

Universidad de las Ciencias Informática



Facultad 2

Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos

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

Autor (es): Danisey Garcia Rodríguez
Alejandro Manuel Salazar Calderín

Tutores: Ing. Mayrín Ramos Maestre
Ing. Alberto Arce Martínez

Co-tutor: DrC. Raúl René Crespo Heredia
Ing. Yisel Valdés Rodríguez

“Año 55 de la Revolución”
Ciudad de la Habana, Julio 20

DECLARACIÓN DE AUTORÍA

Declaramos ser autores de la presente tesis y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

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

Danisey Garcia Rodríguez

Alejandro Manuel Salazar Calderín

Firma del Autor

Firma del Autor

Ing. Mayrín Ramos Maestre

Firma del Tutor

Ing. Alberto Arce Martínez

Firma del Tutor



Datos de Contacto

Tutores:

Ing. Mayrín Ramos Maestre: Ingeniera en Ciencias Informáticas.

Especialista General.

E-mail: mmaestre@uci.cu

Ing. Alberto Arce Martínez: Ingeniero en Ciencias Informáticas.

Especialista General.

E-mail: aarce@uci.cu



AGRADECIMIENTOS

Danisey

A mi familia, mis primos, mis tíos, mis abuelos y demás, por su apoyo y su cariño.

En especial a mis padres Niurka y Quirino que siempre me incitaron a luchar por el futuro, por todo ese amor sin límites, por su apoyo, dedicación, sacrificio, por ser ejemplo para mí y por haber perdido el sueño en el empeño.

A mi abuela María que siempre ha estado ahí para mí y me ha ofrecido su apoyo incondicional, ha sido mi alegría y mi guía.

A mi hermanita y mi primo Yander que los adoro y siempre han estado junto a mí.

A mi novio Jorge Roberto Jova por ofrecirme toda su confianza, amor y apoyo incondicional en todo este tiempo. Gracias por tu ayuda, tus consejos y por hacer de mí alguien diferente.

A mis amistades que siempre estuvieron ahí para mí en todo momento, Saly, Dianky, la Titi, Rafiky, Yany, Rolo, Ángel, Rutminy, Norey y a todos los demás que son innumerables.

Al Joe que más que un profesor de gimnasio y más que un padrino, fue siempre un amigo. Gracias por tus consejos y por situarme en mis mejores momentos.

A mis cotutores Yisel, Crespo y mi tutora Mayrín que siempre estuvieron atentos a nosotros, nos tuvieron confianza y sobre todo paciencia. Gracias por su Tiempo.

A mi compañero de tesis y su novia Yanet que siempre estuvieron dispuestos a ofrecirme su ayuda incondicional.

A Deyler, Celia, Tan, Dany y Pedrito que fueron mis compañeros de cuarto los últimos tiempos y con quienes compartí muchas experiencias.

A todos los que siempre estuvieron pendiente de mí. A TODOS MUCHÍSIMAS GRACIAS.



AGRADECIMIENTOS

Alejandro

Quisiera agradecer en especial a mi mamá Ana Iris, mis hermanos Angel Rilder y Manuel Alejandro por ser un ejemplo para mí, por aconsejarme y apoyarme en todo momento, por haber estado siempre tan pendientes de mí, por ser las personas que siempre me incitaron a luchar por mi futuro y lo más importante por ser ellos la razón de mi existir.

Sobre todo dar las gracias a mi novia Yanet Tolentino, la persona que me ayudó, me apoyó y estuvo siempre ahí para mí durante toda mi carrera. Especialmente por el amor y comprensión que me ha brindado todos estos años y por ser la persona por la cual estoy hoy haciendo mi sueño realidad.

Agradecer a mi suegro José Enrique Tolentino y a mi suegra María del Carmen por haberme acogido como un hijo más. A mí cuñado Williams. A ellos muchas gracias por haberme ayudado y apoyado en todo momento y sobre todo por sus consejos.

A mi compañera de tesis por haber luchado a mi lado en contra de las adversidades y los obstáculos que nos encontramos en el camino.

A mi tutora Mayrín, a mi cotutora Yisel y al cotutor Raúl René Crespo. A ellos por habernos apoyado siempre que lo necesitamos, por ser más que tutores y cotutores ya que nos brindaron su amistad incondicional. Muchas gracias por la confianza y sobre todo la paciencia que tuvieron con nosotros.

A mi familia, amigos(as), compañeros(as) de estudio, en especial a Andrés, Niorges y Asiel y a todas aquellas personas que de una forma u otra contribuyeron con su granito de arena en mi formación profesional.

A mí, porque no, por haber peleado por un sueño que me parecía imposible, el cual logré.



RESUMEN

Los Almacenes y Mercados de Datos en la actualidad juegan un papel fundamental en las empresas. Ambos son repositorios de datos diseñados para facilitar el análisis de los datos en función de apoyar la toma de decisiones, siendo el Mercado de Datos un almacén más pequeño, especializado en un área de negocio.

El objetivo fundamental del presente trabajo es especificar los pormenores del análisis, diseño e implementación del Mercado de Datos para el sistema Front End Localización Automática de Vehículos (FE-AVL). Esta solución centraliza la información referente a las alarmas activadas por unidades ya sea en rejas o rutas, así como su localización geográfica. Para su desarrollo se emplearon varias herramientas de la Suite Pentaho como: Workbench, Diseño de Reportes, Integración de Datos y Servidor de Inteligencia de Negocio, guiado por la metodología de Desarrollo de Almacenes de Datos propuesta por la Facultad 6 de la Universidad de Ciencias Informáticas. Además de la implementación del Mercado de Datos, se probaron los requisitos del mismo a partir de los Casos de Pruebas basados en Casos de Uso, donde se obtuvo un resultado satisfactorio.

La solución propuesta ofrece la posibilidad de crear Vistas de Análisis y Reportes de dos tipos para la visualización de la información relacionada con las alarmas, las flotas y las posiciones geográficas de las mismas, permitiendo así, el análisis de la información en apoyo al proceso de toma de decisiones.

Palabras claves: Mercado de Datos, Almacén de Datos, Localización Automática de Vehículos, Toma de Decisiones, FE-AVL, Vistas de Análisis.



ÍNDICE

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA	9
1.1. Introducción	9
1.2. Conceptos Fundamentales	9
1.2.1. Sistemas Inteligentes de Transporte	9
1.2.2. Sistema FE-AVL	11
1.2.3. Sistemas de Soporte de Decisiones	11
1.2.4. Almacenes de Datos y Mercados de Datos	12
1.2.5. Modelado de Datos	16
1.2.6. Extracción, Transformación y Carga (ETL)	20
1.2.7. Sistemas OLAP y OLTP	21
1.3. Tendencias Actuales	23
1.3.1. Almacenes de Datos y Mercados de Datos	23
1.3.2. Sistemas de Localización automática de vehículos (AVL)	24
1.4. Lenguajes utilizados	26
1.4.1. Lenguaje MDX	26
1.4.2. Lenguaje SQL	26
1.4.3. Lenguaje de Modelado UML	27
1.5. Herramientas utilizadas	27
1.5.1. Herramienta CASE para el modelado: Visual Paradigm para UML	27
1.5.2. Colección de Pentaho para la Inteligencia de Negocio (Pentaho Open BI Suite)	27
1.5.3. Power Architect	30
1.5.4. PostgreSQL	31
1.6. Metodologías de Desarrollo de Almacenes de Datos	32
1.6.1. Metodología de Desarrollo de Almacenes de Datos.	33
1.7. Conclusiones	35
CAPÍTULO 2. ANÁLISIS Y DISEÑO DEL MERCADO DE DATOS	36
2.1. Introducción	36



2.2.	Estudio Preliminar del Negocio.....	36
2.3.	Requerimientos del Mercado de Datos	37
2.3.1	Requisitos de Información	38
2.3.2	Requisitos Funcionales	40
2.3.3	Requisitos No Funcionales.....	41
2.3.1.	Roles y Permisos	44
2.4.	Reglas del Negocio	45
2.5.	Casos de Uso del Sistema	45
2.5.1.	Descripción de Casos de Uso del Sistema	46
2.6.	Arquitectura del Mercado de Datos	48
2.7.	Modelo de Datos Dimensional.....	50
2.8.	Descripción de dimensiones y hechos	53
2.9.	Desarrollo de la Matriz BUS.....	54
2.10.	Modelo de Datos Físico	54
2.11.	Conclusiones Parciales.	55
CAPÍTULO 3. IMPLEMENTACIÓN Y VALIDACIÓN.....		56
3.1.	Introducción.	56
3.2.	Implementación de los subsistemas de integración.....	56
3.3.	Implementación de los trabajos (Jobs).....	59
3.4.	Implementación del subsistema de visualización de datos.....	60
3.4.1.	Cubos OLAP	61
3.4.2.	Vistas de Análisis	63
3.4.3.	Reportes.....	63
3.5.	Validación del Mercado de Datos.....	65
3.5.1.	Diseño de los Casos de Pruebas	66
3.6.	Resultado y Funcionalidades Obtenidas.....	68
3.7.	Conclusiones.	68
CONCLUSIONES GENERALES.....		69



Introducción

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

RECOMENDACIONES.....	70
REFERENCIAS BIBLIOGRÁFICAS	71
GLOSARIO DE TÉRMINOS	74



INTRODUCCIÓN

La necesidad de información constituye una constante en la evolución de la historia humana. Desde antaño los individuos han sentido cierta curiosidad por conocer todo lo que acontece a su alrededor. En ese sentido, el origen de la información permitió el desarrollo y el éxito de determinadas empresas y organizaciones. Actualmente, el mundo se encuentra inmerso en la “Era de la Información”, donde se han engendrado grandes cambios de paradigmas y estrategias establecidas por muchos años, los cuales desarrollaron las nuevas Tecnologías de la Información y las Comunicaciones (TIC), que cada vez son más dinámicas e impredecibles.

Los sistemas empresariales sufren grandes modificaciones, debido en gran medida al crecimiento considerable de la información y a la complejidad de su manipulación, por lo que se hace más difícil la capacidad de generación y almacenamiento de la misma. La estrategia empleada por estos sistemas empresariales es la de fomentar el uso de nuevas metodologías y herramientas que faciliten la toma de decisiones, y que a su vez permitan un rápido acceso a la información manejada por la administración de la empresa.

El empleo de las TIC y los Sistemas de Información (SI) han cambiado la forma de operar en muchas empresas alrededor del mundo, permitiéndoles realizar importantes mejoras a los métodos de administración empleados, como son: la automatización de los procesos operativos y la disponibilidad de una plataforma de información, necesaria para el apoyo en la toma de decisiones. Un Sistema de Información es un conjunto de elementos que interactúan entre sí con el fin de apoyar las actividades de una empresa o negocio. Estos sistemas logran con su implantación ventajas competitivas y la reducción del avance de los competidores. (1)

Uno de los avances que ha impactado desde su surgimiento hasta la actualidad son los Sistemas de Soporte a las Decisiones, debido al gran cúmulo de información que poseen hoy las empresas en todo el mundo, y que son de vital importancia para el funcionamiento y desarrollo de las mismas. Por el valor que posee la información relevante en la Administración Empresarial y por la necesidad de conservación de datos históricos por largos períodos de tiempo, contar con datos personalizados de acuerdo a las necesidades de los usuarios, para el análisis de los mercados y el control de la producción, surgen los Sistemas de Soporte a las Decisiones, Almacenes de Datos (por sus siglas AD) y Mercados de Datos (por

sus siglas MD), que hacen la gestión laboral más fácil a todos los usuarios encargados de manipular la documentación de una empresa o determinada área de la misma.

La toma de decisiones es un proceso que está presente en cada momento de nuestras vidas, donde sea necesario escoger el camino correcto a seguir después de evaluar las diferentes alternativas posibles. En ocasiones resulta muy engorroso cuando se trata de administrar una empresa debido al gran volumen de información generada, razón necesaria para contar con especialistas que sean capaces de analizar profundamente esta información. Las empresas se crean para obtener resultados económicos favorables, en gran medida esto depende de las decisiones acertadas, que a su vez dependen de la disposición de la información en los momentos precisos. Por otra parte la información debe de ser confiable y presentar determinada protección oficial, que no permita alteraciones y falsificaciones. He aquí la importancia de que las entidades puedan disponer de un software para obtener la información necesaria y a su vez agilizar el proceso de toma de decisiones.

Los Sistemas de Soporte a las Decisiones son herramientas administrativas capaces de ayudar a los usuarios encargados de la toma de decisiones de una entidad, a enfrentar problemas de diferentes índoles, mediante una interacción entre la base de datos de la organización y programas de software. Para la ejecución de la función objetiva de estas herramientas, se puede hacer uso de las Bases de Datos (por sus siglas BD), AD y MD de las empresas donde se requiere implementar el sistema.

Los **Almacenes de Datos** y los **Mercados de Datos** son el centro de atención para los grandes sectores empresariales de hoy en día, ya que constituyen uno de los soportes fundamentales para el proceso de toma de decisiones gerenciales; de ahí la importancia de que la información guardada en ellos sea confiable y con calidad. (2)

Cuba no está ajena a este desarrollo y son evidentes los esfuerzos realizados durante los últimos años por distintas instituciones, como son la Oficina Nacional de Estadísticas e Información (ONEI) y la Empresa de Telecomunicaciones de Cuba Sociedad Anónima (ETECSA S.A). Otra de las instituciones que ha jugado un papel primordial en el desarrollo de los Almacenes y Mercados de Datos para la centralización y el control de la información manejada por el sector empresarial es la Universidad de las Ciencias Informáticas (UCI), un centro de altos estudios que tiene como metas desde su creación en el

año 2002 la informatización de la sociedad y el desarrollo de la industria del software en Cuba para contribuir al desarrollo económico del país.

La UCI cuenta con varios centros de desarrollo enfocados a soluciones informáticas en diferentes ramas, como es el Centro de Telemática, cuya misión es el desarrollo de sistemas informáticos en las ramas de la Seguridad Informática y las Telecomunicaciones. En el caso del departamento especializado en las tecnologías de telecomunicación se desarrolló el sistema Front End AVL-Automatic Vehicle Location/Localización Automática de Vehículos (FE-AVL). Este sistema se encarga de la gestión y control de flotas¹, del área funcional de los Sistemas Inteligentes de Transporte (SIT). Los sistemas de gestión y control de flotas tienen como función principal el monitoreo y control de flotas de dispositivos móviles. Además, es necesario que estos sistemas cuenten con funcionalidades que permitan el análisis de información histórica, con el objetivo de apoyar a la toma de decisiones y así reducir gastos y mejorar el desempeño de los negocios.

Actualmente el sistema FE-AVL cuenta con un módulo que ofrece Reportes dinámicos a través de plugins², sin embargo, es necesaria la creación de una estructura consistente, que ofrezca acceso e integración de cualquier fuente de datos. Por otra parte se debe almacenar la información por largos períodos de tiempo, por tanto la visualización de los datos del negocio en múltiples dimensiones y hasta su nivel mínimo de detalle, debe ser lo menos costoso posible en esfuerzo y tiempo; además definir Reportes de uso constante, de modo que se pueda evitar el atraso y la lentitud en la consulta de los datos y en su elaboración. El hecho de contar con la información necesaria influye en el correcto análisis de los datos para apoyar el proceso de la toma de decisiones.

Dada la situación problemática antes planteada surge entonces como **problema científico**:

¿Cómo contribuir a la disponibilidad de la información para facilitar su análisis y apoyar la toma de decisiones en el sistema FE-AVL?

De acuerdo con las características de la investigación se planteó como **objeto de estudio** los Almacenes de Datos y se definió como **campo de acción** Mercado de Datos para el sistema FE-AVL.

¹ **Flota**: Conjunto de entidades que tiene asociado un dispositivo GPS

² **plugin**: Aplicación informática de hardware o software que añade una característica o un servicio específico a un sistema más grande, este es enchufado al sistema ya existente. Su presencia es muy habitual en los navegadores web, en reproductores de música y en sistemas de gestión de contenidos.

El **objetivo general** del presente trabajo de diploma es desarrollar un Mercado de Datos que contribuya al análisis de la información para la toma de decisiones en el sistema FE-AVL.

A raíz del objetivo general anteriormente planteado surge una **idea a defender**:

Si se desarrolla un Mercado de Datos para la gestión y control de las flotas del sistema FE-AVL, entonces se logrará una mayor integración y disponibilidad de la información en beneficio del proceso de toma de decisiones en la dirección de las empresas que deciden hacer uso de los sistemas de localización automática de vehículos.

Para dar cumplimiento al objetivo general se desarrollan las siguientes **tareas de la investigación**:

1. Fundamentación teórica de los Mercados de Datos para la toma de decisiones.
2. Selección de la metodología, herramientas y tecnologías a emplear en el desarrollo del Mercado de Datos.
3. Análisis, Diseño e Implementación del Mercado de Datos para facilitar el análisis de información y apoyar la toma de decisiones en el sistema FE-AVL.
4. Validación del Mercado de Datos elaborado a partir del diseño y aplicación de los Casos de Prueba.

Durante la etapa de investigación fueron empleados los siguientes **métodos de investigación científica**:

Métodos Teóricos

- **Analítico-Sintético:** Posibilita el análisis y procesamiento de la información recopilada durante la investigación sobre los Almacenes y Mercados de Datos, metodologías, tecnologías y herramientas de desarrollo.
- **Histórico-Lógico:** Permite el estudio de los antecedentes y tendencias actuales de los Almacenes de Datos, metodologías, tecnologías y herramientas de desarrollo.
- **Modelación:** Permite realizar la modelación del ciclo de vida del Mercado de Datos para el sistema FE-AVL.

Métodos Empíricos

➤ **Experimento:** Posibilita realizar pruebas al Mercado de Datos para verificar el correcto funcionamiento de las funcionalidades implementadas.

El presente trabajo de diploma consta de 3 capítulos, a continuación se ofrece una breve descripción de su contenido:

Capítulo 1. Fundamentación Teórica: Se abordan los temas relacionados con la tecnología de los Almacenes y Mercados de Datos respaldando el presente trabajo, abarcando definiciones, ventajas y desventajas de este tipo de soluciones, así como de las herramientas y metodología a emplear.

Capítulo 2. Análisis y Diseño del Mercado de Datos: Se realiza el análisis del negocio, se definen los requerimientos del sistema y las estructuras de almacenamiento de datos. Se definen además, los roles y permisos del sistema, se diseña el diagrama de Casos de Uso (CU) del sistema, el modelo de datos y se especifican las reglas del negocio.

Capítulo 3. Implementación y Validación del Mercado de Datos: Se implementan los subsistemas de integración y visualización además de realizar la validación del MD a través de los Casos de Prueba.

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

1.1. Introducción

El desarrollo del Mercado de Datos para el sistema FE-AVL conlleva la realización de una investigación, la cual está guiada fundamentalmente en dos direcciones: los Sistemas Inteligentes de Transporte (SIT) y los Sistemas de Soporte de Decisiones.

En el presente capítulo se exponen los principales detalles de la investigación consumada acerca de los Sistemas Inteligentes de Transporte y sus principales características, en especial, el sistema FE-AVL, los Almacenes y Mercados de Datos. Además se plasman las definiciones necesarias, se describen las herramientas, lenguajes y metodología a emplear para el desarrollo de la propuesta de Mercado de Datos para el sistema FE-AVL.

1.2. Conceptos Fundamentales

1.2.1. Sistemas Inteligentes de Transporte

Los Sistemas Inteligentes de Transporte (SIT), son un conjunto integrado de soluciones tecnológicas de las telecomunicaciones, la electrónica y la informática, diseñadas para mejorar la operación y seguridad del transporte ya sea por mar, por tierra o por aire. La implementación de los SIT se lleva a cabo mediante el empleo de las infraestructuras de redes ya existentes, estos permiten ahorros importantes en las inversiones destinadas a fortalecer sus capacidades y a evitar el colapso por falta de recursos en el transporte.

