



Universidad de Cienfuegos "Carlos Rafael Rodríguez"
Facultad de Informática
Carrera de Ingeniería Informática

*“Implementación de una arquitectura de Inteligencia
de Negocios para la toma de decisiones sobre las
Infracciones del Tránsito en el MININT”*

Trabajo de diploma para optar por el título de Ingeniería en Informática

Autor:

Damián Lima Moreira

Tutores:

My. Manuel Cárdenas

Msc. Eduardo Concepción

Cienfuegos, 2009

Declaración de autoría

Declaro que soy el único autor de este trabajo, el cual será utilizado por la Dirección del MININT para los fines que estime conveniente, tanto de forma parcial como total y que además no podrá ser presentado en eventos ni publicado sin la autorización del MININT.

Para que así conste firmo (firmamos) la presente a los ____ días del mes de J____ del 2009.

Firma del Autor

Damián Lima Moreira

Firma del Tutor

My. Manuel Cárdenas

Firma del Tutor

Msc. Eduardo Concepción

Los abajo firmantes certificamos que el presente trabajo ha sido revisado según acuerdo de la dirección de nuestro centro y el mismo cumple los requisitos que debe tener un trabajo de esta envergadura referente a la temática señalada.

Firma del Tutor

My. Manuel Cárdenas

Firma del Tutor

Msc. Eduardo Concepción

Firma ICT

Firma Vicedecano

Agradecimientos

Agradezco con todo mi corazón a mi madre y a mi padre por brindarme todo su amor y por apoyarme siempre. Todo lo bueno que he logrado hasta hoy se lo debo a ellos.

A mi novia que ha estado a mi lado ayudándome todo este tiempo.

A mi hermanita linda que es la luz de mi casa.

A mis tutores por haberme guiado.

A René y a Rosalina por brindarme su ayuda.

A mis profesores de la Universidad de Cienfuegos.

A mis compañeros del MININT.

A todos los que de una forma u otra me ayudaron en la realización de este trabajo.

Dedicatoria

A mi familia.

Resumen

El MININT (Ministerio del Interior) tiene como misión garantizar el orden y la tranquilidad ciudadana. Pertenecen a esta institución las fuerzas públicas de la PNR (Policía Nacional Revolucionaria), las cuales tienen entre sus funciones velar por el cumplimiento de las leyes en las carreteras con el objetivo de evitar accidentes y preservar las vidas humanas.

Los dirigentes de esta institución necesitan diariamente diferentes reportes para evaluar el comportamiento de las infracciones del tránsito y poder tomar decisiones vitales en esta área, por lo que para suplir este requerimiento de información se desarrolla en el presente trabajo de diploma una arquitectura de Inteligencia de Negocio (Business Intelligence o BI) compuesta de un MD (Mercado de Datos), cargado con información de las infracciones, y una herramienta web para visualizar su contenido, partiendo de la información almacenada por los sistemas de Procesamiento de Transacciones en Línea (OLTP) del MININT.

Para realizar el diseño del MD se utilizó la metodología *Data Warehouse Engineering Process (DWEP)*, la cual está basada en RUP (Rational Unified Process) y UML (Unified Modeling Language), lenguaje de modelado que se usa para documentar artefactos de un sistema de software.

Índice

Introducción	1
Capítulo 1 – Fundamentación teórica.....	7
1.1 Introducción.....	7
1.2 Inteligencia de Negocio.....	7
1.2.1 Sistemas de Soporte a la toma de Decisiones (DSS)	9
1.2.2 Procesamiento de Transacciones en Línea y Procesamiento Analítico en Línea.....	10
1.2.3 Almacén de Datos	12
1.2.4 Mercado de Datos.....	16
1.2.5 Procesos de Extracción, Transformación y Carga (ETL)	16
1.2.6 Modelo Dimensional.....	17
2.1 Metodología para el Desarrollo de un Almacén de Datos.	18
2.1.1 Diseño de un AD.	18
2.1.2 Modelo Conceptual.....	19
2.1.3 Modelo Lógico.....	20
2.1.4 Modelo Físico	22
2.2 Sistemas Gestores de bases de datos que dan soporte a BI.....	22
2.2.1 Microsoft SQL Server 2005	23
2.2.2 Microsoft SQL Server 2008.....	25
2.2.3 Oracle	27
2.2.4 ¿Por qué Oracle?	31
2.3 Oracle SQL Developer	31
2.4 Conclusiones.....	32
Capítulo 2 – Descripción de la Situación Actual	33
2.1 Introducción.....	33
2.2 Descripción del Objeto de Estudio.....	33
2.3 Sistemas Existentes.....	37
2.4 Conclusiones.....	37
Capítulo 3 – Construcción de la solución propuesta	38
3.1 Introducción.....	38
3.2 Descripción del sistema propuesto	38
3.2.1 Requerimientos	38
3.2.2 Diseño del Mercado de Datos	40
3.2.2.2 Nivel 1 (Definición Global del Modelo)	41
3.2.2.3 Nivel 2 (Definición de Estructura de Esquemas Estrella).....	41
3.2.2.4 Nivel 3 (Definición de Estructura de Hechos y Dimensiones)	42
3.2.2.5 Mapeo de Datos.....	53
3.2.2 Modelo Lógico.....	55
3.2.2.1 Procesos ETL	55
3.2.3 Modelo Físico	55
3.2.3.2 Diagrama de Despliegue.....	56
3.4 Validación de la solución Propuesta.....	58
3.4 Conclusiones.....	58
Conclusiones	60

Recomendaciones.....	61
Referencias bibliográficas	62
Bibliografía	64
Glosario de términos.....	66
Anexos.....	67
Anexo 1 – Prototipos de la aplicación.....	68
Anexo 2 – Modelo Lógico del Esquema Estrella	72
Anexo 3: Validación.....	76

Índice de Tablas

Tabla 1: Diferencias entre bases de datos operacionales y almacenes de datos	14
Tabla 2: Estereotipos empleados en el modelo conceptual del AD. [9]	20
Tabla 3: Estereotipos empleados en el modelo de los procesos ETL. [9].....	21

Índice de Figuras

Figura 1: Esquema conceptual del AD. Nivel1.	41
Figura 3. 1 Estructura de los Hechos.....	53
Figura 3. 2 Diagrama de Mapeo de los Datos.	54
Figura 3. 7 Diagrama de Componentes.	56
Figura 3. 18 Modelo Lógico del Esquema Estrella. Parte I.....	73
Figura 3. 19 Modelo Lógico del Esquema Estrella. Parte II.....	74
Figura 3. 20 Modelo Lógico del Esquema Estrella. Parte III.....	75

Introducción

Debido a los beneficios que reportan en la actualidad, las Tecnologías de la Información y las Comunicaciones (TIC) son acogidas en numerosas organizaciones con el fin de mejorar la gestión y el control de sus bienes. En Cuba, esta tendencia ha ido en incremento y el MININT no se ha quedado atrás, siendo el presente trabajo una muestra de ello y del interés de modernizar sus métodos de trabajo con el objetivo de hacer una labor más efectiva.

La situación económica y social que caracteriza a la sociedad moderna genera profundos cambios en las empresas las cuales se preparan para ser más flexibles y establecen estrategias con el objetivo de adaptarse al entorno altamente turbulento en el que desarrollan sus acciones. Ante ambientes tan poco estables y la imposibilidad de actuar a ciegas, los miembros de la organización y, en particular, su alta gerencia necesitan manipular grandes volúmenes de información para cumplir con sus funciones esenciales.

La segunda mitad del Siglo XX fue testigo de algunos hechos concretos que propiciaron grandes cambios, entre ellos: la aparición de las computadoras, el crecimiento acelerado de su velocidad de procesamiento y capacidad para el almacenamiento de datos, la facilidad de interconexión de las computadoras y la aparición de la gran red de redes, Internet. Ello ha posibilitado disponer de servicios de acceso en línea a bases de datos y ha provocado una gran explosión de información que rebasa la capacidad de procesamiento de las organizaciones y la búsqueda de herramientas para el manejo de estos grandes volúmenes de información.

En la evolución de las tecnologías de la información se puede observar como se ha puesto énfasis en hacer más rápidos y eficientes a los sistemas de Procesamiento de Transacciones en Línea (OLTP) también conocidos como Sistemas Operacionales. Estos tienen como objetivos reflejar el estado y el funcionamiento de la empresa, desenvolviéndose muy bien en entornos de gestión puros que manejan transacciones u operaciones diarias sencillas donde son muy importantes la actualización y el tiempo de respuesta. La función de estos sistemas consiste en recoger y almacenar los datos generados a partir de las aplicaciones informáticas que se utilizan en los distintos

departamentos de la organización permitiendo la impresión de algunos informes predefinidos según las necesidades de la compañía. Desde la visión de la dirección estratégica están diseñados para resolver problemas de carácter operativo, por lo que no son útiles para hacer análisis más avanzados. A esto se le suma que los requerimientos de información de las empresas han cambiado, los responsables de los distintos departamentos cada vez precisan de más informes que por lo general son de carácter urgente y los departamentos de informática no son capaces de satisfacer esta creciente demanda de información. Las organizaciones comenzaron a valorar la importancia de los datos y a considerarlos como materia prima de la información, un activo intangible de la empresa que si es bien utilizado puede generar ventajas competitivas reales.

Surgió entonces la necesidad de crear nuevas bases de datos centrales, con información resumida y detallada. Los problemas técnicos que se generaron fueron de diversa índole. En primer lugar, para cargar la nueva base de datos había que centralizar la información que se encontraba dispersa, duplicada y en distintos formatos. En segundo lugar, cómo cargar esa información en la nueva base de datos central de modo que fuera confiable. Y en tercer lugar, cómo acceder a ella de forma confiable.

Las herramientas de Inteligencia de Negocio (Business Intelligence o BI) permiten medir y controlar el desarrollo de las variables importantes del negocio, buscando identificar, proyectar y predecir tendencias a partir de los datos acumulados lo que se traduce en una ventaja competitiva. Estas aplicaciones hacen análisis complejos y llevarlo a cabo sobre un sistema operacional puede resultar en la degradación del mismo, con el consiguiente impacto en la operación del negocio debido al gran volumen de datos sobre el que se actúa y el tiempo de procesamiento que conlleva la devolución de los pedidos, además de que la estructura de las bases de datos que el sistema OLTP utiliza no es la mejor para la obtención de la información deseada. Así aparece un nuevo elemento, el cual se ha convertido en una de las tendencias tecnológicas más significativas en la administración de información y tiene como principal componente el Almacén de Datos (AD o Data Warehouse). Este va a contener una réplica de la información previamente filtrada y procesada de las bases de datos de los sistemas OLTP.

Un AD es una combinación de conceptos y tecnologías que cambian significativamente la manera en que es entregada la información a los usuarios. Su objetivo principal es satisfacer los requerimientos de información internos de la empresa para una mejor gestión, con eficiencia y facilidad de acceso.

Aunque diversas organizaciones y personas individuales logran comprender el enfoque de un Almacén de Datos, la experiencia ha demostrado que existen muchas dificultades potenciales. Reunir los elementos de datos apropiados desde diversas fuentes en un ambiente integral centralizado, simplifica el problema de acceso a la información y en consecuencia, acelera el proceso de análisis, consultas y el menor tiempo de uso de la información.

En este ámbito existe otro concepto importante, el de Mercado de Datos (MD). Este comparte los mismos principios que un AD, está enfocado a un departamento o área específica y es utilizado en algunos contextos donde es necesario que un grupo específico de personas solo accedan a parte de la información de la empresa.

La carga de información en el Mercado o Almacén de Datos mediante los procesos ETL (Extracción, Transformación y Carga) se puede realizar en forma manual, a través de creación de programas, o bien, cuando el volumen de información y dispersión de fuentes lo justifican, utilizando herramientas diseñadas para este fin.

Para poder acceder a la información en forma rápida y flexible se ha creado el modelo dimensional (DM), conocido también como Diagrama en Estrella, el cual permite trabajar con la información en diferentes niveles de detalle, aunque también se puede optar por un almacenamiento multidimensional.

Durante años BI ocupó un segundo plano en las prioridades de la empresa, por su escasa madurez operativa representaba inversión de elevado costo en dinero y tiempo, en este sentido, sobre todo en empresas de nivel mundial, se ha dado un paso importante convirtiéndose en uno de los objetivos principales de las mismas.

Teniendo en cuenta los aspectos anteriores se estableció:

Trabajar con Oracle como sistema gestor de bases de datos y sus herramientas de BI para implementar una arquitectura donde se publique la información íntegra, limpia y

consistente para la toma de decisiones, a partir de los datos referentes a las infracciones del tránsito que se encuentran en sus bases de datos operacionales. La información se almacenará mediante cargas periódicas de datos y al ocurrir cambios en los sistemas OLTP.

Se tiene como **situación problemática** que en el MININT existen dificultades con la visualización de información sobre las infracciones del tránsito debido a diferentes problemas como son: la falta de recepción de algunos datos necesarios y dispersión de los que se recogen debido a la utilización de diferentes esquemas de bases de datos donde se guarda la información relacionada con el proceso, además de que la estructura de dichos esquemas no es la mejor para realizar consultas que involucran grandes volúmenes de datos, pues se afecta el rendimiento de los sistemas operacionales. A estas deficiencias se le añaden la falta de gráficas en los informes que muestren la información desde distintos ángulos del negocio unido a la carencia de una interfaz amigable al usuario.

El **problema** a resolver con este trabajo es el siguiente:

La falta de informes y reportes con gráficas que muestren una visión del negocio desde distintas perspectivas; tras la ausencia de una arquitectura de apoyo a la toma de decisiones en el área de infracciones del tránsito.

El **objeto de estudio** de la presente investigación es la información almacenada en las bases de datos que se genera en los procesos de gestión y control de infracciones en el MININT, delimitándose en el **campo de acción** de Tránsito perteneciente al Órgano de la Policía Nacional Revolucionaria del MININT.

Teniendo en cuenta lo anterior, se define como **objetivo general** de este trabajo el siguiente: Implementar una arquitectura de Inteligencia de Negocios para la toma de decisiones sobre las infracciones del tránsito en el MININT.

De este objetivo general se desprenden los siguientes **objetivos específicos**:

1. Realizar un estudio del proceso de control de infracciones que tiene lugar en el MININT.
2. Diseñar el Mercado de Datos que contendrá la información de las infracciones.
3. Implementar los procesos ETL que cargarán el Mercado de Datos.

