datos prepara y explora los datos para sacar la información oculta en ellos.

1.2.3. Motivos para desarrollar un Almacén de datos

Existen muchas ventajas por las que es recomendable usar un almacén de datos y por las que una empresa debe decidir implantarlo. Entre ellas se pueden citar las siguientes:

- Los almacenes de datos facilitan el funcionamiento de las aplicaciones de los sistemas de apoyo a la decisión tales como informes de tendencia, que muestran los resultados reales frente a los objetivos planteados.
- Los almacenes de datos pueden trabajar de manera conjunta para así aumentar el valor operacional de las aplicaciones empresariales.
- Acceso a toda la información de la empresa. La información que proviene de sistemas de origen diferente se consolida, sin importar si provienen de la misma o varias fuentes.
- Consistencia de la información al consolidarla desde varios departamentos de origen a un solo destino. Esto facilita la posterior toma de decisiones al poder hacer un mejor análisis de la información.
- Beneficios en costos, tiempos y productividad de la empresa (SANZ, 2010).

1.2.4. Estructura de almacenamiento de un almacén de datos

La diferencia fundamental entre los sistemas de Bases de Datos tradicionales y los Almacenes de Datos, es que en estos últimos sus estructuras son representadas a través de un modelo multidimensional, es decir, que almacenan la misma información que el diagrama entidad relación pero la organiza de forma diferente, garantizando de esa manera la velocidad y eficiencia en la recuperación de la misma.

Algunas de las características de dicho modelo son:

- Adecuado para resumir y organizar datos (ej. hojas de cálculo).
- Enfocado para trabajar sobre datos de tipo numérico.
- Más simple, fácil de visualizar y entender que el modelado Entidad Relación (E/R).

En la siguiente figura se muestra la estructura espacial de este tipo de diseño:

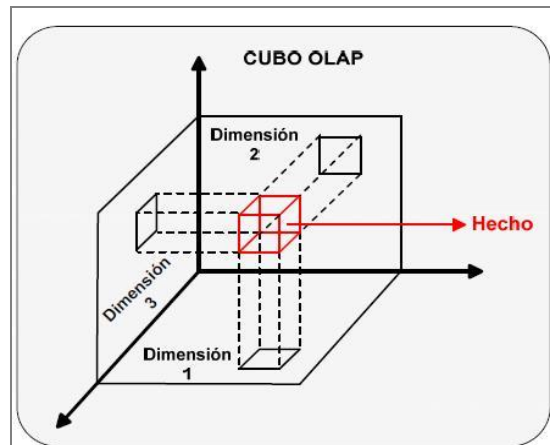


Figura 2: Estructura de un Cubo OLAP.

El diseo multidimensional es un m3todo de diseo de bases de datos basado en el modelo relacional. Est3 compuesto por dos tipos de tablas:

- Varias tablas de dimensiones, cada una formada por tuplas de atributos de la dimensi3n.
- Una tabla de hechos, compuesta por tuplas, una por cada hecho registrado. Este hecho contiene alguna variable o variables medidas u observables y las identifica con punteros a las tablas de dimensiones.

Se denomina “**hechos**” a los indicadores del negocio. Es decir, son todas aquellas medidas num3ricas que se incluyen en un sistema de Inteligencia de Negocio (BI). Son datos instant3neos en el tiempo, que son filtrados, agrupados y explorados a trav3s de condiciones definidas en las tablas de dimensiones.

Tablas de hechos

Las tablas de hechos contienen los hechos, medidas o indicadores que ser3n utilizados por los analistas de negocio para apoyar el proceso de toma de decisiones (BERNABEU, 2007).

Dimensiones

Las **dimensiones**, son los diferentes puntos de vista por los que se desea analizar la informaci3n. Las dimensiones incluyen los diferentes atributos que se quieren analizar, que adem3s se estructuran de forma jer3rquica, conforme a diferentes niveles de detalle (ESPINOSA, 2010).

Tablas de Dimensiones

Las tablas de dimensiones definen cómo están los datos lógicamente organizados y proveen el medio para analizar el contexto del negocio, contienen datos cualitativos. Además representan los aspectos de interés, mediante los cuales los usuarios podrán filtrar y manipular la información almacenada en la tabla de hechos.

Los datos dentro de estas tablas, que proveen información del negocio o que describen alguna de sus características, son llamados datos de referencia (ROBELO, 2010).

1.2.5. Esquemas multidimensionales

Esquema Estrella: consta de una tabla de hechos central y de varias tablas de dimensiones relacionadas con esta, a través de sus respectivas claves.

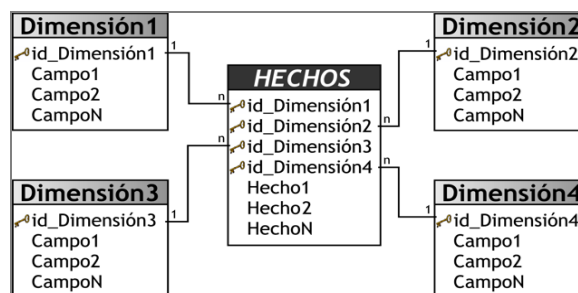


Figura 3: Esquema Estrella.

A continuación se destacarán algunas características de este modelo, que ayudarán a comprender mejor el porqué de sus ventajas:

- Posee los mejores tiempos de respuesta.
- Su diseño es fácilmente modificable.
- Existe paralelismo entre su diseño y la forma en que los usuarios visualizan y manipulan los datos.
- Simplifica el análisis.
- Facilita la interacción con herramientas de consulta y análisis.

Esquema Copo de Nieve: existe una tabla de hechos central que está relacionada con una o más tablas de dimensiones, las que a su vez pueden estar relacionadas o no con una o más tablas de dimensiones.

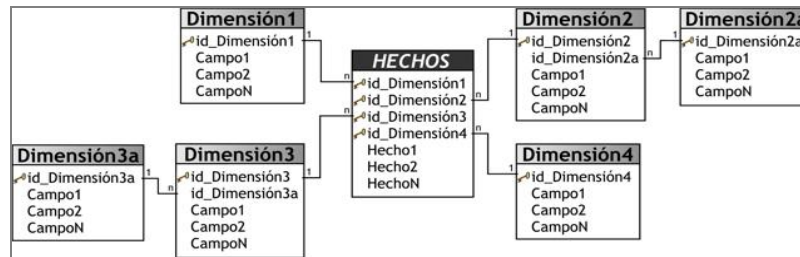


Figura 4: Esquema Copo de Nieve.

Se pueden definir las siguientes características de este tipo de modelo:

- Posee mayor complejidad en su estructura.
- Hace una mejor utilización del espacio.
- Es muy útil en tablas de dimensiones de muchas tuplas.
- Las dimensiones están normalizadas, por lo que requiere menos esfuerzo de diseño.
- Puede desarrollar clases de jerarquías fuera de las dimensiones, que permiten realizar análisis de lo general a lo detallado y viceversa.

A pesar de todas las características y ventajas que trae aparejada la implementación del esquema copo de nieve, existen dos grandes inconvenientes para ello:

- Si se poseen múltiples dimensiones, cada una de ellas con varias jerarquías, se creará un número de dimensiones bastante considerable, que pueden llegar al punto de ser inmanejables.
- Al existir muchas uniones y relaciones entre tablas, el desempeño puede verse reducido.

Esquema Constelación de Hechos: está compuesto por una serie de esquemas de estrella, es decir, una tabla de hechos central con una o más tablas de hechos auxiliares y sus respectivas tablas de dimensiones.

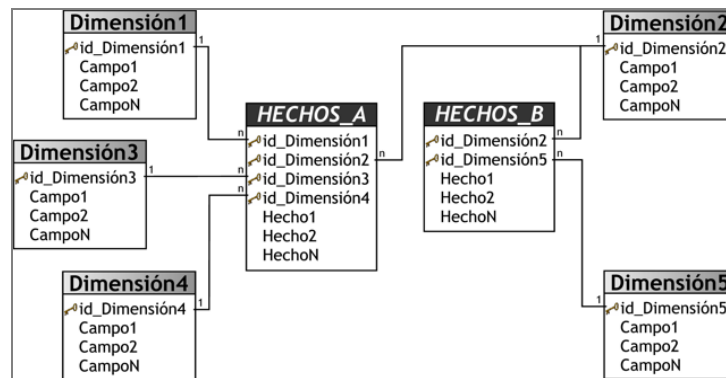


Figura 5: Esquema Constelación de Hechos.

Su diseño y cualidades son muy similares a las del esquema en estrella, pero posee una serie de diferencias con el mismo, que son precisamente las que los caracterizan. Entre ellas se pueden mencionar:

- Permite tener más de una tabla de hechos, por lo cual se podrán analizar más aspectos claves del negocio con un mínimo esfuerzo adicional de diseño.
- Contribuye a la reutilización de dimensiones, ya que una misma dimensión puede utilizarse para varias tablas de hechos.
- No es soportado por todas las herramientas de consulta y análisis.

El esquema estrella es el tipo de modelo que se ajusta de acuerdo a sus características, para la realización del mercado de datos, ya que se cuenta con una sola tabla de hecho y las tablas de dimensiones no se encuentran normalizadas.

1.2.6. Procesamiento Analítico en Línea

Procesamiento Analítico en Línea (OLAP por sus siglas en inglés) es una tecnología que se basa en el análisis multidimensional de los datos y que le permite al usuario tener una visión más rápida e interactiva de los mismos. Posee una gran capacidad para realizar cálculos de múltiples dimensiones, lo que permite gran variedad de informes y análisis de grandes volúmenes de datos.

Las principales características de OLAP son:

- **Rápido:** la primera regla se refiere a que el sistema debe ser capaz de responder de una forma rápida y ágil a la información que le sea solicitada por el usuario, el cual no deberá esperar más de cinco segundos a la hora de resolver peticiones sencillas y no más de veinte segundos en las peticiones complejas.

- **Análisis:** significa que el sistema debe poder reflejar cualquier lógica del negocio para poder responder a las preguntas específicas y necesidades empresariales.
- **Compartido:** el sistema deberá proporcionar herramientas que garanticen la confidencialidad de los datos y la seguridad de acceso por perfiles de los usuarios.
- **Multidimensional:** la herramienta deberá proporcionar soporte a cada una de las múltiples jerarquías que puedan existir dentro de la organización de información.
- **Información:** son todos los datos e información derivada de este proceso de análisis, la cual permitirá la toma de decisiones.

Existen diferentes tipos de sistemas OLAP, entre los que se encuentran:

- Procesamiento Analítico Multidimensional en Línea (MOLAP por sus siglas en inglés).
- Procesamiento Analítico Relacional en Línea (ROLAP por sus siglas en inglés).
- Procesamiento Analítico Híbrido en Línea (HOLAP por sus siglas en inglés).

MOLAP

Los datos fuente del cubo son almacenados junto con sus agregaciones en una estructura multidimensional de alto rendimiento. El almacenaje de MOLAP provee excelente rendimiento y compresión de datos. Como se dice, todo va en el cubo.

Tiene el mejor tiempo de respuesta, dependiendo solamente del porcentaje y diseño de las agregaciones del cubo. En general este método, es muy apropiado para cubos con uso frecuente por su rápida respuesta.

ROLAP

Toda la información del cubo, sus datos, su agregación, sumas, entre otros, son almacenados en una base de datos relacional. ROLAP no almacena copia de la base de datos, accede a las tablas originales cuando necesita responder a preguntas, es generalmente, mucho más lenta que las otras dos estrategias de almacenaje. Típicamente ROLAP se usa, para largos conjuntos de datos que no son frecuentemente buscados, tales como datos históricos de los años más recientes.

HOLAP

Permite un análisis híbrido de la información, es decir que une lo mejor de los dos tipos anteriores. El análisis HOLAP ayudará a reducir costos de hardware ya que necesitará menos espacio en disco que en las bases de datos relacionales.

Entre los sistemas OLAP mencionados anteriormente se utilizará el sistema ROLAP por las ventajas

que brinda, entre las que se encuentran:

- Seguridad e integridad en las bases de datos.
- Es escalable para grandes volúmenes de datos.
- Los datos de los sistemas ROLAP pueden ser compartidos con aplicaciones SQL.

A diferencia de MOLAP Y HOLAP, los datos de ROLAP son accedidos directamente desde el Almacén de Datos u otra fuente de datos relacional y no son almacenados por separados.

1.2.7. Definición de Mercado de Datos

Un Mercado de Datos es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento.

1.3. Inteligencia de Negocio

Inteligencia de negocio (BI por sus siglas en inglés) es el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada (interna y externa a la compañía) en información estructurada, para su explotación directa (reporte, análisis OLAP, etcétera) o para su análisis y conversión en conocimiento, dando así soporte a la toma de decisiones sobre el negocio (SINNEXUS, 2011).