La evolución de estos sistemas es permanente y sus aplicaciones hoy en día son múltiples y variadas. Por ejemplo, los primeros pasos fueron en el transporte aéreo pero en la actualidad los mayores esfuerzos van orientados al transporte terrestre, precisamente por los problemas existentes en todo el mundo por la congestión de tráfico, sobre todo en las grandes ciudades debido en gran medida al crecimiento poblacional, la urbanización y el desarrollo.

Estos sistemas funcionan en tres etapas:

- La obtención de información referente a las unidades de transporte e infraestructuras.
- La integración y análisis de la información obtenida, generando mensajes de respuesta priorizados en función de su relevancia.

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

- La difusión de los mensajes generados mediante infraestructura y sistemas a bordo.(3)

La adquisición, procesamiento, integración y el suministro de la información suponen el alma de los SIT, permitiendo que las autoridades, operadores y viajeros estén mejor informados y mucho más coordinados. De esta manera agiliza y facilita el proceso de toma de decisiones.

Los SIT se clasifican en cinco áreas funcionales:

- **Sistemas Avanzados de Gestión del Tráfico** (por sus siglas en inglés ATMS, Advanced Traffic Management Systems): Se encargan de detectar situaciones del tráfico para luego transmitir las al punto de control a través de las redes de comunicaciones.

- **Sistemas Avanzados de Información para viajeros** (por sus siglas en inglés ATIS, Advanced Traveler Information Systems): Permiten que los usuarios tengan acceso a la información de las carreteras en tiempo real para elegir rutas de viaje y modos de transporte.

- **Operaciones de Vehículos Comerciales** (por sus siglas en inglés CVO, Commercial Vehicle Operations): El sistema incluye principalmente el control automático de vehículos, la gestión de la flota, equipos de programación y pago electrónico. (4)

- **Sistemas Avanzados de Control y Seguridad de Vehículos** (por sus siglas en inglés AVCSS, Advanced Vehicle Control Systems): Estos sistemas aplican tecnologías avanzadas en vehículos y carreteras, y ayudan a los conductores a controlar sus vehículos con el fin de reducir accidentes y mejorar la seguridad del tráfico.(4)

- **Sistemas Avanzados de Transporte Público** (por sus siglas en inglés APTS, Advanced Public Transportation Systems): Estos sistemas incluyen principalmente vigilancia automática de vehículos, VPS, equipos de programación y boletos electrónicos. (4)

La gran gama de posibilidades que ofrecen estos sistemas son más que evidentes, así como la importancia de su continua evolución, especialmente en un mundo donde casi todos los procesos giran en torno a las tecnologías de comunicación y la información.

En el área de las Operaciones de Vehículos Comerciales está enmarcado el sistema FE-AVL, desarrollado en el Centro de Telemática de la facultad dos de la Universidad de las Ciencias Informáticas.

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

1.2.2. Sistema FE-AVL

FE-AVL es un sistema automático de localización vehicular desarrollado para la gestión y control de flotas. Una de las metas más importantes de este sistema es fortalecer el proceso de la toma de decisiones ante determinadas situaciones que se pueden presentar con las flotas.

Este sistema posee una estructura flexible, basada en diferentes módulos, que puede crecer y actualizarse a medida que se necesite, para esto presenta una serie de mecanismos que permiten resolver los problemas y las necesidades de los clientes. A continuación se mencionan los principales módulos con que cuenta esta solución:

- **Módulo Administración:** Es el encargado de realizar las configuraciones necesarias y entre sus funcionalidades se encuentran: la gestión de dispositivos móviles, la gestión de grupos de dispositivos móviles, la gestión de usuarios, la gestión de módulos y la gestión de acceso.
- **Módulo Rastreo:** Permite al usuario conocer en cada momento la posición geográfica que ocupa cada uno de los vehículos pertenecientes a la flota.
- **Módulo Reportes:** Es el encargado de visualizar los datos históricos solicitados por el usuario en forma de tablas o Reportes gráficos de diferentes tipos. Cada reporte es un plugin y el usuario puede crear tantos Reportes como desee y de acuerdo a sus necesidades. Cada módulo se diseñó con una estructura compuesta por dos partes fundamentales, el núcleo y los plugins. (5)

1.2.3. Sistemas de Soporte de Decisiones

Los Sistemas Inteligentes de Transporte son cada vez más especializados. Una de las metas más importantes para los desarrolladores de estos complejos sistemas es fortalecer el proceso de toma de decisiones ante determinados sucesos en los que se ven involucradas las flotas. Hoy en día estas herramientas se conciben además como Sistemas de Soporte de Decisiones, debido a la necesidad de presentarle al usuario la información precisa con mayor rapidez y eficiencia.

Un Sistema de Soporte de Decisiones (DSS por sus siglas en inglés Decision Support System) puede ser definido de una manera general como un sistema de computación diseñado para apoyar al proceso de toma de decisiones (de planeamiento, de administración y operacionales) en un negocio. Un DSS es la ventana del usuario a la información almacenada en el almacén de datos. Este tipo de sistema contiene

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

todos los servicios o procesos para elegir, manipular y analizar datos informativos y presentar los resultados. (6)

1.2.4. Almacenes de Datos y Mercados de Datos

Por el valor que hoy posee la información relevante manejada por las empresas, por la necesidad de conservación de datos históricos por largos períodos de tiempo y contar con datos personalizados de acuerdo a las necesidades de los usuarios, es que surgen los Sistemas de Soporte de Decisiones, y dentro de ellos, los Almacenes y Mercado de Datos.

La aparición de los Almacenes de Datos o Data Warehouse contribuyó en gran medida a dar respuesta a grandes problemas que existían en el sistema empresarial, especialmente en la administración de las grandes corporaciones. Esto se debe al gran volumen de información generado por estas y lo engorroso que resultaba el análisis de todos los datos a la hora de tomar alguna decisión.

Las dos personalidades más reconocidas en el ámbito de los Almacenes de Datos son Bill Inmon³ y Ralph Kimball⁴, éste último es considerado el principal promotor del enfoque dimensional para el diseño de los Almacenes de Datos, aunque existen otros especialistas que emiten sus propios criterios en cuanto a algunos aspectos conceptuales.

En la actualidad se puede afirmar que es una tecnología madura y estable. Por esta y otras razones las empresas se aseguran de disponer de informaciones valiosas con las que se puedan adoptar estrategias empresariales precisas y en tiempo récord, que a su vez les permita situarse en la cima de la competencia. Es una manera real de agilizar la toma de decisiones a partir del análisis de datos complejos que no están a la mano, debido al gran volumen de información generado en estas empresas.

Se conocen dos grandes filosofías en cuanto a la relación existente entre los conceptos de Almacén de Datos y Mercado de Datos, precisamente defendidas por las dos principales personalidades mencionadas

³ **William H. Inmon** (nacido en 1945) Científico especializado en la rama de la informática, reconocido por muchos como el padre de los Almacenes de Datos.

⁴ **Ralph Kimball** (nacido en 1944) Especialista en el campo de los Almacenes de Datos y la Inteligencia de Negocios. Es reconocido como uno de los arquitectos originales del almacenamiento de datos.

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

anteriormente, Inmon y Kimball. Inmon propuso la idea de que los Mercados de Datos se sirven del Almacén de Datos para extraer información, la cual está almacenada en tercera forma normal, en un modelo relacional; mientras que Kimball sostiene que el Almacén de Datos es el resultado de la unión de los Mercados de Datos de la organización.

Almacén de Datos

Un Almacén de Datos (AD) es un almacén de información temática orientado a cubrir las necesidades de aplicaciones de los sistemas de Soporte de Decisiones y de la Información de Ejecutivos (EIS), que permite acceder a la información corporativa para la gestión, control y apoyo a la toma de decisiones. (7)

Entre las facilidades que ofrecen los Almacenes de Datos se encuentran:

- La posibilidad de mejorar la medición de las acciones y los resultados de estas.
- La optimización de los procesos empresariales.
- Establecer prioridades con anterioridad para la toma de decisiones.

Características:

Temático: Los datos son organizados por tema para facilitar el acceso a los mismos y un mejor entendimiento por parte de los usuarios.

Integrado: Los datos son integrados en una estructura consistente.

No volátil: La información almacenada es permanente.

Histórico: Los datos se almacenan por largos períodos de tiempo debido a que se emplean para realizar análisis de tendencias a partir de los distintos valores que puede tomar una variable en el tiempo para permitir realizar determinadas comparaciones.

Estas características en conjunto permiten garantizar la fortaleza de los Almacenes de Datos, ya que facilitan el acceso a los datos por parte de los usuarios; así como el funcionamiento de los Sistemas de Apoyo a las Decisiones, los cuales se nutren enteramente de la información almacenada en estos. A pesar de ser tan eficaces y tan poderosos es necesario señalar que estos son muy costosos y en

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

ocasiones devuelven información sub-óptima ante una petición de información por parte del usuario, es por ello que se debe definir antes de su construcción cual es la información realmente útil.

Estructura de un Almacén de Datos

Estructura básica de la arquitectura de un Almacén de Datos:

- *Datos operacionales.* Origen de datos para el componente de almacenamiento físico del Almacén de Datos.
- *Extracción de datos.* Selección sistemática de datos operacionales usados para formar parte del Almacén de Datos.
- *Transformación de datos.* Procesos para resumir y realizar cambios en los datos operacionales.
- *Carga de datos.* Inserción de datos en el Almacén de Datos.
- *Almacén.* Almacenamiento físico de datos de la arquitectura Almacén de Datos.
- *Herramienta de acceso.* Herramientas que proveen acceso a los datos.(8)

Estructura física de un Almacén de Datos:

La estructura física del AD puede presentar las configuraciones que a continuación se muestran:

- *Arquitectura centralizada.* Todo el almacén de datos se encuentra en un único servidor.
- *Arquitectura distribuida.* Los datos del almacén se reparten entre varios servidores. Asignando cada servidor a uno o varios temas lógicos.
- *Arquitectura distribuida por niveles.* Refleja la estructura lógica del almacén, asignando los servidores en función del nivel de agregación de los datos que contienen. Un servidor está dedicado para los datos de detalle, otro para los resumidos y otro para los muy resumidos.(8)

Ventajas de un Almacén de Datos

Entre las ventajas más notables de un Almacén de Datos se pueden mencionar:

- Integrar datos históricos sobre la actividad de la organización (o negocio) en un único repositorio.
- Analizar los datos del negocio desde la perspectiva de su evolución en el tiempo.
- Prever tendencias de evolución del negocio.
- Identificar nuevas oportunidades de negocio y tomar decisiones estratégicas.

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

- Reducir los costes materiales y humanos en la toma de decisiones.(9)

Mercado de Datos

Un MD proporciona un esquema de almacenamiento que permite realizar consultas complejas, por medio de la interacción e interrelación de múltiples entidades, manteniendo solo un repositorio con la información agregada y ordenada, según las necesidades de los usuarios. Además, provee de una interface de consulta que permite al usuario la posibilidad de análisis de la información para la toma de decisiones. (10) En general, los Mercados de Datos son Almacenes de Datos, pero de forma reducida, es decir, un departamento específico. Un conjunto de mercados conformarían el Almacén de Datos.

Características:

- Se centran en los requisitos de los usuarios asociados a un departamento en un área de negocios concreto.
- No contienen datos operacionales detallados.
- Son más sencillos a la hora de utilizarlos y comprender sus datos, debido a que la cantidad de información que contienen es mucho menor que en los Almacenes de Datos.
- Permiten un acceso a los datos mucho más rápido por parte de los usuarios, debido a la mejora en los tiempos de respuesta de las consultas, por la reducción en el volumen de información a ser accedida.

Estas características convierten al Mercado de Datos en una potente tecnología para el almacenamiento de datos. Por esta y otras razones es que se propone el desarrollo de un Mercado de Datos para el sistema FE-AVL para dar solución a la problemática planteada inicialmente, ya que este proyecto requiere el almacenamiento de datos históricos que contribuyan a fortalecer el proceso de toma de decisiones.

Ventajas de un Mercado de Datos

- Poseen menos volumen de información que los Almacenes de Datos, lo que posibilita mayor rapidez en las respuestas a las consultas y que su implementación sea más sencilla y a la vez más fácil.
- Reflejan rápidamente beneficios a las organizaciones que los utilizan ya que se gana en tiempo y costo a la hora de análisis y entrega de la información completa, correcta y consistente, oportuna y accesible.
- Brinda la facilidad de analizar y explotar toda la información que contiene.

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

➤ Facilita a la administración la obtención de datos concretos y reales que pueden ser decisivos en la toma de decisiones estratégicas y tácticas.(11)

1.2.5. Modelado de Datos

Para comprender el modelado de datos, especialmente en el mundo de los Almacenes de Datos, es necesario conocer primeramente las diferencias existentes entre los dos tipos de modelado: Entidad-Relación y dimensional.

El modelado Entidad/Relación se utiliza habitualmente para crear un único modelo complejo de todos los procesos de una organización. Este enfoque ha demostrado ser efectivo para crear sistemas eficientes de Procesamiento Transaccional en Línea (OLTP por sus siglas en inglés). Por otra parte, el modelado dimensional crea modelos individuales para reflejar procesos discretos de negocio. Este modelado organiza la información en estructuras que usualmente corresponden a la forma en que los analistas realizan sus consultas de los datos del Data Warehouse. (6)

Modelo Entidad/Relación

El modelo de datos de entidad-relación (ER) se basa en una percepción de un mundo real que consiste en un conjunto de objetos básicos llamados entidades y de relaciones entre estos objetos. Se desarrolló para facilitar el diseño de bases de datos permitiendo especificar un esquema empresarial. Este esquema representa la estructura lógica general de la base de datos. (12)

Entre sus características fundamentales se pueden mencionar:

- Maneja la redundancia fuera de los datos.
- Divide los datos en entidades y estas son representadas como tablas en una base de datos.
- Crecen fácilmente. De esta manera cada vez se hacen más complejos.
- Se pueden apreciar muchos caminos para ir de una tabla a otra.
- El diagrama se visualiza simétrico, sin distinguir a primera vista la importancia de unas respecto a otras. Por lo tanto no es fácil de entender tanto para usuarios como para los diseñadores de estos modelos.

Sin embargo, para el propósito de un Almacén de Datos, este modelo presenta los siguientes problemas:

- Legibilidad limitada. Los usuarios finales no son capaces de entender el modelo ER por lo que no pueden navegar por él en busca de información.

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

- Dificultan el acceso a un modelo ER general a través de las herramientas de consulta ya que estas no cuentan con grandes prestaciones cuando trabajan en entornos relacionales que manejan grandes volúmenes de información.
- Su empleo frustra el principal atractivo de los Almacenes de Datos que es la recuperación de información intuitiva y con alto rendimiento.

Modelo Dimensional

Modelo Dimensional es el nombre que se le da a una técnica muy utilizada especialmente en Almacenes de Datos para diseñar el modelo lógico. Existen diferentes apreciaciones en cuanto a su concepto, pero básicamente coinciden en que es un modelo para presentar la información de una manera estándar, sencilla y sobre todo intuitiva para los usuarios, que además permite acceder a la información con mayor rapidez por parte de los manejadores de las bases de datos. (13)

Esquemas Multidimensionales

Existen varios esquemas para el modelado de los datos en un AD siendo los más utilizados:

- Esquema de Estrella
- Esquema de Copo de nieve
- Esquema de Constelación de hechos

El “esquema en estrella” o “star schema”, comúnmente llamado así por los diseñadores de Almacenes de Datos debido a su semejanza con una estrella, está compuesto por una tabla principal llamada hechos y un conjunto de pequeñas tablas alrededor de esta, llamadas dimensiones. Estas últimas contienen información textual descriptiva y sus atributos se emplean como fuente de las restricciones en las consultas al Almacén de Datos. En este modelo las dimensiones no se normalizan, lo cual permite minimizar el número de uniones y a su vez, incrementar el rendimiento de las consultas.

El “esquema copo de nieve” o “snowflake schema” es una variante del esquema en estrella donde se normalizan las dimensiones reflejando las jerarquías en las mismas y conservando lo esencial del modelo en estrella: las tablas de hechos. La ventaja del modelo copo de nieve es eliminar la redundancia de datos y por lo tanto ocupar menos espacio en disco. (14)

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

El esquema llamado “constelación de hechos” es más complejo que las otras arquitecturas debido a que contiene múltiples tablas de hechos. (15)

Una de las grandes virtudes de este esquema es su gran flexibilidad. A pesar de esto, es necesario señalar que su arquitectura puede llegar a ser compleja, y difícil su mantenimiento debido al aumento de las tablas vinculadas entre sí.

Componentes del Modelo Dimensional

A continuación se mencionan los componentes de un modelo dimensional:

- **Hechos:** Son operaciones o actividades. De manera general, es algo que ocurre en el tiempo.
- **Medidas:** Son valores numéricos que describen el hecho que se está analizando. Ayudan a medir el desempeño del negocio. Existen tres clases de medidas: aditivas, semi-aditivas y no aditivas.
- **Dimensiones:** Son categorías que describen el contexto en el cual se analizan las medidas. Proveen un método general para organizar la información corporativa.
- **Jerarquías:** Definen cómo los datos son sumariados desde los niveles más bajos hacia los más altos.
- **Granularidad:** Es el nivel de detalle admitido en una tabla de hechos.

Tabla de Hechos: Es la tabla central de un esquema dimensional la cual contiene los valores de las medidas de negocio. (16)

Medidas aditivas, semi-aditivas y no aditivas

Medida aditiva: Se define como una medida que al aplicarle la función de agregación SUM sobre las dimensiones siempre se obtienen valores semánticamente correctos. (17) Pueden ser combinadas a lo largo de cualquier dimensión.

Medidas semi-aditivas: Es aditiva sobre unas dimensiones y no lo es sobre otras. (17)

Medidas no aditivas: Se define como una medida que al aplicarle la función de agregación SUM sobre las dimensiones los valores obtenidos son incorrectos. No pueden ser combinadas a lo largo de ninguna dimensión.

Dimensiones lentamente cambiantes

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

Las dimensiones lentamente cambiantes son aquellas dimensiones que se mantienen “casi” constantes en el tiempo y que pueden preservar la estructura dimensional independiente del tiempo, con sólo agregados menores relativos para capturar la naturaleza cambiante del tiempo. (18)

Ralph Kimball planteó tres estrategias a seguir cuando se trata de dimensiones lentamente cambiantes:

Tipo 1: Sobrescribir el viejo valor en el registro dimensional y por lo tanto perder la capacidad de seguir la historia.

Este tipo es el más básico y sencillo de implementar, ya que si bien no guarda los cambios históricos, tampoco requiere ningún modelado especial y no necesita que se añadan nuevos registros a la tabla. (18)

Tipo 2: Crear un registro dimensional adicional (con una nueva llave) que permita registrar el cambio presentado por el valor del atributo. De esta forma permanecerían en la base tanto el antiguo como el nuevo valor del registro con lo cual es posible segmentar la historia de la ocurrencia.

Tipo 3: Crear un campo “actual” nuevo en el registro dimensional original el cual almacene el valor del nuevo atributo, manteniendo el atributo original también. Cada vez que haya un nuevo cambio en el atributo, se modifica el campo “actual” solamente. No se mantiene un registro histórico de los cambios intermedios. (18)

Ventajas del Modelo Dimensional

Este modelo presenta importantes ventajas de las que carece el modelo relacional. Entre las que se encuentran:

- El marco predecible del esquema estrella resiste a los cambios inesperados en el comportamiento del usuario.
- El diseño lógico puede realizarse independientemente de los patrones de consultas esperados, siendo consideradas de la misma forma tanto las interfaces de usuario como las estrategias de consulta, así como el lenguaje de consulta generado contra el modelo dimensional.
- Flexibilidad. Los nuevos elementos de datos y las nuevas decisiones de diseño son fácilmente adaptables. Todas las tablas pueden modificarse simplemente agregando nuevos registros de datos o se pueden incluir nuevas dimensiones al modelo sin necesidad de volver a cargar los datos posteriormente.