4. Implementar una presentación web que muestre la información del Mercado de Datos.
5. Validar el sistema.

Para dar cumplimiento a estos objetivos se definieron las siguientes **tareas**:

1. Entrevista a funcionarios responsables del trabajo con las infracciones para identificar los principales procesos que se desarrollan.
2. Estudio de tecnologías para el desarrollo de un Mercado de Datos.
3. Estudio de propuestas de UML para el desarrollo de un Mercado de Datos.
4. Definición de los requerimientos de información que debe brindar el producto final.

Con la realización de este trabajo se defiende la siguiente **idea**:

Si se Implementa una arquitectura de Inteligencia de Negocios para la toma de decisiones sobre las infracciones del tránsito en el MININT se contará con una herramienta de apoyo a la toma de decisiones que permitirá analizar el comportamiento de las infracciones del tránsito.

El valor práctico de este proyecto radica en que al comenzar a utilizarse en el MININT será una herramienta que facilitará el trabajo de los directivos de esta institución al permitir generar gráficas y reportes sobre el comportamiento de las infracciones del tránsito mediante una interfaz amigable y sencilla. Además de establecer las bases para la realización de estudios futuros que realicen el descubrimiento de conocimiento oculto, que eleve la calidad de la información que sirve de apoyo a la toma de decisiones.

El presente trabajo, estructurado en 3 capítulos, resume la siguiente información:

Capítulo 1. Fundamentación teórica. Contiene un estudio del estado del arte donde se explican los conceptos fundamentales vinculados a los sistemas para el soporte a la toma de decisiones, además de las tendencias y tecnologías actuales seleccionadas para el desarrollo de este proyecto.

Capítulo 2. Descripción de la situación actual. Se describen los sistemas existentes que son utilizados actualmente por los funcionarios del MININT y el flujo de trabajo que se utiliza para el control de las infracciones.

Capítulo 3. Construcción de la solución propuesta. Se desarrolla la propuesta y se documentan el despliegue de la arquitectura mediante los componentes de software desarrollados.

Capítulo 1 – Fundamentación teórica

1.1 Introducción

En el presente capítulo se hace un análisis sobre las principales tendencias, tecnologías y herramientas que se deben tener en cuenta a la hora de desarrollar un producto de Inteligencia de Negocios.

1.2 Inteligencia de Negocio

Hoy en día la información es vital para obtener una ventaja competitiva en el mundo de los negocios. Para mantener competitiva una empresa, los dirigentes de esta requieren de un acceso rápido y fácil a información útil y valiosa de la organización.

Con los avances de las tecnologías de la informática y las comunicaciones, los datos pueden encontrarse dispersos en muchas fuentes. Integrar estas para generar información de manera que recoja todo aquello de importancia para el negocio y colocarlo de una manera sencilla para ser analizada y trabajada por los dirigentes es la labor de los sistemas de Inteligencia de Negocio (Business Intelligence o BI). La BI propicia que las organizaciones puedan traducir sus objetivos en indicadores de estudio, y que estos puedan ser analizados desde diferentes perspectivas, con el fin de encontrar información que no solo se encargue de responder a preguntas de lo que está sucediendo o ya sucedió, sino también, que posibilite la construcción de modelos, mediante los cuales se podrán predecir eventos futuros.

Business Intelligence es una amplia categoría de métodos, tecnologías y herramientas para coleccionar, almacenar, estructurar, analizar y acceder a los datos con el objeto de tomar mejores decisiones de negocio. Las aplicaciones de BI incluyen el soporte de actividades de decisión, consulta e informe, procesamiento analítico en línea (OLAP), análisis estadísticos, previsión y minería de datos. También se puede decir que BI es un proceso que permite acumular y analizar la información de bases de datos y almacenes de datos. [1]

Entre los beneficios más importantes que BI proporciona a las organizaciones, vale la pena destacar los siguientes:

- Reduce el tiempo mínimo que se requiere para recoger toda la información relevante de un tema en particular, ya que la misma se encontrará integrada en una fuente única de fácil acceso.
- Automatiza la asimilación de la información, debido a que la extracción y carga de los datos necesarios se realizará a través de procesos predefinidos.
- Proporciona herramientas de análisis para establecer comparaciones y tomar decisiones.
- Cierra el círculo que hace pasar de la decisión a la acción.
- Permite a los usuarios no depender de reportes o informes programados, porque los mismos serán generados de manera dinámica.
- Posibilita la formulación y respuesta de preguntas que son claves para el desempeño de la empresa.
- Permite acceder y analizar directamente los indicadores de éxito.
- Se pueden identificar cuáles son los factores que inciden en el buen o mal funcionamiento de la empresa.
- Se podrán detectar situaciones fuera de lo normal.
- Permitirá predecir el comportamiento futuro con un alto porcentaje de certeza, basado en el entendimiento del pasado.
- El usuario podrá consultar y analizar los datos de manera sencilla e intuitiva. [2]

Algunos de los elementos más importantes que se pueden encontrar en una arquitectura de BI son:

- Almacenes de Datos y Mercados de Datos
- OLAP (Procesamiento Analítico en Línea)

- Modelo Dimensional [3]

1.2.1 Sistemas de Soporte a la toma de Decisiones (DSS)

Los Sistemas de Soporte a la toma de Decisiones son herramientas de Inteligencia de Negocio enfocadas al análisis de los datos de una organización. En principio, puede parecer que el análisis de datos es un proceso sencillo, y fácil de conseguir mediante una aplicación hecha a medida. Sin embargo, no es así: la aplicaciones de los sistemas operacionales suelen disponer de una serie de informes predefinidos en los que presentan la información de manera estática, pero no permiten profundizar en los datos, navegar entre ellos, manejarlos desde distintas perspectivas, etc.

Entre las características principales de los DSS se encuentran las siguientes:

Informes dinámicos, flexibles e interactivos, de manera que el usuario no tenga que ceñirse a los listados predefinidos que se configuraron en el momento de la implantación, y que no siempre responden a sus dudas reales.

No requiere conocimientos técnicos. Un usuario no técnico puede crear nuevos gráficos e informes y navegar entre ellos. Por tanto, para examinar la información disponible o crear nuevas métricas no es imprescindible buscar auxilio en el departamento de informática.

Rapidez en el tiempo de respuesta, ya que la base de datos subyacente suele ser un almacén o un mercado de datos corporativo, con modelos de datos en estrella o copo de nieve. Este tipo de bases de datos están optimizadas para el análisis de grandes volúmenes de información.

Integración entre los sistemas/departamentos de la compañía. El proceso ETL previo a la implantación de un Sistema de Soporte a la Decisión garantiza la calidad y la integración de los datos entre las diferentes unidades de la empresa.

Cada usuario dispone de información adecuada a su perfil. No se trata de que todo el mundo tenga acceso a toda la información, sino de que tenga acceso a la información que necesita para que su trabajo sea lo más eficiente posible.

Disponibilidad de información histórica. En estos sistemas está a la orden del día comparar los datos actuales con información de otros períodos históricos de la compañía, con el fin de analizar tendencias, fijar la evolución de parámetros de negocio, etc. [4]

1.2.2 Procesamiento de Transacciones en Línea y Procesamiento Analítico en Línea

En dependencia de las funciones que realizan los sistemas en una empresa se pueden clasificar en sistemas de Procesamiento de Transacciones en Línea (OLTP) y sistemas de Procesamiento Analítico en Línea (OLAP).

Los sistemas OLTP se encuentran en la mayoría de las organizaciones y son aplicaciones que ejecutan y respaldan las operaciones diarias de la empresa, como compras, ventas e inventarios, mediante consultas y actualizaciones a las bases de datos.

Algunas de las características más comunes de este tipo de transacciones son:

- Realizan altas, bajas o modificaciones
- Consultas rápidas, escuetas y predecibles
- Poco volumen de información
- Transacciones rápidas
- Baja redundancia de datos [5]

El OLAP sustenta el estudio del comportamiento del negocio. Permite el análisis de los datos consolidados de la empresa (mediante operaciones de consulta solamente) y ayuda al usuario a sintetizar la información de la empresa a través de vistas personalizadas, análisis históricos y pronósticos.

Son aplicaciones que se encargan de analizar datos del negocio para generar información táctica y estratégica que sirve de soporte para la toma de decisiones. Mientras que las transacciones OLTP utilizan Bases de Datos Relacionales u otro

tipo de archivos, OLAP logra su máxima eficiencia y flexibilidad operando sobre Bases de datos Multidimensionales.

Podemos nombrar las siguientes características como las más sobresalientes de estas aplicaciones:

- Estructura de datos transparente al usuario
- Solo consulta. Trabajan sobre la información operacional generada por los sistemas OLTP
- Consultas sobre grandes volúmenes de datos no predecibles
- Información histórica
- Alta redundancia de datos para facilitar la generación de consultas y obtener buenos tiempos de respuesta
- Trabaja con resúmenes de miles de registros condensados en una sola respuesta

Mientras que las aplicaciones OLTP se caracterizan por estar actualizadas constantemente por varios usuarios a través de transacciones operacionales sobre datos individuales, las aplicaciones OLAP son utilizadas por personal de niveles ejecutivos que requieren datos con alto grado de agregación y desde distintas perspectivas (dimensiones). [5]

Los sistemas operacionales tienen interfaces muy especializadas para un usuario que realiza una operación rutinaria y los usuarios estratégicos realizan consultas variadas, por lo que los sistemas deben ser sencillos y con toda la información disponible que cubra cualquier consulta requerida, por lo que el software final está orientado a un usuario en particular y, por ende, debe adecuarse al conocimiento que tenga sobre el tema.

Las preguntas que responde un sistema operacional son referentes a las transacciones que se realizan diariamente y a nivel registro o suma de registros de un solo tipo. Un usuario de un sistema OLTP realiza frecuentemente preguntas sobre registros como pueden ser el estado actual de una factura, movimientos de un cliente, cantidad surtida por un proveedor, fecha del último movimiento de un distribuidor, etc. Las preguntas de un ejecutivo pueden también ser específicas,

pero se orientan más a agrupamientos de datos como pueden ser totales por zona, promedios de clientes, tendencias de ventas e incluso pronósticos. Toda esta información se encuentra de alguna forma en las bases de datos de los sistemas operacionales, pero lanzar una consulta como las ventas totales del año anterior puede implicar hasta días en resolverse y otro tiempo para publicar los datos.

Los sistemas OLAP deben cumplir las características de la prueba FASMI (*Análisis Rápido de Información Multidimensional Compartida*) sin importar la forma en que fueron implementados.

A continuación se explica el significado de los términos de FASMI:

- **Rápido:** El sistema debe proporcionar la información al usuario a una velocidad constante. La mayoría de las peticiones deben responderse en segundos.
- **Análisis:** El sistema debe realizar análisis estadísticos y numéricos básicos de los datos ya sean, predefinidos por el desarrollador de la aplicación a través de una gran población de usuarios.
- **Compartida:** El sistema debe implementar los requerimientos de seguridad necesarios para compartir datos potencialmente confidenciales a través de una gran población de usuarios.
- **Multidimensional:** El sistema debe permitir trabajar con múltiples dimensiones de los datos que es la característica esencial del OLAP.
- **Información:** El sistema debe brindar acceso a todos los datos, así como a la información necesaria y relevante para la aplicación, independientemente y de su ubicación y su volumen. [6]

1.2.3 Almacén de Datos

“Un Almacén de Datos es una colección de datos integrada, orientada a temas, variante en el tiempo y no volátil, utilizada como apoyo para los procesos de toma de decisión”. [7]

Estas características significan:

- Integrada: Contiene una base de datos centralizada y consolidada que integra datos derivados de toda la organización. Los datos se almacenan en un formato consistente y existe un único esquema de representación.
- Orientada a temas: los datos se organizan y se resumen por temas lo cual facilita su acceso y entendimiento por parte de los usuarios finales. Un AD se enfoca a las entidades del negocio mientras los sistemas operacionales se orientan a los procesos.
- Variante en el tiempo: los datos almacenados se asocian a un instante o período de tiempo, quedando almacenada información histórica de los datos.
- No volátil: Una vez almacenados los datos no son modificados.

Desarrollar un AD significa integrar los datos en un único depósito desde el cual los usuarios finales puedan fácilmente ejecutar consultas, confeccionar reportes y realizar análisis. Este refuerza los datos almacenados en diferentes fuentes, organizándolos y poniéndolos a disposición de los encargados de tomar decisiones en la empresa, sin importar su nivel de habilidad técnica o las plataformas de trabajo.

La siguiente tabla muestra una comparación entre las bases de datos operacionales y los almacenes de datos.

Tabla 1: Diferencias entre bases de datos operacionales y almacenes de datos

Parámetros	Base de Datos Operacional	Almacén de Datos
Propósito	Soportar operaciones diarias	Recuperar información
Características de los datos	Datos actuales, cambiantes	Datos estables, históricos
Redundancia de datos	No	Si
Usuarios concurrentes	Muchos	Pocos
Operación predominante	Actualización	Consulta
Tiempo de respuesta	Critico	No crítico
Estructura de base de datos	Complejas	Simple

Entre las principales ventajas que reporta el uso de los Almacenes de Datos se encuentran las siguientes:

- Menor coste en la toma de decisiones en términos de tiempo y fiabilidad de la información obtenida.
- Mayor flexibilidad a la hora de realizar análisis.
- Crea una arquitectura de datos única para todas las aplicaciones analíticas.
- Resuelve los problemas de integridad y calidad de datos.
- Permite la toma de decisiones estratégicas y tácticas.
- Logra un impacto positivo sobre los procesos de toma de decisiones ya que cuando los usuarios tienen acceso a una mejor calidad de información, la empresa puede aprovechar el enorme valor potencial de sus recursos de información y transformarlo en valor verdadero; permitir al usuario adquirir mayor confianza acerca de sus propias decisiones y de las del resto, y lograr así, un mayor entendimiento de los impactos ocasionados.

- Mejora la entrega de información, es decir, información completa, correcta, consistente, oportuna y accesible. Información que los usuarios necesitan, en el momento adecuado y en el formato apropiado.
- Transforma datos orientados a las aplicaciones en información orientada a la toma de decisiones.
- Elimina la producción y el procesamiento de datos que no son utilizados ni necesarios, producto de aplicaciones mal diseñadas o que ya no son utilizadas.