1.4. Metodología de desarrollo de Almacenes de Datos

La metodología es el estudio analítico y crítico de los métodos de investigación que se siguen para alcanzar objetivos en una ciencia.

1.4.1. Metodologías para los almacenes de datos

En el diseño de un almacén de datos, se destacan un conjunto de metodologías que definen y guían todo el ciclo de vida del desarrollo de dicho almacén. Estas diferentes metodologías se pueden englobar dentro de dos grandes bloques: descendente o top-down y ascendente o bottom-up, que se corresponden con las metodologías propuestas por Bill Inmon y Ralph Kimball respectivamente. Estos autores merecen una especial atención porque, en muchos aspectos, se consideran los precursores del Almacén de Datos y sus opiniones son muy valoradas en la industria.

Metodología de Inmon

La metodología de Inmon se basa en un enfoque descendente (top-down), propone construir primero el almacén de datos y a partir de este los mercados de datos, plantea la creación de un repositorio de datos corporativo como fuente de información consolidada, persistente, histórica y de calidad. Al ser construido descendentemente los mercados de datos se nutren del almacén de datos corporativo, convirtiéndose en un complejo empresarial de bases de datos relacionales.

Inmon afirma que la creación de una base de datos relacional con una leve normalización necesita ser la base para los mercados de datos. Por lo que no se crean los mercados de datos directamente desde los Sistemas de Procesamiento de Transacciones en Línea (OLTP) a través de un área de ensayo. En lugar de ello, se crean a partir de la arquitectura relacional de los datos corporativos.

Metodología Kimball

Kimball se basa en dividir el mundo de Inteligencia de Negocio entre los hechos y las dimensiones, esta es eficaz y conduce a una solución completa en un corto período de tiempo. Además, tiene abundante documentación y se puede encontrar una respuesta a casi todas las preguntas que se puedan tener. Se puede empezar de cero y dar al usuario una primera información sobre sus datos en cuestión.

Entre sus características principales está que su arquitectura es ascendente (bottom-up), plantea que se debe crear por cada departamento un conjunto de mercados de datos independientes orientados a los temas que estén relacionados con él.

La principal diferencia que existe entre ambas metodologías está basada en la forma de enfrentar el problema. Para Bill Inmon el almacén de datos es una parte de un sistema de inteligencia de negocio (BI) dentro de una empresa que tiene un almacén de datos y los mercados de datos obtienen su información a partir de este almacén. Por otro lado Ralph Kimball plantea que el almacén de datos es un conglomerado de todos los mercados de datos dentro de una empresa y la información siempre se almacena de acuerdo al Modelo Dimensional. Esto implica que, para Bill Inmon el desarrollo del almacén de datos debe ser completo para su correcto funcionamiento mientras que la visión de Ralph Kimball permite desarrollar mercados de datos particulares que contengan la lógica de negocio en la que se está interesado profundizar, permitiendo de esta manera que no se tenga que realizar el desarrollo de todas las problemáticas del negocio.

1.4.2. Elección de la metodología para el desarrollo del mercado de datos EMGEF

El Centro de Tecnología de Gestión de Datos (DATEC) propone una metodología para este tipo de soluciones basada en el enfoque planteado por Kimball, pero se adapta a las características

particulares de la UCI y del centro DATEC, denominándose esta: Propuesta de metodología para el desarrollo de almacenes de datos en DATEC. El centro definió trabajar con una adaptación del ciclo de vida de la metodología de Kimball por los siguientes elementos:

- Identifica la tabla de hechos y la tabla de dimensiones, lo cual agiliza el proceso de desarrollo y con ello la toma de decisiones.
- Propone la construcción de Mercados de Datos departamentales y después el AD, esto trae como ventaja, que la creación y la puesta en marcha de los Mercados de Datos se produce en un lapso de tiempo corto y después se valora si se construye o no el Almacén de Datos.
- Existe amplia documentación de la misma, así cualquier duda que exista puede ser atendida rápidamente.

Es una metodología madura y reconocida por los usuarios dedicados al tema, además de tener bien definidas sus etapas, actividades, roles y artefactos. Como complemento de este modelo y por las características de trabajo de la UCI se tomó lo planteado por Leopoldo Zenaido Zepeda en su tesis de doctorado en conjunto con un colectivo de autores, que proponen incluir los casos de uso para guiar el proceso de desarrollo y de esta forma estar más alineados a las tendencias y normas de la universidad.

Flujos de trabajo

Entre los flujos de trabajo que presenta esta metodología se encuentran los siguientes:

- **Estudio preliminar o Planeación:** se realiza el estudio de la entidad cliente para determinar lo que se desea construir y las condiciones que existen para el desarrollo de la misma, la Planeación del proyecto, se definen los objetivos, el alcance preliminar, los costos estimados y otras series de actividades.
- **Requerimientos:** se realiza en dos direcciones, una, identificando las necesidades de información y reglas del negocio; y la otra con un levantamiento detallado de las fuentes de datos a integrar. Es aquí donde se definen los requerimientos a través de la comparación de las necesidades y las reglas del negocio.
- **Arquitectura y diseño:** aquí se definen las estructuras de almacenamiento, se diseñan las reglas de extracción, transformación y carga, así como la arquitectura de información que regirá el desarrollo de la solución.
- **Implementación:** se lleva a cabo el diseño físico.

1.5. Herramientas para el desarrollo del mercado de datos

Actualmente existe un incremento de las tecnologías de información, creándose una constante competitividad por parte de los productores de estas herramientas, lo cual genera grandes exigencias en cuanto a productividad y calidad de los productos.

1.5.1. Herramienta de modelado

Las herramientas CASE (Ingeniería de Software Asistida por Computadora), hoy en día son de gran utilidad para la modelación de proyectos, flujos de información, entidades de datos así como requerimientos del sistema, facilitando el proceso de Planeación. Mucha de la información que se captura durante esta fase se usará también durante el desarrollo y el mantenimiento del ciclo de vida del sistema.

Visual Paradigm 8.0

Herramienta profesional de modelado CASE, que soporta el ciclo de vida completo del desarrollo del software, comprendido en: análisis y diseño orientado a objetos, construcción, pruebas y despliegue. Este software de modelado UML (Lenguaje Unificado de Modelado) ayuda a una rápida construcción de aplicaciones de calidad, sin un elevado costo. Permite elaborar diagramas de clases, así como documentación.

Es una herramienta propietaria que posee varias versiones, como son: Enterprise, Professional, Standard, Modeler, Personal y Community (que es gratuita), además sus distintas ediciones tienen compatibilidad. Soporta aplicaciones Web aunque las imágenes y reportes generados no son de muy buena calidad. Está disponible en varios idiomas. Posee generación de código para Java en Lenguaje de Marcado de Hipertexto (HTML) y es muy fácil de instalar y actualizar (TECHNOLOGIES).

Para el modelado de este sistema se utilizará esta herramienta, debido a las grandes posibilidades que este brinda; en la UCI existe una licencia para la utilización del mismo, lo cual posibilita su utilización.

1.5.2. Sistema Gestor de Bases de Datos

Un Sistema Gestor de Bases de Datos (SGBD) es una colección de programas cuyo objetivo es servir de interfaz entre la base de datos, el usuario y las aplicaciones. Se compone de un lenguaje de definición de datos, de un lenguaje de manipulación de datos y de un lenguaje de consulta. Un SGBD permite definir los datos a distintos niveles de abstracción y manipular dichos datos, garantizando la seguridad e integridad de los mismos (CAVSI).

PostgreSQL 9.1

PostgreSQL es un sistema de gestión de base de datos relacional orientada a objetos y libre, publicado bajo la licencia BSD, es la licencia de software otorgada principalmente para los sistemas BSD (Berkeley Software Distribution). Es más completo que MySQL ya que permite métodos almacenados, restricciones de integridad, vistas, etcétera.

Principales características de este gestor de bases de datos:

- Implementación del estándar SQL92/SQL99.
- Soporta distintos tipos de datos: además del soporte para los tipos base, también soporta datos de tipo fecha, monetarios, elementos gráficos, datos sobre redes, cadenas de bits, etc. También permite la creación de tipos propios.
- Incorpora una estructura de datos de arreglo (array).
- Incorpora funciones de diversa índole: manejo de fechas, geométricas, orientadas a operaciones con redes, etcétera.
- Permite la declaración de funciones propias, así como la definición de disparadores.
- Soporta el uso de índices, reglas y vistas.
- Incluye herencia entre tablas.
- Permite la gestión de diferentes usuarios, como también los permisos asignados a cada uno de ellos.

Ventajas

- Por su arquitectura de diseño, posee una gran escalabilidad. Es capaz de ajustarse al número de CPU y a la cantidad de memoria que posee el sistema de forma óptima, haciéndole soportar una mayor cantidad de peticiones simultáneas de manera correcta.
- Implementa el uso de reversiones (*rollback's*), subconsultas y transacciones, haciendo su funcionamiento mucho más eficaz.
- Tiene la capacidad de comprobar la integridad referencial, así como también la de almacenar procedimientos en la propia base de datos.
- Tiene mejor soporte para los disparadores (*triggers*) y procedimientos en el servidor (SCRIBD).

Esta es la herramienta a utilizar para el desarrollo del mercado de datos, ya que permite independencia tecnológica, por ser una propuesta de Código Abierto (*Open Source*) que sobrepasa a muchas propietarias y por su gran potencialidad y adaptabilidad al problema en cuestión.

1.5.3. Administrador de bases de datos

Existen múltiples herramientas para la administración de bases de datos, para la realización de este trabajo se utilizó PgAdmin.

PgAdmin 1.14: Es una de las herramientas más populares para administrar las bases de datos en PostgreSQL y presenta las siguientes características:

- Es un software libre.
- Accede a todos los objetos del PostgreSQL y responde las necesidades de todos los usuarios, desde escribir consultas SQL simples hasta desarrollar bases de datos complejas.
- Permite que se pueda usar en Linux, FreeBSD, Solaris, Mac OS X y Windows (TOOLS, 2010).

1.5.4. Herramienta de Perfilado

DataCleaner es una aplicación de código abierto para el perfilado, la validación y comparación de datos. Estas actividades ayudan a administrar y supervisar la calidad de los datos con el fin de garantizar que la información es útil y aplicable a su situación de negocio. Es una de las aplicaciones más fáciles de usar para la calidad de los datos. Normalmente se utiliza antes, durante y después del proceso de Extracción, Transformación y Carga (ETL).

- Antes, para profundizar en los orígenes de los datos que serán usados en el trabajo.
- Durante, en caso de existir cualquier desajuste inesperado durante el proceso ETL.
- Después de asegurar la coherencia y la calidad en la fuente de datos que han poblado.

DataCleaner puede acceder y analizar prácticamente cualquier almacén de datos, incluyendo:

- Base de datos como Oracle, Microsoft SQL Server, PostgreSQL, MySQL, OpenOffice (ODB), entre otras.
- Hojas de cálculo Excel.
- Archivos XML (DATACLEANER, 2012).

1.5.5. Herramienta de desarrollo

Suite Pentaho: La Open BI Suite de Pentaho, provee un completo espectro de funcionalidades de Business Intelligence (BI, Inteligencia de Negocios), incluyendo reportes, análisis, tableros de control, minería de datos, integración de datos y una plataforma de BI que la han convertido en la suite de código abierto más popular del mundo (CORAIL, 2012).

Algunas de las funcionalidades integradas con la plataforma de Inteligencia de Negocios (BI) de Pentaho son:

- Pentaho Data Integration.
- Pentaho Schema Workbench.
- Mondrian OLAP Server.
- Pentaho BI Server.

Pentaho Data Integration es una de las herramientas utilizadas en el proceso de Extracción, Transformación y Carga (ETL) de código abierto que reúne un conjunto de componentes que permiten modelar y ejecutar transformaciones sobre flujos de datos.

Presenta las siguientes características:

- Puede funcionar sobre varias plataformas a través de un sistema que soporte Java 1.4 o una versión superior.
- Exige solamente alrededor de 128 MB de RAM.
- Provee un Lenguaje de programación Java para la Conectividad de Base de Datos (JDBC)² que permite la conexión con cualquier base de datos sin tener que instalar un cliente adicional.
- Se integra con ficheros de Microsoft Office, Servicios Web y cubos MOLAP.
- Esta herramienta incluye procesamiento optimizado de los ficheros planos.
- Brinda soporte para metadatos e incorpora operaciones de transformación.
- Permite operar con los campos en el flujo de datos, renombrando, calculando campos en función de otros y realizando búsquedas auxiliares en bases de datos.
- Corrección de errores (GRAVITAR, 2012).