1.2.5.1. Justificación del modelo de almacenamiento de datos a utilizar

Para la realización del Mercado de Datos para el sistema FE-AVL se definió el empleo del modelo dimensional para realizar la modelación de los datos. Esto se debe principalmente a la mejora que ofrece

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

en cuanto a rendimiento y velocidad con respecto a otros tipos de modelación de datos. De los esquemas mencionados del modelo dimensional se seleccionó el Esquema de Estrella debido a su simplicidad y velocidad para ser usado en análisis multidimensionales como Procesamiento Analítico en Línea, OLAP, por sus siglas en inglés. Además, este esquema permite la implementación de la funcionalidad de una base de datos dimensional empleando una base de datos relacional.

1.2.6. Extracción, Transformación y Carga (ETL)

Son las siglas en inglés de *Extraer*, *Transformar* y *Cargar* (Extract, Transform and Load). Es el proceso que permite mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, Mercado o Almacén de Datos para analizar, o en otro sistema operacional para apoyar un proceso de negocio.(19)

Extraer: Es el primer paso del proceso ETL y consiste en extraer los datos de los sistemas de origen.

Transformar: Esta fase aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados. A continuación se mencionarán algunas de las transformaciones que se le pueden aplicar a los datos:

- Seleccionar solo ciertas columnas para su carga (por ejemplo, que las columnas con valores nulos no se carguen).
- Traducir códigos (por ejemplo, si la fuente almacena una “H” para Hombre y “M” para Mujer pero el destino tiene que guardar “1” para Hombre y “2” para Mujer).
- Obtener nuevos valores calculados (por ejemplo, búsquedas, combinaciones, etc.).
- Generación de campos clave en el destino. (19)

Carga: Esta fase es el momento en el que los datos de la fase de transformación son cargados en el sistema destino. Al realizar esta operación se aplicarán todas las restricciones y triggers⁵ que se hayan

⁵ **trigger:** Es un bloque PL/SQL (Procedural Language/Structured Query Language) asociado a una tabla, que se ejecuta como consecuencia de una determinada instrucción SQL (una operación DML: INSERT, UPDATE o DELETE) sobre dicha tabla.

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

definido en la misma y si están bien definidos contribuyen a garantizar la calidad de los datos en el proceso ETL.

1.2.7. Sistemas OLAP y OLTP

1.2.7.1. Procesamiento Analítico en Línea (OLAP)

OLAP es el acrónimo en inglés de *Procesamiento Analítico en Línea (On-Line Analytical Processing)*. Es una solución utilizada en el campo de la *Inteligencia de Negocios (Business Intelligence)*, la cual consiste en consultas a estructuras multidimensionales (o *Cubos OLAP*) que contienen datos resumidos de grandes Bases de Datos o Sistemas Transaccionales (*OLTP*). Se usa en informes de negocios de ventas, marketing, informes de dirección, minería de datos y áreas similares. La principal característica que potencia a OLAP, es que es lo más rápido a la hora de hacer Selecciones (SELECTS), en contraposición con OLTP que es la mejor opción para Insertar (INSERTS), Actualizar (UPDATES) y Eliminar datos (DELETES). (20)

A continuación se relacionan algunas clasificaciones en cuanto al tipo de motor que se emplea para el almacenamiento de los datos entre las soluciones OLAP.

➤ **ROLAP (Relational OLAP):** Este tipo de implementación OLAP almacena los datos en un motor *relacional*. La arquitectura está compuesta por un servidor de base de datos relacional y un servidor donde se encuentra el motor OLAP. La principal ventaja de esta arquitectura es que permite el análisis de una gran cantidad de datos.

➤ **MOLAP (Multidimensional OLAP):** Almacena los datos en una base de datos *multidimensional*. En este caso el resumen de la información es calculado por adelantado con el objetivo de optimizar los tiempos de respuesta. Estos valores pre calculados son la base de las ganancias de desempeño de este sistema.

➤ **HOLAP (Hybrid OLAP):** Esta implementación OLAP almacena algunos datos en un motor relacional y otros en una base de datos multidimensional.

➤ **DOLAP (Desktop OLAP):** Es un OLAP orientado a equipos de escritorio. Este trae toda la información que necesita analizar desde la base de datos relacional y luego la guarda en el escritorio. Por tanto, todas las consultas y análisis son hechas contra los datos almacenados en el escritorio. (20)

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

Cubos OLAP

Un cubo OLAP es una base de datos multidimensional, la cual guarda los datos físicos a través de un vector multidimensional. El pionero en este tipo de bases de datos y precursor del diseño del cubo OLAP fue Edgar Frank Codd⁶. En general su idea era guardar los datos en vectores con el objetivo de realizar un análisis mucho más rápido, estos vectores serían los cubos. De esta manera se resolvía el problema de las bases de datos relacionales, las cuales no podían analizar grandes cantidades de datos instantáneamente. Una de las principales características de la potencia de estos cubos es la rapidez en que ejecutan las sentencias SQL de tipo SELECT.

1.2.7.2. Procesamiento Transaccional en Línea (OLTP)

OLTP (por sus siglas en inglés On Line Transaction Processing / Procesamiento Transaccional en Línea) o también conocido como procesamiento operacional es un tipo de proceso especialmente rápido en el que las solicitudes de los usuarios son resueltas de inmediato; naturalmente, ello implica la concurrencia de un mecanismo que permite el procesamiento de varias transacciones a la vez. Este sustenta las operaciones diarias de la empresa y describe los requerimientos operacionales del sistema. Estos sistemas son óptimos en el manejo de datos que constantemente están cambiando y usualmente tienen un gran número de usuarios que están ejecutando transacciones simultáneamente y que actualizan ó modifican los datos en tiempo real.

Soporta las actividades de investigación y navegación del usuario terminal, además de contar con una síntesis de la información de la organización a través de vistas personalizadas, análisis históricos y pronósticos. (21)

1.2.7.3. Diferencias entre un sistema OLTP y OLAP

En el caso de las aplicaciones OLTP es importante resaltar que se caracterizan por la creación de muchos usuarios, actualizaciones o recuperación de registros individuales. Las bases de datos OLTP se perfeccionan para la actualización de transacciones. En cambio las aplicaciones OLAP son usadas por analistas y gerentes que frecuentemente quieren tener una vista de datos de nivel superior.

A continuación se muestra una tabla con un resumen de las características más importantes que diferencian a las aplicaciones OLAP y OLTP. (22)

⁶ **Edgar Frank Codd**: Científico informático inglés (23-08-1923/18-04-2003), conocido por sus aportes a la teoría de bases de datos relacionales.

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

Característica	OLTP	OLAP
Tamaño de la BD	GB	GB a TB
Origen de los Datos	Interno	Interno y Externo
Actualización	On-Line	Batch
Períodos	Actual	Histórico
Consultas	Predecibles	Ad-Hoc
Actividad	Operacional	Analítica

Tabla 1. Diferencia entre las aplicaciones OLAP y OLTP.

1.2.7.4. Justificación de la técnica de almacenamiento de datos a utilizar

Posteriormente al análisis de algunos aspectos referentes a los sistemas OLTP y OLAP, se concluye que el sistema empleado para el desarrollo del MD para el sistema FE-AVL será el OLAP, debido a las facilidades que nos brinda, entre las que se encuentran: la gran capacidad de almacenamiento de datos históricos y la posibilidad de consultas a estructuras multidimensionales. De las clasificaciones de los sistemas OLAP en cuanto al motor de almacenamiento de datos se seleccionó el ROLAP, ya que permite el acceso a los datos almacenados en el MD para proporcionar el análisis OLAP, esto posibilita que los usuarios finales puedan ejecutar sus análisis multidimensionales y soporta técnicas de optimización de accesos para acelerar las consultas, ejemplo de ello la desnormalización y Joins múltiples.

1.3. Tendencias Actuales

1.3.1. Almacenes de Datos y Mercados de Datos

En la actualidad las diversas empresas se han dado a la tarea de actualizar sus métodos para el almacenamiento de la información y de esta manera obtener una posición ventajosa con respecto a la competencia.

La tecnología de Almacenes de Datos es muy popular hoy en día entre las empresas ya que permite convertir los datos en información útil. Algunos de los empleos a nivel mundial donde se ha empleado esta tecnología son:

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

- En la Consejería de Vivienda del Gobierno de la Rioja, España. Donde además se implementó un Mercado de Datos para la Consejería de Asuntos Sociales.
- En el Instituto Nacional de Estadística de España.
- En el Instituto Nacional de Estadística de Chile.

Cuba no está exenta de los avances tecnológicos a nivel mundial, por lo que se ha estado trabajando en asimilar estas tecnologías para su empleo en diferentes empresas que manejan gran volumen de datos, donde a veces se dificulta el análisis de información, como por ejemplo: la Oficina Nacional de Estadísticas e Información (ONEI), la red nacional de Genética Médica, entre otros.

La Empresa de Telecomunicaciones de Cuba S.A (ETECSA) es una de las entidades más interesadas en este tema y que ha obtenido buenos resultados en los últimos años con el empleo de Almacenes de Datos en el Departamento Comercial y Mercadotecnia, para gestionar los macro-procesos de Facturación y Cobros, Ventas, Mercadotecnia-Posventa y Asistencia al usuario.

La Empresa de Proyectos de Arquitectura e Ingeniería (EMPAI) de Matanzas posee un Mercado de Datos para el control de la información relevante del proceso de negocio de Gestión del Capital Humano para elevar la efectividad de la organización. (23)

1.3.2. Sistemas de Localización automática de vehículos (AVL)

Son varias las aplicaciones informáticas que proporcionan servicios de localización automática de vehículos en diferentes países del mundo, cada una de ellas con características muy particulares y diferentes a las demás aplicaciones. Algunos de estos ejemplos son:

- **IP AVL:** Es un servicio de origen Argentino, basado en el acceso a un portal de internet situado en sus servidores, en el que el usuario podrá visualizar por cada uno de sus vehículos la siguiente información: reportes de posición, velocidad, activar o desactivar salidas, entre otras. (24) El sistema contempla entre sus funcionalidades ofrecer información histórica durante cierto período. El almacenamiento de esta información no se realiza a través de un MD, sino en una Base de Datos Operacional, con tablas definidas para este fin con el objetivo de no sobrecargar las tablas que poseen la información actual; al cumplirse el período de tiempo establecido para el almacenamiento, esta información es borrada.

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

➤ **MPS AVL:** Es un sistema escalable de origen Español, destinado a ofrecer en uno o más puestos de operador el acceso, comunicación y gestión del conjunto de recursos móviles o fijos empleados (vehículos, helicópteros, barcos, personal, puntos fijos de monitorización, etc.). Funciona sobre la base de una estructura cliente-servidor, en la que se contemplan las bases de datos del sistema, de recursos, de servicios y de información corporativa, así como la cartografía y las comunicaciones móviles en diferentes estándares y protocolos. Entre las funcionalidades disponibles encontrará: Localización Automática de Vehículos (AVL) y Gestión de rutas (TMS). (25) Sin embargo no contempla entre sus funcionalidades la de ofrecer información histórica para el apoyo a la toma de decisiones.

➤ **Software AVL Fenix10:** Es una nueva plataforma Argentina de rastreo satelital AVL (localización vehicular automatizada) con autogestión. (26) Entre sus principales funcionalidades se encuentran los recorridos históricos y el control de combustible por software.

Algunas de sus características más notables son:

- Soporte Multiplataforma para todos los navegadores.
 - Gestión de usuarios, equipos, flota y niveles de permisos para usuarios.
 - Localización en mapa y en texto con refresco automático de última tecnología (no se refresca la pantalla completa del navegador, solo las unidades y eventos acorde a su desplazamiento en el mapa.)
 - Control de Combustible por Software (se especifica consumo cada 100km y tipo de combustible)
- Recorridos históricos con filtro: fecha, hora inicio, hora fin, eventos y velocidad.
- Exportación de información a Excel. Control de Km total recorrido. (26)

En Cuba, específicamente en el centro de Telemática de la UCI se realizan importantes esfuerzos en cuanto al desarrollo de aplicaciones informáticas para la localización automática de vehículos, ejemplo de esto es el sistema FE-AVL cuyas principales características fueron abordadas en epígrafes anteriores. Dicho sistema es una solución para el monitoreo y control de flotas, posee la ventaja de contar con varios módulos que le permiten adaptarse a diferentes entorno. En el caso del módulo de Reporte permite la generación de Reportes dinámicos.

A partir de una encuesta realizada por el equipo de desarrollo se determinaron las necesidades de información las cuales deben ser los posibles requisitos, de acuerdo a esto y a lo descrito anteriormente los ejemplos internacionales no tratan la información que se desea manejar, el coste de la adquisición del

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

mismo sería muy alto, y la información que gestionan algunas de estas soluciones no es la que se desea guardar en el MD. La solución que más se asemeja a nuestro objetivo es la del módulo de Reportes del sistema FE AVL; sin embargo, en la propuesta se desea hacer uso de la Inteligencia de Negocios y las técnicas OLAP para definir Vistas de Análisis (en lo adelante VA) que permitan el desglose de la información hasta un nivel mínimo de detalle. Este propósito se alcanzará con la utilización de las herramientas que posee la Suite de Pentaho, la cual es usada por proporcionar un amplio espectro de capacidades de Inteligencia de Negocio. Por otra parte, el sistema fuente, en este caso la Base de Datos Operacional del sistema FE-AVL ha sido actualizada, por lo que en el diseño e implementación del mercado se tendrán en cuenta dichas actualizaciones.

1.4. Lenguajes utilizados

1.4.1. Lenguaje MDX

MDX es un acrónimo de Multidimensional Query eXpression. Es un lenguaje de consulta más cercano al lenguaje natural que el SQL sin embargo tiene funciones y fórmulas que lo hacen muy potente para el análisis de datos. (27)

Es el lenguaje estándar para las bases de datos multidimensionales conocidas habitualmente como OLAP. Fue diseñado con el objetivo de realizar consultas sobre los *cubos OLAP* en general y en *Analysis Services* (traducido en Servicios de Análisis) en particular. Este devuelve un conjunto de celdas como resultado de tomar un subconjunto de las celdas de un cubo original.

Este lenguaje es usado en la herramienta Pentaho Schema Workbench con el objetivo de crear expresiones para manipular los datos, ordenarlos, filtrarlos o agruparlos, realizar cálculos, consultar datos multidimensionales o para crear expresiones MDX para su uso en un cubo.

1.4.2. Lenguaje SQL

SQL (Structured Query Language) es un lenguaje de programación diseñado para almacenar, manipular y recuperar datos almacenados en bases de datos relacionales. (28)

SQL permite especificar diferentes tipos de operaciones sobre estas bases de datos relacionales, por ejemplo: insertar datos en una tabla.

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

Dicho lenguaje es empleado mediante la herramienta PostgreSQL, la cual es utilizada para la creación y manipulación de datos almacenados en la base de datos.

1.4.3. Lenguaje de Modelado UML

El lenguaje para modelamiento unificado (UML), es un lenguaje para la especificación, visualización, construcción y documentación de los artefactos de un proceso de sistema intensivo. (29)

Incluye algunos aspectos conceptuales como son los procesos de negocio y funciones del sistema, además de aspectos concretos como es el caso de las expresiones de lenguajes de programación, los esquemas de bases de datos y componentes reutilizables. Este lenguaje se emplea a través de la herramienta Visual Paradigm para realizar la modelación del negocio.

1.5. Herramientas utilizadas

1.5.1. Herramienta CASE para el modelado: Visual Paradigm para UML

Visual Paradigm para UML es una herramienta UML profesional que soporta el ciclo de vida completo del desarrollo de software. Entre sus cualidades más notables se pueden mencionar: que permite una mayor calidad en poco tiempo y a un menor costo. (30) Además integra el diseño visual, la generación de código fuente, bases de datos, documentación y realización de ingeniería directa o inversa, etc.

El Visual Paradigm fue la herramienta seleccionada para realizar la modelación del negocio del MD para el sistema FE-AVL ya que la metodología seleccionada para el desarrollo del mismo la propone. Esta herramienta permite modelar lógicamente la estructura del MD, representar la jerarquía con los niveles asociados, generar el modelo físico a partir del modelo lógico, generar un XML con el diseño preliminar de los esquemas OLAP a partir del diseño del Mercado de Datos y generar finalmente el script de BD para cargar el diseño en el gestor de BD.

1.5.2. Colección de Pentaho para la Inteligencia de Negocio (Pentaho Open BI Suite)

La solución de código abierto de Pentaho para la Inteligencia de Negocio pretende ser una alternativa a las soluciones propietarias tradicionales más completas, como es el caso de: Business Objects, Microstrategy, Microsoft, etc. Por esta razón incluye todos los componentes con que cuentan estas soluciones BI propietarias y que a continuación se mencionan: (31)

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

- Reportes (Reporting)
- Análisis
- Dashboards
- Workflow
- Minería de Datos (Data Mining)
- Extracción, Transformación y Carga (ETL)
- Single Sign-On. Ldap
- Auditoría de uso y rendimiento
- Planificador
- Notificador
- Seguridad. Perfiles

Esta potente solución posee las siguientes características:

- Posee Licencia **GNU GPL**⁷. Esta licencia está orientada a proteger la libre distribución, modificación y uso de software.
- Aplicación web desarrollada en J2EE.
- Es Modular.
- Cuenta con un Servidor Web, el cual puede correr en servidores compatibles con J2EE como JBOSS AS, WebSphere, Tomcat, WebLogic y Oracle AS.
- La Base de datos puede ser manipulada desde cualquiera de estos gestores de bases de datos: vía JDBC, DB2, Servidor SQL de Microsoft, MySQL, Oracle, PostgreSQL, Firebird, etc.
- Es Multiplataforma.
- Posee un repositorio de datos basado en XML.

⁷ **GPL** del inglés General Public License/Licencia Pública General de GNU, ampliamente usada en el mundo del software, es una licencia libre y gratuita que garantiza a los usuarios la libertad de usar, modificar y copiar el software.

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

- Todos sus componentes están expuestos vía Servicios Web (Web Services).
- Incluye: ETL, Reportes, Procesamiento Analítico en Línea, Minería de Datos, Tablero de Instrumentos, etc.

Las herramientas de la colección de Pentaho para el desarrollo del Mercado de Datos para el sistema FE-AVL son:

- **Pentaho BI Server CE(versión 3.8.0):** Es el servidor de Pentaho, el cual viene configurado por defecto con Apache Tomcat, incluye la consola de Usuario (PUC) y la consola administrativa (*Pentaho Administration Console*).
- **Pentaho Schema Workbench (versión 3.3.0):** Se emplea para describir y publicar esquemas y cubos.
- **Pentaho Metadata Editor (versión 4.1.0):** Se emplea para describir y publicar modelos de negocio.
- **Pentaho Report Designer (versión 3.8.3):** Es un editor basado en eclipse con prestaciones profesionales y de calidad, que cuenta con una capacidad de personalización de informes de acorde a las necesidades de negocio destinado a desarrolladores. Está estructurado de forma que los desarrolladores puedan acceder a todas sus prestaciones con rapidez. Además, incluye un editor de consultas que facilita la confección de los datos que serán utilizados en un informe.
- **Pentaho Data Integration (versión 4.2.0):** Es una parte de la suite Pentaho Open Source Business Intelligence, la cual incluye software para todas las áreas de soporte a la hora de tomar decisiones de negocio.

Es la herramienta con la cual se diseñan y despliegan transformaciones y tareas ETL, constituye un pilar importante en el éxito de Pentaho.

Este es uno de los componentes más importantes de esta plataforma de inteligencia de negocios y parece ser la más estable y confiable. Es una herramienta muy intuitiva y fácil de utilizar. Su uso permite evitar grandes cargas de trabajo manual frecuentemente difícil de mantener y de desplegar.

A continuación se mencionarán las principales funcionalidades de Pentaho Data Integration:

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

- Control de la extracción de los datos y su automatización.
- Acceso a diferentes tecnologías.
- Proporcionar la gestión integrada de Almacenes de Datos y Mercados de Datos, integrando la extracción, transformación y carga para la construcción de los Almacenes de Datos corporativos y de los Mercados de Datos.
 - Uso de la arquitectura de metadatos, facilitando la definición de los objetos de negocio y las reglas de consolidación.