Algunas desventajas son:

- Alta complejidad.
- La puesta en funcionamiento requiere de un largo plazo.
- Coste elevado. Requiere una gran inversión, debido a que su correcta construcción no es tarea sencilla y consume muchos recursos, además, su misma implementación implica desde la adquisición de herramientas de consulta y análisis, hasta la capacitación de los usuarios.
- Existe resistencia al cambio por parte de los usuarios.
- Los beneficios del almacén de datos son apreciados a mediano y largo plazo. Este punto deriva del anterior, y básicamente se refiere a que no todos los usuarios confiarán en el AD en una primera instancia, pero sí lo harán una vez que comprueben su efectividad y ventajas. Además, su correcta utilización surge de la propia experiencia.
- Si se incluyen datos propios y confidenciales de clientes, proveedores, entre otros, el depósito de datos atentará contra la privacidad de los mismos, ya que cualquier usuario podrá tener acceso a ellos.
- Infravaloración del esfuerzo necesario para su diseño y creación.
- Incremento continuo de los requerimientos del usuario.
- Subestimación de las capacidades que puede brindar la correcta utilización del AD y de las herramientas de BI en general.

1.2.4 Mercado de Datos

Es un Almacén de Datos centrado en un departamento o un área específica. Tiene las mismas características de un AD y brinda las mismas facilidades, pero no está orientado a satisfacer las necesidades de toda la empresa. Se utilizan debido a que el acceso a todos los datos a veces no es conveniente, o necesario, para determinados usuarios que solo necesitan un subconjunto de estos. En muchos casos, los AD comienzan siendo MD con el objetivo de minimizar los riesgos para luego ir ampliando su alcance gradualmente.

Algunas de las ventajas de la construcción de un Mercado de Datos son:

- Almacenar pequeños conjuntos de datos y en consecuencia, menor necesidad de recursos.
- Encontrarse más rápidamente las necesidades de los departamentos.
- Permitir consultas más rápidas debido al menor volumen de datos que maneja.
- Dividir los datos para imponer estrategias de control de acceso.
- Comparado con un AD se implementan más fácilmente, además de que son menores el tiempo de construcción y de puesta en marcha.
- Reducen la demanda del depósito de datos.

1.2.5 Procesos de Extracción, Transformación y Carga (ETL)

Uno de las tareas fundamentales en la creación y carga de Almacén de Datos son los llamados procesos ETL. Son los procesos que se encargan de mover datos desde múltiples fuentes, reformatearlos, limpiarlos y cargarlos en el mercado o almacén de datos. Los procesos ETL pueden hacer una copia idéntica o una transformación de los orígenes, dependiendo de su propósito.

Para que tengan éxito estos procesos se deben analizar los datos corporativos para descubrir inexactitudes, anomalías y otros problemas para eliminarlas a la hora de la integración.

Es necesaria la limpieza y transformación de los datos para evitar redundancia e inconsistencia, estandarizar medidas, formatos, fechas, tratar valores nulos, etc.

Existen numerosos desafíos para implementar unos procesos ETL eficaces y fiables como son los siguientes:

- Los volúmenes de datos crecen de forma exponencial, y los procesos ETL tienen que procesar grandes cantidades de datos granulares (productos vendidos, llamadas telefónicas, transacciones bancarias, entre otros). Algunos sistemas de BI se actualizan simplemente de manera incremental, mientras que otros requieren una recarga completa en cada iteración.
- A medida que los sistemas de información crecen en complejidad, también aumenta la disparidad de las fuentes. Los procesos ETL necesitan una extensa conectividad a las aplicaciones, bases de datos, archivos, servicios Web, etc.
- Las transformaciones implicadas en los procesos ETL pueden ser muy complejas. Los datos necesitan agregarse, analizarse, procesarse estadísticamente, etc.

1.2.6 Modelo Dimensional

El modelo más usado en la implementación de Almacenes de Datos es el Dimensional el cual no tiene en cuenta la tercera forma normal a diferencia de los sistemas OLTP. La forma en que este empaqueta los datos busca facilitar la comprensión por parte del usuario, además de la ejecución rápida y eficiente de las consultas.

Este divide los datos en dos grupos:

Medidas, generalmente numéricas que son almacenadas en las **Tablas de Hechos**.

Descripciones del entorno de estas medidas, que son almacenadas en las **Tablas de Dimensiones**.

Los esquemas más usados en el diseño de Almacenes o Mercados de Datos son:

Esquema de Estrella: Es el más sencillo de los usados en el modelado dimensional. Se llama así debido a que su diagrama se asemeja a una estrella. El centro de la estrella consiste de una o más Tablas de Hechos que pueden estar conectadas a una o varias Tablas de Dimensiones.

Esquema de Copo de Nieve: Un tipo de Esquema de Estrella cuyas tablas de dimensión se encuentran normalizadas de forma parcial o completa. [8]

2.1 Metodología para el Desarrollo de un Almacén de Datos.

En la actualidad existen varias propuestas de modelos y métodos para llevar a cabo el diseño de almacenes de datos aunque ninguna de estas propuestas ha sido aceptada como un modelo estándar para realizar el modelado multidimensional, pues no cubren todas las fases y transformaciones necesarias. Debido a esta gran variedad de modelos utilizados en las fases de diseño de los AD se desarrolló un método que proporciona guías de diseño para crear y transformar estos modelos durante la fase de desarrollo del almacén datos, el *Data Warehouse Engineering Process (DWEPE)*, propuesto en la tesis de Sergio Luján Mora [9], el mismo es un método orientado a objetos, independiente de cualquier implementación específica, ya sea relacional, multidimensional, orientado a objetos, etc. permite la representación de todas las etapas del diseño de un AD. Está basada en UML (Lenguaje Unificado de Modelado) y RUP (Proceso Unificado de Desarrollo de Software).

Como el DWEPE es una instancia de RUP para el desarrollo de almacenes de datos establece al igual que este, el ciclo de vida de un proyecto en cuatro fases: Inicio, Elaboración, Construcción y Transición, así como cinco flujos de trabajo fundamentales: Requerimientos, Análisis y Diseño, Implementación y Prueba, además adiciona dos nuevas actividades: Mantenimiento y Revisión Post-Desarrollo. [9]

2.1.1 Diseño de un AD.

El desarrollo de un AD se puede estructurar en un marco integrado por cinco etapas y tres niveles que definen los diferentes diagramas empleados para modelarlo.

Etapas del proceso:

- Origen: se definen las fuentes operacionales de datos de las que se va a nutrir el AD. Ejemplo: sistemas OLTP, ficheros, etc.

- Integración: se definen los procesos de mapeo de datos entre las fuentes de origen y el almacén de datos (Procesos ETL).
- Almacén de Datos: se realiza el diseño de la estructura del almacén de datos.
- Adaptación: se definen los procesos de mapeo entre el almacén de datos y las estructuras utilizadas por el cliente.

En cada una de las etapas descritas anteriormente se analizan tres niveles:

- Conceptual: se definen las estructuras correspondientes a la etapa desde el punto de vista conceptual, o sea, con un mayor nivel de abstracción, conteniendo solo los elementos más significativos.
- Lógico: incluye los aspectos lógicos del diseño de las estructuras correspondientes a la etapa. Ejemplo: definición de atributos, llaves primarias, procesos ETL, etc.
- Físico: abarca todos los aspectos referentes a la implementación física de las estructuras analizadas y diseñadas en los niveles anteriores. [9]

En los diferentes niveles se diseñan diferentes tipos de diagramas que se basan en una extensión de UML adaptada al dominio específico de diseño de almacenes de datos.

2.1.2 Modelo Conceptual

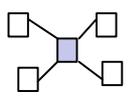
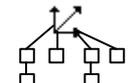
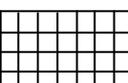
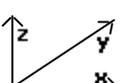
En esta etapa el proceso de modelado se divide en tres niveles:

Nivel 1: Se realiza a partir de la representación de esquemas de estrella como paquetes y relaciones. Las relaciones indican que los esquemas implicados comparten al menos una dimensión. Debe seleccionarse adecuadamente el paquete principal con el objetivo de minimizar las dependencias y evitar ciclos.

Nivel 2: Se define la estructura de cada una de las estrellas. Un paquete representa un hecho o una dimensión de un esquema estrella. Las relaciones entre dimensiones implican que las dimensiones comparten, al menos, un nivel de sus jerarquías (esquema copo de nieve).

Nivel 3: Se define la estructura de todos los hechos y dimensiones. Se representan empleando un conjunto de clases referentes a cada nivel jerárquico en las dimensiones y en el caso de los hechos mediante el esquema de estrella completo. [9]

Tabla 2: Estereotipos empleados en el modelo conceptual del AD. [9]

Estereotipo	Descripción	Icono
StarPackage	Paquete que representa el esquema de estrella, empleado en el nivel 1 para diseñar el modelo conceptual global del AD.	
FactPackage	Paquete que representa los hechos, empleado en el nivel 2 para diseñar la estructura de cada uno de los esquemas de estrella.	
DimensionPackage	Paquete que representa las dimensiones, empleado en el nivel 2 para diseñar la estructura de cada uno de los esquemas de estrella.	
Fact	Clase que representa los hechos, empleado en el nivel 3 para diseñar la estructura de cada uno de los paquetes de hechos.	
Dimension	Clase que representa las dimensiones, empleado en el nivel 3 para diseñar la estructura de cada uno de los paquetes de dimensiones.	
Base	Clase base que representa los niveles jerárquicos en cada dimensión, empleado en el nivel 3.	

2.1.3 Modelo Lógico

Como su nombre lo indica es una continuación lógica de los diagramas realizados en el Modelo Conceptual. Se realiza un mapeo entre los artefactos desarrollados en el modelo anterior y los correspondientes a este. Es decir, a partir de las clases y relaciones se obtienen tablas, columnas, tipos de datos, llaves, índices,

dependencias, relaciones y vistas. El proceso de mapeo no se realiza a todos los elementos representados de forma conceptual, solo a los que lógicamente pueden ser representados en la base de datos. Ejemplo: los atributos calculables no se incluyen en la base de datos. En este modelo también se incluye la definición de los procesos ETL.

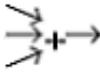
Los procesos ETL son los responsables de la extracción de los datos de diferentes fuentes operacionales, de su transformación (conversión, normalización, limpieza, etc.) y su carga en el AD.

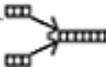
Para poder realizar estos procesos hay que tener en cuenta una serie de aspectos. Estos son:

1. Selección de las fuentes de datos de extracción.
2. Realización de transformaciones: se analiza la calidad de los datos provenientes de las fuentes. A continuación se transforman los datos para ajustarlos a las necesidades y reglas establecidas para el sistema. Entre las principales transformaciones están: filtrado, conversión de tipos de datos, calcular valores derivados, generación de secuencias, etc.
3. Unir las fuentes de datos que están relacionadas a través de uno o más atributos. Lograr unificar los datos que posteriormente serán cargados.
4. Seleccionar el destino de los datos. Pueden ser varios.
5. Realizar el mapeo de los atributos de los datos fuentes hacia los atributos de la estructura destino.
6. Cargar los datos. [9]

La metodología de diseño de almacenes de datos DWEP propone una extensión de UML que permite el modelado de los procesos ETL. Para ello se ha definido el siguiente conjunto de estereotipos:

Tabla 3: Estereotipos empleados en el modelo de los procesos ETL. [9]

Estereotipo	Descripción	Icono
Aggregation	Agrega los datos en base a algún criterio. (SUM, AVG, MAX/MIN, COUNT, etc.)	

Conversion	Genera atributos derivados a partir de datos existentes: cambio de formato o tipo de datos, datos calculables, etc.	A→B
Filter	Filtra los datos a partir de restricciones.	
Incorrect	Redirige los registros incorrectos a un destino separado para someterlos a transformaciones.	
Join	Une dos fuentes de datos que están relacionadas por sus atributos.	
Loader	Carga los datos en las estructuras del AD.	
Merge	Integra los datos de varias fuentes con atributos compatibles.	
Surrogate	Genera una clave sustituta.	123→
Log	Registra las operaciones realizadas en el proceso ETL. Se utiliza fundamentalmente para monitorear las transformaciones.	

2.1.4 Modelo Físico

DWEP propone emplear los diagramas de componentes y de despliegue de UML para modelar el nivel físico del AD.

La modelación del diseño físico de un AD en las primeras etapas del proyecto permite tratar con importantes aspectos de la implementación antes de comenzar este proceso y por ende permite reducir el tiempo total desarrollo del proyecto.

2.2 Sistemas Gestores de bases de datos que dan soporte a BI

Los sistemas gestores de bases de datos dan soporte a los sistemas OLTP para sustentar las transacciones diarias de las empresas y están perfeccionándose continuamente, pero no todos han dado soporte a los sistemas

OLAP para la exploración de datos y la extracción de conocimiento. En este epígrafe se hace un estudio de algunos de los gestores más utilizados hoy en día.

2.2.1 Microsoft SQL Server 2005

Microsoft SQL Server 2005 es una plataforma de base de datos que se utiliza en el procesamiento de transacciones en línea (OLTP) a gran escala, el almacenamiento de datos y las aplicaciones de comercio electrónico; es también una plataforma de Business Intelligence para soluciones de integración, análisis y creación de informes de datos.

Brinda la herramienta Microsoft SQL Server 2005 **Analysis Services** (SSAS), la cual permite llevar a cabo un rápido análisis de datos empresariales y proporciona funciones de procesamiento analítico en línea (OLAP) y de minería de datos en aplicaciones de Business Intelligence. Esta combina el análisis tradicional basado en OLAP y la elaboración de informes basada en relaciones al permitir a los programadores definir un único modelo de datos, denominado Unified Dimensional Model (UDM), a partir de uno o más orígenes de datos físicos. Todas las consultas de usuario final desde aplicaciones OLAP, de elaboración de informes y de BI personalizadas obtienen acceso a los orígenes de datos subyacentes a través del modelo UDM, que proporciona una única vista empresarial de estos datos relacionales.

Microsoft SQL Server 2005 brinda una plataforma para crear soluciones de integración de datos de alto rendimiento que se llama Microsoft SQL Server 2005 **Integration Services** (SSIS) que sustituye a los Servicios de transformación de datos (DTS) del SQL Server anterior. Permite resolver complejos problemas empresariales mediante la copia o descarga de archivos, el envío de mensajes de correo electrónico como respuesta a eventos, la actualización de almacenes de datos, la limpieza y la administración de objetos y datos de SQL Server. Los paquetes pueden funcionar por separado o conjuntamente con otros paquetes para hacer frente a las complejas necesidades de la empresa. Integration Services puede extraer y transformar datos de muchos orígenes distintos, como archivos de

datos XML, archivos planos y orígenes de datos relacionales, y, posteriormente, cargarlos en uno o varios destinos.