Pentaho Schema Workbench: es un entorno visual para el desarrollo y prueba de cubos OLAP Mondrian. Provee un mecanismo para buscar datos con rapidez y tiempo de respuesta uniforme independientemente de la cantidad de datos en el cubo o la complejidad del procedimiento de búsqueda. Permite la ejecución de consultas de Expresiones Multidimensionales (MDX) contra el esquema y la base de datos y la navegación por la base de datos subyacente (GRAVITAR, 2012).

Mondrian OLAP Server: es un servidor OLAP de código abierto que gestiona la comunicación entre la aplicación OLAP y la base de datos. Permite crear cubos de información para análisis

² Java Data base Connectivity (JDBC), permite la ejecución de operaciones sobre bases de datos desde el lenguaje de programación Java.

multidimensional. Este proporciona la conexión a la base de datos y ejecuta las sentencias SQL (GRAVITAR, 2012).

Pentaho BI Server: La aplicación más conocida de la Plataforma de BI es el servidor de BI de Pentaho, que funciona como una red basada en sistema de gestión de informes, el servidor de integración de aplicación es un motor de flujo de trabajo ligero (secuencias de acción.) Está diseñado para integrarse fácilmente en cualquier proceso de negocio (EASY, 2010).

1.5.6. Servidor de Aplicaciones

Apache Tomcat: herramienta utilizada como servidor web.

Algunas de sus características son:

- Es el servidor web más utilizado para trabajar en entornos web.
- Es una implementación completamente funcional de los estándares de JSP y servlets.
- Puede especificarse como el manejador de las peticiones de JSP y servlets recibidas por servidores Web populares.
- Está integrado en la implementación de referencia Java 2 Enterprise Edition (J2EE) de Sun Microsystems (CRISTIAN, 2006).

1.6. Conclusiones del capítulo

A partir del estudio realizado se concluyó lo siguiente:

- La metodología seleccionada permite guiar todo el ciclo de vida del desarrollo del mercado.
- Las herramientas y tecnologías seleccionadas permiten el diseño e implementación de un sistema para el área de Planeación de la EMGEF.

CAPÍTULO 2 ANÁLISIS Y DISEÑO

En el presente capítulo se realiza una descripción detallada del negocio y los temas de análisis. Se definen las reglas del negocio, los requisitos de información, los funcionales, no funcionales y los casos de uso del sistema. Además, se realiza la matriz BUS y el modelo de dato.

2.1. Análisis

2.1.1. Definición del negocio

La Empresa de Mantenimiento de Grupos Electrógenos Fuel-Oil (EMGEF) recoge una serie de datos diarios provenientes de las provincias del país donde se encuentran ubicados los grupos electrógenos. Toda central tiene la obligación de reportar un parte diario del funcionamiento de los grupos electrógenos de su área a través del sistema estadístico con que esta cuenta. La información que se maneja en esta empresa está relacionada con el estado y funcionamiento de los motores de los diferentes grupos electrógenos.

2.1.2. Tema de análisis

Un tema de análisis es la división o clasificación de la información de una organización de acuerdo a las diferentes temáticas que esta contenga. En el caso de la investigación centrada en la base de datos perteneciente al sistema estadístico para el Mantenimiento de Grupos Electrógenos Fuel Oil se define como tema de análisis: las ocurrencias presentadas por los motores.

Dentro de este tema se recogen los siguientes datos:

- Problemas reportados.
- Problemas resueltos.

2.1.3. Reglas del negocio

Las reglas del negocio son las transformaciones que se le deben realizar a ciertos datos para obtener a partir de ellos otros datos.

RN 1: No pueden existir campos nulos.

RN 2: Los identificadores deben ser únicos.

RN 3: Las cadenas de caracteres no pueden contener números.

RN 5: Una vez cargados los datos en el almacén, no pueden existir campos nulos.

RN 6: El identificador de las dimensiones no puede ser nulo.

RN 7: Variación = Cantidad de motores del año actual – Cantidad de motores del año anterior.

2.1.4. Descripción de los actores del sistema

Tabla 1: Descripción de los actores del sistema.

Actor	Objetivo
Administrador	Responsable de administrar los usuarios del sistema, asignarles sus respectivos roles, a la vez que debe administrar dichos roles y los reportes.
Administrador ETL	Responsable de llevar a cabo los procesos de extracción, transformación y carga de los datos.
Analista	Responsable de inicializar los casos de uso relacionados con la visualización de los reportes del mercado de datos.

2.1.5. Necesidades de los usuarios

Después del análisis del negocio que tiene lugar en la EMGEF se define como necesidad de los usuarios el análisis de la información relacionada con el área de Planeación, que se recopila a través de la base de datos del Sistema Estadístico para el Mantenimiento de Grupos Electrónos Fuel Oil.

2.1.6. Requisitos de información

Los requisitos de información describen qué información debe almacenar el sistema para satisfacer las necesidades de clientes y usuarios.

RI 1: Obtener cantidad de ocurrencias por Unidades Empresariales de Base (UEB) en un periodo de tiempo.

RI 2: Obtener cantidad de ocurrencias por causa en un periodo de tiempo.

RI 3: Obtener cantidad de ocurrencias por acción en un periodo de tiempo.

2.1.7. Requisitos funcionales

Los requisitos funcionales de un sistema describen la funcionalidad o los servicios que se espera que este provea.

RF 1: Autenticar usuario.

RF 2: Adicionar roles.

RF 3: Eliminar roles.

RF 4: Adicionar usuarios.

- RF 5: Eliminar usuarios.
- RF 6: Insertar reportes.
- RF 7: Modificar reportes.
- RF 8: Eliminar reportes.
- RF 9: Extraer información.
- RF 10: Realizar transformación y carga.
- RF 11: Abrir navegador OLAP.
- RF 12: Mostrar editor MDX.
- RF 13: Mostrar Padres.
- RF 14: Ocultar repeticiones.
- RF 15: Intercambiar ejes.
- RF 16: Mostrar gráfico.
- RF 17: Configurar gráfico.
- RF 18: Configurar impresión.
- RF 19: Exportar a PDF.
- RF 20: Exportar a Excel.
- RF 21: Mostrar propiedades.
- RF 22: Suprimir filas.
- RF 23: Detallar miembros.
- RF 24: Entrar en detalles.
- RF 25: Mostrar datos de origen.

2.1.8. Requisitos no funcionales

Con el propósito de satisfacer al máximo las necesidades del cliente, así como la calidad de la herramienta se definió un listado de las propiedades o cualidades que el producto debe tener, las cuales se describen a continuación:

2.1.8.1. Disponibilidad

RNF 1: Asegurar la disponibilidad del sistema.

El sistema debe estar disponible durante el horario de trabajo, de 8:00am a 5:00pm.

RNF 2: Asegurar la recuperación ante un fallo.

El sistema debe ser capaz de recuperarse ante un fallo, teniendo en cuenta la complejidad y naturaleza de este. El tiempo para su correcta recuperación fluctúa entre 10 minutos y 72 horas.

RNF 3: Garantizar la persistencia de la información.

Se debe realizar un respaldo total de los datos del almacén de datos con una frecuencia semanal. Esta información se almacenará en el edificio correspondiente a la EMGEF de La Habana y será responsabilidad del grupo de informática de la EMGEF. Este requisito también contribuye a la seguridad del sistema.

2.1.8.2. Concurrencia

RNF 4: Garantizar la conexión de múltiples usuario al mismo tiempo.

El sistema debe permitir que existan al menos 3 usuarios conectados de forma simultánea.

2.1.8.3. Restricciones de diseño

RNF 5: Utilizar el Sistema Gestor de Base de Datos definido durante la investigación.

El gestor de base de datos que se utilizará es PostgreSQL 9.1 y como interfaz de administración de dicho gestor PgAdmin 1.14.

RNF 6: Utilizar las herramientas para la implementación de la capa de inteligencia de negocios definidas durante la investigación.

De la suite Pentaho, se usarán los siguientes componentes:

- Schema Workbench 3.2.0.
- Pentaho BI Server 3.10.

Para el uso de las herramientas anteriores se requiere la instalación de la máquina virtual de java (Java Virtual Machine 6.0).

2.1.8.4. Requisitos de Software

RNF 7: Instalar en las estaciones de trabajo el software necesario para el correcto funcionamiento del sistema.

Las configuraciones de software de las máquinas clientes deben contar con los siguientes elementos:

- Windows XP, GNU/Linux. Ubuntu 11.04 o versiones superiores respectivamente.

- Navegador web.
- Adobe Flash Player.

Las configuraciones de software para el servidor de base de datos deben contar con los siguientes elementos:

- Windows server 2003, GNU/Linux. Ubuntu 11.04 o versiones superiores respectivamente.
- PostgreSQL 9.1 como Sistema Gestor de Base de Datos (SGBD).
- PgAdmin 1.14 como interfaz de administración del SGBD.

Las configuraciones de software el servidor de aplicaciones deben contar con los siguientes elementos:

- Windows XP, GNU/Linux. Ubuntu 11.04 o sus versiones superiores respectivamente.
- Pentaho BI Server 3.10.
- Java Virtual Machine 6.0.
- Apache Tomcat 6.0.
- Navegador web.
- Adobe Flash Player.

2.1.8.5. Requisitos de Hardware

RNF 8: Definir las interfaces de hardware que soportará el sistema.

El sistema podrá interactuar solamente con una interfaz de hardware: la impresora. Esta interacción se ocasionará cuando se necesite imprimir un reporte en formato físico. El acceso a la impresora será mediante el protocolo TCP/IP a través de la interfaz que ofrece el hardware.

RNF 9: Proporcionar características mínimas de hardware a las estaciones de trabajo.

Características de un cliente ligero.

RNF 10: Proporcionar características mínimas de hardware a los servidores.

Para lograr una explotación aceptable del sistema los servidores deben contar con los siguientes requerimientos de hardware:

- 1 GB RAM.
- 80 GB de disco duro.

2.1.9. Casos de uso del sistema

Para la confección del diagrama de caso de uso, se agruparon los 25 requisitos funcionales del sistema en 7 casos de uso, los 3 requisitos de información en 1 caso de uso de información y se definieron las relaciones existentes entre ellos y los actores del sistema identificados.

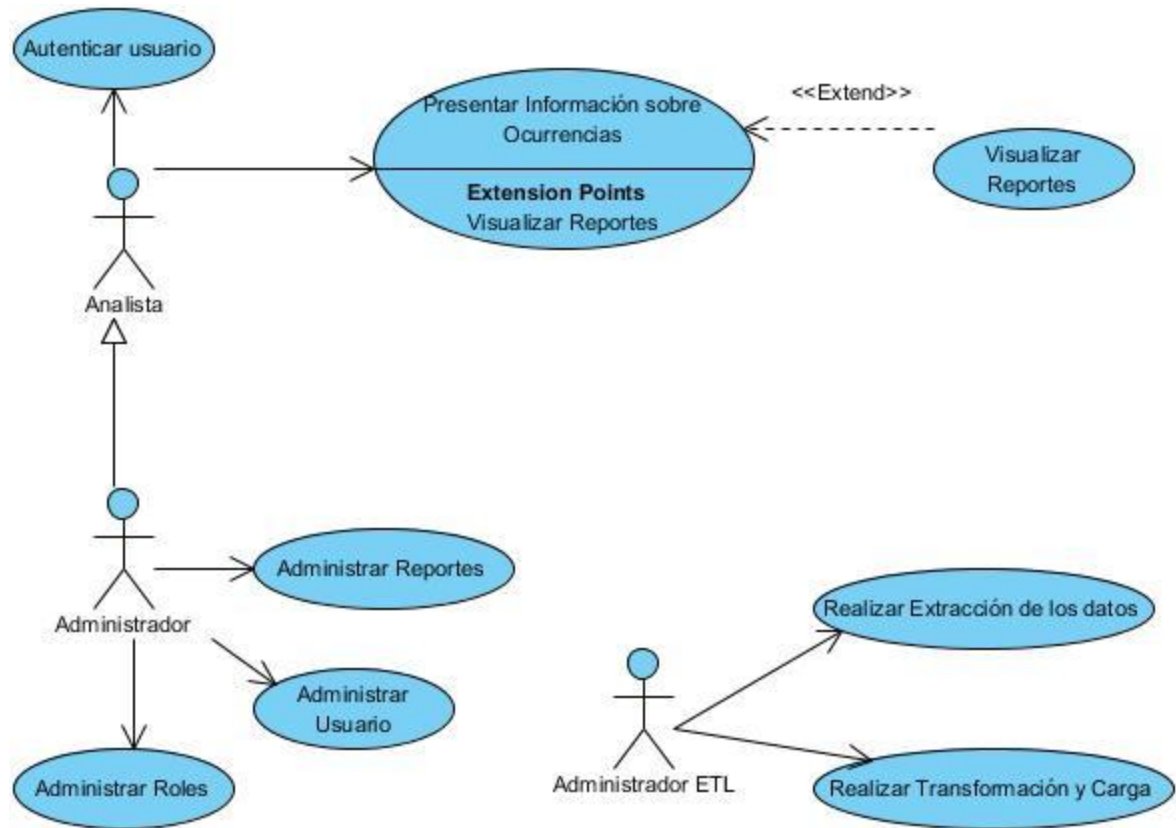


Figura 6: Diagrama de casos de uso del sistema.