Para el desarrollo del Mercado de Datos del sistema FE-AVL fue seleccionada la versión 4.2.0 de Pentaho Data Integration, la cual incluye las siguientes ventajas:

- Facilita la comunicación entre los interesados y los desarrolladores.
- Reduce riesgos y costos de implementación.
- Permite probar de forma empírica y temprana la arquitectura de la aplicación de inteligencia de negocios. (32)

1.5.3. Power Architect

Power Architect es una herramienta que permite realizar perfiles de datos en bases de datos de origen, ingeniería inversa de bases de datos existentes y además genera los metadatos de ETL de forma automática. (33)

Algunas de las características de Power Architect que lo convierten en una herramienta poderosa son:

- Conexión con múltiples bases de datos al mismo tiempo.
- Compara modelos de bases de datos e identifica discrepancias.
- Dibuja y borra tablas y columnas en su visualizador.
- Genera Código SQL a partir del esquema en el visualizador.
- Genera Informes visuales.
- Realiza Ingeniería Inversa (recuperar modelos de una base de datos existente).
- Guarda el código y la estructura en el proyecto para poder trabajar remotamente. (34)

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

1.5.4. PostgreSQL

PostgreSQL es un sistema de gestión de bases de datos objeto-relacional. (35) A pesar de ser de código abierto como muchos otros proyectos, su desarrollo es dirigido por una comunidad de desarrolladores denominada el PGDG (Grupo de desarrollo global de PostgreSQL). Es un sistema objeto-relacional debido a que incluye características propias de la orientación a objetos, como por ejemplo la herencia, tipos de datos, funciones, restricciones, disparadores, reglas e integridad transaccional. Aunque posee estas características es necesario aclarar que no es un SGBD (Sistema Gestor de Base de Datos) puramente orientado a objetos.

Actualmente es uno de los Sistemas Gestores de Bases de Datos (SGBD) más populares y más poderosos, esto se debe en gran medida a las ventajas que le ofrece a los desarrolladores de bases de datos, como son:

- Puede ser empleado en los principales sistemas operativos que existen: Linux, Unix, Mac OS, Windows.
- Soporte para los lenguajes más populares: PHP, C, C++, Perl, Python, entre otros.
- Soporte de protocolo de comunicación encriptado por SSL (Protocolo de Capa de Conexión Segura).
- Extensiones para alta disponibilidad, nuevos tipos de índices, datos espaciales y minería de datos.
- Aprovecha la potencia de sistemas multiprocesadores, gracias a su implementación multihilo.

Y para muchos quizás la más importante, que puede ser utilizado, modificado y distribuido por cualquier persona gratuitamente, para cualquier propósito sea comercial, académica o privada.

Esta herramienta fue seleccionada como Sistema Gestor de Base de Datos, con el objetivo de crear, modificar y acceder a la base de datos en su versión PostgreSQL 9.1.

PgAdmin

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

Aplicación gráfica para gestionar el gestor de bases de datos PostgreSQL siendo la más completa y popular con licencia de código abierto. (36) Esta aplicación es usada para desarrollar la Base de Datos del mercado y realizar todo tipo de consultas sobre la misma.

1.6. Metodologías de Desarrollo de Almacenes de Datos

Actualmente existen diferentes metodologías para el desarrollo de Almacenes de Datos y Mercados de Datos, ejemplo de ellas son; Hefesto, Inmon, Kimball, entre otras; las cuales tienen como función principal guiar el proceso de desarrollo de un Almacén de Datos.

Kimball

La metodología de Kimball es una de las más usadas en el mundo debido a su efectividad para desarrollar soluciones de construcción de un almacén de datos. Esta apunta a una solución completa que puede implementarse en poco tiempo, dar resultados rápidos para demostrar el valor de la solución al negocio y no perder nunca la confianza de los involucrados.

Inmon

La metodología Inmon por el contrario puede tener una implementación mucho más tardada, y es recomendada cuando se hace demasiado difícil representar el modelo a través de dimensiones y la complejidad de la solución se hace demasiado grande. Ambas metodologías pueden implementarse en un mismo Almacén de Datos; si es necesario utilizar esta metodología, se recomienda hacerlo en iteraciones ya avanzadas y siempre empezar con Kimball. (37)

Hefesto

La metodología Hefesto entre sus principales directrices plantea que la construcción e implementación de un Almacén de Datos puede adaptarse muy bien a cualquier ciclo de vida de desarrollo de software. Lo que se debe tener muy en cuenta, es no utilizar metodologías que requieran fases extensas de reunión de requerimientos y análisis, fases de desarrollo monolítico que conlleve demasiado tiempo y fases de despliegue muy largas. Lo que se busca, es entregar una primera implementación que satisfaga una parte de las necesidades, para demostrar las ventajas del Almacén de Datos y motivar a los usuarios. (38)

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

No solo son usadas las metodologías reconocidas a nivel internacional sino que estas pueden ser adaptadas en las entidades según las necesidades de las mismas en el desarrollo de sus productos, ejemplo de esto es la que a continuación se expone.

1.6.1. Metodología de Desarrollo de Almacenes de Datos.

Esta metodología fue elaborada por integrantes del Centro de Tecnologías de Gestión de Datos (DATEC) de la facultad 6 en la Universidad de Ciencias Informáticas y tiene sus cimientos sobre la base de la Metodología de Kimball fundamentalmente para definir aspectos específicos del desarrollo de Almacenes de Datos.

La Metodología de Desarrollo de Almacenes de Datos toma como basa la metodología de Kimball para definir los aspectos específicos del desarrollo de Almacenes de Datos. Para lograr la alineación con el modelo de desarrollo LPS (Línea de Producción de Software) utiliza las áreas prácticas definidas por el SEI⁸ y algunas consideraciones de la propuesta de modelo de desarrollo de LPS elaborado para los centros de producción. Se hereda de CMMI⁹ algunos de sus enfoques, forma de documentar los procesos y actividades asociadas a estos procesos como parte de su contenido. Esta metodología no incluye todas las actividades asociadas a las áreas prácticas de las LPS, debido a que algunas están más relacionadas con gestión de la organización y no del proyecto.

La elección de un ciclo de vida adecuado para cada desarrollo está relacionada con las características del producto a obtener y el entorno en el cual debe desarrollarse. Para definir las fases del ciclo de vida de la metodología se tuvieron en cuenta las fases propuestas por la Metodología de Kimball y los procesos que plantea CMMI llevado a cabo en la UCI. Las fases propuestas por Kimball están directamente relacionadas con los desarrollos de Almacenes de Datos y se basan en los hitos más importantes de dichos desarrollos. Por su parte CMMI define fases genéricas para todos los desarrollos de software, pero que no permiten destacar características propias de las soluciones de Almacenes de Datos. (39)

Fases del ciclo de vida de esta metodología

⁸ Software Engineering Institute (SEI)/ Instituto de Ingeniería de Software que ayuda a los principios de Ingeniería de Software avanzadas y prácticas.

⁹ Modelo de Madurez de Capacidad Integrado (**CMMI** del inglés Capability Maturity Model Integration), creado para evaluar las capacidades de las organizaciones de ingeniería de sistemas, ingeniería de software, además del desarrollo integrado del producto y del proceso

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

- **Estudio preliminar y planeación:** En esta fase se realiza un estudio detallado de la entidad cliente y se llevan a cabo las tareas de planeación del proyecto.
- **Requisitos:** En esta fase se realizan las entrevistas al cliente para determinar los requisitos de información.
- **Arquitectura:** En esta fase se definen las vistas arquitectónicas de la solución, como son: subsistemas y componentes a utilizar.
- **Diseño e Implementación:** En esta fase se define el diseño de las estructuras de almacenamiento de datos, los procesos de integración de datos, así como el diseño gráfico de la aplicación definido por el cliente. Posteriormente se implementan cada uno de los subsistemas.
- **Prueba:** En esta fase se realizan las pruebas que validan la calidad del producto final.
- **Despliegue:** Esta fase consta de dos etapas, primero el despliegue piloto y luego la carga de los datos históricos. Se realiza además, la capacitación y transferencia tecnológica de la solución a los clientes.
- **Soporte y Mantenimiento:** Esta fase inicia cuando la solución está implantada y en explotación. Se brindan variados servicios de soporte como es el caso del soporte en línea, vía telefónica, web, correo u otros y el acompañamiento al cliente.
- **Gestión del proyecto:** Esta fase se ejecuta a lo largo de todo el ciclo de vida del proyecto. Es donde se controla, gestiona y chequea todo el desarrollo, los gastos, las utilidades, los recursos, las adquisiciones, los planes y cronogramas entre otras actividades relacionadas con la gestión de proyectos.

Como puede evidenciarse el ciclo de vida de la metodología está organizado por fases, algunas de ellas podrán ser implementadas de forma paralela como es el caso de la fase de Requisitos y Arquitectura, además durante la fase de Diseño e Implementación podrán desarrollarse varios componentes al mismo tiempo. (39)

Esta metodología fue seleccionada para llevar a cabo el desarrollo del Mercado de Datos para el sistema FE-AVL debido en gran medida a que mantiene las principales estrategias a seguir elaboradas por Kimball y además, por haber sido concebida teniendo en cuenta las características propias de la producción de software en la UCI. Además provee una guía para la organización del proceso de desarrollo de un AD, así como las principales actividades a realizar para garantizar la correcta realización

Capítulo 1. Fundamentación Teórica

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

del proyecto. También facilita la adaptación y formación de los especialistas y estudiantes que por primera vez se enfrenten al desarrollo de un proyecto de desarrollo de AD.

1.7. Conclusiones.

En el presente capítulo se abordaron diferentes conceptos asociados al tema en cuestión, como son: los Sistemas Inteligentes de Transporte en especial el Sistema Automático de Localización Vehicular (FE-AVL), los Sistemas de Soporte de Decisiones y los Almacenes y Mercados de Datos. Teniendo en cuenta la investigación realizada se decide desarrollar el MD para el sistema FE-AVL, para lo cual se selecciona UML como lenguaje de modelado, SQL y MDX como lenguajes para la confección de consultas y como SGBD a PostgreSQL en su versión 9.1. Se seleccionó además a Pentaho Open Source BI como la herramienta a emplear para el trabajo con el MD, el tipo de motor que se emplea para el almacenamiento de los datos entre las soluciones OLAP fue ROLAP, ya que acumula los datos en un motor relacional y permite el análisis de gran cantidad de información. Para realizar la modelación de la base de datos fue empleado el Power Architect y Visual Paradigm para la modelación de todos los procesos involucrados en el desarrollo de la solución. Para guiar el proceso de construcción del MD se seleccionó la Metodología de Desarrollo de Almacenes de Datos creada en el centro DATEC. El correcto empleo de las herramientas seleccionadas ofrece la posibilidad de desarrollar un MD para el sistema FE-AVL que cumpla con los requisitos propuestos inicialmente.

CAPÍTULO 2. ANÁLISIS Y DISEÑO DEL MERCADO DE DATOS

2.1. Introducción

En el presente capítulo se muestra el análisis del negocio y el diseño del MD con las características necesarias para satisfacer las necesidades planteadas por el cliente. Además de exponer los elementos fundamentales que ayudan al desarrollo de la propuesta; se describen los requisitos informativos, funcionales y no funcionales, así como los Casos de Uso del sistema y las reglas del negocio.

2.2. Estudio Preliminar del Negocio

El sistema FE-AVL tiene como objetivo el control y monitorización de flotas, ya sea en los sectores públicos, de emergencias, entre otros, pues cuenta con gran variedad de servicios gracias a los estándares digitales que hoy se implementan, dicho sistema gestiona la información que se tiene en el receptor GPS (Sistema de Posicionamiento Global / Global Positioning System) de un dispositivo y la transmite a un centro de control donde se encuentra instalado un SIG -Sistema de Información Geográfica / Geographic Information System - para monitorizar los recursos. En otras palabras, controla la ubicación geográfica de las unidades a través de información GPS que es enviada a una base de datos, la cual registra no solo las rutas a transitar de una unidad, su posición en tiempo real, o la reja en la que debe estar transitando, sino que controla las alarmas disparadas por las unidades ya sea en una reja ¹⁰o en una ruta¹¹. Las alarmas no son más que alertas que se envían en tiempo real cuando ocurre un evento, y pueden ser de diferentes tipos, establecidas previamente por la persona que utilice el sistema FE-AVL. De dicho sistema se desea guardar históricamente las alarmas que han sido disparadas en relación con las unidades, las rejadas y las rutas; además de la localización de una determinada unidad en un tiempo dado y los puntos GPS que componen una reja o una ruta. El hecho de contar con un histórico de los datos que se gestionan en el sistema FE-AVL permite el análisis de información, información como: la fecha donde se dispararon más alarmas, que unidad disparó más alarmas, en que ruta o reja se dispararon mayor cantidad de alarmas, en qué lugar se encontraba una unidad en un momento determinado. Sin embargo al ser el MD propuesto un mercado genérico, y que aún no cuenta con un usuario para el cual se esté

¹⁰ **reja:** Término utilizado para identificar conjunto de puntos geográficos que determinan un área específica en metros cuadrados.

¹¹ **ruta:** Término utilizado para identificar un camino definido para el paso de los vehículos de determinada Flota. Está conformada por un conjunto de puntos geográficos y poseen un tamaño determinado en km.

Capítulo 2. Análisis y Diseño

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

desarrollando en un negocio específico, las posibles decisiones quedan a merced de un futuro cliente, ya que estas influirán en el funcionamiento de la organización, generando repercusiones positivas o negativas según su elección, de ahí su importancia en determinar la mejor para su propósito.

Teniendo en cuenta los procesos del negocio y la información que se desea almacenar se propone agrupar en un área de análisis lo relacionado a las alarmas, en otra lo relacionado a los puntos GPS y en la última lo relacionado a la localización geográfica, quedando de la siguiente forma:

Áreas de análisis:

- Análisis de los indicadores relacionados con las Alarmas: A través de esta Área de Análisis, se puede determinar la cantidad de alarmas activadas por determinados indicadores como son la hora, la ruta, la reja, puntos GPS, día, mes, flota y tiempo.

- Análisis de los indicadores relacionados con los puntos GPS de las rutas y las rejas: A través de esta Área de Análisis se puede determinar la cantidad de puntos GPS que conforman una reja o una ruta, además de dado un punto GPS, determinar a qué reja o ruta pertenece.

- Análisis de los indicadores relacionados con la localización por puntos GPS de las flotas: A través de esta Área de Análisis se puede determinar la flota, el grupo, las unidades que pertenecen a un grupo, los puntos GPS, el mes, el día, la hora, el año, el semestre y la semana por determinados indicadores como son, unidad, puntos GPS, mes, año, semana y semestre.

2.3. Requerimientos del Mercado de Datos

En la presente investigación se desea contribuir al almacenamiento de la información histórica para el apoyo a la toma de decisiones de los indicadores referentes a las alarmas, los puntos GPS y la localización geográfica de las unidades, de ahí la importancia de identificar las necesidades de información como base para un correcto diseño del Mercado de Datos. Además para ello se identifican los requerimientos de la solución y el análisis de los requisitos definidos, que permiten identificar las estructuras bases del modelo lógico dimensional. (39)

2.3.1 Requisitos de Información

Los Requisitos de Información describen los datos que el sistema debe proveer o debe acceder. Estos se definen a partir de las necesidades de información identificadas en el negocio, que permiten el análisis del comportamiento de los indicadores a medir según los objetivos y metas de la organización. Se definen dos aristas:

- ✓ La información que debe presentar el sistema a través de su capa de presentación de datos para ser accedida por parte de los usuarios del negocio.
- ✓ La información a la que debe acceder el sistema para satisfacer las necesidades de información, o sea, define la información que deben ofrecer las fuentes de datos operativas.

Los requisitos de información se describen con sus variables de entrada y salida. Las variables de entrada representan las aristas de análisis por las que se desea filtrar un indicador y las variables de salida representan los indicadores del negocio que se desean analizar. (39)

RI1. Obtener cantidad de alarmas por tiempo y alarma.

RI2. Obtener cantidad de alarmas por tiempo, hora y alarma.

RI3. Obtener cantidad de alarmas por tiempo, ruta y alarma.

RI4. Obtener cantidad de alarmas por tiempo, reja y alarma.

RI5. Obtener cantidad de alarmas por tiempo, flota y alarma.

RI6. Obtener cantidad de alarmas por tiempo, hora, ruta y alarma.

RI7. Obtener cantidad de alarmas por tiempo, hora, reja y alarma.

RI8. Obtener cantidad de alarmas por tiempo, hora, puntos GPS y alarma.

RI9. Obtener cantidad de alarmas por tiempo, hora, flota y alarma.

RI10. Obtener cantidad de alarmas por tiempo, ruta, puntos GPS y alarma.

RI11. Obtener cantidad de alarmas por tiempo, hora, ruta, puntos GPS y alarma.

RI12. Obtener cantidad de alarmas por tiempo, reja, puntos GPS y alarma.

RI13. Obtener cantidad de alarmas por tiempo, hora, reja, puntos GPS y alarma.

RI14. Obtener cantidad de alarmas por tiempo, flota, puntos GPS y alarma.

RI15. Obtener cantidad de alarmas por tiempo, hora, flota, puntos GPS y alarma.

RI16. Obtener cantidad de alarmas por mes y alarma.

Capítulo 2. Análisis y Diseño

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

- RI17. Obtener cantidad de alarmas por mes, hora y alarma.
- RI18. Obtener cantidad de alarmas por mes, ruta y alarma.
- RI19. Obtener cantidad de alarmas por mes, reja y alarma.
- RI20. Obtener cantidad de alarmas por mes, flota y alarma.
- RI21. Obtener cantidad de alarmas por mes, hora, ruta y alarma.
- RI22. Obtener cantidad de alarmas por mes, hora, reja y alarma.
- RI23. Obtener cantidad de alarmas por mes, hora, puntos GPS y alarma.
- RI24. Obtener cantidad de alarmas por mes, hora, flota y alarma.
- RI25. Obtener cantidad de alarmas por mes, ruta, puntos GPS y alarma.
- RI26. Obtener cantidad de alarmas por mes, hora, ruta, puntos GPS y alarma.
- RI27. Obtener cantidad de alarmas por mes, reja, puntos GPS y alarma.
- RI28. Obtener cantidad de alarmas por mes, hora, reja, puntos GPS y alarma.
- RI29. Obtener cantidad de alarmas por mes, flota, puntos GPS y alarma.
- RI30. Obtener cantidad de alarmas por mes, hora, flota, puntos GPS y alarma.
- RI31. Obtener cantidad de alarmas por día y alarma.
- RI32. Obtener cantidad de alarmas por día, hora y alarma.
- RI33. Obtener cantidad de alarmas por día, ruta y alarma.
- RI34. Obtener cantidad de alarmas por día, reja y alarma.
- RI35. Obtener cantidad de alarmas por día, flota y alarma.
- RI36. Obtener cantidad de alarmas por día, hora, ruta y alarma.
- RI37. Obtener cantidad de alarmas por día, hora, reja y alarma.
- RI38. Obtener cantidad de alarmas por día, hora, puntos GPS y alarma.
- RI39. Obtener cantidad de alarmas por día, hora, flota y alarma.
- RI40. Obtener cantidad de alarmas por día, ruta, puntos GPS y alarma.
- RI41. Obtener cantidad de alarmas por día, hora, ruta, puntos GPS y alarma.
- RI42. Obtener cantidad de alarmas por día, reja, puntos GPS y alarma.
- RI43. Obtener cantidad de alarmas por día, hora, reja, puntos GPS y alarma.
- RI44. Obtener cantidad de alarmas por día, flota, puntos GPS y alarma.
- RI45. Obtener cantidad de alarmas por día, hora, flota, puntos GPS y alarma.
- RI46. Obtener cantidad de puntos GPS por ruta.

Capítulo 2. Análisis y Diseño

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

RI47. Obtener cantidad de puntos GPS por reja.

RI48. Obtener cantidad de puntos GPS por ruta y puntos GPS.