En versiones anteriores, DTS sólo proporcionaba un conjunto limitado de tareas y transformaciones. Para el programador resultaba difícil crear paquetes con flujos de trabajo complejos y repetitivos, así como aplicar diferentes tipos de transformaciones de nivel de columna a los datos. El diseñador gráfico que se utilizaba para crear paquetes se combinaba con el flujo de trabajo y el flujo de datos en una única superficie de diseño y ofrecía unas opciones de control de flujo limitadas. Integration Services resuelve muchas de las dificultades y limitaciones de DTS. Las mejoras de SSIS incluyen una nueva arquitectura extensible, un nuevo diseñador de paquetes y multitud de nuevas tareas, estructuras de bucle y transformaciones, así como mejoras en la implementación, administración y el rendimiento de los paquetes.

Integration Services contiene un variado conjunto de tareas y transformaciones integradas, herramientas para la creación de paquetes y el servicio Integration Services para ejecutar y administrar los paquetes. Las herramientas gráficas de Integration Services se pueden usar para crear soluciones sin escribir una sola línea de código. También se puede programar el amplio modelo de objetos de Integration Services para crear paquetes mediante programación y codificar tareas personalizadas y otros objetos de paquete.

También cuenta con SQL Server 2005 **Reporting Services** (SSRS) la cual es una plataforma de creación de informes basada en servidor que ofrece una extensa gama de informes de datos de orígenes de datos relacionales y multidimensionales. Reporting Services contiene componentes de procesamiento, un completo conjunto de herramientas que sirven para crear y administrar informes, y una interfaz de programación de aplicaciones (API) con la que los programadores pueden integrar o ampliar el procesamiento de datos e informes en aplicaciones personalizadas.

Con Reporting Services, se pueden crear informes interactivos, tabulares o de formato libre que recuperan los datos a intervalos programados o a petición, cuando el usuario abre un informe. Los usuarios también pueden crear informes basados en modelos predefinidos y examinar de manera interactiva los datos del modelo. Todos los informes se pueden representar con formatos de escritorio o basados en Web. Se puede elegir entre diversos formatos de visualización para representar los informes a petición en los formatos preferidos para la manipulación o impresión de datos.

Reporting Services es una solución basada en servidor y, por lo tanto, proporciona un modo de centralizar el almacenamiento y la administración de informes, proporcionar acceso seguro a informes, modelos y carpetas, controlar el procesamiento y la distribución de informes, y normalizar el uso de los informes en la empresa. [10]

2.2.2 Microsoft SQL Server 2008

Esta versión de Microsoft SQL Server presenta nuevas características y mejoras en sus productos que aumentan la calidad de sus servicios.

La plataforma **Analysis Services** presenta mejoras en el diseño de agregaciones que hacen que sea más fácil examinar y modificar los diseños de agregaciones que se muestran agrupados por grupo de medida. Nuevos asistentes que permiten modificar a la vez las configuraciones de almacenamiento para las agregaciones en una o varias particiones y establecer con facilidad los valores de uso de las agregaciones, además de poder adjuntar nuevas agregaciones a una agregación existente. También cuenta con nuevos mensajes de advertencia que alertan a los usuarios cuando se desvían de las prácticas recomendadas del diseño de agregaciones.

Se optimiza el diseño de cubos al simplificar y mejorar el asistente para cubos posibilitando crear mejores cubos en menos pasos. Para el diseño de dimensiones aparece un nuevo diseñador de relaciones de atributos que permite examinar y

modificar las relaciones de atributos con mayor facilidad, nuevas advertencias que avisan a los usuarios cuando se desvían de los procedimientos de diseño recomendados o cometen errores lógicos en el diseño de bases de datos, asistente para dimensiones simplificado y mejorado que detecta automáticamente jerarquías de elementos primarios y secundarios, proporcionando una configuración de error predeterminada más segura, permite modificar atributos y jerarquías con mayor facilidad.

La funcionalidad de copia de seguridad y restauración de Analysis Services tiene una nueva estructura de almacenamiento y un rendimiento mejorado en todos los escenarios de copia de seguridad y restauración.

Cuenta con una nueva estructura de almacenamiento que proporciona un repositorio más robusto para la base de datos archivada con la cual no hay ningún límite práctico para el tamaño del archivo de base de datos, ni tampoco existe límite para el número de archivos que puede tener una base de datos.

Las extensiones de personalización de Analysis Services permiten a los programadores crear nuevos objetos y funcionalidades de Analysis Services, y proporcionarlos dinámicamente en el contexto de la sesión de usuario. Los programadores no tienen que crear especificaciones detalladas sobre dónde o cómo buscar la funcionalidad extendida. En su lugar, los programadores pueden compartir inmediatamente estos nuevos objetos y funcionalidades tanto con usuarios finales como con otros programadores.

Integration Services presenta nuevas características y mejoras para la instalación, componentes, administración de datos, rendimiento y solución de problemas. Estas características nuevas y mejoras aumentan la eficacia y productividad de los programadores, los administradores y los trabajadores del conocimiento que desarrollan soluciones de transformación de datos. Las mejoras en el rendimiento de la transformación Búsqueda incluyen una carga en memoria caché más rápida y operaciones de búsqueda más eficaces.

Microsoft SQL Server 2008 Reporting Services (SSRS) presenta numerosas características y nuevas mejoras que aumentan las capacidades de generación de informes de los desarrolladores de soluciones de informes en los siguientes aspectos:

- Creación de informes
- Procesamiento y la representación de informes
- Herramientas y la arquitectura del servidor
- Programación de informes [11]

2.2.3 Oracle

Oracle Ofrece una administración eficiente, confiable y segura de los datos para todo tipo de aplicaciones, desde sistemas de alto volumen de transacciones en línea hasta aplicaciones de consulta intensiva. No sólo soporta las necesidades de la compleja administración de datos, sino que también brinda las herramientas para administrar los sistemas, ofrece la flexibilidad para distribuir efectiva y eficientemente los datos a los usuarios y la escalabilidad para alcanzar el rendimiento óptimo de todos los recursos de computación disponibles.

Oracle Business Intelligence Suite Enterprise Edition Plus (OBIEE) es una completa suite de productos BI de la empresa, entrega toda la gama de capacidades de BI incluyendo tableros interactivos, alertas, presentación de informes financieros y empresariales, inteligencia predictiva en tiempo real, y mucho más. Además de proporcionar la gama completa de funcionalidad de BI, la plataforma Oracle Business Intelligence Suite EE se basa en un resultado moderno de Arquitectura Web Orientada a Servicios que ofrece la próxima generación de verdadera capacidad de BI.

Algunos de los componentes de OBIEE son:

- Oracle BI Server
- Oracle BI Administration Tool
- Oracle BI Answers
- Oracle BI Interactive Dashboards

Oracle Business Intelligence Server: Un verdadero servidor de BI diseñado para ser altamente escalable, optimizando concurrencia y paralelismo para que las aplicaciones de BI estén disponibles para la mayor cantidad de audiencia posible. Proporciona acceso a los datos centralizados y el cálculo, en lo esencial la creación de una gran tubería a través de la cual cualquier persona puede consumir toda la información en cualquier forma en cualquier parte de la empresa. El servidor de BI es fundamental para todos los procesos empresariales que consumen información, incluidos los tableros, las consultas ad hoc, la capacidad de interacción inteligente, el espíritu empresarial y la producción de informes, presentación de informes financieros, análisis OLAP, minería de datos, y otros servicios Web. Todas estas aplicaciones requieren el acceso a la rica y amplia serie de datos a través de la empresa, y todos ellos requieren la sofisticada infraestructura de cálculo y de agregación que ofrece la plataforma para entregar valor.

La plataforma soporta un completo complemento de acceso, análisis, información y opciones de entrega, todo ello en un entorno web totalmente integrado. Cada uno de estos componentes sirve a diferentes públicos en la organización que tienen distintos intereses con respecto a los mismos datos y que necesitan acceder a ellos de diferentes maneras. Pero, a diferencia de otras herramientas de BI, todos los componentes están integrados en una arquitectura común, sin fisuras y que permite una experiencia de usuario intuitiva.

Oracle Business Intelligence Administration Tool

Esta herramienta permite la conexión a una o varias fuentes de datos para manipular su información y prepararla para darle sentido de forma que sea comprensible para el usuario. Su objetivo es crear un repositorio que organiza el paso de los datos por tres capas:

La Capa Física. Define las fuentes de datos a las que el Oracle BI Server presenta las consultas y las relaciones físicas entre bases de datos y otras fuentes de datos

que son usadas para procesar múltiples consultas. Las fuentes de datos pueden ser de la misma o diferentes variedades. Puede importar esquemas o parte de ellos desde fuentes de datos existentes. Además, se pueden crear objetos en la capa física manualmente.

Al importar, muchas de las propiedades de las fuentes de datos se configuran automáticamente con base en la información recopilada durante el proceso de importación. Después de la importación, también se pueden definir otras características físicas de las fuentes de datos, tales como relaciones de unión, que tal vez no existen en la fuente de datos. No puede haber una o más fuentes de datos en la capa física, incluyendo bases de datos, hojas de cálculo y documentos XML.

La Capa Modelo del Negocio y de Mapeo. En esta se define el negocio o lógica, el modelo de los datos y especifica el mapeo entre los esquemas del modelo de negocio y la capa física. Aquí es donde los esquemas físicos son simplificados para formar la base de la presentación de los datos que verán los usuarios. La Capa Modelo del Negocio y de Mapeo del Administration Tool puede contener uno o más objetos de modelo de negocio. Un objeto del modelo de negocio contiene las definiciones y asignaciones del modelo de negocio y mapeos de tablas lógicas a físicas.

El propósito principal del modelo de negocio es captar como los usuarios piensan de su negocio utilizando su propio vocabulario. El modelo del negocio simplifica el esquema físico. La mayoría del vocabulario se traduce en columnas lógicas en el modelo de negocio. Colecciones de columnas forman tablas lógicas.

Capa de presentación. La capa de presentación se construye a partir de las capas física y Modelo de Negocios y de Mapeo. Esta añade un nivel de abstracción sobre la capa Modelo del Negocio y de Mapeo. Es el punto de vista de los datos vistos por los usuarios finales en herramientas de cliente y aplicaciones, como Oracle BI Answers. La Capa de Presentación proporciona un medio para simplificar aún más o personalizar la Capa Modelo del Negocio y de Mapeo para

los usuarios finales. La simplificación de la vista de los datos para los usuarios hace que sea más fácil hacer consultas en base a las necesidades de los usuarios porque se pueden exponer sólo los datos que son significativos para los usuarios, organizarlos en una forma relacionada con el modo en que los usuarios piensan acerca de ellos, cambiar el nombre de los datos necesarios para el conjunto de usuarios. [12]

Oracle Business Intelligence Dashboards

Oracle BI Dashboards también llamado Oracle BI Interactive Dashboards ofrece a cualquier trabajador con el conocimiento intuitivo, acceso interactivo a información que es dinámicamente personalizada en función de la persona y la función de la identidad. En el ambiente Oracle BI Dashboards, el usuario final está trabajando con reportajes en vivo, anuncios, gráficos, tablas, gráficos, símbolos y en una pura arquitectura web. El usuario tiene plena capacidad para profundizar, navegar, modificar e interactuar con estos resultados. Oracle BI Dashboards también puede agregar contenido de una amplia variedad de fuentes, incluida la Internet, servidores de archivos compartidos, y repositorios de documentos. [13]

¿Qué es un Business Intelligence Dashboard?

El término significa tablero de control de BI y es una pantalla de visualización de la información más importante, necesaria para la realización de uno o más objetivos, consolidados y organizados en una sola presentación de modo que la información se pueda controlar con un vistazo. Suele contener un resumen de alto nivel para que los usuarios puedan entender rápidamente las grandes tendencias que afectan a la organización, con la capacidad de profundizar a mayor nivel de detalle, según sea necesario. Es importante que el tablero muestre la información de forma clara, concisa, de manera intuitiva y que la visualización de la información se pueda personalizar para los requisitos de un usuario particular. [14]

Oracle Business Intelligence Answers

Oracle BI proporciona capacidades especiales de respuestas orientadas al usuario final en la web. Los usuarios interactúan con una vista lógica de la información - completamente oculta la complejidad de la estructura de datos- y puede crear

fácilmente diagramas, informes y tableros totalmente interactivos y profundizables que pueden ser guardados y compartidos, todo ello en una ambiente visualmente atractivo. Los resultados son nuevos niveles de autosuficiencia para el usuario de negocios, en un ambiente completamente seguro. [15]

2.2.4 ¿Por qué Oracle?

Teniendo en cuenta las características de los Gestores de Bases de Datos anteriormente expuestos, se utilizará Oracle debido a la probada eficiencia en el manejo de grandes volúmenes de datos, pues ha demostrado ser un gestor estable al cual se le han corregido sus errores. El mismo es capaz de manejar eficientemente todo el volumen de información de nuestro Mercado de Datos y brindarnos todo el soporte necesario para su construcción, mantenimiento y visualización de la información al proveer una suite integrada de productos de BI de alta calidad

Además, independientemente de todas las potencialidades que reúne Oracle, es necesario destacar que se ha utilizado en el Ministerio del Interior por más de 10 años y que toda la información referente a las infracciones del tránsito se encuentra almacenada sobre servidores de bases de datos Oracle, además de que constituyó un requerimiento del cliente el uso de éste como sistema gestor de base de datos.

2.3 Oracle SQL Developer

Oracle SQL Developer es un producto de libre acceso de Oracle para el manejo de base de datos. El mismo posee una potente interfaz gráfica, la cual hace fácil el trabajo de manipular datos, como así también navegar y administrar objetos de bases de datos, todo lo cual es manejado con un mínimo de esfuerzo.

Oracle SQL Developer no sólo maneja base de datos de Oracle, sino que además posee la capacidad de conectar y manejar MySQL, Microsoft SQL Server y hasta bases de datos en Microsoft Access. Pero su fortaleza está en las bases de datos de Oracle junto con la edición, ejecución y corrección de código en PL/SQL. [16]

2.4 Conclusiones

Contar con una arquitectura de BI que permita mostrar el comportamiento de las infracciones del tránsito para apoyar el proceso de toma de decisiones es importante y necesario, por lo que después de realizar un estudio de las tendencias, metodologías y tecnologías actuales se propone una arquitectura de BI compuesta de un Mercado de Datos y una herramienta web para mostrar su contenido.

Las tecnologías seleccionadas para su implementación son:

- Oracle como sistema gestor de bases de datos.
- Oracle Business Intelligence Suite Enterprise Edition Plus como solución integradora de las herramientas necesarias para implementar la arquitectura de Inteligencia de Negocio.

La metodología que se utilizará para la documentación del sistema será DWEP, basada en la metodología RUP y en la notación UML por las ventajas antes expuestas.

Capítulo 2 – Descripción de la Situación Actual

2.1 Introducción

En el presente capítulo se describe como se realiza el proceso de control de infracciones en el MININT, además de caracterizar los sistemas existentes para este fin que son utilizados en esta institución.