2.1.10. Descripción de los casos de uso críticos

Tabla 2: CU Realizar extracción de datos.

Actores	Administrador de ETL.
Resumen	El CU inicia cuando el actor desea realizar la extracción de los datos correspondientes a las fuentes de información. Se extraen los datos de la fuente. El CU finaliza una vez que los datos seleccionados por el actor son extraídos.
Complejidad	Alta.
Prioridad	Alta.
Precondiciones	Disponibilidad de las fuentes.

Poscondiciones	Datos disponibles para transformar.	
Flujo de eventos		
Flujo básico Realizar extracción de datos.		
	Actor	Sistema
1.	El actor interactúa con la herramienta Pentaho Data Integration (PDI) para realizar la extracción de los datos.	
2.		El sistema muestra el área de trabajo y le da al actor la posibilidad de cargar la transformación.
3.	Selecciona la transformación a cargar o realiza los pasos para una nueva transformación.	
4.	Configura los parámetros de entrada de la transformación.	
5.		El sistema almacena los datos en el área temporal.
6.		El sistema muestra los datos existentes en esta área.
8.	El actor realiza la acción de aceptar y finaliza.	
Flujos Alternos		
	Actor	Sistema
6.1		El sistema muestra un mensaje de error.
Relaciones	CU Incluidos	No aplica.
	CU Extendidos	No aplica.
Asuntos pendientes	[Posibles mejoras al caso de uso].	

Tabla 3: CU Realizar Transformación y Carga de los datos.

Actores	Administrador de ETL.	
Resumen	El CU inicia cuando el actor desea realizar la transformación y carga de los datos correspondientes a las fuentes de información ya extraídas. Se transforman y cargan los datos extraídos. El CU finaliza una vez que los datos han sido transformados y cargados a la base de datos.	
Complejidad	Alta.	
Prioridad	Alta.	
Precondiciones	Los datos se encontraron correctamente extraídos en el área temporal y las estructuras del mercado de datos se encontraron disponibles para su uso. En la base de datos debe existir una estructura para almacenar la información.	
Poscondiciones	Datos transformados y cargados.	
Flujo de eventos		
Flujo básico Transformación y Carga de los datos.		
	Actor	Sistema
1.	El administrador ETL interactúa con la herramienta PDI para realizar la extracción.	
2.		El sistema levanta el área de trabajo y le brinda la posibilidad al usuario de cargar la transformación.
3.	El administrador ETL carga la transformación a ejecutar.	
4.	El administrador ETL configura las transformaciones definidas en las reglas del negocio.	
5.		El sistema realiza las transformaciones pertinentes.
6.		El sistema almacena los datos en la Base de Datos.
7.		El sistema muestra un mensaje de satisfacción en cada una de las operaciones realizadas.
8.	El administrador ETL cierra la ventana de trabajo.	

Flujos Alternos		
	Actor	Sistema
7.1.		El sistema muestra un mensaje de error y regresa al paso 3.
Relaciones	CU Incluidos	No aplica.
	CU Extendidos	No aplica.
Asuntos pendientes	Posibles mejoras al caso de uso.	

Tabla 4: CU Presentar información sobre ocurrencias.

Actores	Analista.	
Resumen	El CU inicia cuando el analista selecciona presentar información sobre ocurrencias, luego se realiza la selección de la información a mostrar. Finaliza cuando se muestran los resultados.	
Complejidad	Alta.	
Prioridad	Alta.	
Precondiciones	Actor autenticado, completitud del mercado de datos y carga de los datos.	
Postcondiciones	Disponibilidad de opciones (cruce de variables) de reportes relacionados con el CU Presentar información sobre Ocurrencias.	
Flujo de eventos		
Flujo básico Presentar información sobre Ocurrencias.		
	Actor	Sistema
1.	El actor accede a la aplicación.	
2.		El sistema muestra la interfaz principal.
3.	El actor se autentica en el sistema.	
4.		El sistema muestra la interfaz con el área de análisis.
5.	El actor selecciona área de análisis A.A. Ocurrencias.	
6.		El sistema muestra los libros de trabajo que

		están contenidos dentro del A.A. Ocurrencias.
7.	El actor selecciona un libro de trabajo.	
8.		El sistema muestra los reportes contenidos dentro del L.T seleccionado.
9.	El actor selecciona el reporte deseado.	
10.		EL sistema muestra el contenido del reporte dentro del área de trabajo y brinda opciones al actor de realizar cambios en el reporte para su mejor análisis. Ir al caso de uso Visualizar reporte . Finaliza el caso de uso.
Opciones de reportes del comportamiento de los servidores		
Perspectivas de análisis		Posibles resultados
		Medidas Periodicidad
Variables de entradas disponibles relacionadas con el CU Presentar información sobre Ocurrencias :	Variables de salidas relacionadas con el hecho:	Rango de tiempo en que se solicitan las variables de salida Visualizar información de trámites:
<ul style="list-style-type: none"> ➤ UEB ➤ Tiempo ➤ Causa ➤ Accion 	<ul style="list-style-type: none"> ➤ cant_ocurrencia(CO) 	<ul style="list-style-type: none"> ➤ Diario ➤ Mensual ➤ Anual
Relaciones	CU Incluidos	No aplica
	CU Extendidos	El CU Presentar información sobre

		Ocurrencias se extiende del CU Visualizar Reportes
Asuntos pendientes	Posibles mejoras al caso de uso.	

2.2. Diseño

El proceso de diseño de un almacén define el modelo conceptual, lógico, físico y de visualización (BI) del mismo. En el modelo conceptual se especifican los requerimientos del usuario a través de las dimensiones, hechos y medidas que componen el almacén; en el modelo lógico se genera el modelo de datos; en el modelo físico se diseñan las transformaciones, y en el de visualización se diseñan reportes candidatos por cada uno de los libros de trabajo definidos en el área de análisis.

Dimensiones

Las dimensiones son las características de un hecho, que permiten su posterior análisis en el proceso de toma de decisiones; una entidad de negocio respecto a la cual se deben calcular las métricas.

Las dimensiones definidas son las siguientes:

- dim_ueb: describe el nombre de todas las Unidades Empresariales de Base (ueb) y centrales de la empresa.
- dim_causa: contiene todos los tipos y subtipos de causa que existen.
- dim_acción: describe el tipo de acción que puede tener una ocurrencia.
- dim_tiempo: Esta es la dimensión más común e importante en los diseños de mercados de datos ya que define una línea de tiempo específica.

2.2.3. Medidas

Las medidas son los valores de datos numéricos que se analizan, son las variables de salida en el diseño, de ahí que representen lo contable que se necesita conocer. Para el desarrollo de este mercado se definieron las siguientes medidas:

- cant_ocurrencia: Cantidad de ocurrencias.

2.2.4. Tabla de hechos

Las tablas de hechos contienen las dimensiones y las medidas asociadas al mismo. La tabla de hechos tiene una clave primaria compuesta por las claves foráneas de las tablas dimensiones asociadas a él.

Para el desarrollo de este mercado se definió la siguiente tabla de hecho:

- hecho_ocurrencia: en este hecho se guarda la información de los problemas reportados y resueltos de los diferentes grupos electrógenos.

Tabla 5: Tabla de hechos.

Hecho	Dimensiones asociadas	Medidas
hecho_ocurrencia	dim_ueb	cant_ocurrencia
	dim_causa	
	dim_acción	
	dim_tiempo	

Matriz BUS

La Matriz Bus es la representación de la relación que existe entre los hechos y las dimensiones. Se define como la habilidad para describir y seguir la vida tanto de una dimensión como de un hecho, la cual permite determinar el impacto que provocaría un cambio durante el desarrollo del sistema.

Tabla 6: Matriz BUS.

DIMENSIONES	HECHO
	hecho_ocurrencia
dim_ueb	X
dim_causa	X
dim_acción	X
dim_tiempo	X

2.2.5. Modelo de Datos

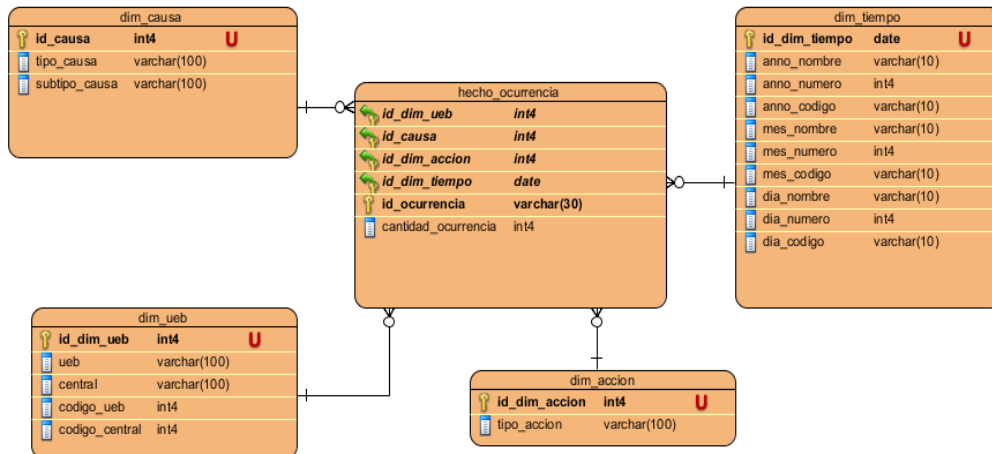


Figura 7: Modelo de datos.

2.2.6. Diseño de las transformaciones

Las transformaciones son realizadas para el proceso de extracción, transformación y carga de los datos. Una vez ejecutada dicha transformación, se valida un grupo de restricciones definidas por el analista. En caso de no cumplirse una de estas restricciones, los datos se almacenan en ficheros Excel para darles tratamiento; en caso contrario, todos los datos se insertan en la tabla que tiene como salida dicha transformación.

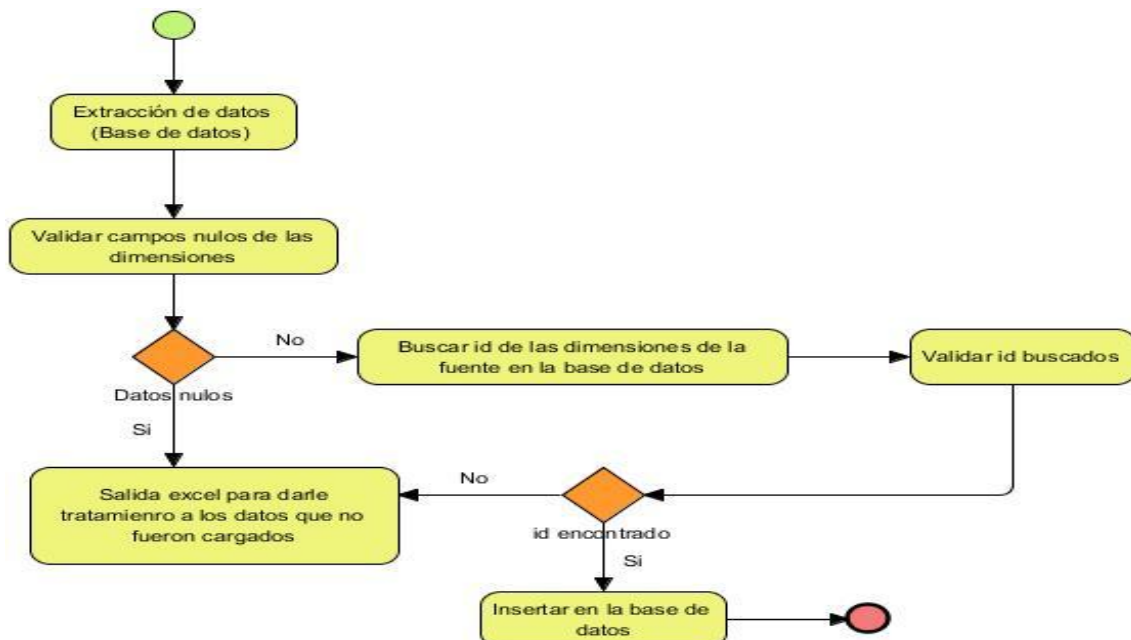


Figura 8: Diseño de las transformaciones.

2.2.7. Políticas de seguridad

2.2.7.1. Roles y permisos

Para el acceso al mercado de datos se define un usuario por cada uno de los roles existentes en el sistema, con el objetivo de garantizar un control de acceso basado en roles y así cada usuario, opera en el sistema según los permisos que se le definan al rol.