RI49. Obtener cantidad de puntos GPS por reja y puntos GPS.

RI50. Obtener ruta a la que pertenece por puntos GPS.

RI51. Obtener reja a la que pertenece por puntos GPS.

RI52. Obtener las flotas, los grupos que tiene, las unidades de los grupos, puntos GPS, el mes, el día y la hora por semestre y año.

RI53. Obtener las flotas, los grupos que tiene, las unidades de los grupos, puntos GPS, el mes, el día y la hora por trimestre y año.

RI54. Obtener las flotas, los grupos que tiene, las unidades de los grupos, puntos GPS, el día y la hora por semana, mes y año.

RI55. Obtener las flotas, los grupos que tiene, las unidades de los grupos, puntos GPS y la hora por día, mes y año.

RI56. Obtener las flotas, los grupos que tiene, las unidades de los grupos, puntos GPS, el día y la hora por mes y año.

RI57. Obtener las flotas, los grupos que tiene, las unidades de los grupos, la hora, el día, el mes y el año por puntos GPS.

RI58. Obtener la flota, el grupo al que pertenece, puntos GPS, la hora, el día, el mes y el año por unidad.

RI59. Obtener la flota, el grupo al que pertenece, puntos GPS, la hora, día, semana, mes y año por unidad.

RI60. Obtener la flota, el grupo al que pertenece, puntos GPS, la hora, el día, el mes, trimestre y año por unidad.

RI61. Obtener la flota, el grupo al que pertenece, puntos GPS, la hora, el día, el mes, semestre y año por unidad.

2.3.2 Requisitos Funcionales

Los Requisitos Funcionales (RF) son capacidades o condiciones que debe cumplir el sistema, o sea, no son más que acciones que debe ser capaz de realizar el sistema para satisfacer las necesidades del cliente expuestas anteriormente. Además especifican comportamientos de Entrada/Salida. Estos requisitos son identificados durante la etapa de análisis, e incluyen las funcionalidades que se deben



Capítulo 2. Análisis y Diseño

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

implementar en los tres subsistemas que se desarrollan en soluciones de Almacenes y Mercados de Datos. A continuación se muestran los RF del Mercado de Datos FE-AVL.

- RF1.** Autenticar usuario.
- RF2.** Adicionar usuario.
- RF3.** Eliminar usuario.
- RF4.** Crear Vistas de Análisis.
- RF5.** Eliminar Vistas de Análisis.
- RF6.** Crear Reportes.
- RF7.** Eliminar Reportes.
- RF8.** Extraer datos.
- RF9.** Transformar y cargar datos.
- RF10.** Adicionar rol.
- RF11.** Eliminar rol.

2.3.3 Requisitos No Funcionales

Los Requisitos No Funcionales (RNF) son aquellas cualidades o propiedades que debe presentar el sistema, o sea, son aquellos requerimientos que presentan restricciones físicas sobre los funcionales. Este tipo de requerimientos cubren necesidades del sistema como son rendimiento, seguridad, usabilidad, soporte, entre otras. Después de un exhaustivo análisis fueron definidos 14 requisitos no funcionales para el Mercado de Datos FE-AVL. A continuación se muestran algunos de ellos.

➤ **Usabilidad**

RNF 1. Cumplir con las pautas de diseño de las interfaces.

El sistema debe tener una interfaz gráfica amigable y uniforme que incluya diferentes opciones. Las pautas de diseño se realizarán siguiendo de la estructura de información la cual será descrita posteriormente en el epígrafe 3.4 del capítulo 3.

RNF 2. Mostrar todos los textos de la aplicación (mensajes, títulos y demás) en idioma español.

Los títulos de los componentes de la interfaz, los mensajes para interactuar con los usuarios y los mensajes de error, deben ser en idioma español, tener una apariencia uniforme en todo el sistema y ser lo suficientemente informativos para el mejor entendimiento del usuario.

Capítulo 2. Análisis y Diseño

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

RNF 3. Agilizar el acceso a los Reportes del Mercado de Datos mediante la distribución de la información por áreas de análisis. El usuario podrá acceder de manera rápida a la información que solicita en el área correspondiente de acuerdo al objetivo de su solicitud.

RNF 4. Diseñar un reporte del Mercado de Datos de manera sencilla y ágil. Un usuario con conocimientos básicos de sistema podrá diseñar un reporte del Mercado de Datos de manera ágil, sencilla y dinámica.

➤ Seguridad

RNF 5. Garantizar la persistencia de la información.

Se debe realizar un respaldo total de los datos del Mercado de Datos con una frecuencia trimestral, asegurando que la información esté disponible para el usuario en todo momento.

RNF 6. Garantizar que solo tengan acceso a la información aquel personal que esté autorizado.

➤ Para acceder al MD se debe poseer permiso para conectarse al servidor donde se encuentra implementado.

➤ Todo aquel que tenga acceso al servidor del MD podrá interactuar con él en dependencia del rol que posea.

➤ Todos los usuarios deben contar con un rol específico de acuerdo a su especialidad en el MD y con ello los permisos pertinentes.

➤ Restricciones de diseño

RNF 7. Utilizar los lenguajes de programación definidos durante la investigación.

Como lenguaje dentro del Sistema Gestor de Base de Datos (SGBD) para la programación en el almacén de datos se utilizará SQL. En la implementación de los procesos de integración de datos se utilizará el lenguaje Java Script. También se hará uso del lenguaje MDX para realizar las consultas.

RNF 8. Utilizar el SGBD definido durante la investigación.

El SGBD que se utilizará es PostgreSQL y como interfaz de administración de dicho gestor PgAdmin.

RNF 9. Utilizar la herramienta de integración de datos definida durante la investigación.

Para el proceso de integración de datos se usará la herramienta Pentaho Data Integration.



Capítulo 2. Análisis y Diseño

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

RNF 10. 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: Herramienta gráfica que se utiliza para construir el esquema multidimensional que soportará la creación de los Reportes multidimensionales.

- Pentaho BI Server: Servidor que se encarga de visualizar los Reportes, tableros de control digital, controlar el acceso a la información y unificar en una solución de inteligencia de negocios el uso de las demás herramientas que componen la suite.

- Pentaho Administration Console: Herramienta para administrar el Pentaho BI Server, que permite la administración de las conexiones a las bases de datos, tareas programadas, así como los roles y usuarios.

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

➤ Interfaz

RNF 11. Acceso al sistema.

El usuario deberá acceder a la aplicación a mediante el protocolo HTTP, usando preferiblemente el navegador web Firefox en su versión 2.0 o posterior.

Interfaces de hardware

RNF 12. 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 o una Vista de Análisis en formato físico. El acceso a la impresora será mediante el protocolo TCP/IP a través de la interfaz que ofrece el hardware.

RNF 13. 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.
- 1 Microprocesador Core2Duo.

Interfaces de software

RNF 14. 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 al menos con:

- Firefox 2.0 o superior.
- Java Virtual Machine 6.0 y Schema Workbench 3.2.1 en caso de que un usuario capacitado requiera la construcción de esquemas multidimensionales para el diseño de nuevos Reportes.

2.3.1. Roles y Permisos

Los roles y los permisos en los sistemas de información juegan un papel primordial, puesto que estos contribuyen a la protección contra acciones que puedan afectar la integridad de los datos. Por tal motivo, el Mercado de Datos posee como parte de su seguridad la autenticación de los usuarios, a los cuales se les asigna un rol con un grupo de permisos específicos según el tipo de usuario para interactuar con la aplicación. A continuación se muestran los diferentes roles y permisos que posee cada uno de estos roles con que cuenta el Mercado de Datos del sistema FE-AVL.

Programador ETL: Posee un acceso total a todas las Áreas de Análisis (AA) del Mercado de Datos.

Administrador: Posee un acceso total a todas las Áreas de Análisis. Es capaz de cambiar las Vistas de Análisis y hacer nuevas, además de poder crear y visualizar los Reportes, pues administra todo el sistema.

Usuario: Posee un acceso de sólo lectura a todas las AA del Mercado de Datos.

El MD del sistema FE-AVL además de las pautas de seguridad definidas anteriormente cuenta con la seguridad que nos proporciona la Plataforma Pentaho BI, la cual dispone de un mecanismo de seguridad compuesto por cuatro áreas las cuales se mencionan a continuación:

- Seguridad de acceso a datos de objetos: Este mecanismo incluye usuarios y contraseñas, autorizaciones permitidas, recursos web y protección a datos.
- Autenticación: Este tiene que ver con el procesamiento de la información interactiva al iniciar sesión donde ésta es comparada con los datos del usuario que se autentica en ese momento.

Capítulo 2. Análisis y Diseño

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

➤ Autorización de recursos web (URL): Esta brinda protección a las URL para responder a cada usuario si puede acceder o no a una determinada página. Esto lo decide el administrador de recursos web, quien brinda a cada usuario autenticado un permiso de seguridad específico, delimitando las páginas a las que tiene acceso y a las que no.

2.4. Reglas del Negocio

Las Reglas del Negocio (en lo adelante RN), no son más que un conjunto de normas, operaciones, definiciones y políticas que deben cumplirse o restricciones y condiciones que deben satisfacerse; todas estas regulan algún proceso de negocio. Estas normas o restricciones son identificadas y definidas en el levantamiento de información y en el análisis de las fuentes por los especialistas del sistema FE-AVL, de acuerdo a las necesidades de información. A continuación se exponen las RN identificadas para el Mercado de Datos FE-AVL.

RN1: Cuando en la BD hay campos nulos, estos deben ser cargados como nulos, no como cero, de modo que pueda quedar el MD estandarizado y se pueda saber cuándo una unidad no posea grupo o flota.

RN2: Cuando la unidad no posee ni grupo ni flota hay que definirle un código, para poder ser identificada.

RN3: Para borrar el contenido de la BD, debe existir un intervalo mínimo de tiempo de 3 días después de haberse realizado la carga del almacén.

RN4: Cuando se desea adicionar una unidad a un grupo o flota específica, se debe registrar la fecha en que se realizó, o sea, insertarlo como una nueva tupla en la dimensión definida para la flota.

2.5. Casos de Uso del Sistema

Los CU del Sistema (CUS) son definidos durante la fase de análisis y diseño y están enfocados al punto de vista del usuario. Estos tienen como objetivo describir la interacción existente entre un CU y un usuario. Un CU del sistema, agrupa un conjunto de requisitos que se definen teniendo en cuenta las necesidades del cliente.

A continuación se expone el diagrama de CUS definido para el Mercado de Datos del sistema FE-AVL.

Capítulo 2. Análisis y Diseño

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

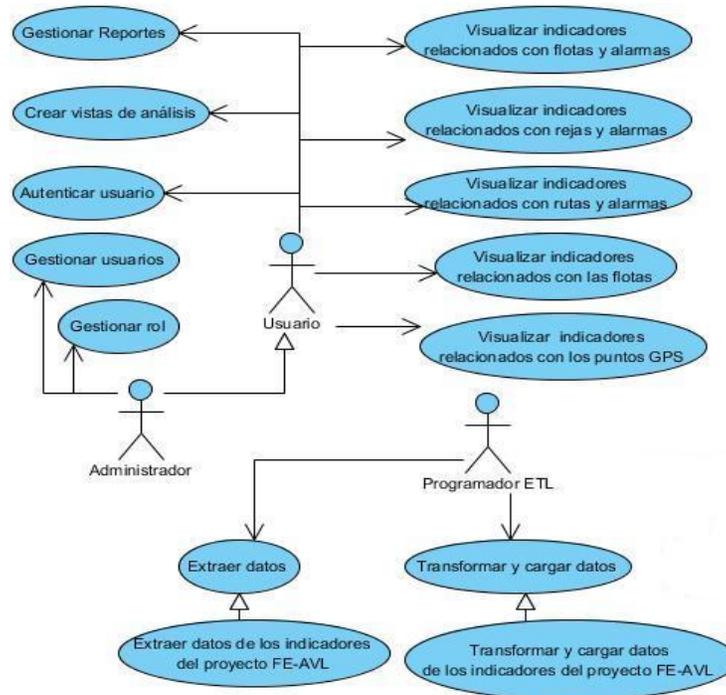


Ilustración 1. Diagrama de Casos de Uso del Sistema.

2.5.1. Descripción de Casos de Uso del Sistema

CU. Visualizar los indicadores relacionados con Flotas y Alarmas: Comprende toda la información relacionada con las flotas y las alarmas.

CU. Visualizar los indicadores relacionados con Rejas y Alarmas: Comprende toda la información relacionada con las Rejas y las Alarmas.

CU. Visualizar los indicadores relacionados con Rutas y Alarmas: Comprende toda la información relacionada con las Rutas y las Alarmas.

CU. Visualizar los indicadores relacionados con los puntos GPS: Comprende toda la información relacionada con los puntos GPS.

CU. Visualizar los indicadores relacionados con las Flotas: Comprende toda la información relacionada con las Flotas.

CU. Autenticar usuario: Este CU consiste en todo un proceso por el cual deben pasar los usuarios para interactuar con el sistema, el proceso de autenticación.

Capítulo 2. Análisis y Diseño

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

CU. Gestionar usuario: Este CU consiste en un proceso mediante el cual el administrador del sistema puede realizar diferentes operaciones sobre los usuarios como son, eliminar, insertar o modificar usuario.

CU. Gestionar Rol: Este CU consiste en un proceso mediante el cual el administrador del sistema puede realizar diferentes operaciones con los usuarios como son, eliminar, insertar o modificar rol.

CU. Gestionar Vista de Análisis: Este CU consiste en todo un proceso mediante el cual se pueden crear y eliminar vistas de análisis del sistema.

CU. Gestionar reporte: Este CU consiste en todo un proceso mediante el cual se puede crear, eliminar o visualizar un reporte en el sistema.

CU. Extraer datos: Consiste en todo un proceso mediante el cual se extraen la información de la fuente de datos hacia el MD.

CU. Realizar transformaciones y carga de datos: Este CU consiste en todo un proceso mediante el cual se puede realizar la transformación y carga de los datos correspondientes a las fuentes de información ya extraídas.

Descripción del Caso de Uso: Visualizar los indicadores relacionados con Flotas y Alarmas.

Objetivo	Presentar información determinada relacionada con los indicadores específicos de las flotas y las alarmas.	
Actores	Usuario Administrador	
Resumen	El CU inicia cuando el actor desea hacer un análisis de la información relacionada con los indicadores específicos de las flotas y las alarmas. El actor selecciona la Vista de Análisis deseada, el sistema muestra la información de la misma y puede detallar la información según lo desee. El CU finaliza cuando el actor termina el análisis de la información.	
Complejidad	Alta	
Prioridad	Media	
Precondiciones	El usuario se autenticó correctamente. El Mercado de Datos ha sido poblado. Las Vistas de Análisis han sido creadas.	
Postcondiciones	La Vista de Análisis fue consultada.	
Flujo de eventos		
Flujo básico Visualizar los indicadores relacionados con Flotas y Alarmas.		
	Actor	Sistema
1.	Se autentica	
2.		Muestra la interfaz de trabajo del BI.
3.	Selecciona en el AA "Alarmas".	

Capítulo 2. Análisis y Diseño

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

4.		Muestra los Libros de Trabajo correspondientes
5.	Selecciona el Libro de Trabajo "Indicadores relacionados con Flotas y Alarmas".	
6.		Muestra las Vistas de Análisis de este libro.
7.	Selecciona la Vista de Análisis que desee consultar.	
8.		Muestra la Vista de Análisis seleccionada
9.	Visualiza la Vista de Análisis	
Flujos alternos		
2a. Los datos de autenticación son incorrectos		
	Actor	Sistema
1.		Muestra un mensaje de error. Vuelve al paso 1 del flujo básico.
Opciones de Visualizar los indicadores relacionados con Flotas y Alarmas.		
Perspectivas de análisis		Posibles resultados
		Medidas
		Periodicidad
Variables de entrada relacionadas con el CU Presentar información relacionada con los indicadores específicos de flotas y alarmas: <ul style="list-style-type: none"> • dim_alarma • dim_flota • dim_tiempo • dim_hora • dim_puntosgps 		Variables de salida disponibles en el hecho hech_alarma_flota: <ul style="list-style-type: none"> • cantidad_alarmas_flota Rango de tiempo en que se solicitan las variables de salida: Semanal
Flujos alternos		
Relaciones	CU Incluidos	No aplica
	CU Extendidos	No aplica
Requisitos funcionales	no	Sección: "3.2 Requisitos no funcionales" del documento.

Tabla 2. Descripción del CU: Visualizar los indicadores relacionados con Flotas y Alarmas.

El resto de las descripciones de los Casos de Uso del Sistema se encuentran en los Anexos del 1 al 10.

2.6. Arquitectura del Mercado de Datos

La arquitectura del Mercado de Datos FE-AVL cuenta con diferentes subsistemas los cuales se mencionan a continuación:



Capítulo 2. Análisis y Diseño

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

- Fuente Origen: En este componente se encuentran los datos operacionales que serán extraídos y transformados para poblar el Mercado de Datos FE-AVL. En este caso se encuentra la Base de Datos Operacional del proyecto FE-AVL. Al ser este subsistema de donde se obtiene la información para poblar el MD, el mismo se relaciona con el subsistema de Integración, donde se realiza todo el proceso de ETL.
- Proceso de ETL: En este proceso se realizan todas las actividades relacionadas con la Extracción, Transformación y Carga de los datos obtenidos a partir de la fuente origen, además del almacenamiento de los mismos en el Mercado de Datos FE-AVL. Una vez terminado todo este proceso de ETL del MD, se da paso a la ejecución del siguiente subsistema, subsistema de visualización, refiriéndose a las herramientas utilizadas por el usuario para acceder a la información almacenada en el MD.
- Herramienta de Acceso: Esta herramienta es la encargada de ofrecer el acceso a los datos almacenados en el MD FE-AVL, así como a los metadatos publicados para la realización de los reportes Ad-hoc. Este proceso es llevado a cabo a través del servidor de Inteligencia de Negocio Pentaho BI Server.

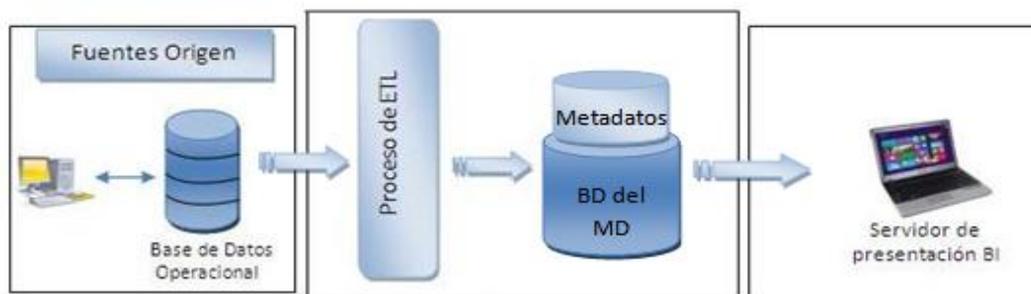


Ilustración 2. Arquitectura del MD FE-AVL

En la ilustración siguiente se muestra el diseño de la Base de Datos Operacional del proyecto FE-AVL, dicha base de datos es genérica, o sea, que actualmente por ser un sistema desarrollado a partir de una línea de investigación, contiene información general sobre lo necesario para la gestión de flotas, la cual fue usada como fuente origen para poblar el MD. De esta se seleccionaron los datos de importancia histórica para su posterior análisis. La Base de Datos Operacional cuenta con un conjunto de tablas, pero de ellas se escogen solo algunos de sus atributos, por ejemplo: de la tabla *unit* los atributos *name* e *idunit*, en correspondencia con esta tabla también se guarda el atributo *description* de la tabla *unitype*, el atributo *name* de la tabla *flotegrup* y el atributo *name* de la tabla *flote*, de la tabla *alarm* el atributo *alarmdescription*, de la tabla *pointgps* los atributos *latitude* y *longitude*, y de la tabla *gaterute* el atributo *name*. Además se

almacenan los identificadores de las tablas mencionadas anteriormente y de la tabla *gpsinfo* se tiene en cuenta la fecha.