2.2 Descripción del Objeto de Estudio

Las reglas que regulan el comportamiento en las carreteras están indicadas en la Ley No. 60, Código de Vialidad y Tránsito. Su última edición está actualizada con las modificaciones puestas en vigor por el Decreto-Ley No. 231, de 12 de diciembre del 2002, “Modificativo de la Ley No. 60, de 28 de septiembre de 1987, Código de Vialidad y Tránsito”[17].

A continuación se explican algunos de los aspectos que aparecen en la Ley 60:

Licencia de conducción: es el documento que acredita que su titular está autorizado para conducir vehículos de motor, según las categorías que en la misma se señalen.

Permiso para el aprendizaje: documento otorgado por el Ministerio del Interior a los aspirantes a la obtención de la licencia de conducción.

Se establecen tres clases de licencia de conducción:

- 1) Nacional, autoriza a su titular a conducir, en todo el territorio nacional.
- 2) Especial militar, se expide únicamente a los miembros de las instituciones armadas que se encuentran en servicio activo, a solicitud del jefe de la unidad o jefe de nivel jerárquico superior correspondiente. Los titulares de esta licencia sólo están autorizados a conducir los vehículos que en la misma se especifique, dentro de la ruta trazada por sus responsables y sujeto al entrenamiento y a las restricciones que establezca el jefe del nivel jerárquico superior de la unidad.
- 3) Internacional o permiso internacional de conducción expedido por la República de Cuba, se atiene a los requisitos aprobados por la Conferencia de Viena, de 1968, sobre Circulación por Carreteras.

El Ministerio del Interior podrá disponer la suspensión de la licencia de conducción por un período de un mes hasta un año:

1) A los titulares que dentro de un año natural cometan más de dos infracciones de las regulaciones del tránsito calificadas como “muy peligrosas”, o que alcancen la puntuación que al efecto se determine por el Ministerio del Interior.

2) A los titulares que, en un año natural, cometan más de una infracción del tránsito conduciendo un vehículo de uso personal, después de haber ingerido bebidas alcohólicas.

También podrá cancelar la licencia de conducción o permiso de aprendizaje:

1) A los titulares que en los dos últimos años naturales se les suspenda la licencia de conducción, administrativa o judicialmente, en más de dos ocasiones.

2) A los titulares que, durante un año natural después de haberseles suspendido la licencia de conducción administrativa o judicialmente, y hayan cumplido el período dispuesto para dicha medida, incurran en una nueva infracción calificada como “muy peligrosa”.

3) A los titulares que, durante el tiempo de suspensión de la licencia de conducción, circulen por la vía conduciendo un vehículo de motor. En este caso, la cancelación se hará efectiva después de cumplido el tiempo de la suspensión.

4) A los titulares, conductores de vehículos de carga o de transporte colectivo de pasajeros, o conductores profesionales que actúen como tales, que en más de una ocasión, durante un período de tres años naturales, se les suspenda la licencia de conducción por conducir después de haber ingerido bebidas alcohólicas, lo cual debe estar avalado por el dictamen de la autoridad competente.

5) A los aspirantes que durante el período de vigencia del permiso de aprendizaje circulen por la vía conduciendo un vehículo, incumpliendo las regulaciones establecidas para el aprendizaje.

6) A los titulares que en más de una ocasión, dentro de un año natural, conduzcan un vehículo sin las correspondientes chapas de identificación, expedidas por el Registro de Vehículos.

Las **infracciones** de las regulaciones del tránsito de vehículos establecidas en el Código de Vialidad y Tránsito se sancionarán administrativamente con multas, que serán impuestas y cobradas del modo y en las cuantías correspondientes al grado de peligrosidad de estas.

Las multas por infracciones de las regulaciones del tránsito son impuestas por los miembros de la Policía Nacional Revolucionaria encargados de esta función, los cuales las notifican a los infractores mediante boletas de notificación en las que se consignan, entre otros, los datos siguientes:

1. Nombre y apellidos del infractor.
2. Domicilio del infractor.
3. Número de la licencia de conducción.
4. Fecha en que se cometió la infracción.
5. Lugar en que se cometió la infracción.
6. Número de la chapa de identificación del vehículo.
7. Precepto legal infringido.
8. Cuantía de la multa.
9. Datos del agente actuante.

En el mismo acto de imposición de la multa se entregará al infractor un comprobante de la boleta de notificación, y éste deberá firmar la boleta en presencia de la autoridad que la impone, como constancia de dicho acto.

De acuerdo a su peligrosidad las multas se clasifican en:

Grupo 1: Muy Peligrosas, equivalen a \$30.00 y acumulan 12 puntos.

Grupo 2: Peligrosas, equivalen a \$15.00 y acumulan 8 puntos.

Grupo 3: Menos Peligrosas, equivalen a \$10.00 y acumulan 6 puntos.

Después de notificada una multa a un infractor éste tiene un periodo de diez días hábiles para reclamar la validez de esta en la Unidad Municipal de la PNR (Policía Nacional Revolucionaria) del Municipio en el que se cometió la infracción. Dicha reclamación consiste en una simple exposición de sus consideraciones, de forma escrita o verbal.

El jefe de la Unidad Municipal de la PNR resolverá la reclamación dentro de los veinte días naturales siguientes a su interposición, y se le notifica al reclamante en el término de veinte días naturales siguientes a la fecha de la decisión.

El pago de las multas se efectuará en las oficinas habilitadas para ese fin, dentro de los treinta días naturales siguientes a la notificación de la infracción. Al efectuar dicho pago se presentará el comprobante de la boleta de notificación. En este tiempo no se le tienen en cuenta los puntos que le suma la multa ni la deuda de esta. En caso de haber reclamado y el resultado haya sido “con lugar” (a favor del reclamante), le es rembolsado el importe de la multa (si ya había sido pagada) y no se le tienen en cuenta los puntos ni la deuda que esta le acumulaba (en el caso contrario si le cuentan).

Si el pago se efectúa después de transcurrir el plazo de treinta días naturales, sin exceder de sesenta, el importe de la multa se duplica.

Un conductor que transcurrido el plazo de sesenta días naturales sin haber abonado el importe de la multa en la cuantía que corresponda, se puede sancionar por un período de tres meses a un año. Además, después de los sesenta días el talón de pago pierde su valor y se debe dirigir a un Centro de Licencia para recibir un Comprobante de Deuda, documento necesario para poder hacer efectivo el pago.

Si transcurrido el término de la sanción, aún no se ha hecho efectivo el pago, automáticamente la sanción administrativa de suspensión se convierte en una cancelación de licencia, por un período que puede oscilar entre tres y cinco años.

Las multas caducarán transcurrido un año contado a partir de la notificación. Este término se interrumpirá si se ha hecho algún requerimiento al infractor para que pague, o por la realización de cualquier otro acto de la administración dirigido al cobro de la multa.

2.3 Sistemas Existentes

Actualmente existen varios sistemas mediante los cuales se puede acceder a la información de las infracciones aunque no todos brindan las mismas facilidades.

Infracciones y ExpCond son parecidos en cuanto a la función que realizan pues ambos se utilizan para gestionar la información de las infracciones. Es decir, mediante estos se insertan los datos de las matrices en la base de datos, se actualizan y se muestran. Su debilidad consiste en que solo permiten visualizar los datos de un solo individuo y a la hora de tomar decisiones es necesario tener una perspectiva general del negocio, no detalles de un solo elemento.

El sitio web Expediente del Conductor permite mostrar información general pero carece de una interfaz amigable, de gráficos que faciliten el análisis, de muchos de los reportes requeridos y de flexibilidad a la hora de realizar consultas.

2.4 Conclusiones

Después de realizar un estudio de la situación actual en que se ejecuta el proceso de control de infracciones se identificaron problemas vinculados con la visualización de reportes y la ejecución de consultas complejas, debido a que el diseño de datos no es el mejor.

En estos momentos los dirigentes del MININT no cuentan con una herramienta que automatice adecuadamente la visualización de los resultados, por lo que se requiere el empleo de técnicas y arquitecturas de Inteligencia de Negocio que permitan el modelado y automatización del flujo de trabajo correspondiente. Específicamente se hace necesario la implementación de un Mercado de Datos, con los respectivos procesos ETL requeridos para poblarlo y la posterior visualización de los resultados en un sitio web.

Capítulo 3 – Construcción de la solución propuesta

3.1 Introducción

En este capítulo se realiza una descripción detallada del proceso de diseño del MD, específicamente abordando la realización de las diferentes fases de diseño utilizando la metodología DWEP a través de varios diagramas creados en tres niveles: conceptual, lógico y físico.

3.2 Descripción del sistema propuesto

El sistema que se propone se centra en la implementación de una arquitectura de apoyo a la toma de decisiones para los dirigentes involucrados en el análisis de la información referente a las infracciones del tránsito.

Para el manejo de la información se decidió implementar un Mercado de Datos debido a que un Almacén se orienta a la recopilación de la información de toda una organización y su costo de desarrollo en recursos y tiempo es mucho mayor, a diferencia de un MD que tiene carácter departamental y mantiene las mismas características. Como patrón de diseño se decidió utilizar el modelo multidimensional basado en el esquema de estrella.

La población del MD se realiza a partir de la información proveniente de las Bases de Datos de infracciones, de personas y de licencias de conducción. A partir de estas fuentes se realiza la extracción, transformación y carga de los datos mediante los procesos de ETL.

El sistema que se propone posibilita realizar tareas de análisis sobre los datos almacenados mediante reportes de varios tipos como tablas, gráficos, indicadores, etc.

3.2.1 Requerimientos

3.2.1.1 Requerimientos Funcionales

1. El sistema permitirá la transformación de las estructuras del MD a estructuras más sencillas y con significado para el usuario con pocos conocimientos sobre dichas estructuras.

2. El sistema tendrá que extraer, transformar y cargar datos desde los sistemas operacionales.
3. El sistema permitirá a los usuarios la presentación de la información de acuerdo a los criterios establecidos.
4. El sistema permitirá activar o desactivar dimensiones en las consultas.

3.2.1.2 Requerimientos no Funcionales

Apariencia o interfaz externa

El sistema debe mostrar una interfaz amigable, sencilla, agradable, legible y de fácil uso, informando siempre al usuario dónde está y qué puede hacer desde allí. El contenido será plasmado de forma clara. Se evitarán imágenes y animaciones almacenadas en archivos de gran tamaño.

Usabilidad

Los usuarios podrán interactuar con el sistema, aún teniendo pocos conocimientos del manejo de la computadora y del ambiente Web. Para su manejo se documentarán a través de la ayuda.

Rendimiento

El tiempo medio de respuesta a consultas simples echas al MD deberá ser de no más de 1 minuto y para las consultas complejas podría ser de 1 a 3 minutos.

Ayuda y documentación en línea

El sistema contará con ayuda. En ella se describirán las funcionalidades de la aplicación, con el fin de garantizar el buen desempeño de los usuarios a la hora de interactuar con el mismo.

Hardware

Se requiere disponer del servidor Web y de base de datos, con los requerimientos de hardware que necesitan Oracle BI Server y Oracle9i. Las terminales de la empresa sólo requerirán estar conectadas a la red.

Software

La aplicación necesita para su ejecución, Oracle9i como sistema gestor de base de datos y Oracle BI Server como servidor Web. En las terminales clientes sólo se necesitará un navegador Web.

Seguridad

El sistema está definido para operar el sistema del control de acceso mediante el BI Administration Tool y el BI Interactive Dashboards, garantizando de esta forma el acceso a los datos a las personas debidamente autorizadas a través de contraseñas y niveles de acceso.

Precisión de los datos

El sistema debe proceder a la actualización de los datos desde las fuentes cada vez que ocurran cambios en ellas.

3.2.2 Diseño del Mercado de Datos

El diseño de un AD se estructura en niveles o capas por las que transitan los datos, para ello, es necesario crear tres modelos: conceptual, lógico y físico. A continuación, se describen la creación cada uno de los modelos.

3.2.2.1 Modelo Conceptual

Para seleccionar la estrategia de diseño del MD se tomaron en consideración aspectos como:

- El diseño debe basarse en la estructura y la información almacenada en las diferentes bases de datos que constituyen la fuente de ingreso del MD, incluyendo solo la información que sea relevante.
- El conjunto de datos debe poseer un nivel de granularidad que permita extraer información referente a las infracciones de modo que a menor granularidad mayor nivel de detalle en la información.
- Debe ser flexible y debe permitir posibles ampliaciones sin traer grandes complicaciones en el modelo.

Para el diseño del MD se decidió emplear el esquema de estrella por definir una estructura sencilla y no existir necesidad de crear jerarquías en las dimensiones.

La metodología que se está siguiendo divide el proceso de diseño conceptual del AD en tres niveles para una mejor comprensión:

3.2.2.2 Nivel 1 (Definición Global del Modelo)

A continuación se muestra la definición del esquema de estrella correspondiente al MD.

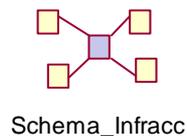


Figura 1: Esquema conceptual del AD. Nivel1.

3.2.2.3 Nivel 2 (Definición de Estructura de Esquemas Estrella)

En el diseño de un Almacén o Mercado de Datos no puede faltar la dimensión tiempo, la cual es fundamental debido a que da cumplimiento al registro de información histórica. Para el diseño de esta propuesta se creó esta dimensión de forma que pueda resumir datos por: días, meses, trimestres, semestres y años.

La figura 2 muestra el esquema de estrella correspondiente al MD (nivel 2), donde se pueden observar en el centro con forma de cubos las tablas de hecho y a su alrededor las tablas de dimensiones con las que está relacionada.