Tabla 7: Roles y permisos.

Roles	Permisos	
	Lectura	Escritura
Sobre la Base de Datos		
Administrador	X	X
Administrador ETL	X	x
Analista	X	
Sobre la aplicación		
Administrador	X	x
Analista	X	

2.3. Conclusiones del capítulo

En este capítulo se crearon todos los artefactos necesarios para la implementación del mercado de datos:

- Se definieron los requisitos de información y funcionales necesarios para la obtención de un modelo lógico que sustentará los datos contenidos en la fuente.
- Se diseñó el diagrama de caso de uso del sistema, lo que permitió representar la relación entre los 8 casos de uso definidos y los 3 actores que interactúan con la aplicación.
- Con el diseño del modelo de datos se representó la correspondencia existente entre el hecho y las 4 dimensiones identificadas en el negocio.

CAPÍTULO 3 IMPLEMENTACIÓN

El presente capítulo está encaminado a la implementación de los diferentes aspectos relacionados con los procesos de integración de datos, con el propósito de brindar una mayor comprensión de las estrategias y procedimientos utilizados. Además, se abordan elementos relacionados con la implementación de la capa de inteligencia de negocio, incluyendo la creación de las estructuras necesarias para la navegación y el análisis de los datos.

3.1. Implementación del modelo de datos

Con el objetivo de garantizar una adecuada organización en el almacén de datos y un mayor entendimiento, se definió un solo esquema, en el cual se encuentran agrupados los hechos y las dimensiones específicas del almacén, teniendo en cuenta que existe un solo hecho el esquema queda de la siguiente manera:

Tabla 8: Esquemas y tablas del mercado.

Esquemas	Tablas
mart_emgef	dim_ueb
mart_emgef	dim_causa
mart_emgef	dim_acción
mart_emgef	dim_tiempo
mart_emgef	hecho_ocurrencia

3.2. Implementación de los procesos de Extracción, Transformación y Carga

3.2.1. Implementación de las transformaciones

La transformación diseñada se aplicó a cada uno de los hechos del mercado. El componente más utilizado en cada una de ellas fue: la búsqueda en base de datos para encontrar los identificadores de las dimensiones y poder realizar la carga de los hechos. En la figura 9 se muestra la transformación del hecho ocurrencia.

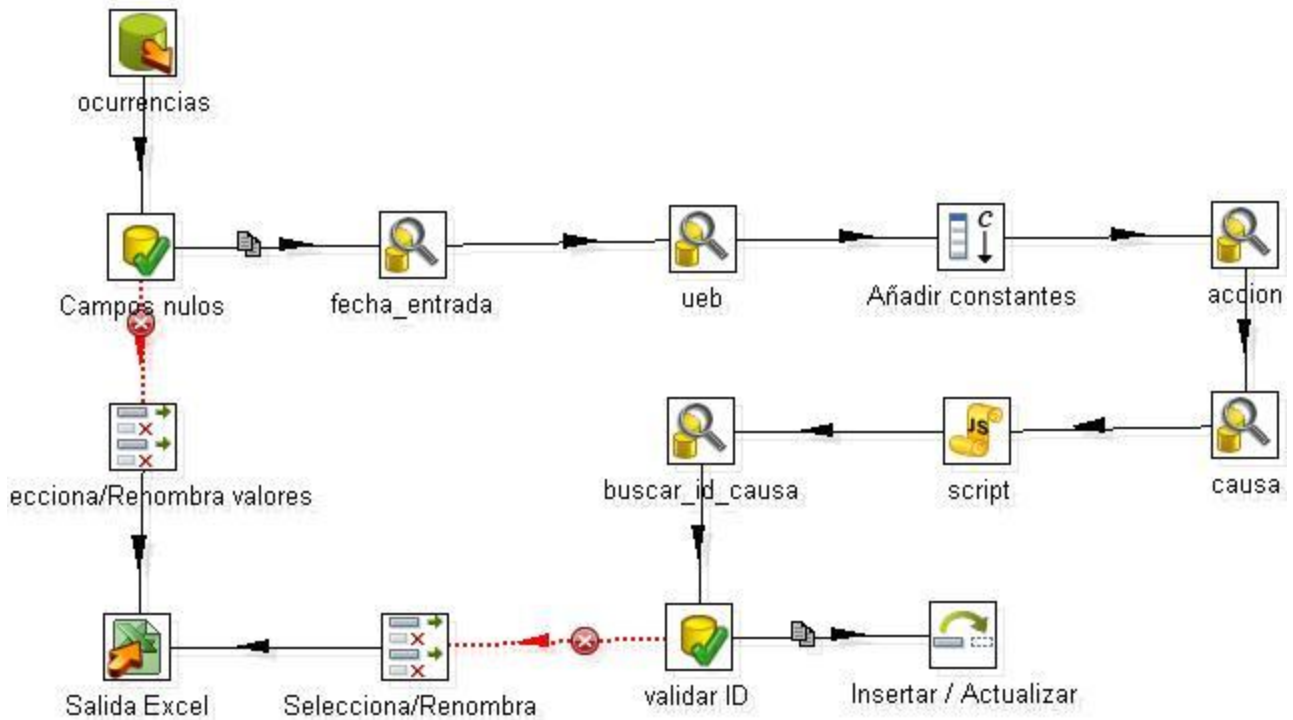


Figura 9: Transformación del hecho ocurrencia.

3.2.2. Implementación de los trabajos

Al concluir la realización de todas las transformaciones necesarias para la carga de los datos, se realiza la implementación del trabajo, el cual se encarga de ejecutar todas las transformaciones en un orden lógico definido, primero las dimensiones y después el hecho (Ver figura 10). Al ejecutar el trabajo, este primeramente verifica la conexión con el servidor de base de datos y luego ejecuta cada una de las transformaciones definidas.

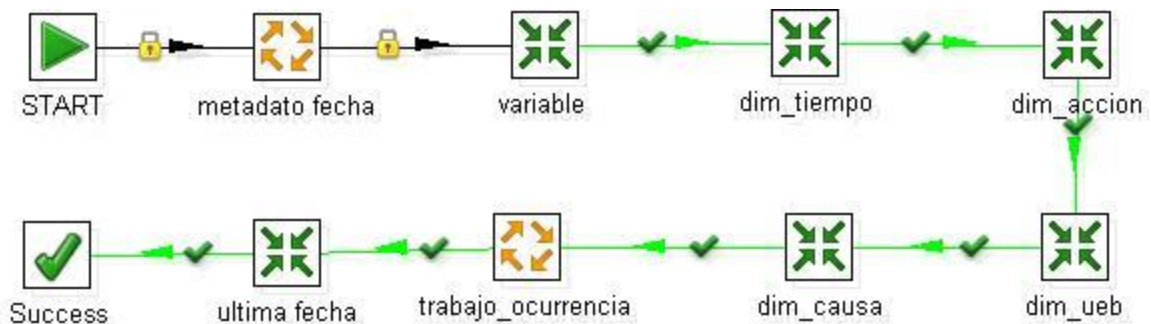


Figura 10: Implementación del trabajo general.

3.3. Implementaci3n de los procesos de Inteligencia de Negocio

3.3.1. Implementaci3n del cubo OLAP

Para la implementaci3n del subsistema de visualizaci3n, es necesaria la creaci3n de los cubos multidimensionales en los cuales se definen las dimensiones, los niveles de jerarquía de las dimensiones y las medidas, esto se realiza a trav3s de la herramienta Pentaho Schema Workbench, donde la misma genera el fichero XML, el cual contendr3 todos los cubos OLAP.

Durante la implementaci3n del cubo, primero se definieron las dimensiones compartidas entre ellos, y luego se implement3 un cubo correspondiente al hecho con que se cuenta; el cual est3 formado 4 dimensiones y una medida (Ver figura 11).

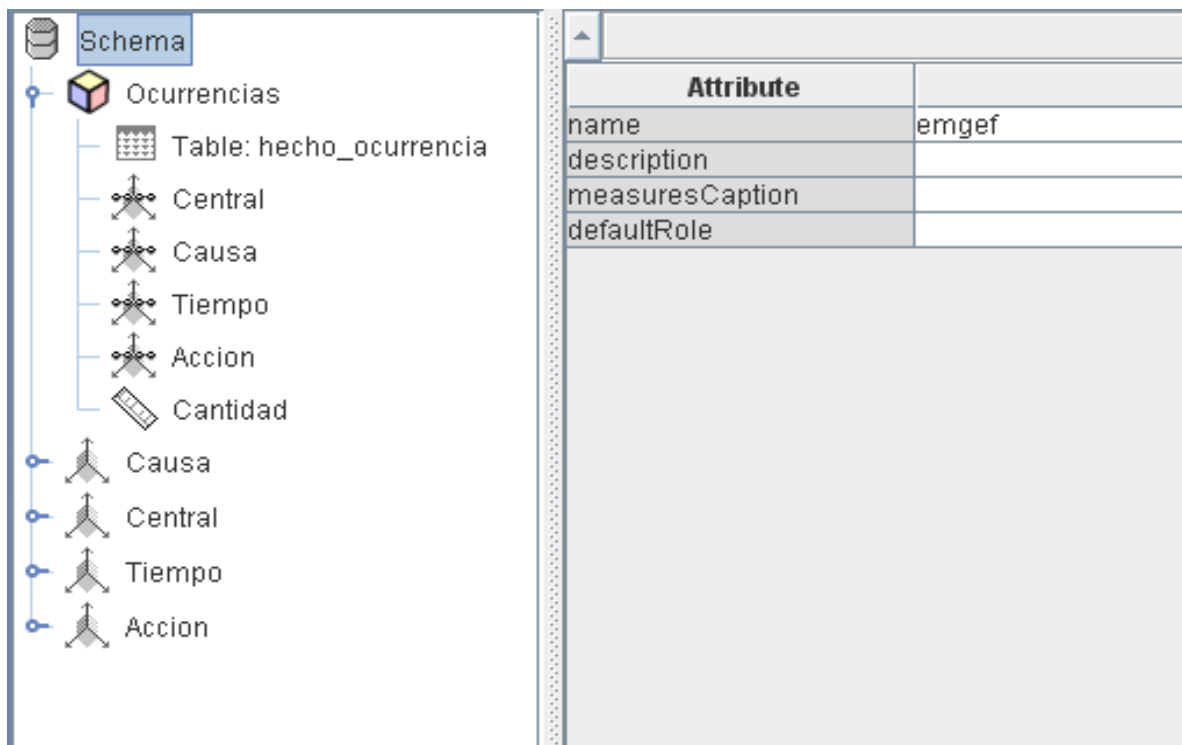


Figura 11: Implementaci3n del cubo OLAP.

3.3.2. Mapa de Navegación

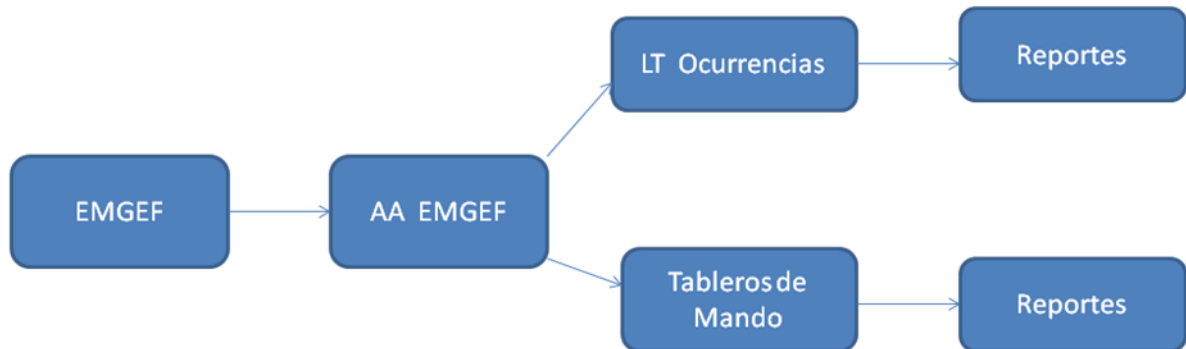


Figura 12: Mapa de navegación.

Elementos del Sistema de Navegación

Nivel Principal

- EMGEF (Empresa de Mantenimiento de Grupos Electrónicos Fuel-Oil): se refiere al entorno en el que se desarrolla la aplicación.

Área de Análisis

- A.A EMGEF: se refiere al área donde estarán contenidos todos los L.T que contienen la agrupación de la información referente a los indicadores a medir en cada uno de ellos.

Libro de Trabajo

- L.T Ocurrencias.

Reporte del libro de trabajo: Ocurrencias.