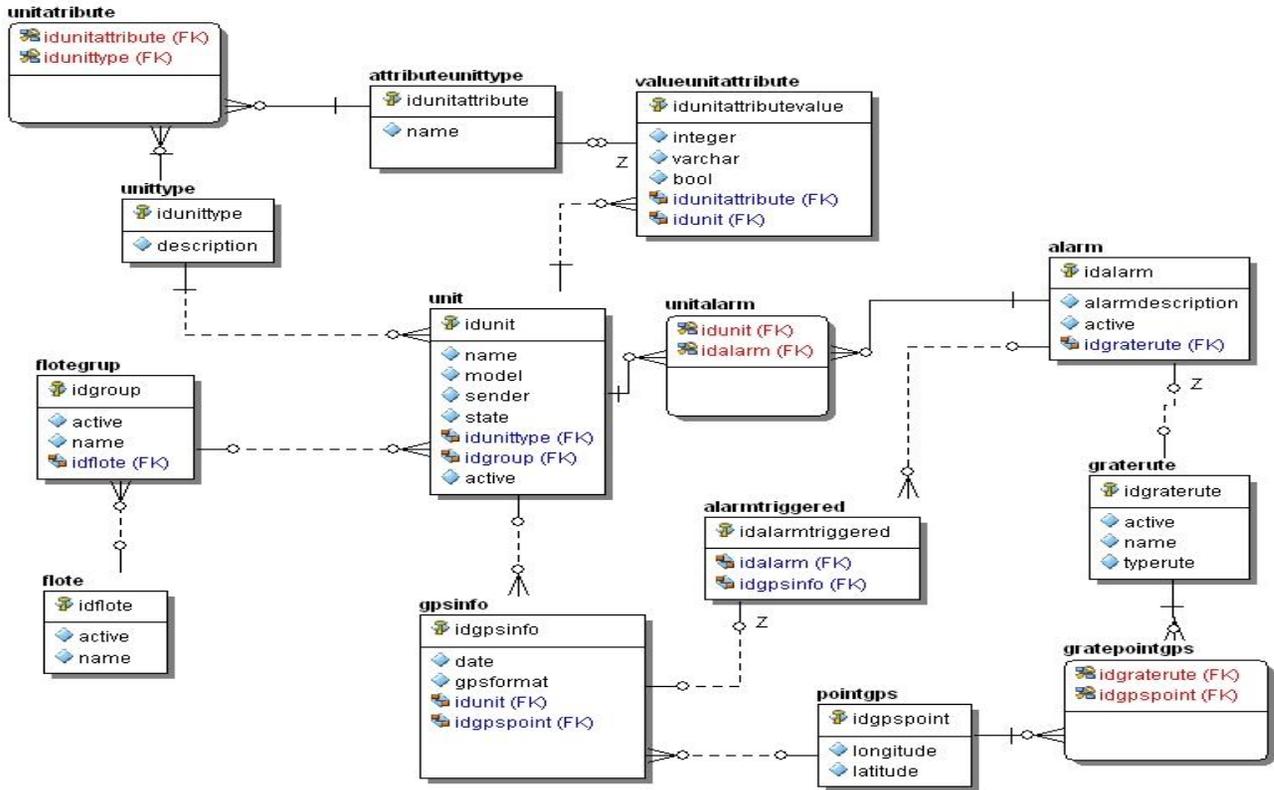


Ilustración 3. Modelo de la Base de Datos Operacional del proyecto FE-AVL

2.7. Modelo de Datos Dimensional

El modelo dimensional busca presentar la información de una manera estándar, sencilla y sobre todo intuitiva para los usuarios, además de que permite accesos a la información mucho más rápida por parte de los manejadores de bases de datos. (40)

El modelo dimensional propuesto está compuesto por varias tablas de hechos y tablas de dimensiones, las cuales se relaciona entre sí, dicho modelo fue desglosado en secciones para su mejor entendimiento. A continuación se muestran estas secciones del MD del sistema FE-AVL.

Capítulo 2. Análisis y Diseño

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

La siguiente ilustración describe la sección relacionada con la tabla de hechos “*alarma-flota*”, en la misma se puede observar la relación existente entre las tablas de dimensiones tiempo, flota, alarma, hora y punto GPS, refiriéndose a la posición geográfica, con la tabla de hechos mencionada anteriormente.

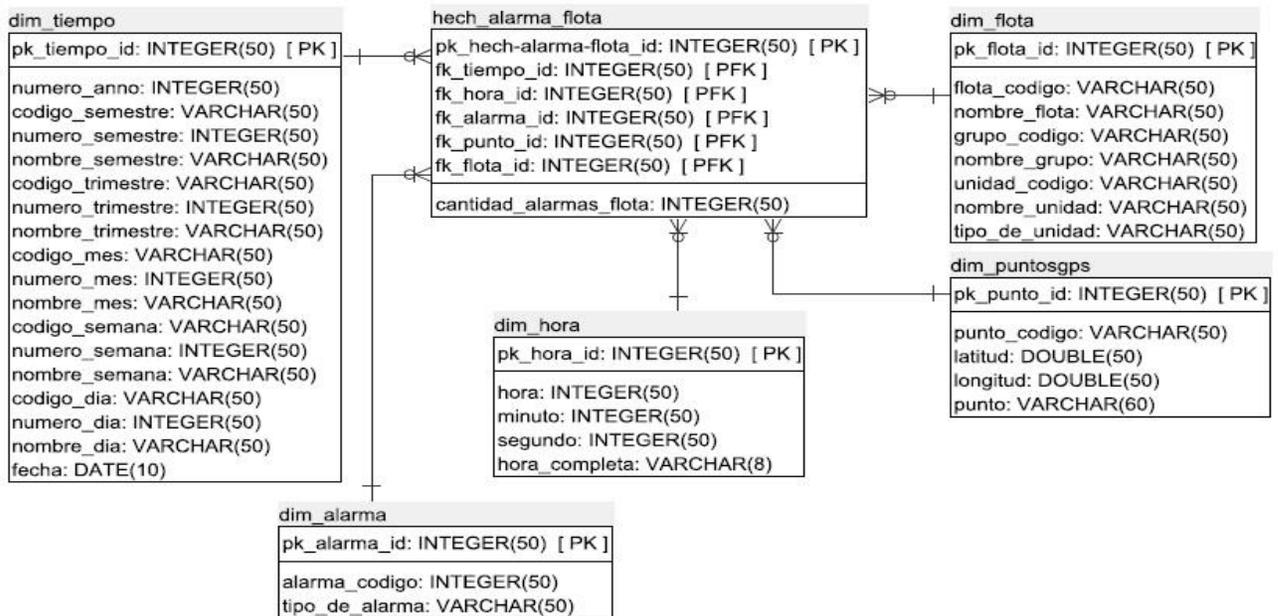
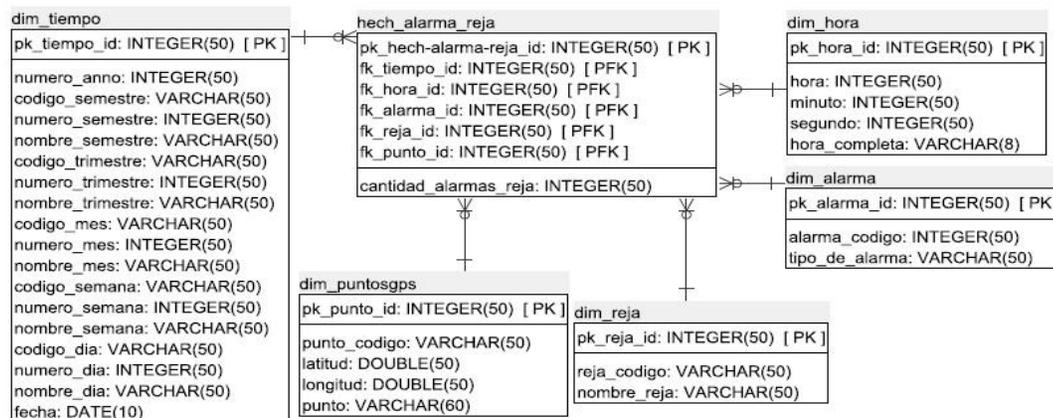


Ilustración 4. Tabla de hechos relacionada con las alarmas y las flotas.

La ilustración 5 muestra otra de las secciones del modelo dimensional, esta se relaciona con el hecho “*alarma-reja*”. En la misma se puede observar la relación entre dicha tabla de hecho y las dimensiones tiempo, hora, alarma, reja y punto GPS.



Capítulo 2. Análisis y Diseño

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

Ilustración 5. Tabla de hechos relacionada con las alarmas y las rejas.

La siguiente ilustración refleja otra de las secciones del modelo dimensional del MD FE-AVL, en la misma se observa la relacion existente entre la tabla de hecho “*alarma-ruta*” y las dimensiones tiempo, ruta, alarma, hora y punto GPS.

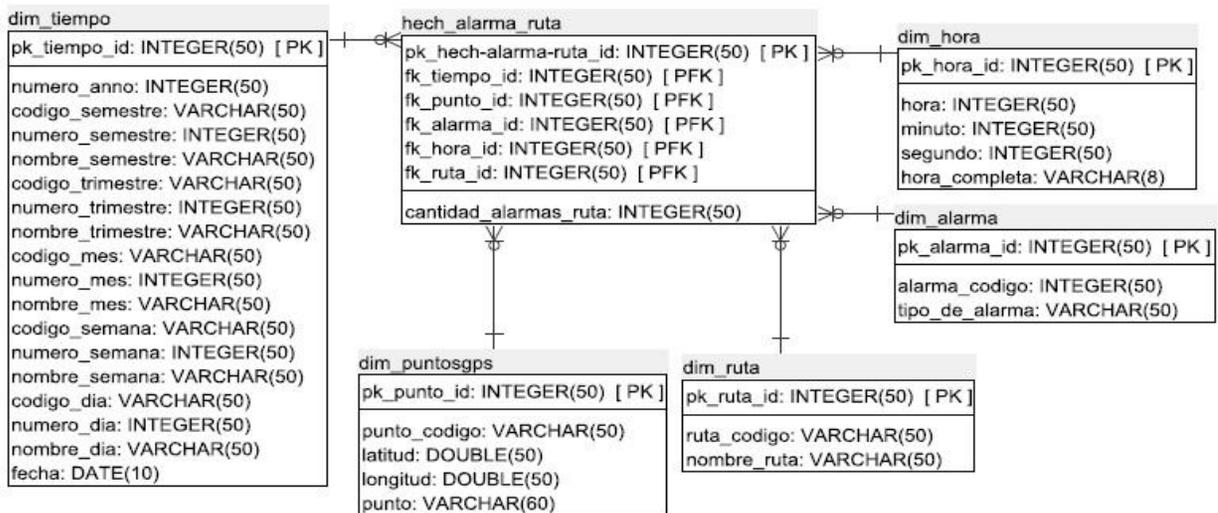


Ilustración 6. Tabla de hechos relacionada con las alarmas y las rutas.

A continuacion se presenta otra de las secciones del modelo dimensional correspondiente al MD FE-AVL. Esta seccion muestra la relacion existente etre las tablas de hechos “*reja-puntosGPS*” y “*ruta-puntosGPS*” con las dimensiones ruta, reja y puntos GPS.

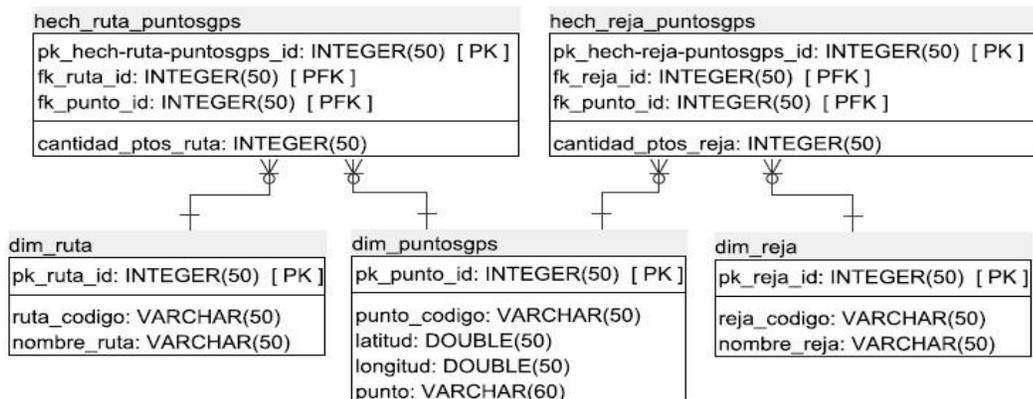


Ilustración 7. Tablas de hechos relacionadas con las rejas, las rutas y los puntos GPS.

2.8. Descripción de dimensiones y hechos

Luego de un estudio realizado sobre toda la información que se recoge en el sistema FE-AVL, fueron identificados los siguientes hechos y dimensiones:

Lista de Hechos:

hech_alarma_flota: Contiene la cantidad de alarmas disparadas por flotas y los identificadores de las dimensiones que se relacionan con este hecho.

hech_alarma_reja: Contiene la cantidad de alarmas disparadas en las rejillas y los identificadores de las dimensiones que se relacionan con este hecho.

hech_alarma_ruta: Contiene la cantidad de alarmas disparadas en las rutas y los identificadores de las dimensiones que se relacionan con este hecho.

hech_reja_puntosgps: Contiene la cantidad de puntos GPS que componen una rejilla y los identificadores de las dimensiones que se relacionan con este hecho.

hech_ruta_puntosgps: Contiene la cantidad de puntos GPS que componen una ruta y los identificadores de las dimensiones que se relacionan con este hecho.

Medidas Aditivas: cantidad_alarmas_flota, cantidad_alarmas_ruta, cantidad_alarmas_reja, cantidad_ptos_ruta y cantidad_ptos_reja.

Lista de Dimensiones:

dim_tiempo: Contiene la información distribuida en niveles de tiempo, año, semestre, trimestre, mes, semana y día.

dim_flota: Contiene la información relacionada con las flotas, los grupos en que se divide y las unidades que pertenecen a estos grupos.

dim_puntosgps: Contiene la información relacionada con los puntos GPS, latitud y longitud.

dim_hora: Contiene la hora, minuto, segundo.

dim_alarma: Contiene el tipo de alarma.

dim_ruta: Contiene el nombre de las rutas.

dim_reja: Contiene el nombre de las rejillas.

Dimensiones lentamente cambiantes: dim_alarma, dim_flota, dim_reja, dim_ruta, dim_puntosgps.

Capítulo 2. Análisis y Diseño

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

Para el manejo de estas dimensiones se utiliza la estrategia de tipo 2 propuesta por Kimball y explicada en el capítulo anterior. Dicha estrategia plantea, crear un registro dimensional adicional que permita registrar el cambio presentado por el valor del atributo.

2.9. Desarrollo de la Matriz BUS

La Matriz BUS o matriz dimensional, representa la relación existente entre las tablas de hechos y las tablas de dimensiones con las que éstas se relacionan. A continuación, se muestra la matriz correspondiente al Mercado de Datos para el sistema FE-AVL:

Matriz dimensional					
Dimensiones	Proceso de negocio				
	hech_alarma _flota	hech_ruta_ puntosgps	hech_reja _puntosgps	hech_alarma _ruta	hech_alarma _reja
dim_tiempo	x			x	x
dim_flota	x				
dim_puntosgps	x	x	x	x	x
dim_hora	x			x	x
dim_alarma	x			x	x
dim_ruta		x		x	
dim_reja			x		x

Tabla 3. Desarrollo de la Matriz BUS.

2.10. Modelo de Datos Físico

El modelo físico del MD FE-AVL está conformado por tres esquemas para lograr una mayor organización en el contenido del mismo. En el esquema “dimensiones” se encuentran las tablas de dimensiones del MD, en este caso son siete dimensiones. EL esquema “*mart_FE-AVL*” se ubican todas las tablas de hechos que se relacionan con las tablas de dimensiones del MD FE-AVL, son un total de cinco tablas de hechos, y por último el esquema “*mart_errores*”, éste esquema cuenta con una tabla de error asociada a

Capítulo 2. Análisis y Diseño

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

cada tabla de hecho, donde en caso de ocurrir un error en la carga de los datos, se almacena en la tabla de error correspondiente al hecho los atributos necesarios para identificarlo.

Esquemas	Tablas que lo componen
dimensiones	dim_alarma dim_flota dim_puntosgps dim_ruta dim_reja dim_tiempo dim_hora
mart_FE-AVL	hech_alarma_flota hech_alarma_reja hech_alarma_ruta hech_reja_puntosgps hech_ruta_puntosgps
mart_errores	tabla_de_error_hech_alarma_flota tabla_de_error_hech_alarma_reja tabla_de_error_hech_alarma_ruta tabla_de_error_hech_reja_puntosgps tabla_de_error_hech_ruta_puntosgps

Tabla 4. Modelo de Datos Físico

2.11. Conclusiones Parciales.

En el presente capítulo se detalla la información manejada en el sistema FE-AVL donde a partir de la identificación de las necesidades de los usuarios se definieron 61 requisitos de información, 11 requisitos funcionales y 15 requisitos no funcionales. Además se describen las reglas del negocio y los roles definidos hasta el momento, estos son: administrador, usuario y programador ETL, cada uno con los permisos correspondientes. Por otra parte se identificaron los hechos y las dimensiones que conforman el modelo dimensional, la matriz BUS que permite obtener la relación existente entre las tablas de hechos y las dimensiones; además el modelado tanto lógico como físico, donde en este último se observan los esquemas definidos y las tablas que lo componen, para de esta manera generar el script necesario y llevar a cabo la propuesta de solución. De este modo queda definido el diseño del Mercado de Datos para el sistema FE-AVL en pos de facilitar su posterior implementación.

Capítulo 3. Implementación y Validación

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

CAPÍTULO 3. IMPLEMENTACIÓN Y VALIDACIÓN

3.1. Introducción.

El presente capítulo está dirigido a la implementación de los aspectos relacionados con el proceso de integración de datos, inteligencia de negocio y las estructuras necesarias para la navegación y análisis de los datos, a través de los Reportes y las Vistas de Análisis; esperando que queden plasmados claramente los procedimientos utilizados. Además, para la validación del MD se tendrán en cuenta los Casos de Pruebas basados en Casos de Uso, con el objetivo de evaluar la calidad del producto.

3.2. Implementación de los subsistemas de integración

El proceso de ETL permite la integración consistente de los datos, mover datos desde múltiples fuentes, reformatearlos y limpiarlos para ser cargados en una BD, Mercado o Almacén de Datos. Posteriormente a este proceso los datos quedarán limpios. A continuación se seleccionan los campos a usar en el MD que se ejemplifica en la ilustración, los cuales tienen un nombre estándar ya que un mismo campo puede ser referido por nombres diferentes en varias tablas o fuentes de origen (Ver Ilustración 8).

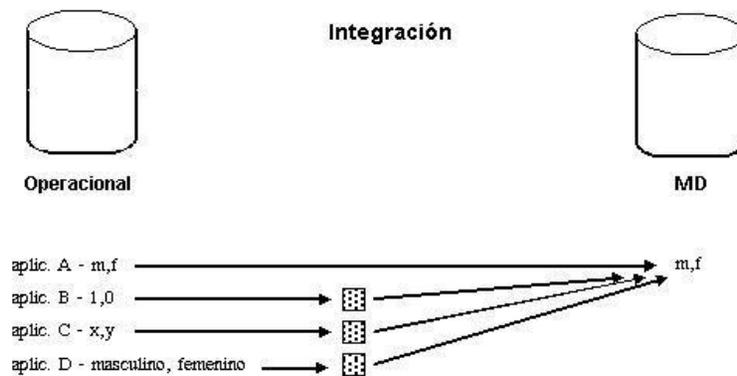


Ilustración 1. Integración de Datos en el proceso de ETL.