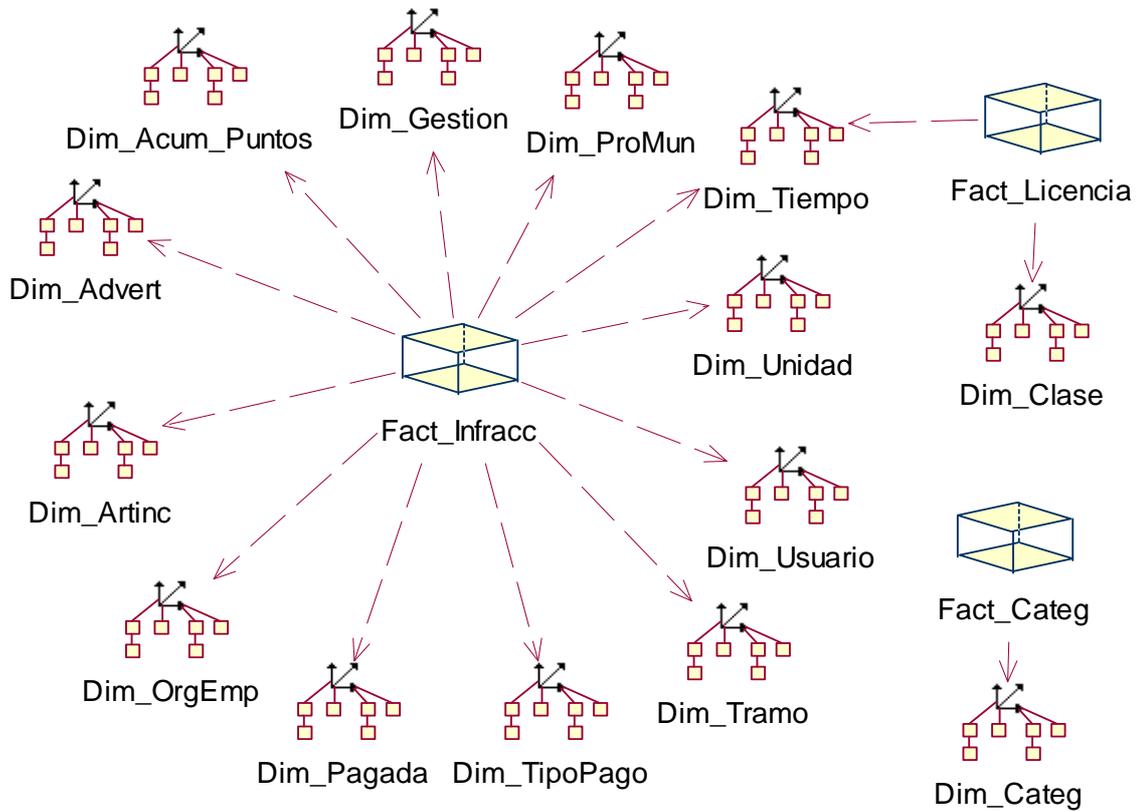


Figura 2: Estructura del Esquema de Estrella

3.2.2.4 Nivel 3 (Definición de Estructura de Hechos y Dimensiones)

En este nivel se definen las propiedades estructurales de las dimensiones y hechos, especificando los atributos y relaciones.

A continuación, se muestra la definición de las estructuras de las dimensiones:

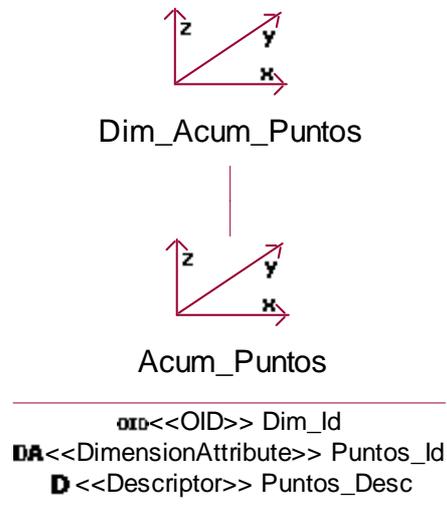


Figura 3: Estructura de la Dimensión: Dim_Acum_Puntos

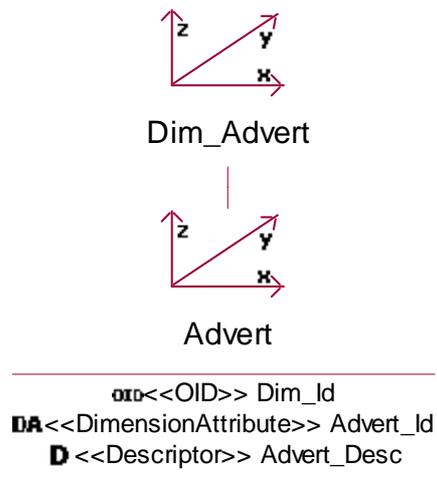


Figura 4: Estructura de la Dimensión: Dim_Acum_Puntos

Figura 1: Estructura de la Dimensión: Dim_Advert

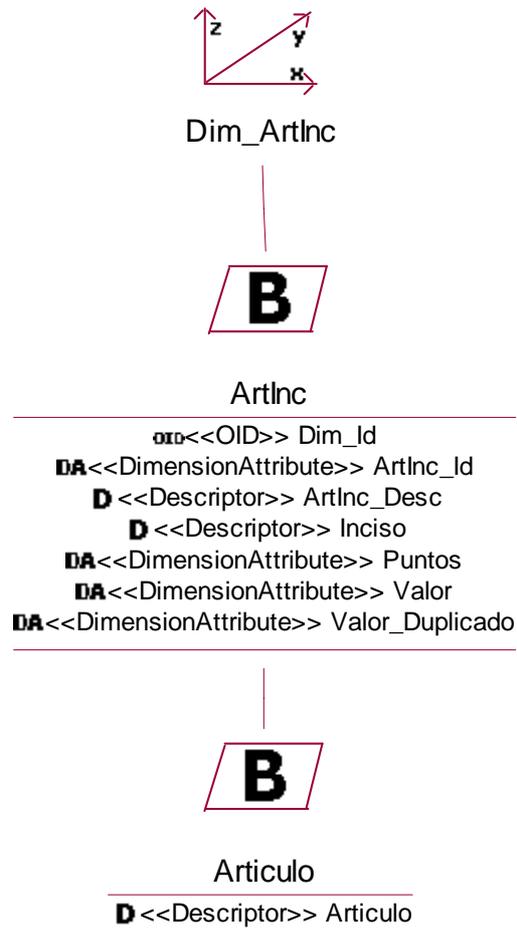


Figura 2: Estructura de la Dimensión: Dim_ArtInc.