- Cantidad de ocurrencias por acción, causa externa y tiempo.
- Cantidad de ocurrencias por acción, central y tiempo.
- Cantidad de ocurrencias por acción, mantenimiento y tiempo.
- Cantidad de problemas reportados por causa, central y tiempo.
- Cantidad de problemas solucionados por causa, central y tiempo.

Tableros de Mando

- Tableros de mando

Reportes de los Tableros de mando

- Ocurrencias por causas.

➤ Ocurrencias por causas y ueb.

3.3.3. Arquitectura de la información

La arquitectura de información del mercado de datos para el área de Planeación de la Empresa de Mantenimiento de Grupos Electrónicos Fuel-Oil está definida por un libro de trabajo correspondiente al hecho con que cuenta el mercado, en el cual van a estar los reportes asociados al mismo, además de dos Tableros de mando realizados en el mercado. A continuación se muestra la vista de análisis del mercado de datos para el área de Planeación de EMGEF.

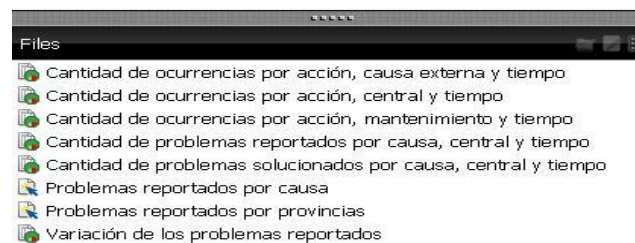


Figura 13: Vista de análisis.

3.3.4. Implementación de los reportes

En la implementación de los reportes se utilizaron consultas **mdx**. A continuación se muestra la consulta para calcular la variación, obteniendo como resultado la diferencia entre un año y otro con respecto a la cantidad de problemas reportados.

```
with member [Measures].[Variación] as '((([fecha].CurrentMember, [Measures].[cantidad]) - ([fecha].PrevMember, [Measures].[cantidad]))'
```

```
select NON EMPTY {[Measures].[cantidad], [Measures].[Variación]} ON COLUMNS,
```

```
NON EMPTY Hierarchize(Union({[fecha].[All fechas]}, [fecha].[All fechas].Children)) ON ROWS
```

```
from [Problemas reportados]
```

Figura 14: Consulta mdx.

3.3.5. Implementación de los Tableros de Mando

El Tablero de Mando es una herramienta de control empresarial que permite establecer y monitorizar los objetivos de una empresa y de sus diferentes áreas o unidades. También se puede considerar como una aplicación que ayuda a una compañía a expresar los objetivos e iniciativas necesarias para cumplir con su estrategia, mostrando de forma continuada cuándo la empresa y los empleados alcanzan los resultados.

A continuación se muestran algunas imágenes de los Tableros de Mandos realizados en el mercado de datos para el área de Planeación de la EMGEF.

En la Figura 15 se representa un Tablero de Mando en el cual dada la acción (problemas reportados y problemas resueltos) se muestra la cantidad de problemas por año, al seleccionar unos de esos años se representa la cantidad de ocurrencias por los tipos de causas existentes.

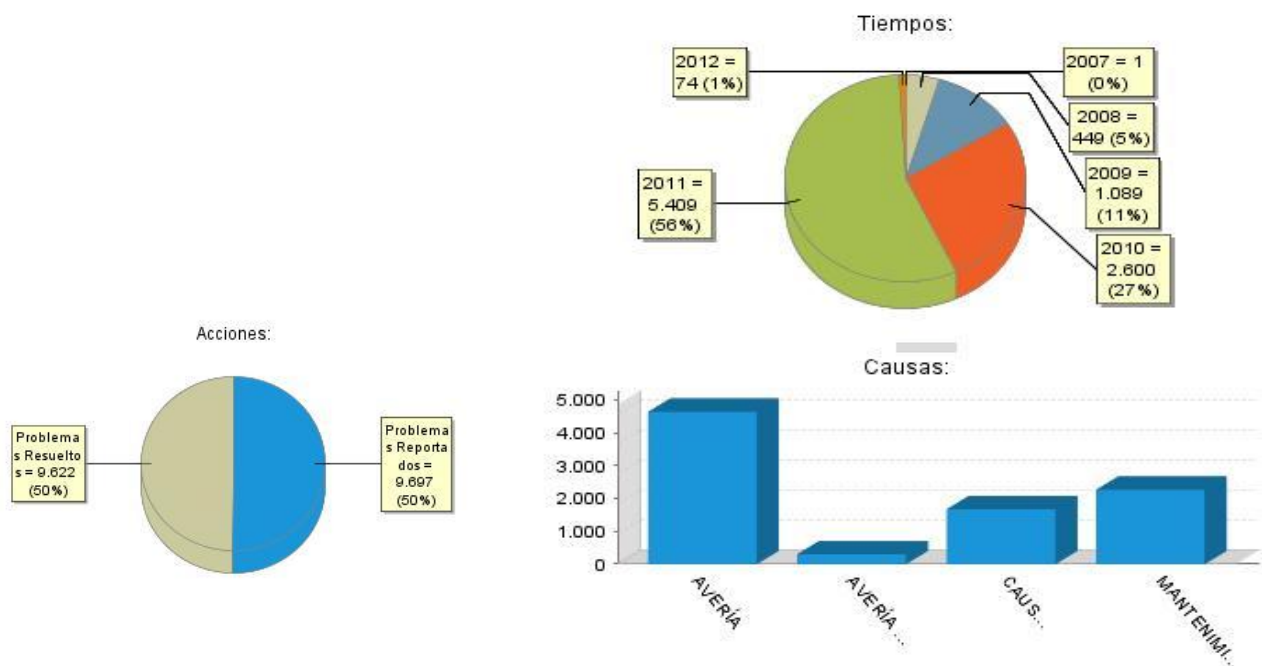


Figura 15: Tablero de mando: Ocurrencias por causas.

En la Figura 16 se muestra otro Tablero de Mando realizado en el Mercado de datos para el área de Planeación de EMGEF, en el cual al seleccionar un determinado año se muestran las ocurrencias por meses, al escoger uno de esos meses se representa la cantidad de problemas reportados y resueltos para dicho mes, al indicar un tipo de acción específica se muestran las causas que dieron origen a esos problemas, dada la causa seleccionada se muestran las Unidades Empresariales de Base (UEB)

responsables de dichos reportes y finalmente se realiza un análisis general de las ocurrencias con el objetivo de ver el estado en que se encuentran las UEB.

Año: 2009

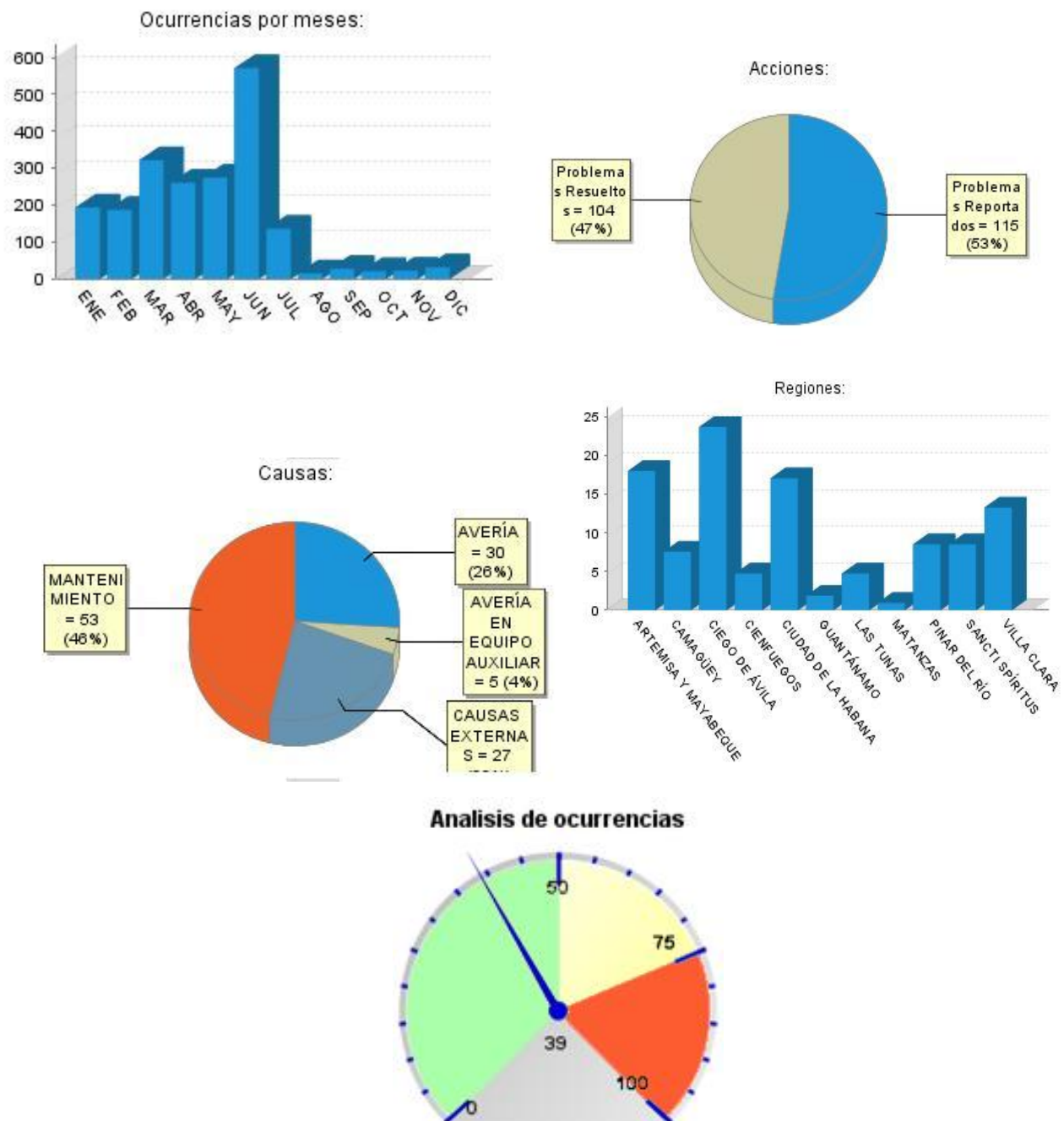


Figura 16: Tablero de mando: Ocurrencias por causas y UEB.

3.4. Conclusiones del capítulo

En el capítulo se describieron los elementos de implementación para la construcción del mercado de datos.

- Como resultado de la implementación del proceso ETL, se logró poblar el mercado de manera satisfactoria.
- Se implementaron los cubos OLAP necesarios para satisfacer los requisitos de información del cliente.
- Se desarrolló el subsistema de visualización con los reportes básicos solicitados por el cliente, así como los cuadros de mandos necesarios para satisfacer las necesidades del cliente.

CAPÍTULO 4 PRUEBAS DE SOFTWARE

Una vez culminado el análisis, diseño e implementación del mercado de datos para el área de Planeación de la EMGEF, se debe probar la solución mediante las listas de chequeo y los casos de prueba para así verificar que el sistema cumpla con los requerimientos necesarios que garanticen al usuario final del sistema la confiabilidad de los datos cargados en el mercado de datos.

4.1. Pruebas de la solución

La evaluación del producto es la comprobación de que la solución está acorde a las necesidades y exigencias de los clientes. Para aprobar la propuesta de solución se realizó un encuentro con el cliente: Mario Pedroso Caballero, quedando conforme con la propuesta mostrada y satisfecho con el cumplimiento de los requisitos. Dando paso a la realización de las pruebas.

Para lograr obtener un producto con calidad existen diferentes tipos de pruebas que pueden ser aplicadas:

- **Prueba unitaria:** se enfocan en un programa o un componente que desempeña una función específica que puede ser probada y que se asegura que funcione tal y como lo define la especificación del programa. Los programadores siempre prueban el código durante el desarrollo, por lo que las pruebas unitarias son realizadas solamente después de que el programador considera que el componente está libre de errores.
- **Prueba de integración:** su objetivo es identificar errores introducidos por la combinación de programas o componentes probados unitariamente, además, verificar que las especificaciones de diseño sean alcanzadas. Componentes individuales son combinados con otros componentes para asegurar que la comunicación, enlaces y los datos compartidos ocurran apropiadamente. No son verdaderamente pruebas de sistema porque los componentes no están implementados en el ambiente operativo.
- **Prueba de sistema:** son usualmente conducidas para asegurar que todos los módulos trabajan como sistema, sin error. Es similar a la prueba de integración pero con un alcance mucho más amplio. Las pruebas del sistema examinan qué tan bien el sistema cumple con los requerimientos de la organización y su utilidad, seguridad y desempeño. También se realizan estas pruebas a la documentación del sistema.

- **Prueba de aceptación:** son realizadas principalmente por los usuarios con el apoyo del equipo del proyecto. El propósito es confirmar que el sistema está terminado, que desarrolla puntualmente las necesidades de la organización y que es aceptado por los usuarios finales.