En este proceso se combinan datos provenientes de múltiples fuentes y brindan a los usuarios una vista unificada de los mismos. La implementación de los subsistemas de integración es un paso fundamental dentro de la implementación de un MD. La unificación minuciosa de los datos ayuda a evitar o reducir todo tipo de redundancia e inconsistencias entre los mismos.

Capítulo 3. Implementación y Validación

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

Para llevar a cabo el proceso de ETL es necesario realizar un grupo de actividades. Primero se extraen los datos de las fuentes origen que pueden ser una BD, ficheros de texto, archivos .dbf¹² o un fichero Excel para llenar las dimensiones y los hechos definidos. Posteriormente se procede a la validación de la información teniendo en cuenta las reglas del negocio. En caso de que se encuentren errores, estos son almacenados en una tabla error del MD, en correspondencia con la tabla de hecho en la cual se ha generado el error. En caso que los datos puedan obtenerse correctamente, son sometidos a un conjunto de búsquedas y transformaciones para llevar a cabo la inserción en las tablas del MD. A continuación se muestran ilustraciones de este proceso en el Mercado de Datos FE-AVL.

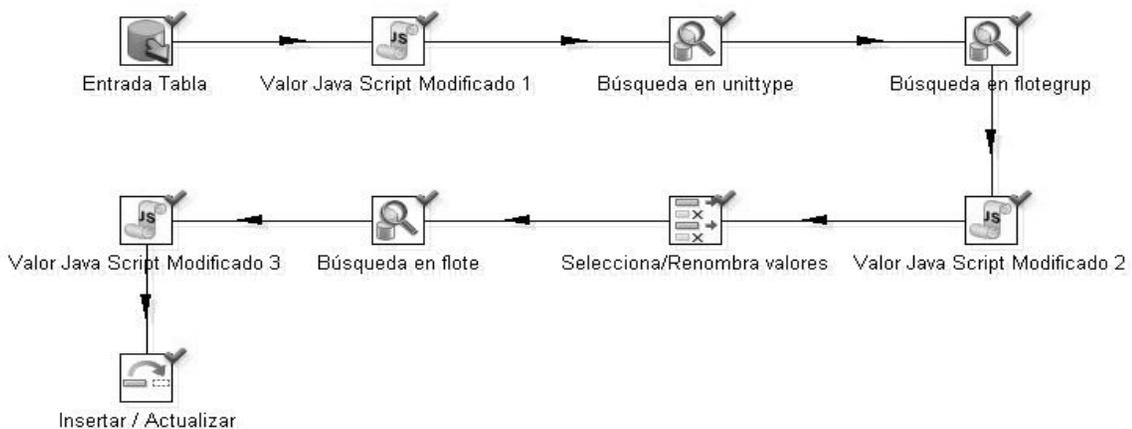


Ilustración 2. Proceso de transformación de la dimensión Flota

En la ilustración anterior se muestra el proceso de transformación de la dimensión **Flota**, donde primeramente fueron cargados desde la BD Operacional los datos de interés de las tablas correspondientes. Luego se realizan las búsquedas para validar a través de un código Java Script que los campos (*nombre de la Flota, código de la Flota, nombre del Grupo, código del Grupo, nombre de la Unidad, tipo de unidad y código de la Unidad*) no estén vacíos; en caso que el nombre del grupo y/o el código sean nulos, el código toma el valor “#9999” y el nombre el valor “No definido”. En el caso de las unidades sus atributos nunca serán nulos. Por último se insertan o actualizan los valores en la dimensión flota del MD, teniendo en cuenta la estrategia definida para las dimensiones lentamente cambiantes.

¹² **Archivos.dbf:** Archivos generados por el sistema de gestión de **BD** DBase (primer sistema de gestión de base de datos usado ampliamente para microcomputadoras, publicado por la compañía de software Ashton-Tate).

Capítulo 3. Implementación y Validación

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

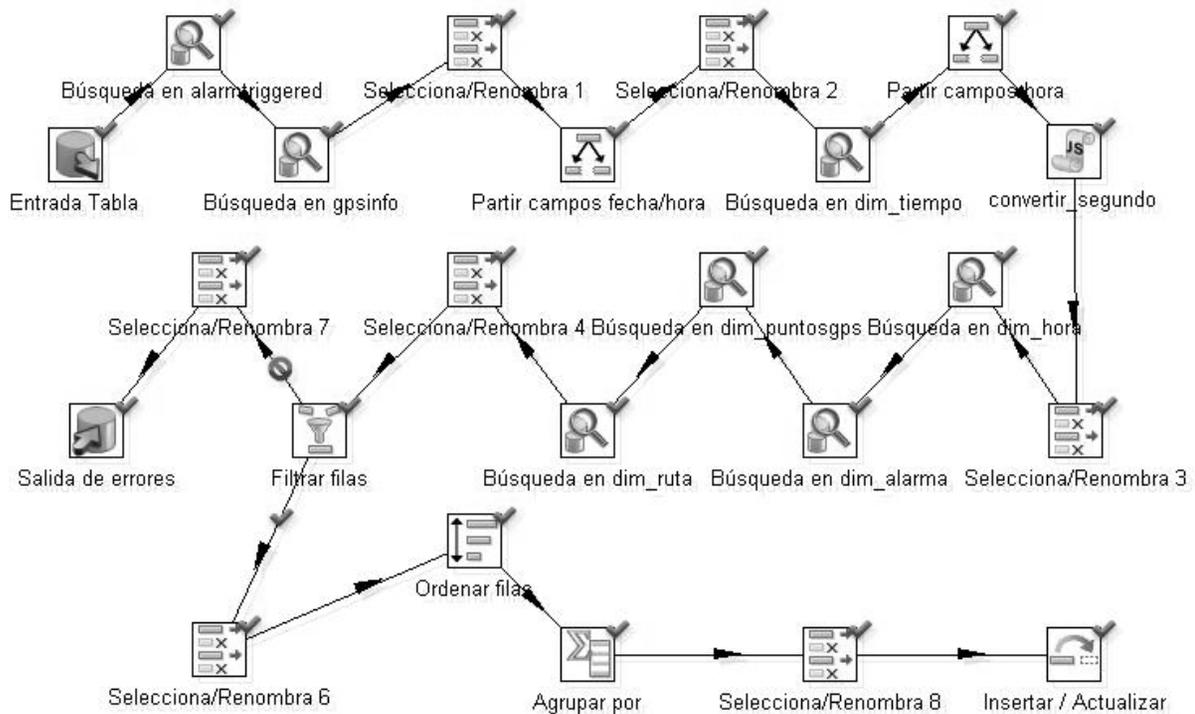


Ilustración 10. Proceso de transformación del hecho Alarma-Ruta

En la ilustración anterior se aprecia la transformación del hecho **Alarma-Ruta**. Primeramente se cargan desde la BD Operacional las tablas necesarias, luego se procede a la realización de la búsqueda en la tabla *alarmtriggered* y *gpsinfo* para obtener el campo *Date*, el cual es dividido para obtener de forma separada la *fecha* y la *hora*. Una vez obtenida la fecha, se hace una búsqueda en la dimensión tiempo para comparar la fecha de la dimensión con la fecha obtenida anteriormente. Terminado este proceso se dividió el campo hora para obtener los atributos hora, minuto y segundo por separado, para luego ser comparado con la hora el minuto y los segundos ubicados en la dimensión hora. Realizadas las diferentes búsquedas si coinciden los datos que son comparados por sus códigos, entonces se procede al filtrado de las filas que valida que los datos no sean nulos; en caso de que sean nulos, se almacenan en una tabla del MD con una descripción del error. En caso de no ser nulos se ordenan las filas y se agrupan los *id* de las dimensiones y se selecciona el *id* por el que se desean contar la cantidad de alarmas. Por último se insertan o actualizan los datos en la tabla de hecho alarma-ruta.

Capítulo 3. Implementación y Validación

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

3.3. Implementación de los trabajos (Jobs)

Los Jobs son tareas asignadas para que sean ejecutadas en un determinado momento. El Job es el encargado de ejecutar las transformaciones realizadas a los datos que serán cargados en el MD. Una vez creada la conexión al MD, todo se encuentra listo para poblar el mismo, donde serán los Jobs los encargados de ejecutar las transformaciones correspondientes a los hechos y dimensiones. Una vez que hayan sido realizadas todas las transformaciones, tanto de las dimensiones como de los hechos, se crea un Job por cada tabla de hecho y un Job general, donde este último se encarga de ejecutar los Jobs de cada uno de los hechos y las transformaciones de las dimensiones, este Job permite ejecutar todas las transformaciones, en vez de una a una. A continuación se muestran ejemplos de estos Jobs o trabajos.



Ilustración 31. Job para ejecutar la transformación del hecho Alarma-Flota

En la ilustración anterior se muestra el Job del hecho Alarma- Flota. Para ejecutar el Job primeramente se inicia el trabajo, esta operación se realiza en la opción *START*, luego en el paso llamado *obtener_fecha* se obtiene la fecha que se tiene en la tabla metadato, esta es la fecha de la última carga, posteriormente se ejecuta el paso llamado *carga_hech_alarma_flota* donde se ejecuta la transformación del hecho, en caso de existir valores nuevos, se insertan en la tabla de hecho, después de haber almacenado los datos nuevos se ejecuta el paso llamado *actualizar_fecha* para almacenar la fecha actual, la cual debe ser mayor que la última fecha almacenada en tabla metadato; por último se hace uso del componente *Success* para finalizar el Job, el cual adquiere el color verde al ejecutarse todos los pasos satisfactoriamente.

Capítulo 3. Implementación y Validación

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

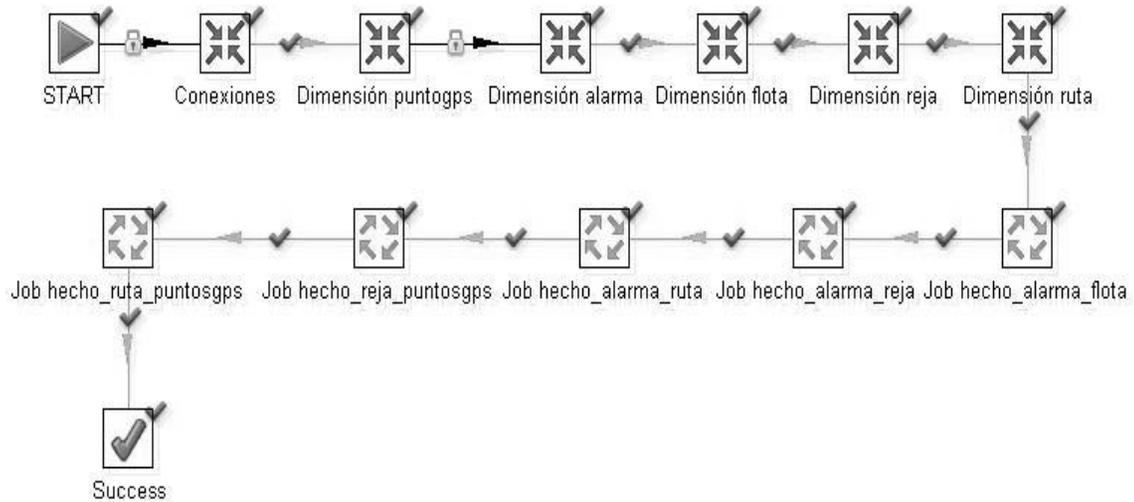


Ilustración 12. Job general para las transformaciones de la Base de Datos

En la ilustración anterior se muestra el Job general. Para ejecutar el Job primeramente se inicia el trabajo, esta operación se realiza en la opción *START*, luego en el paso llamado *Conexiones* se establece la fuente y el destino de la información, para ello se especifica el nombre predeterminado de la Base de Datos Operacional y del Mercado de Datos, posteriormente se ejecutan una a una las transformaciones correspondientes a las dimensiones y los Job de cada hecho; por último se hace uso del componente *Success* para finalizar el Job, el cual adquiere el color verde al ejecutarse todos los pasos satisfactoriamente.

3.4. Implementación del subsistema de visualización de datos

Para llevar a cabo una adecuada visualización de los datos es necesario clasificar y organizar la información, de modo que al navegar por la aplicación el usuario sea capaz de identificar fácilmente el contenido y las funcionalidades con que cuenta el sistema con solo leer el nombre. La Arquitectura de la Información es definida como la disciplina “que se ocupa del diseño estructural de los sistemas de información, su problema central es la organización, recuperación y presentación de información mediante el diseño de ambientes intuitivos”. (41)

A continuación se muestra el mapa de navegación del Mercado de Datos para el sistema FE-AVL:

Capítulo 3. Implementación y Validación

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

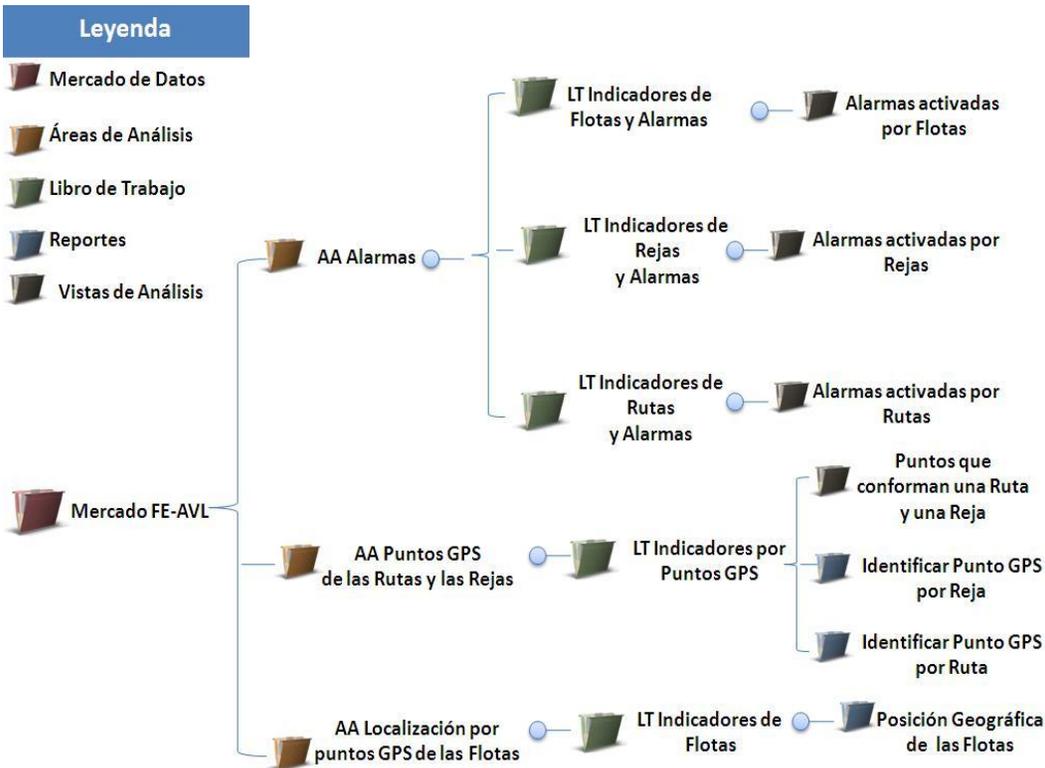


Ilustración 13. Mapa de navegación.

En correspondencia con los Requisitos de Información de dicho MD se definieron un total de 12 Reportes y 49 Vistas de Análisis, agrupadas en cinco Libros de Trabajo (LT) y ubicadas en tres Áreas de Análisis (AA). Por dichas áreas se analizan los Reportes, y los LT definen las categorías a las que puede pertenecer un determinado reporte (ver ilustración anterior).

3.4.1. Cubos OLAP

Los cubos OLAP son realizados a través de la herramienta Schema Workbench. El total de cubos del MD para la visualización de la información es definido en dependencia de la cantidad de tablas de hechos con que cuente el modelo dimensional. Para el MD fueron definidos 11 cubos OLAP, 5 de ellos correspondientes a las cinco tablas de hechos del modelo dimensional y el resto son definidos por el uso de diferentes niveles en la jerarquía de la dimensión donde se almacena el tiempo. A continuación se

Capítulo 3. Implementación y Validación

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

muestra una ilustración del esquema general de los cubos OLAP y otra del desglose de uno de estos cubos.

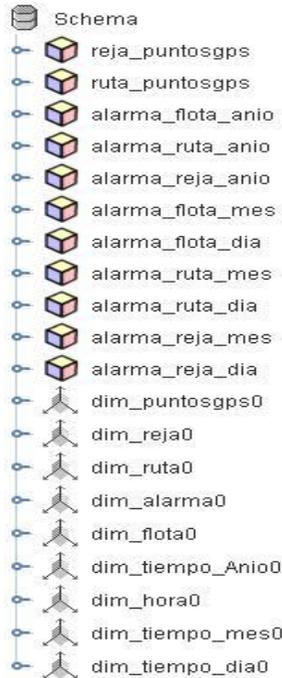


Ilustración 144. Esquema de cubos OLAP correspondiente a los hechos.



Ilustración 15. Cubo OLAP correspondiente al hecho (hech_alarma_reja).

Capítulo 3. Implementación y Validación

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

3.4.2. Vistas de Análisis

Las vistas de análisis pueden ser creadas o consultadas por el usuario en la herramienta de Inteligencia de Negocio de la colección de Pentaho después de publicados los cubos OLAP. Estas tienen como objetivo agrupar determinada información estadística que servirá de apoyo al proceso de toma de decisiones. Estas vistas permiten visualizar las distintas formas de análisis e interpretación de los datos desglosando la información hasta un nivel mínimo de detalle. A continuación se muestra un ejemplo de una de estas Vistas de Análisis.

Dimensión.Ruta		Medidas
	Dimensión.Punto GPS	● Cantidad de puntos GPS
<input type="checkbox"/> Rutas	<input type="checkbox"/> Punto GPS	55
Habana - Mariel	<input type="checkbox"/> Punto GPS	3
	Lat:22.798; Lng:-82.59	1
	Lat:23.546; Lng:-83.9	1
	Lat:22.645; Lng:-85.888	1
Santiago de Cuba - Habana	<input type="checkbox"/> Punto GPS	3
Habana - Matanzas	<input type="checkbox"/> Punto GPS	4
Ciego de Avila - Habana	<input type="checkbox"/> Punto GPS	4
Bauta - Mariel	<input type="checkbox"/> Punto GPS	4
Tunas - Habana	<input type="checkbox"/> Punto GPS	4
Habana - Baracoa	<input type="checkbox"/> Punto GPS	6
Vedado - Cerro	<input type="checkbox"/> Punto GPS	6
La Lisa - Marianao	<input type="checkbox"/> Punto GPS	4
Matanzas - Pinar del Rio	<input type="checkbox"/> Punto GPS	3
Caimito - Playa Baracoa	<input type="checkbox"/> Punto GPS	10
Guantanamo - Holguin	<input type="checkbox"/> Punto GPS	4

Ilustración 56. Vista de Análisis de los puntos GPS que componen una ruta.

3.4.3. Reportes

Los Reportes que son creados para el análisis de la información se realizan a través de la herramienta Diseño de Reportes (Report Designer) y el Servidor de Inteligencia de Negocio (Bi Server CE). Estos Reportes pueden ser estáticos o Ad-hoc (dinámicos). Los Reportes estáticos se realizan en la herramienta de Diseño de Reportes y son predefinidos por el cliente, mientras que los Ad-hoc pueden ser creados por

Capítulo 3. Implementación y Validación

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

los usuarios a través de los metadatos en la herramienta de Inteligencia de Negocio con la información que desee. Estos son llamados *reportes inmediatos*. A continuación se muestran ejemplos de ambos tipos de reportes.

Reportes estáticos

En el siguiente reporte se observa la localización de las unidades semanalmente, mostrando la flota y el grupo a la que pertenece cada unidad, así como la hora y su posición geográfica.