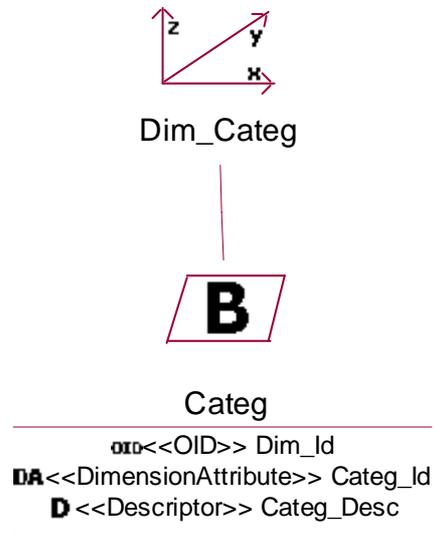


Figura 3: Estructura de la Dimensión: Dim_Categ

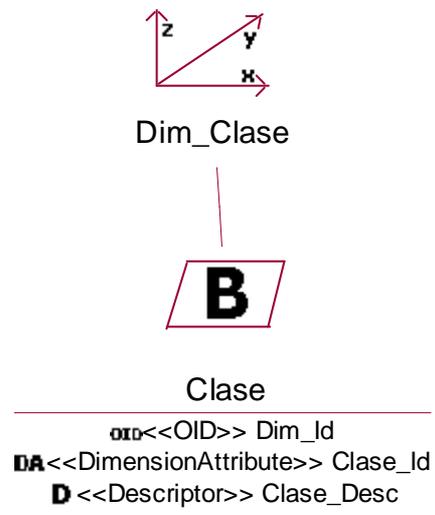


Figura 4: Estructura de la Dimensión: Dim_Clase

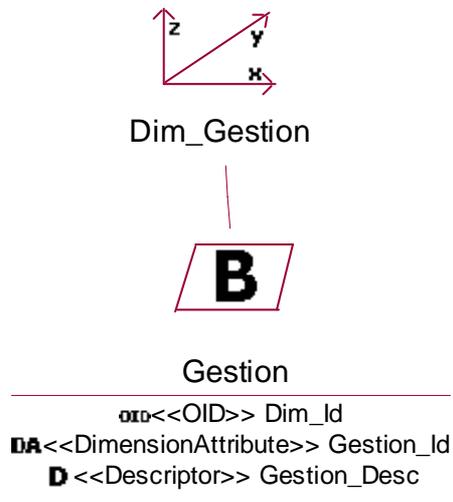


Figura 5: Estructura de la Dimensión: Dim_Gestion

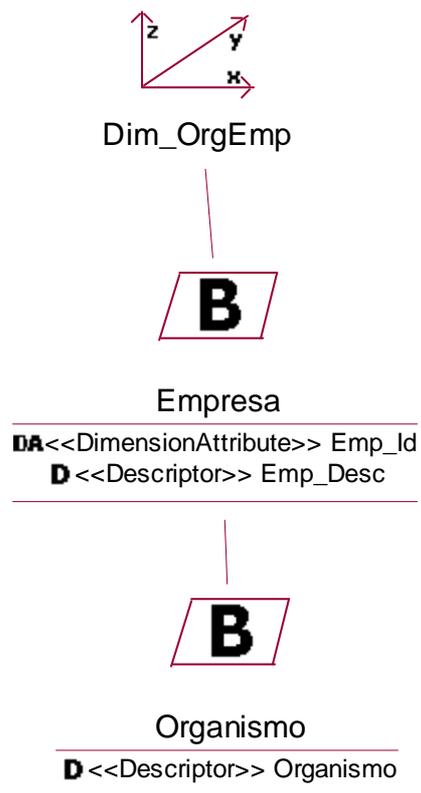


Figura 6: Estructura de la Dimensión: Dim_OrgEmp

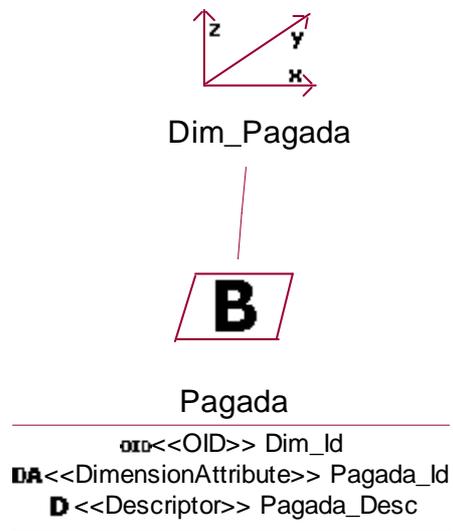


Figura 7: Estructura de la Dimensión: Dim_Pagada

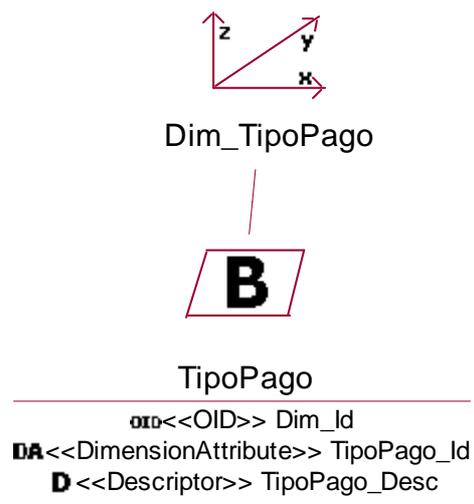


Figura 8: Estructura de la Dimensión: Tipo de Pago

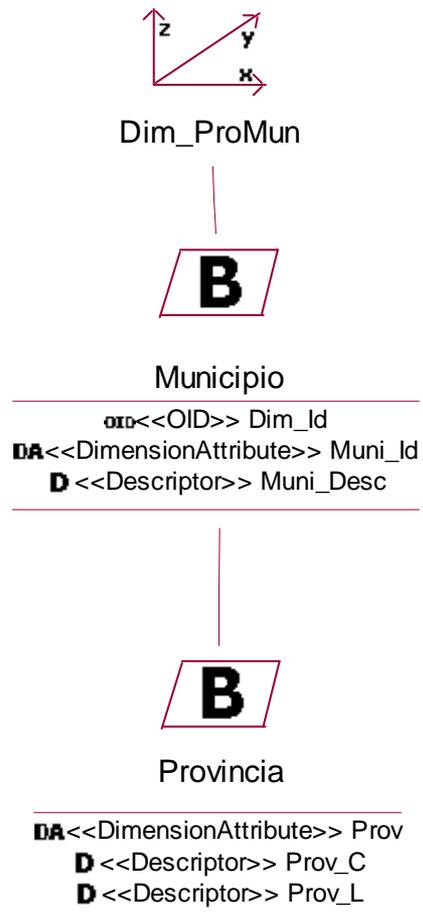


Figura 9: Estructura de la Dimensión: Dim_ProMun

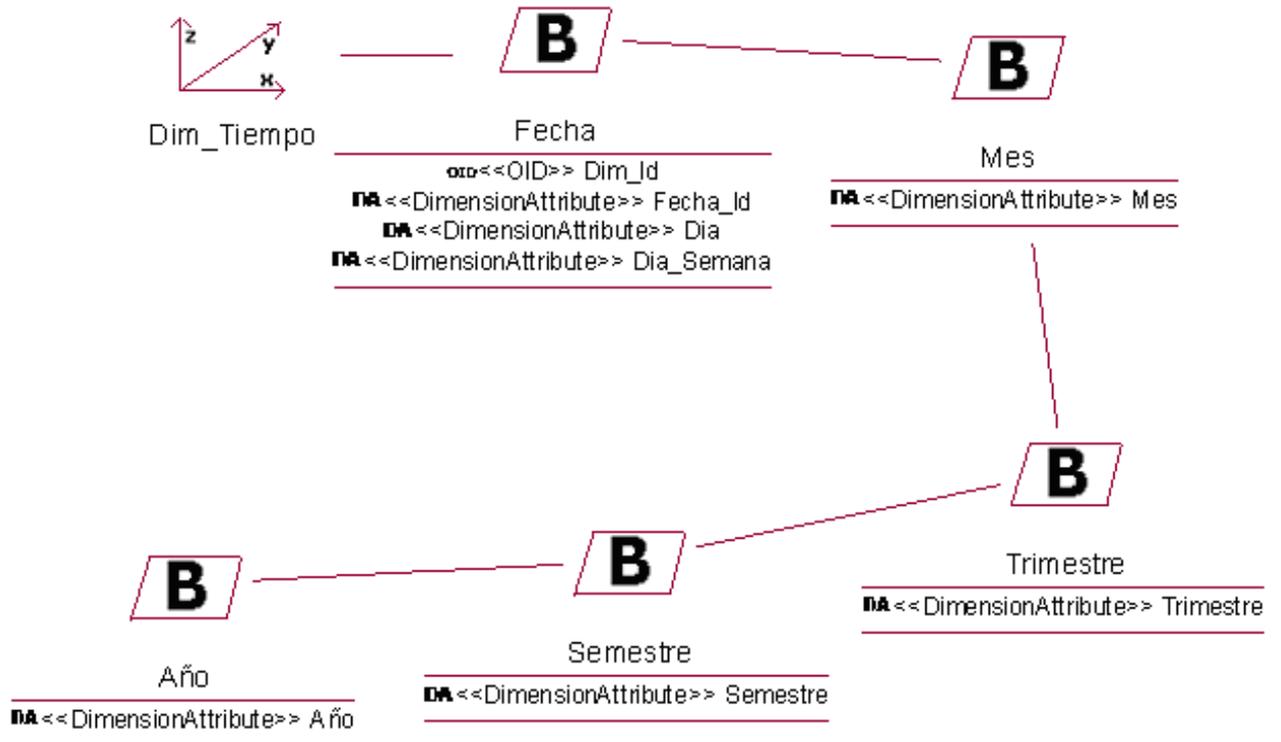


Figura 10: Estructura de la Dimensión: Dim_Tiempo

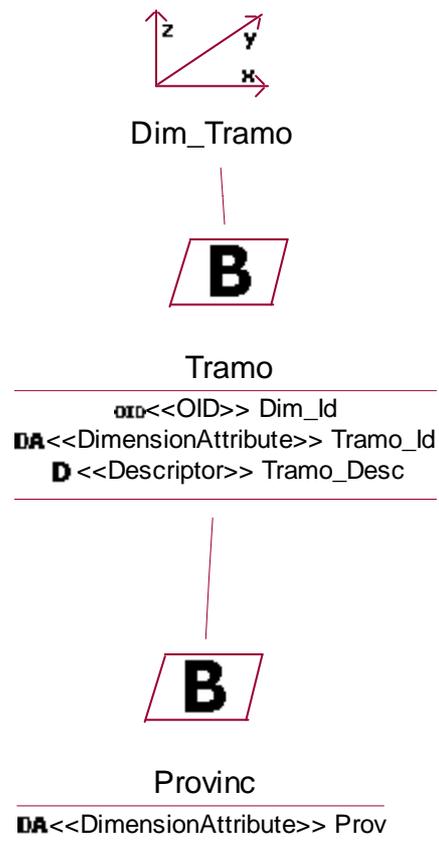


Figura 11: Estructura de la Dimensión: Tramo

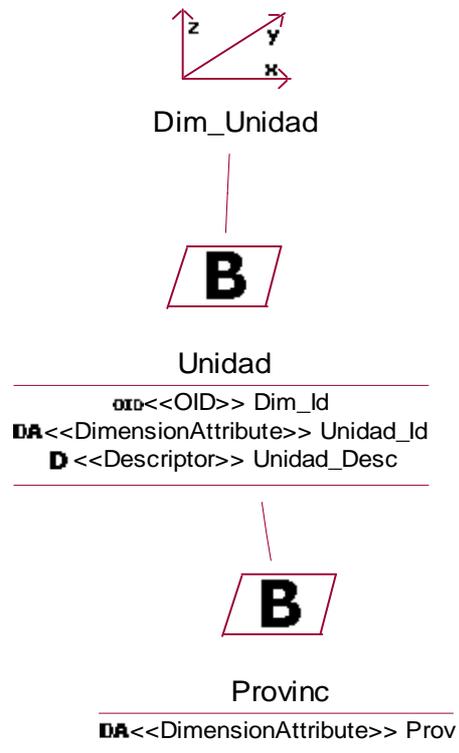


Figura 12: Estructura de la Dimensión: Unidad

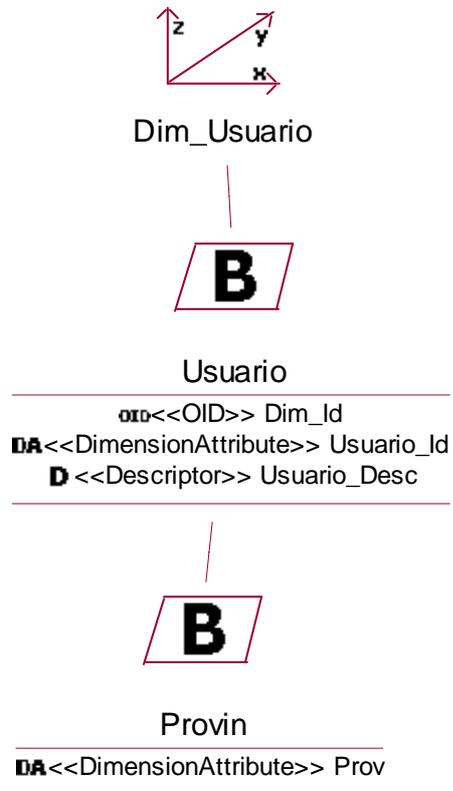


Figura 13: Estructura de la Dimensión: Usuario

A continuación, se muestra la definición de las estructuras de hechos:

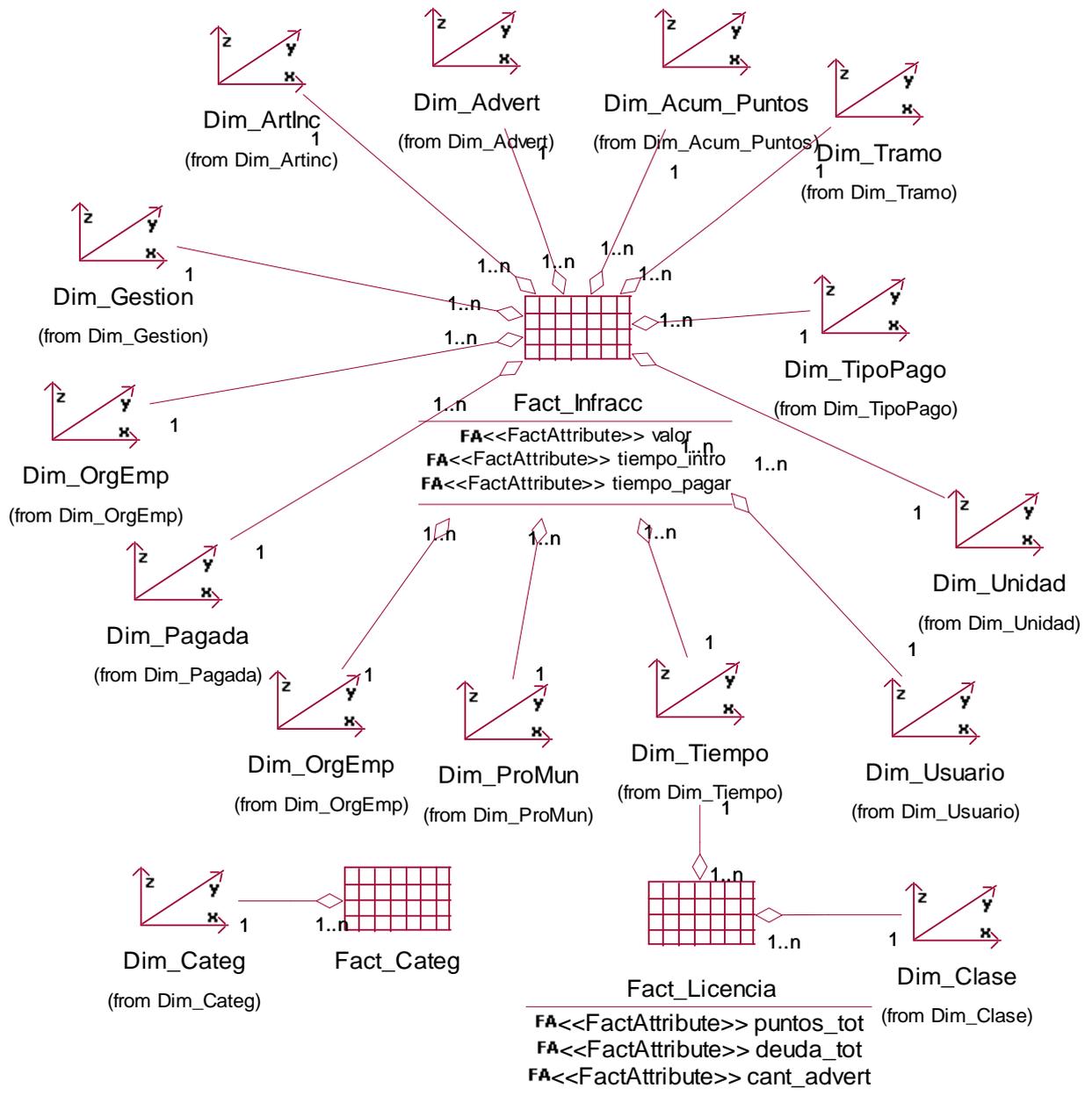


Figura 3. 2 Estructura de los Hechos.

3.2.2.5 Mapeo de Datos

En toda solución de Inteligencia de Negocio que incluya la construcción de un AD o MD es de vital importancia la ejecución de procesos de extracción, limpieza y carga de los datos (ETL) provenientes de las diferentes fuentes operacionales. Para ello, se emplean diagramas de mapeo que representan el flujo de los datos.

En el modelo intervienen tres elementos fundamentales:

- Proveedor: encargado de generar los datos (esquema, tabla o atributo).
- Consumidor: encargado de recibir los datos.
- Emparejamiento: forma y tareas (transformaciones) realizadas durante el mapeo de datos.

En esta propuesta de solución los datos que nutren el MD provienen de varias bases de datos.

El diagrama de mapeo de los datos quedaría de la siguiente forma (Figura 3.53):

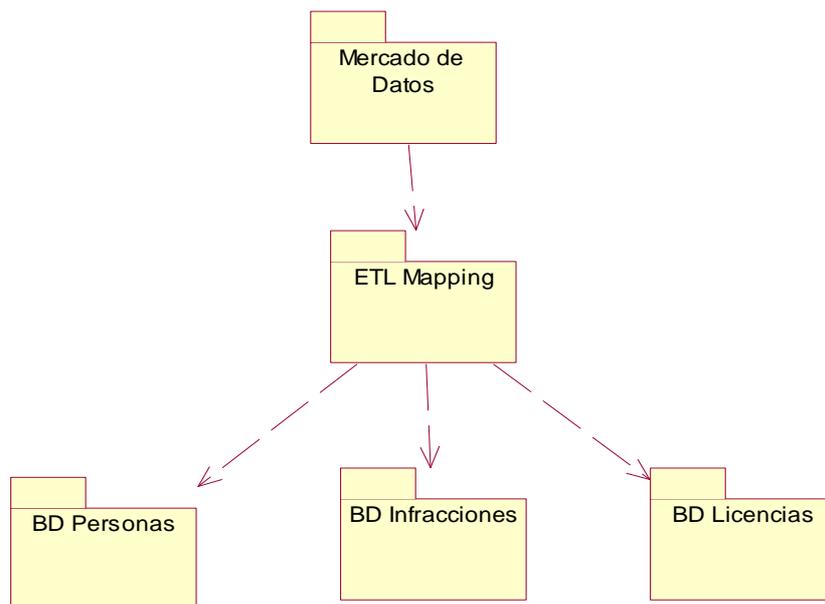


Figura 3. 3 Diagrama de Mapeo de los Datos.

3.2.2 Modelo Lógico

Basados en los artefactos elaborados en el modelo anterior, se realizó un mapeo a los elementos que lógicamente podrían ser representados en la base de datos. O sea, a partir de las clases y relaciones modeladas se derivaron: tablas, columnas, tipos de datos, llaves, índices, dependencias y relaciones.

En los anexos, se muestran los diagramas correspondientes a los esquemas de estrella modelados, con sus respectivas tablas de hechos y dimensiones. Es un solo diagrama que se encuentra dividido para su mejor visualización.

3.2.2.1 Procesos ETL

Los datos que pueblan el Mercado de Datos provienen de varias fuentes operacionales. Para implementar los procesos ETL primeramente se debe analizar la calidad de los datos provenientes de las fuentes, para posteriormente definir las transformaciones necesarias con el propósito de lograr consistencia e integración de la información. Estos se implementaron en el Oracle SQL Developer mediante PL/SQL.

3.2.3 Modelo Físico

Un componente en UML es una parte de un sistema (archivos de configuración, interfaces, bases de datos, tablas, etc.) que encapsula implementación, brinda una interfaz de comunicación y puede contener otros componentes. [9]

En un diagrama los componentes se encuentran conectados por medio de relaciones de dependencia. Los componentes más empleados en este tipo de soluciones poseen como estereotipo: Database, Tablespace y Table.

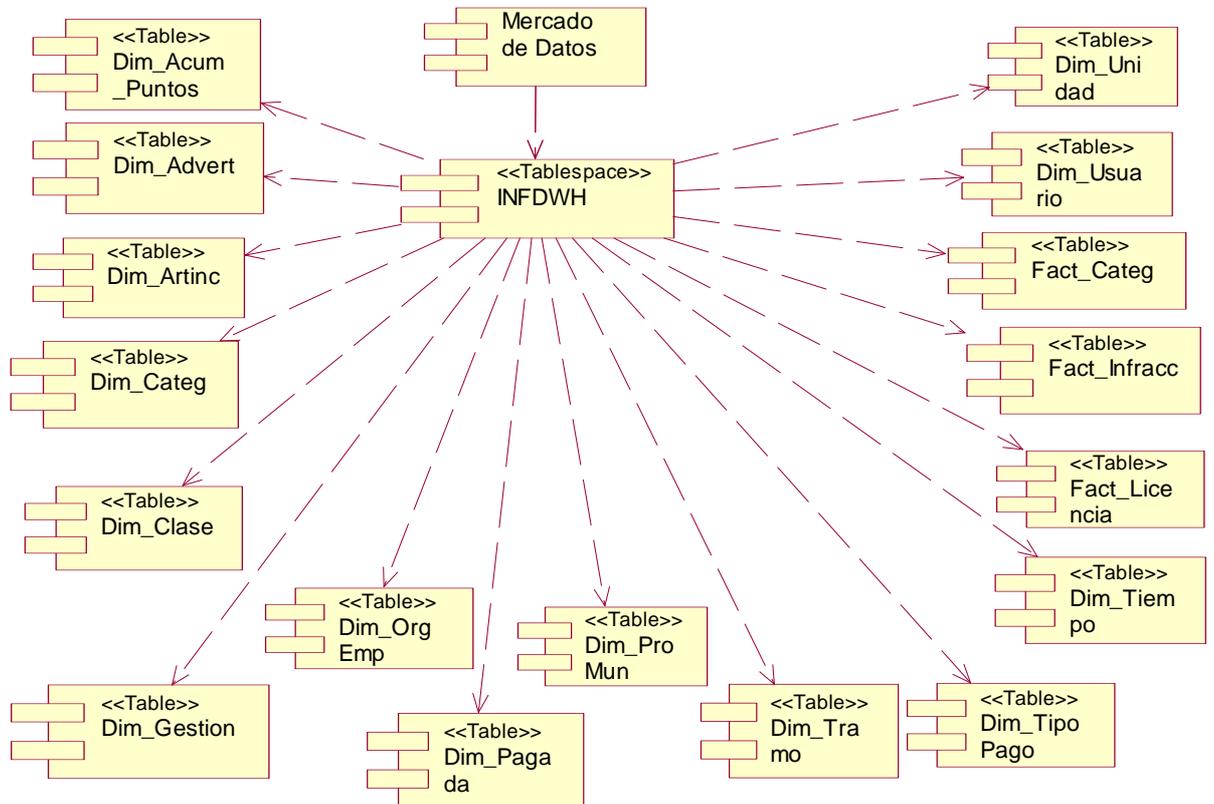


Figura 3. 4 Diagrama de Componentes.