A parte de las pruebas existentes y que se pueden aplicar, el centro realiza un grupo de pruebas encaminadas a la culminación del producto, estas pruebas realizadas son:

- **Pruebas internas:** son realizadas por el equipo de desarrollo y especialistas de los departamentos que participan en la solución.
- **Pruebas de liberación:** son realizadas por los especialistas de calidad, donde en un primer momento son desarrolladas por los especialistas de calidad de DATEC y luego por los especialistas del Centro de Calidad para Aplicaciones Tecnológicas (CALISOFT).

Dentro de las herramientas utilizadas para que se apliquen los distintos tipos de pruebas se tienen los casos de pruebas y las listas de chequeo.

4.2. Diseño de casos de prueba

Para lograr la calidad del producto de software es necesario realizar un conjunto de evaluaciones durante todo el proceso de desarrollo que implique al cliente y desarrollador. Para ello se diseñaron casos de prueba basados en los casos de uso y a su vez en los requerimientos funcionales, comparando cada funcionalidad implementada con la descrita, para verificar hasta qué punto cumplía con las necesidades del cliente.

Para el mercado de datos para el área de Planeación de la EMGEF se diseñó un caso de prueba, a continuación se muestra un escenario del diseño de caso de prueba correspondiente al caso de uso: Presentar información sobre ocurrencia.

Descripción general					
Caso de Uso(CU) "Presentar información sobre ocurrencias" recoge la información referente a las ocurrencias presentadas por los motores.					
Condiciones de ejecución					
Completitud del almacén. Carga de los Datos					
SC Reportes Candidatos					
Escenario	Descripción	Variables de entrada	Variable de salida	Respuesta del sistema	Flujo central
EC 1.1 1-Cantidad de ocurrencias por acción, causa externa y tiempo	Permite visualizar los reportes con las variables presentes en los mismos.	Acción, Causa, Tiempo	Cantidad	El sistema muestra todas las variables disponibles para la cantidad de ocurrencias por acción, causa externa y tiempo.	Se abre la aplicación. Se autentica. Se entra al sistema. Se selecciona en la parte superior izquierda el área de análisis (AA) EMGEF. Se selecciona el libro de trabajo (LT) Ocurrencia. Se selecciona en el área inferior izquierda el reporte: "Cantidad de ocurrencias por acción, causa externa y tiempo".
EC 1.2 2-Cantidad de ocurrencias por acción, central y tiempo	Permite visualizar los reportes con las variables presentes en los mismos.	Acción, Central, Tiempo	Cantidad	El sistema muestra todas las variables disponibles para la cantidad de ocurrencias por acción, central y tiempo.	Se abre la aplicación. Se autentica. Se entra al sistema. Se selecciona en la parte superior izquierda el área de análisis (AA) EMGEF. Se selecciona el libro de trabajo (LT) Ocurrencia. Se selecciona en el área inferior izquierda el reporte: "Cantidad de ocurrencias por acción, central y tiempo".

Figura 17: Caso de prueba Presentar información sobre ocurrencias.

Para mayor entendimiento de este caso de prueba ver Anexo 1.

4.3. Lista de Chequeo

La lista de chequeo es un documento que tiene un conjunto de parámetros a medir sobre un aspecto determinado, dígase documentación o aplicación. Es un instrumento de medición y evaluación que consiste básicamente en un formulario de preguntas referentes al atributo de calidad que se está probando y de las características del documento en el caso de la documentación. Cada pregunta tiene asociada una evaluación en una escala que da una medida del grado de cumplimiento y disponibilidad de la propiedad evaluada, de esta manera se determina la evaluación del elemento probado.

La lista de chequeo contiene diferentes indicadores a evaluar los cuales se encuentran distribuidos en tres secciones fundamentales:

- **Estructura del documento:** abarca todos los aspectos definidos por el expediente de proyecto o el formato establecido por el proyecto.
- **Indicadores definidos por la metodología:** abarca todos los indicadores a evaluar durante la etapa de desarrollo del mercado.
- **Semántica del documento:** contempla todos los indicadores a evaluar respecto a la ortografía,

redacción y demás.

A cada uno de los indicadores anteriores los describe los siguientes elementos:

- **Peso:** Define si el indicador a evaluar es crítico o no.
- **Indicadores a evaluar:** Son los indicadores a evaluar en las secciones Estructura del documento, semántica del documento e indicadores definidos por las diferentes etapas.
- **Evaluación:** Es la forma de evaluar el indicador en cuestión. El mismo se evalúa de 1 en caso de que exista alguna dificultad sobre el indicador y 0 en caso de que el indicador revisado no presente problemas.
- **No procede:** Se usa para especificar que el indicador no es necesario evaluarlo en ese caso.
- **Cantidad de elementos afectados:** Especifica la cantidad de errores encontrados sobre el mismo indicador.
- **Comentario:** Especifica los señalamientos o sugerencias que quiera incluir la persona que aplica la lista de chequeo.

En esta investigación se aplicaron las siguientes listas de chequeo a los artefactos de los procesos Extracción Transformación y Carga (ETL) con el fin de evaluar y verificar el potencial de cada uno de ellos midiendo la confiabilidad y seguridad de los datos cargados:

- Lista de chequeo del Mapa Lógico de Datos.
- Lista de chequeo del Diccionario de Datos.
- Lista de chequeo de Registro de Sistemas Fuentes.

Una vez aplicada la lista de chequeo se detectan los indicadores evaluados de mal y con el objetivo de darles solución se especifican en una **tabla de no conformidades** (VALENCIA), la cual presenta la siguiente estructura:

- **No:** es un número consecutivo e indica la cantidad de no conformidades identificadas.
- **Elemento de evaluación:** se refiere a un número que identifica al elemento de evaluación para el cual se corresponden los indicadores identificados.
- **No conformidad:** especifica la no conformidad a la que se refiere.
- **Fase correspondiente:** especifica la fase del procedimiento a la que corresponde la no conformidad encontrada.
- **Significación:** especifica si la no conformidad es o no significativa, dependiendo si el indicador

es o no crítico.

- **Recomendación:** especifica si la no conformidad es una recomendación, es decir que no es de obligatorio cumplimiento que se solucione por parte de los diseñadores.
- **Estado NC:** especifica el estado de solución en que se encuentra la no conformidad, puede ser: pendiente o solucionada.
- **Respuesta del equipo de desarrollo:** si es necesario se especifica la respuesta que le da el equipo de desarrollo a la no conformidad.

4.4. Aplicación de la Lista de chequeo

Tabla 9: Aplicación de la Lista de chequeo para el artefacto “Mapa Lógico”.

Estructura del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
crítico	1. ¿El alcance del proyecto describe correctamente los datos de las dimensiones y hechos del mercado de datos?	0			
crítico	2. ¿El objetivo expresa correctamente el propósito del documento?	0			
crítico	3. ¿Existe una adecuada correspondencia entre el origen de los datos y los atributos del mercado?	0			
	4. ¿Se hace un uso adecuado del control del documento?	0			
	5. ¿En la sección de acrónimos se definen todos los acrónimos utilizados en el	1			No se define el acrónimo UEB

	documento?				
Indicadores definidos en el desarrollo					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
	1. ¿Se utilizó un lenguaje cuyas sentencias son expresables mediante una sintaxis bien definida?	0			
Semántica del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
crítico	1. ¿Se han identificado errores ortográficos en el documento?	1			Se identificaron errores ortográficos en el documento.
crítico	2. ¿Se entiende claramente lo que se ha especificado en el documento?	0		0	
	3. ¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?	0		0	

4.5. Resultados de las pruebas

Los resultados obtenidos en las pruebas realizadas a la solución son los siguientes:

- **Pruebas internas:** Realizadas por los especialistas del departamento y miembros del equipo de desarrollo, las cuales arrojaron los siguientes resultados.
 - Aplicación: 2 No Conformidades (NC)

- Reportes candidatos: 0 NC
- Casos de prueba: 7 NC

Para un total de 9 NC.

➤ **Lista de chequeo:** Después de aplicada, se genera el siguiente gráfico de barra, donde se visualiza el comportamiento de los 9 indicadores identificados, de los cuales 5 son críticos. Ver Figura 18.

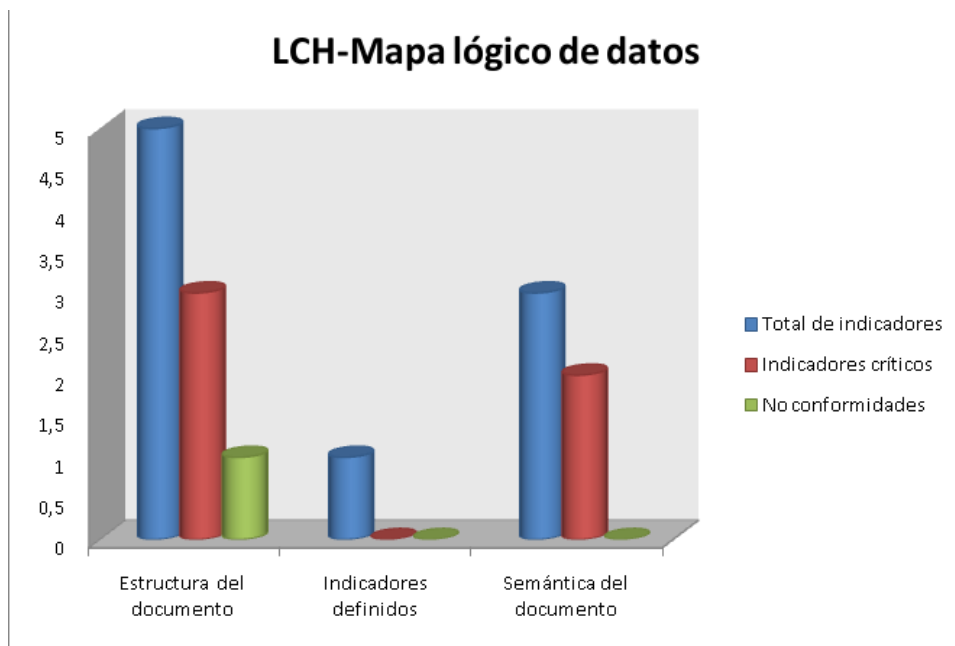


Figura 18: Resultado de la lista de chequeo.

- **Pruebas de aceptación:** Se obtuvo la carta de aceptación del producto por parte del cliente.

Todas las NC anteriormente mencionadas fueron resueltas satisfactoriamente en el período de tiempo establecido.

4.6. Conclusiones del capítulo

En este capítulo se realizaron las pruebas al mercado de datos

- Se aplicaron los casos de prueba correspondientes a los casos de uso de información del sistema, así como la lista de chequeo elaborada, evaluando los resultados de los procesos de análisis e integración de datos.
- Se logró un acercamiento al grado de conformidad y aceptación por parte de los usuarios.

CONCLUSIONES

Al concluir la solución, se logró cumplir con todos los objetivos trazados y las tareas de investigación propuestas, desarrollándose una solución que ofrece soporte a la toma de decisiones mediante un mercado de datos para el área de Planeación de la Empresa de Mantenimiento de Grupos Electrógenos Fuel-Oil. Lo anteriormente planteado se ve demostrado a través de:

- La realización de los fundamentos teóricos donde fueron abordados los principales conceptos relacionados con los mercados de datos, escogiéndose una metodología que permitió organizar de manera estructurada el proceso de desarrollo del mercado de datos y las herramientas que cuentan con todas las características necesarias para el desarrollo correcto de la solución propuesta.
- Se realizó el análisis y diseño del mercado de datos, logrando un mejor entendimiento del proceso de negocio del área de Planeación de la Empresa de Mantenimiento de Grupos Electrógenos Fuel-Oil, lo que posibilitó la implementación de una solución robusta.
- El mercado de datos obtenido en la implementación contribuye al proceso de toma de decisiones, garantizando además la integridad y validez de la información.
- Se evaluó la solución mediante la aplicación de los casos de pruebas y listas de chequeo obteniéndose resultados satisfactorios.

RECOMENDACIONES

- Poner en explotación el mercado de datos para probar las funcionalidades del mismo y detectar nuevas funcionalidades.
- Realizar un estudio acerca de técnicas de optimización, que puedan ser aplicadas al proceso de extracción, transformación y carga e Inteligencia de Negocios desarrollado.