FEAVL
Front End Localización Automática de Vehículos

REPORTE HISTÓRICO

Flotas	Grupos	Unidades	Pos. Geográfica	Hora		
Estacion MMA	MMA-1G	Aston Mac Dragon	Lat:22.987; Lng:-90.77	5	55	1
Estacion MMA	MMA-1G	Concept Spyder	Lat:22.2; Lng:-87.5	1	2	4
Estacion MMA	MMA-1G	Wrangh	Lat:22.798; Lng:-82.59	2	50	41
Estacion MMA	MMA-3H	Style Coupe	Lat:22.645; Lng:-85.888	4	12	12

Ilustración 67. Reporte estático de la localización de unidades semanal.

Reportes Ad-hoc

En el siguiente reporte se presenta la información seleccionada por un usuario, referente a la cantidad de alarmas disparadas en las rejillas en el año 2013, dado un mes, una rejilla y la posición geográfica de la misma.

Capítulo 3. Implementación y Validación

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

Número año: 2013

Nombre del Mes	Nombre de la Reja	Cantidad de Alarmas por Reja	Punto
Enero	Artemisa	1	Lat:10.25; Lng:-52.23
Enero	Bauta	1	Lat:10.5; Lng:-53.54
Enero	Habana	1	Lat:22.2; Lng:-87.5
Enero	Holguin	1	Lat:22.798; Lng:-82.59
Enero	Las Tunas	1	Lat:11.2; Lng:-54.5
Enero	Maríel	1	Lat:10.23; Lng:-51.2
Enero	Playa	1	Lat:15.369; Lng:-60.258
Enero	Villa Clara	1	Lat:11.897; Lng:-55.2

Ilustración 7. Reporte Ad-hoc del total de alarmas por Rejas, punto GPS y mes.

Como ha quedado expuesto anteriormente estos Reportes Ad-hoc se crean haciendo uso de los metadatos, que no son más que datos predefinidos en el sistema a través de la herramienta Pentaho Metadata Editor. Para la creación de estos reportes fueron construidos cinco metadatos asociados a las cinco tablas de hechos del MD FE-AVL, los mismos se muestran a continuación.



Ilustración 89. Metadatos del Mercado de Datos FE-AVL

3.5. Validación del Mercado de Datos

Una de las fases más importantes dentro del ciclo de vida de un software es la fase de prueba. A lo largo de esta etapa se ejecutan un conjunto de acciones que se le aplican al sistema bajo determinadas condiciones de ejecución, que responden al funcionamiento del mismo y contribuyen al aseguramiento de la calidad del software, algunas de ellas se mencionan a continuación.

- **Pruebas unitarias:** Permiten probar el correcto funcionamiento de un componente o subsistema específico y son desarrollados por los propios desarrolladores durante la implementación.
- **Pruebas de integración:** Permite verificar la correcta integración de los componentes y

Capítulo 3. Implementación y Validación

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

subsistemas que conforman la solución. Pone a prueba la vista arquitectónica del sistema de definida en una infraestructura de desarrollo. Estas pruebas son ejecutadas por los arquitectos de software.

➤ **Pruebas de sistema (validación):** Permiten validar el cumplimiento de los requisitos de información y funcionales definidos por los clientes. Son las pruebas más cercanas a la realidad del cliente debido a que los probadores utilizan el sistema de la misma manera que será usado por los clientes. Estas son las pruebas constituyen las actividades fundamentales de la fase de Prueba. (39)

Durante la fase de prueba es donde generalmente se aciertan la mayor cantidad de errores. Es por esta razón que se realizan diferentes instancias de pruebas con el objetivo de entregarle al cliente un producto con la calidad requerida. Estas instancias son:

- **Pruebas internas:** Son desplegadas por el equipo de desarrollo o por un grupo de personas encargadas de probar la calidad de los sistemas realizados.
- **Pruebas de liberación:** Son realizadas generalmente por un grupo de personas o entidad externa al equipo de desarrollo con el fin de probar el sistema en su totalidad y validar el correcto funcionamiento del mismo.
- **Pruebas de aceptación:** Pruebas realizadas por el propio cliente para verificar que sus necesidades han sido satisfechas y que queda conforme con el producto. (39)

3.5.1. Diseño de los Casos de Pruebas

Para verificar el correcto funcionamiento del MD FE-AVL y posteriormente ser entregado al proyecto Front End AVL, es necesario realizar diferentes validaciones que corroboren los requisitos planteados inicialmente. En esta ocasión serán utilizados los Casos de Pruebas basados en Casos de Uso según plantea la metodología seleccionada. Para dicho Mercado de Datos se definieron cinco Casos de Pruebas. A continuación se muestra un ejemplo: el Caso de Pruebas basados en el CU *Visualizar indicadores relacionados con rejás y alarmas*. Los demás Casos de Pruebas se pueden encontrar en los Anexos del 11 al 15.

Escenario	EC 1.1.1. Obtener cantidad de alarmas por tiempo, hora, reja y alarma.
Descripción	Muestra la cantidad de alarmas disparadas dado un tiempo, una hora, una reja y una alarma específica.

Capítulo 3. Implementación y Validación

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

Variables de Entrada	hora, reja, tiempo, alarma
Variables de Salida	alarma
Respuesta del Sistema	Se muestra la información correspondiente al escenario.
Flujo Central	<ol style="list-style-type: none"> 1. En la parte superior izquierda se selecciona el nivel principal Mercado FE-AV. 2. Se selecciona el Área de Análisis Alarmas. 3. Selección del LT Indicadores de Rejas y Alarmas. 4. Se selecciona la Vista de Análisis correspondiente al escenario específico.

Tabla 5. Escenario de Caso de prueba del CU Visualizar indicadores relacionados con rejas y alarmas.

Dentro de las pruebas internas del sistema además de realizar los Casos de Pruebas, también fueron probados los procesos ETL para garantizar que los datos responden al negocio modelado, o sea, se verificó que los datos presentes en la Base de Datos Operacional fueron guardados correctamente en el MD respetando las Reglas del Negocio. Fueron realizadas tres iteraciones durante el proceso de prueba a través de los Casos de Pruebas basados en Casos de Uso. En las dos primeras iteraciones se detectaron un conjunto de No Conformidades (NC) que respondían en su mayoría a uso de caracteres extraños que no eran reconocidos por la herramienta de Inteligencia de Negocio. Estas NC fueron solucionadas estandarizando todos los nombres de las Áreas y Vistas de Análisis. Para una tercera iteración no se detectaron no conformidades.

Número de iteraciones	Cantidad de No Conformidades (#NC)	Descripción de las NC
1	30	La mayoría de las NC fueron referentes a la utilización de caracteres tales como tildes y ñ, el desglose de la jerarquía en la herramienta Workbench y el duplicado de nombres.
2	10	Problemas de visualización de la información debido a los permisos de los usuarios y estructura de los datos para su análisis.
3	0	No se detectaron NC.

Tabla 6. Resultado de la aplicación de los Casos de Pruebas al MD.

Capítulo 3. Implementación y Validación

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

3.6. Resultado y Funcionalidades Obtenidas

El MD para el sistema FE-AVL es una solución que propicia un grado adecuado en la certeza de la información utilizada a la hora de tomar decisiones importantes en una entidad, ya que permite almacenar por largos períodos de tiempo gran cantidad de datos y garantiza un rápido acceso a la información solicitada por el usuario. Como resultado de la propuesta se tiene en su versión 1.0 un MD que forma parte del sistema FE-AVL desarrollado en el centro de Telemática de la Universidad de las Ciencias Informáticas. Dentro de las funcionalidades que presenta dicho sistema de información se encuentran:

- Creación de Vistas de Análisis.
- Creación de Reportes estáticos.
- Creación de Reportes Dinámicos o Ad-hoc.
- Visualizar información de:
 - Las flotas y las alarmas.
 - Las rutas y las alarmas
 - Las rejillas y las alarmas
 - Los puntos GPS de las rutas y las rejillas.

3.7. Conclusiones.

En el presente capítulo se abordaron temas relacionados con la implementación del subsistema de integración, fueron creadas 13 transformaciones para las tablas de hechos y dimensiones, además se realizaron 6 Jobs que permiten la ejecución de estas transformaciones. Por otra parte fueron tratados temas referentes a la estructura de la información, la cual permite una mayor organización del contenido del MD y facilita la navegación al usuario. Fueron creados 12 Reportes y 49 Vistas de Análisis que dan respuesta a los requisitos planteados inicialmente, estas últimas fueron elaboradas a través de 11 Cubos OLAP. Se implementaron 5 metadatos que le permiten al usuario el diseño de Reportes Ad-hoc. Además se llevaron a cabo un conjunto de Casos de Pruebas basados en Casos de Uso para la detección y corrección de errores en la implementación del MD, obteniendo así un producto validado y de una calidad aceptable.

CONCLUSIONES GENERALES

Con la culminación del presente trabajo de diploma se obtuvo un Mercado de Datos, el cual posee como fuente de extracción el sistema FE-AVL del centro de Telemática de la Facultad 2 de la Universidad de las Ciencias Informáticas. Dicho MD contempla la información referente a la Localización Automática de Vehículos en el Control y Gestión de Flotas, lo que permite el almacenamiento histórico para su posterior consulta a través de Reportes o VA, y la generación de nuevos reportes o vista según el usuario lo desee, los cuales posibilitan el análisis de la información manejada para el apoyo a la toma de decisiones.

La investigación cumplió los objetivos planteados y se arribaron a las siguientes conclusiones:

- El estudio del arte propició la correcta selección de la tecnología de almacenamiento, los lenguajes y las herramientas utilizadas en la solución.
- El análisis de la solución permitió enfocar el desarrollo hacia las necesidades objetivas planteadas.
- El diseño lógico aseguró una correcta estructura dimensional.
- La solución obtenida garantizó un almacenamiento estructurado y funcional de la información.
- Las pruebas realizadas validaron la solución en función de los Casos de Uso obteniéndose un nivel adecuado de calidad.

RECOMENDACIONES

- Incitar la investigación de los Almacenes y Mercados de Datos para hacer uso de la Inteligencia de Negocio en las soluciones a problemas empresariales.
- Desplegar el Mercado de Datos para el sistema FE-AVL ajustado a un negocio específico.
- Una vez desplegado el MD mantener actualizada la información ante cambios que puedan ocurrir.

REFERENCIAS BIBLIOGRÁFICAS

1. **Bricenno Vega, Edgar Armando.** Gestipolis. *Gestipolis*. [En línea] Junio de 2005. [Citado el: 16 de Octubre de 2012.] <http://www.gestipolis.com/Canales4/mkt/simparalas.htm>.
2. **Olay, Raúl.** Universidad de Yacambú. *Universidad de Yacambú*. [En línea] [Citado el: 15 de Enero de 2013.] <http://www.oocities.org/es/sirdoochard/electivas/benchmarking/Trabajo2/trabajo2.html>.
3. **Rodríguez, Milton.** Universidad ORT Uruguay. *Universidad ORT Uruguay*. [En línea] Septiembre de 2008. [Citado el: 29 de Noviembre de 2012.] <http://www.ort.edu.uy/fi/pdf/miltonrodriguezlicsisistemasort.pdf>.
4. Centro de Investigación de las Telecomunicaciones. *Centro de Investigación de las Telecomunicaciones*. [En línea] Diciembre de 2010. [Citado el: 29 de Noviembre de 2012.] <http://www.interactic.org.co/index.php>.
5. *FE-AVL sistema automático de localización vehicular.* **Arce Martínez, Alberto, y otros.** La Habana : Revista Cubana de Ciencias Informáticas, 2012, Vol. 6.
6. **Bigatti, Cristian G y Grasso, Mónica.** *Data warehouse*. 2010.
7. **Herederó, Carmen de Pablo, Albarrán Lozano, Irene y Castilla Alcalá, Guillermo.** DIALNET. *DIALNET*. [En línea] 2010. [Citado el: 29 de Noviembre de 2012.] <http://dialnet.unirioja.es/descarga/articulo/187759.pdf>.
8. **Velasco, Roberto Hernando.** [En línea] [Citado el: 20 de octubre de 2011.] <http://www2.rhernando.net/modules/tutorials/doc/bd/dw.html>.
9. **García Gerardo, Clemente.** UNIVERSIDAD POLITÉCNICA DE VALENCIA. *UNIVERSIDAD POLITÉCNICA DE VALENCIA*. [En línea] Junio de 2008. [Citado el: 29 de de Noviembre de 2012.] <http://riunet.upv.es/manakin/bSITream/handle/10251/2505/tesisUPV2842.pdf>.
10. Mercado de datos: conceptos y metodologías de desarrollo. **Chinchilla Arley, Ricardo.** 3, s.l. : Tecnología en Marcha, 2011, Vol. 24.
11. Greenhouse. *Greenhouse*. [En línea] 2010. [Citado el: 29 de Noviembre de 2012.] <http://www.swgreenhouse.com/Productos/Hi-Spins/DataMart.html>.
12. Facultad de Ciencias Agrarias. Universidad Nacional del Litoral. Facultad de Ciencias Agrarias. Universidad Nacional del Litoral. [En línea] [Citado el: 8 de Enero de 2013.] www.fca.unl.edu.ar/agromatica/Docs/09-ModeloEntRel.PDF.
13. UNIVERSIDAD TECNOLÓGICA NACIONAL. *UNIVERSIDAD TECNOLÓGICA NACIONAL*. [En línea] 2010. [Citado el: 30 de Noviembre de 2012.] <http://www.edutecne.utn.edu.ar/sist-gestion-II/Apunte%2520BI.pdf>.
14. ETL – Tools.info. *ETL – Tools.info*. [En línea] [Citado el: 10 de Enero de 2013.] http://etl-tools.info/es/bi/almacenedatos_esquema-conste.

Referencias Bibliográficas

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

15. **León Guzmán, Elizabeth.** *Modelamiento Dimensional*. Bogotá : s.n., 2009.
16. **Zepeda Sánchez, Leopoldo Zenaido.** *Metodología para el Diseño Conceptual de Almacenes de Datos*. Valencia : s.n., 2008.
17. **Wolff, Carmen Gloria.** *Dimensiones lentamente cambiantes*. 2010.
18. Dataprix. *Dataprix*. [En línea] Marzo de 2010. [Citado el: 12 de Diciembre de 2012.] <http://www.dataprix.com/blogs/bernabeu-dario/dimensiones-lentamente-cambiantes>.
19. Ryrhermanos. *Ryrhermanos*. [En línea] 2010. [Citado el: 16 de Noviembre de 2012.] http://correo.ryrhermanos.com/home/jrivera@ryrhermanos.com/Publico/PDI_Parte_1.pdf.
20. Computer Audio Visual Systems Integrator. *Computer Audio Visual Systems Integrator*. [En línea] 2010. [Citado el: 13 de Enero de 2013.] <http://www.cavsi.com/preguntasrespuestas/que-es-procesamiento-analitico-en-liea-olap>.
21. Wikispaces. *Wikispaces*. [En línea] 2009. [Citado el: 28 de Noviembre de 2012.] <http://oltp.wikispaces.com/home>.
22. Wikispaces. *Wikispaces*. [En línea] 2009. [Citado el: 28 de Noviembre de 2012.] <http://oltp.wikispaces.com/Diferencia+entre+OLTP+y+OLAP>.
23. Empai. *Empai*. [En línea] 2008. [Citado el: 16 de Octubre de 2012.] <http://www.empai-matanzas.co.cu/revista/Vol.%202%20%20%20No.%2>.
24. AVL-Plus. *AVL-Plus*. [En línea] [Citado el: 8 de Enero de 2013.] <http://www.avl-plus.com.ar/servicios.htm>.
25. Knosos. *Knosos*. [En línea] [Citado el: 8 de Enero de 2013.] <http://www.knosos.es/espanol/MPSAVL.php>.
26. Sistemas MP SRL. Software de Monitoreo GPS. *Sistemas MP SRL. Software de Monitoreo GPS*. [En línea] [Citado el: 8 de Enero de 2013.] <http://www.sistemasmp.com/software-avl-rastreo-satelital-gps-tracking-central-de-monitoreo/>.
27. **Enric Biosca.** Dataprix. *Dataprix*. [En línea] [Citado el: 8 de Enero de 2013.] www.dataprix.com/files/Tutorial_MDX.pdf.
28. 1KeyData. *1KeyData*. [En línea] [Citado el: 8 de Enero de 2013.] <http://www.1keydata.com/es/sql/>.
29. Magma Soft. *Magma Soft*. [En línea] [Citado el: 12 de Enero de 2013.] www.magma.com.ni/~jorge/upoli_uml/refs/Que_es_UML.doc.
30. Sitio de descargas de software. *Sitio de descargas de software*. [En línea] 2010. [Citado el: 13 de Enero de 2013.] http://www.freedownloadmanager.org/es/downloads/Paradigma_Visu.
31. BI Business Intelligence. Pentaho [En línea] 28 de Abril de 2009. [Citado el: 05 de Enero de 2013.] <http://bi-businessintelligence.blogspot.com/2009/04/pentaho.html>.

Referencias Bibliográficas

“Propuesta de Mercado de Datos para el proyecto Front End Localización Automática de Vehículos”

32. Cognus. *Cognus*. [En línea] 2011. [Citado el: 8 de Enero de 2013.] <http://www.cognus.cl/content/view/794163/Pentaho-Data-Integration-4-0.html>.
33. SQL Power Software. *SQL Power Software*. [En línea] [Citado el: 3 de Enero de 2013.] <http://www.sqlpower.ca/page/architect>.
34. Dataprix. *Dataprix*. [En línea] 2010. [Citado el: 3 de Enero de 2013.] <http://www.dataprix.com/empresa/productos/sql-power-software>.
35. PostgreSQL-es. Portal en español de PostgreSQL. *PostgreSQL-es. Portal en español de PostgreSQL*. [En línea] [Citado el: 15 de Enero de 2013.] http://www.postgresql.org.es/sobre_postgresql.
36. **Aliaga Ibarra, Antonio y Miani Flores, Marcos Agustin**. I.E.S. San Vicente. *I.E.S. San Vicente*. [En línea] 21 de Enero de 2008. [Citado el: 16 de Febrero de 2013.] <https://iessanvicente.com/colaboraciones/postgresql.pdf>.
37. KLE, Transforming Knowledgeintoaction! BI en la práctica. *KLE, Transforming Knowledgeintoaction! BI en la práctica*. [En línea] 2010. [Citado el: 9 de Marzo de 2013.] <http://www.siskle.com/spanish/articulo04.html>.
38. **Bernabeu, Ricardo Dario**. *Hefesto: Metodología propia para la construcción de un DataWarehouse*. Buenos Aires : s.n., 2007.
39. **González Hernández, Yanisbel**. *Propuesta de Metodología para el desarrollo de Almacenes de Datos en DATEC*. La Habana : s.n., 2010.
40. Inteligencia de Negocio para Todos. *Inteligencia de Negocio para Todos*. [En línea] 2010. [Citado el: 23 de Febrero de 2013.] <http://inparatodos.blogspot.com/2010/09/que-es-un-modelo-dimensional.html>.
41. El profesional de la información. **BAEZA-YATES, Ricardo, RIVERA LOAIZA, Cuauhtémoc y VELASCO MARTÍN, Javier**. 3, 2004, Vol. 13.

GLOSARIO DE TÉRMINOS

A

AVL: Sigla que proviene del inglés Automatic Vehicle Location, término utilizado para hacer alusión a la Localización Automática de Vehículos.

F

Front End: Término utilizado para nombrar una aplicación de escritorio para la localización automática de vehículos, diseñada en la Universidad de las Ciencias Informáticas. Dicho sistema tiene como principal característica que debido a su arquitectura es adaptable en diferentes entornos.

Flota: Las flotas están compuestas por un conjunto de unidades agrupadas en diferentes grupos según una clasificación.

G

Grupo de las Flotas: Clasificación para agrupar las unidades por un determinado criterio.

R

Reja: Término utilizado para identificar conjunto de puntos geográficos que determinan un área específica en metros cuadrados.

Ruta: Término utilizado para identificar un conjunto de puntos geográficos y poseen un tamaño determinado en km.

Reportes Pentaho: Motor de presentación, capaz de generar informes sobre la base de un archivo de definición XML.

V

Vista de Análisis: Vista estática de determinada información bien detallada que permite su posterior análisis para apoyar el proceso de toma de decisiones.