REFERENCIAS BIBLIOGRÁFICAS

- BERNABEU, R. D. *DATA WAREHOUSING: Investigación y Sistematización de Conceptos–HEFESTO: Metodología propia para la Construcción de un Data Warehouse*. Córdoba: 2007, 124 p. Disponible en: <http://es.scribd.com/doc/55234594/45/Esquema-Constelacion>.
- CAVSI. *¿Qué es un Sistema Gestor de Bases de Datos o SGBD?* Disponible en: <http://www.cavsi.com/preguntasrespuestas/que-es-un-sistema-gestor-de-bases-de-datos-o-sgbd/>.
- CORAIL, M. C. *Pentaho Open BI*. Las Condes: 2012, [Consultado el: 20/01/2012]. Disponible en: <http://www.cognus.cl/content/view/271452/Pentaho-Open-BI.html>.
- CRISTIAN. *El servidor Apache Tomcat*. Colombia: 2006, [Consultado el: 20/10/2011]. Disponible en: <http://casidiablo.net/el-servidor-apache-tomcat/>.
- DATA CLEANER. *DataCleaner news* Disponible en: <http://datacleaner.eobjects.org/>.
- EASY, P. P. A. M. *Pentaho BI Platform and Server* Disponible en: http://community.pentaho.com/projects/bi_platform/.
- ESPINOSA, R. *El Rincon del BI* [Consultado el: 10/01/2012] Disponible en: <http://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modelado-dimensional/>.
- GRAVITAR. *Características* Disponible en: <http://www.gravitar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>.
- KIMBALL, R. y ROSS, M. *The Data Warehouse Toolkit*. 2da ed. 2002, Disponible en: <http://eng.ui.ac.ir/~m.rezaei/dataWarehouse/kim1.pdf>. ISBN 0-471-20024-7.
- LÓPEZ, I. A. *Data warehouse de soporte a datos de GSA*. 2008.
- ROBELO, B. *Conceptos sobre Datawarehouse* 2010, Disponible en: <http://www.bernardorobelo.com/2010/09/conceptos-sobre-datawarehouse.html>.
- SANZ, M. R. *ANÁLISIS Y DISEÑO DE UN DATAMART PARA EL SEGUIMIENTO ACADÉMICO DE ALUMNOS EN UN ENTORNO UNIVERSITARIO*. Diplomado, Universidad Carlos III de Madrid Escuela Politécnica Superior, 2010.
- SCRIBD. *PostgreSQL* Disponible en: <http://es.scribd.com/doc/36570462/postgreSQL-investigacion>.
- SINNEXUS. *¿Qué es Business Intelligence?* [Consultado el: 12/11/2011] Disponible en: http://www.sinnexus.com/business_intelligence/.

SOTO, L. *Sistemas De Apoyo Toma De Decisiones* Ensenada: Disponible en: <http://www.mitecnologico.com/Main/SistemasDeApoyoTomaDeDecisiones>.

TECHNOLOGIES, B. P. W. I. A. *Visual Paradigm for UML - UML tool for software application development* Disponible en: <http://www.visual-paradigm.com/product/vpum/>.

TOOLS, P. P. *Latest news* Disponible en: <http://www.pgadmin.org/>.

VALENCIA, R. M. *¿QUÉ SON LOS EIS?* 1969, 7 p. Disponible en: <http://www.timogo.com.mx/articulos/EIS.pdf>.

BIBLIOGRAFÍA

1. BERNABEU, R. D. DATA WAREHOUSING: Investigación y Sistematización de Conceptos–HEFESTO: Metodología propia para la Construcción de un DataWarehouse. Córdoba: 2007, 124 p. Disponible en: <http://es.scribd.com/doc/55234594/45/Esquema-Constelacion>.
2. CAVSI. ¿Qué es un Sistema Gestor de Bases de Datos o SGBD? Disponible en: <http://www.cavsi.com/preguntasrespuestas/que-es-un-sistema-gestor-de-bases-de-datos-o-sgbd/>.
3. CORAIL, M. C. Pentaho Open BI. Las Condes: 2012, [Consultado el: 20/01/2012]. Disponible en: <http://www.cognus.cl/content/view/271452/Pentaho-Open-BI.html>.
4. CRISTIAN. El servidor Apache Tomcat. Colombia: 2006, [Consultado el: 20/10/2011]. Disponible en: <http://casidiablo.net/el-servidor-apache-tomcat/>.
5. DATACLEANER. DataCleaner news Disponible en: <http://datacleaner.eobjects.org/>.
6. EASY, P. P. A. M. Pentaho BI Platform and Server Disponible en: http://community.pentaho.com/projects/bi_platform/.
7. EASY, P. P. A. M. Pentaho Products Disponible en: <http://www.pentaho.com/explore/products/>.
8. EASY, P. P. A. M. Thread: Data Cleaner 2. 2011, nº [Consultado el: 21/10/2011]. Disponible en: <http://forums.pentaho.com/showthread.php?80965-Data-Cleaner-2>.
9. ESPINOSA, R. El Rincon del BI [Consultado el: 10/01/2012 Disponible en: <http://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modelado-dimensional/>.
10. GRAVITAR. Características Disponibles en: <http://www.gravitar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>.
11. HAAS, L. Beauty and the Beast: The Theory and Practice of Information Integration. 2007, Disponible en: <http://www.cs.uoregon.edu/Classes/08W/cis607sii/papers/Haas07.pdf>.
12. HANIK, F. INTRODUCTION TO APACHE TOMCAT 6. publicado el: 15/02/2012 de 2007, última actualización: 15/02/2012. Disponible en: http://www.springsource.com/files/uploads/all/pdf_files/news_event/Intro_to_Tomcat6.pdf.

13. KIMBALL, R. y ROSS, M. The Data Warehouse Toolkit. 2da ed. 2002, Disponible en: <http://eng.ui.ac.ir/~m.rezaei/dataWarehouse/kim1.pdf>. ISBN 0-471-20024-7.
14. LÓPEZ, I. A. Data warehouse de soporte a datos de GSA. 2008.
15. PONNIAH, P. Data Warehousing Fundamentals. 1ra ed. Canadá: 2001, [Consultado el: 28/09/2011]. 511 p. Disponible en: http://www.amazon.com/Data-Warehousing-Fundamentals-Comprehensive-Professionals/dp/0471412546#reader_0471412546. ISBN 0-471-41254-6.
16. POSTGRESQL-ES. Disponible en: http://www.postgresql.org.es/sobre_postgresql.
17. ROBELO, B. Conceptos sobre Datawarehouse 2010, Disponible en: <http://www.bernardorobelo.com/2010/09/conceptos-sobre-datawarehouse.html>.
18. SANZ, M. R. ANÁLISIS Y DISEÑO DE UN DATAMART PARA EL SEGUIMIENTO ACADÉMICO DE ALUMNOS EN UN ENTORNO UNIVERSITARIO. Diplomado, Universidad Carlos III de Madrid Escuela Politécnica Superior, 2010.
19. SCRIBD. PostgreSQL Disponible en: <http://es.scribd.com/doc/36570462/postgreSQL-investigacion>.
20. SINNEXUS. Datawarehouse Disponible en: http://www.sinnexus.com/business_intelligence/datawarehouse.aspx.
21. SINNEXUS. Persistencia MOLAP, ROLAP, HOLAP Disponible en: http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx.
22. SINNEXUS. ¿Qué es Business Intelligence? [Consultado el: 12/11/2011] Disponible en: http://www.sinnexus.com/business_intelligence/.
23. SOTO, L. Sistemas De Apoyo Toma De Decisiones Enseñada: Disponible en: <http://www.mitecnologico.com/Main/SistemasDeApoyoTomaDeDecisiones>.
24. TAD, U.-. 7 formas de integración de datos. Disponible en: http://www.muycomputerpro.com/2010/02/05/centro-de-conocimientowhite-papers7-formas-de-integracion-de-datos_we9erk2xxdazdqyeebujg7or6shmcwyoivks1jrr48ecaalib5aehoygizgw2zmr/.
25. TECHNOLOGIES, B. P. W. I. A. Visual Paradigm for UML - UML tool for software application development Disponible en: <http://www.visual-paradigm.com/product/vpum/>.
26. TOMCAT, A. Apache Tomcat Disponible en: <http://tomcat.apache.org/>.

27. TOOLS, P. P. Latest news Disponible en: <http://www.pgadmin.org/>.
28. VALENCIA, R. M. ¿QUÉ SON LOS EIS? 1969, 7 p. Disponible en:
<http://www.timogo.com.mx/articulos/EIS.pdf>.

ANEXOS

Anexo 1 Caso de prueba de caso de uso: “Presentar información sobre ocurrencias”.

Descripción general					
Caso de Uso(CU) "Presentar información sobre ocurrencias" recoge la información referente a las ocurrencias presentadas por los motores.					
Condiciones de ejecución					
Completitud del almacén.					
Carga de los Datos					
SC Reportes Candidatos					
Escenario	Descripción	Variables de entrada	Variable de salida	Respuesta del sistema	Flujo central
EC 1.1 1-Cantidad de ocurrencias por acción, causa externa y tiempo	Permite visualizar los reportes con las variables presentes en los mismos.	Acción, Causa, Tiempo	Cantidad	El sistema muestra todas las variables disponibles para la cantidad de ocurrencias por acción, causa externa y tiempo.	Se abre la aplicación. Se autentica. Se entra al sistema. Se selecciona en la parte superior izquierda el área de análisis (AA) EMGEF. Se selecciona el libro de trabajo (LT) Ocurrencia. Se selecciona en el área inferior izquierda el reporte: "Cantidad de ocurrencias por acción, causa externa y tiempo".
EC 1.2 2-Cantidad de ocurrencias por acción, central y tiempo	Permite visualizar los reportes con las variables presentes en los mismos.	Acción, Central, Tiempo	Cantidad	El sistema muestra todas las variables disponibles para la cantidad de ocurrencias por acción, central y tiempo.	Se abre la aplicación. Se autentica. Se entra al sistema. Se selecciona en la parte superior izquierda el área de análisis (AA) EMGEF. Se selecciona el libro de trabajo (LT) Ocurrencia. Se selecciona en el área inferior izquierda el reporte: "Cantidad de ocurrencias por acción, central y tiempo".
EC 1.3 3- Cantidad de ocurrencias por acción, mantenimiento y tiempo	Permite visualizar los reportes con las variables presentes en los mismos.	Acción, Causa, Tiempo	Cantidad	El sistema muestra todas las variables disponibles para la cantidad de ocurrencias por acción, mantenimiento y tiempo.	Se abre la aplicación. Se autentica. Se entra al sistema. Se selecciona en la parte superior izquierda el área de análisis (AA) EMGEF. Se selecciona el libro de trabajo (LT) Ocurrencia. Se selecciona en el área inferior izquierda el reporte: "Cantidad de ocurrencias por acción, mantenimiento y tiempo".
EC 1.4 4-Cantidad de problemas reportados por causa, central y tiempo	Permite visualizar los reportes con las variables presentes en los mismos.	Causa, Central, Tiempo	Cantidad	El sistema muestra todas las variables disponibles para la cantidad de problemas reportados por causa, central y tiempo.	Se abre la aplicación. Se autentica. Se entra al sistema. Se selecciona en la parte superior izquierda el área de análisis (AA) EMGEF. Se selecciona el libro de trabajo (LT) Ocurrencia. Se selecciona en el área inferior izquierda el reporte: "Cantidad de problemas reportados por causa, central y tiempo".

EC 1.5 5-Cantidad de problemas solucionados por causa, central y tiempo	Permite visualizar los reportes con las variables presentes en los mismos.	Causa, Central, Tiempo	Cantidad	El sistema muestra todas las variables disponibles para la cantidad de problemas solucionados por causa, central y tiempo.	Se abre la aplicación. Se autentica. Se entra al sistema. Se selecciona en la parte superior izquierda el área de análisis (AA) EMGEF. Se selecciona el libro de trabajo (LT) Ocurrencia. Se selecciona en el área inferior izquierda el reporte: "Cantidad de problemas solucionados por causa, central y tiempo".
EC 1.6 6-Variación de los problemas reportados	Permite visualizar los reportes con las variables presentes en los mismos.	Tiempo	Cantidad	El sistema muestra todas las variables disponibles para la variación de los problemas reportados	Se abre la aplicación. Se autentica. Se entra al sistema. Se selecciona en la parte superior izquierda el área de análisis (AA) EMGEF. Se selecciona el libro de trabajo (LT) Ocurrencia. Se selecciona en el área inferior izquierda el reporte: "Variación de los problemas reportados".

GLOSARIO DE TÉRMINOS

Cubo: colección de dimensiones y medidas en un área temática particular.

Código abierto: término con el que se conoce al software distribuido y desarrollado libremente.

Datamart: mercado de datos.

Datawarehouse: almacén de datos.

EMGEF: Empresa de Mantenimiento de Grupos Electrógenos.

HOLAP: Procesamiento Analítico en Línea Híbrido.

MOLAP: Procesamiento Analítico en Línea Multidimensional.

Open source: Código abierto.

ROLAP: Procesamiento Analítico en Línea Relacional.

Rollback: retroceso.

Triggers: disparadores.