3.2.3.2 Diagrama de Despliegue

Un diagrama de despliegue en UML es un artefacto que muestra las características de distribución de los diferentes componentes de la aplicación y los sistemas legados asociados, en tiempo de ejecución. Los elementos involucrados, a los que se les denomina nodos, pueden representar elementos de hardware o de software (Servidores Web, Bases de Datos, etc.) y se relacionan en consecuencia con la comunicación que existe entre ellos. [9]

En el diagrama de despliegue son especificados los diferentes aspectos relativos a la configuración del software y del hardware, por otro lado la distribución de las estructuras lógicas definidas previamente también son representadas. En la figura

17 se puede observar la configuración del servidor que alberga al DM. El disco1 hospeda a las tablas hechos y dimensiones.

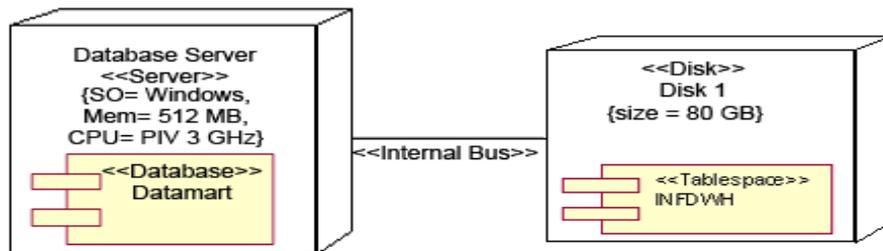


Figura 14: Diagrama de Despliegue del Sistema

3.3 Beneficios Tangibles e Intangibles

Uno de los puntos más importantes a tener en cuenta en el momento de decidir implementar una arquitectura de BI son es el costo que trae aparejado.

El sistema presenta un valor práctico bien definido, ya que está orientado a dotar a los dirigentes del MININT de una herramienta de apoyo a la toma de decisiones sobre las infracciones del tránsito al permitir mostrar la información desde diferentes puntos de vista.

En el caso particular de esta propuesta, los beneficios identificados son intangibles, debido a que no constituye un medio de ingreso económico para la entidad y a que son difíciles de medir, debido a que se necesitaría analizar el valor de la información brindada y determinar si su aprovechamiento provoca algún beneficio cuantificable.

A continuación se argumentan los beneficios identificados:

Proveer al MININT de una herramienta para la recolección, preparación, procesamiento y visualización de la información sobre las infracciones del tránsito que se encarga de extraer los datos de distintos esquemas de bases de datos, limpiarlos para erradicar errores e inconsistencias y cargarlos en una sola estructura centralizada con un diseño que hace más fácil al usuario la

comprensión de la información por lo que permite un mayor aprovechamiento de la información. De esta forma se llega a la conclusión de que es factible el proyecto.

3.4 Validación de la solución Propuesta

Se realizó una encuesta (Ver Anexo 3.1) a los usuarios que se beneficiarán de la interacción con el sistema resultado de este trabajo para conocer su criterio sobre el mismo. Teniendo en cuenta que los compañeros involucrados son pocos no se utilizó ningún tipo de muestreo y se encuestó a toda la población.

Se utilizó el paquete estadístico SPSS V.15 para procesar la información de las encuestas en las que se usó una presentación sencilla, preguntas claras y simples, secuencia lógica y se evitó la fraseología negativa.

Se encuestaron un total de 4 personas que emitieron sus valoraciones sobre el contenido del sistema, las cuales otorgaron puntuaciones promedio de 4 puntos (Muy Alto) con un rango de variación de 0 a 1 puntos demostrando de forma general que el grado de satisfacción con el sistema es elevado, además de que no se encontraron criterios negativos. Para más información ver Anexo de encuestas.

3.4 Conclusiones

En este capítulo se realizó la descripción del sistema propuesto para dar solución a la problemática planteada. Para ello se desarrollaron las etapas de construcción y despliegue de una arquitectura de Business Intelligence, con el empleo de la metodología de diseño basada en UML y en el Proceso Unificado: *Data Warehouse Engineering Process* (DWEPE).

La descripción de la solución basada en DWEPE transitó por la construcción de varios modelos con diferentes niveles de detalle: modelo conceptual, modelo lógico y modelo físico. Esta estructura permitió ir analizando de forma incremental los requisitos a tener en cuenta para la implementación final del repositorio de datos.

Para finalizar se analizaron los beneficios que trae consigo la realización de un proyecto de este tipo y se llegó a la conclusión que tomando en cuenta el ahorro de tiempo y la confiabilidad de la información, el proyecto resulta factible.

Conclusiones

En este trabajo se ha puesto de manifiesto la importancia que tiene la Inteligencia de Negocios en el proceso de toma de decisiones. Teniendo en cuenta los objetivos planteados, se arriba a las siguientes conclusiones:

1. Como resultado del trabajo realizado se logró el análisis, diseño e implementación de una arquitectura de BI que permite gestionar y visualizar información relacionada con las infracciones del tránsito.
2. Para lograr la definición del diseño propuesto se realizó un estudio de los antecedentes de la Inteligencia de Negocios, y se analizaron las tendencias y tecnologías actuales del campo de la Informática que posibilitaron diseñar y desarrollar el sistema propuesto.
3. Se diseñó un mercado de datos para el almacenamiento de la información de las infracciones siendo seleccionado Oracle como sistema gestor de bases de datos y como proveedor de herramientas de desarrollo y de BI necesarias para desarrollar el sistema.
4. Se implementaron los procesos de extracción, transformación y carga de datos desde las diferentes fuentes operacionales al MD.
5. Se aplicaron técnicas de construcción y despliegue de un Almacén de Datos, empleando la metodología de diseño: Data Warehouse Engineering Process (DWEP).
6. Se emplearon productos de BI de Oracle en la construcción de reportes que permiten observar el comportamiento de las infracciones del tránsito.
7. Se realizaron pruebas que permitieron validar el sistema y determinar los elementos que deben ser perfeccionados para una mejor utilización del mismo por parte de los usuarios finales del mismo.
8. Se establecen las bases para la realización de estudios que permitan el descubrimiento de conocimiento oculto que servirá de apoyo a la toma de decisiones con niveles de objetividad muy superiores a los actuales.

Recomendaciones

Incrementar el mercado de datos de infracciones con información sobre los accidentes del tránsito y otros temas relacionados de modo que se puedan aplicar técnicas de Minería de Datos que permitan la extracción de conocimiento oculto que permitan descubrir relaciones entre las infracciones y los accidentes del tránsito junto a otros elementos.

Referencias bibliográficas

- [1]“Sistemas de Información para ejecutivos, Inteligencia de Negocio y Cuadro de Mando Integral”; isg.enmollina.com.
- [2]“Beneficios de la BI en Dataprix.com”; <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/1-business-intelligence/14-beneficios>.
- [3] Christian Reyes Meleán, “Una breve introducción a las tecnologías de la información para la gestión del conocimiento”; upcommons.upc.edu.
- [4] “Sistemas de Soporte a la Decisión (DSS)”; http://www.sinnexus.com/business_intelligence/sistemas_soporte_decisiones.aspx.
- [5] M.A. Mártire, A., “Fundamentos de Data Warehouse”; <http://www.doaj.org/doaj?func=abstract&id=273286>.
- [6] R.C. Nigel Pendse, “The OLAP Report”; <http://www.olapreport.com>.
- [7] Inmon, W. H. , *Building the Data Warehouse*, Wiley Publishing, Inc, .
- [8] P.L. Viv Schupmann , “Oracle9i Data Warehousing Guide Release 2 (9.2) March 2002 Part No. A96520-01.”
- [9] S.L. , “Diseño de Almacenes de Datos con UML.”
- [10]“Libros en línea de Microsoft SQL Server 2005”; [http://msdn.microsoft.com/es-es/library/ms203721\(SQL.90\).aspx](http://msdn.microsoft.com/es-es/library/ms203721(SQL.90).aspx).
- [11]“Libros en línea de Microsoft SQL Server 2008”; <http://msdn.microsoft.com/es-es/library/bb418432%28SQL.10%29.aspx>.
- [12]“Oracle BI Administration Tool”; http://www.oracle.com/technology/obe/obe_bi/bi_ee_1013/bi_admin/biadmin.html#t.
- [13]“Componentes de OBIEE”; <http://www.oracle.com/technology/products/bi/enterprise-edition-platform-components.html>.
- [14]“BI Dashboards de OBIEE”; http://www.oracle.com/technology/pub/articles/rittman_dash.html].
- [15]“Oracle BI Answers”; <http://www.oracle.com/technology/products/bi/enterprise-edition.html>.
- [16]“Oracle SQL Developer”; <http://www.osalt.com/es/sql-developer>.

[17]“Ley No. 60, Código de Vialidad y Tránsito.”

Bibliografía

- “Sistemas de Información para ejecutivos, Inteligencia de Negocio y Cuadro de Mando Integral”; isg.enmollina.com.
- “Beneficios de la BI en Dataprix.com”; <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/1-business-intelligence/14-beneficios>.
- Christian Reyes Meleán, “Una breve introducción a las tecnologías de la información para la gestión del conocimiento”; upcommons.upc.edu.
- “Sistemas de Soporte a la Decisión (DSS)”; http://www.sinnexus.com/business_intelligence/sistemas_soporte_decisiones.aspx.
- M.A. Mártire, A., “Fundamentos de Data Warehouse”; <http://www.doaj.org/doaj?func=abstract&id=273286>.
- R.C. Nigel Pendse, “The OLAP Report”; <http://www.olapreport.com>.
- Inmon, W. H. , *Building the Data Warehouse*, Wiley Publishing, Inc, .
- P.L. Viv Schupmann , “Oracle9i Data Warehousing Guide Release 2 (9.2) March 2002 Part No. A96520-01.”
- S.L. , “Diseño de Almacenes de Datos con UML.”
- “Libros en línea de Microsoft SQL Server 2005”; [http://msdn.microsoft.com/es-es/library/ms203721\(SQL.90\).aspx](http://msdn.microsoft.com/es-es/library/ms203721(SQL.90).aspx).
- “Libros en línea de Microsoft SQL Server 2008”; <http://msdn.microsoft.com/es-es/library/bb418432%28SQL.10%29.aspx>.
- “Oracle BI Administration Tool”; http://www.oracle.com/technology/obe/obe_bi/bi_ee_1013/bi_admin/biadmin.html#t.
- “Componentes de OBIEE”; <http://www.oracle.com/technology/products/bi/enterprise-edition-platform-components.html>.
- “BI Dashboards de OBIEE”; http://www.oracle.com/technology/pub/articles/rittman_dash.html.
- “Oracle BI Answers”; <http://www.oracle.com/technology/products/bi/enterprise-edition.html>.

- “Oracle SQL Developer”; <http://www.osalt.com/es/sql-developer>.
- “Ley No. 60, Código de Vialidad y Tránsito.”

Glosario de términos

Listado de términos ordenados alfabéticamente, que son usados en el trabajo de diploma y cuyo significado es poco conocido. Se indica para cada término su significado.

Almacén de datos: en inglés *Data Warehouse*, es una colección de datos orientada a un dominio, integrada, no volátil y que varía en el tiempo.

Business Intelligence: Conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento mediante el análisis de datos existentes en una organización o empresa.

Carga: Es una de las etapas de los procesos ETL. Es el momento en el cual los datos de la fase anterior son cargados en el destino. Dependiendo de los requerimientos de la organización, este proceso puede abarcar una amplia variedad de procesos diferentes (actualizaciones, mantenimiento de la historia de los datos, etc.).

Consultas ad-hoc: consultas no predefinidas por el sistema, son creadas en tiempo de ejecución por los usuarios según sus necesidades.

Dashboard (Cuadros de Mando): Recurso que agrupa gran cantidad de información en entornos visuales muy llamativos y prácticos. Pueden incluir mapas, gráficos y componentes visuales conformando un entorno intuitivo para el usuario, consiguiendo la representación de complejos modelos de datos, relaciones entre variables y fórmulas de manera sencilla.

Mercado de Datos: Es un almacén de datos departamental, o sea, orientado a un sector determinado de la organización. Es también conocido: mercado de datos.

Extracción: Primera etapa de un proceso ETL, consiste en extraer los datos desde las fuentes y convertirlos en un formato preparado para iniciar el proceso de transformación.

Minería de Datos: Conjunto de técnicas para la inducción de conocimiento útil a partir de masas muy grandes de datos. Tiene, por lo tanto, un solapamiento importante con otras disciplinas como la estadística tradicional, el reconocimiento de patrones, la inteligencia artificial, etc.

Anexos

Documentos, imágenes, tablas, gráficos que complementan el contenido o conocimiento del trabajo de diploma, pero que por tamaño en hojas del cuerpo del documento no es posible incluirlo.

Todos los anexos deben ser referenciados en el cuerpo del documento.

Anexo 1 – Prototipos de la aplicación.



Figura 15: Pantalla de Autenticación

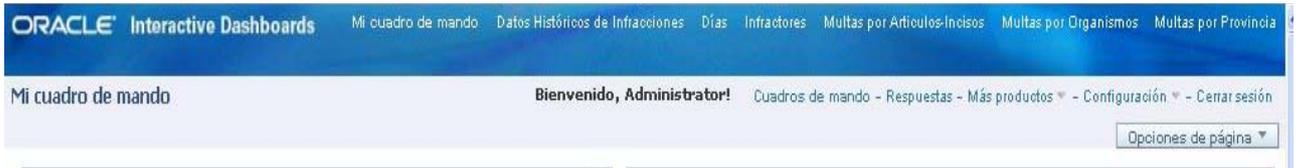


Figura 16: Menu

Capítulo 3. Construcción de la solución propuesta

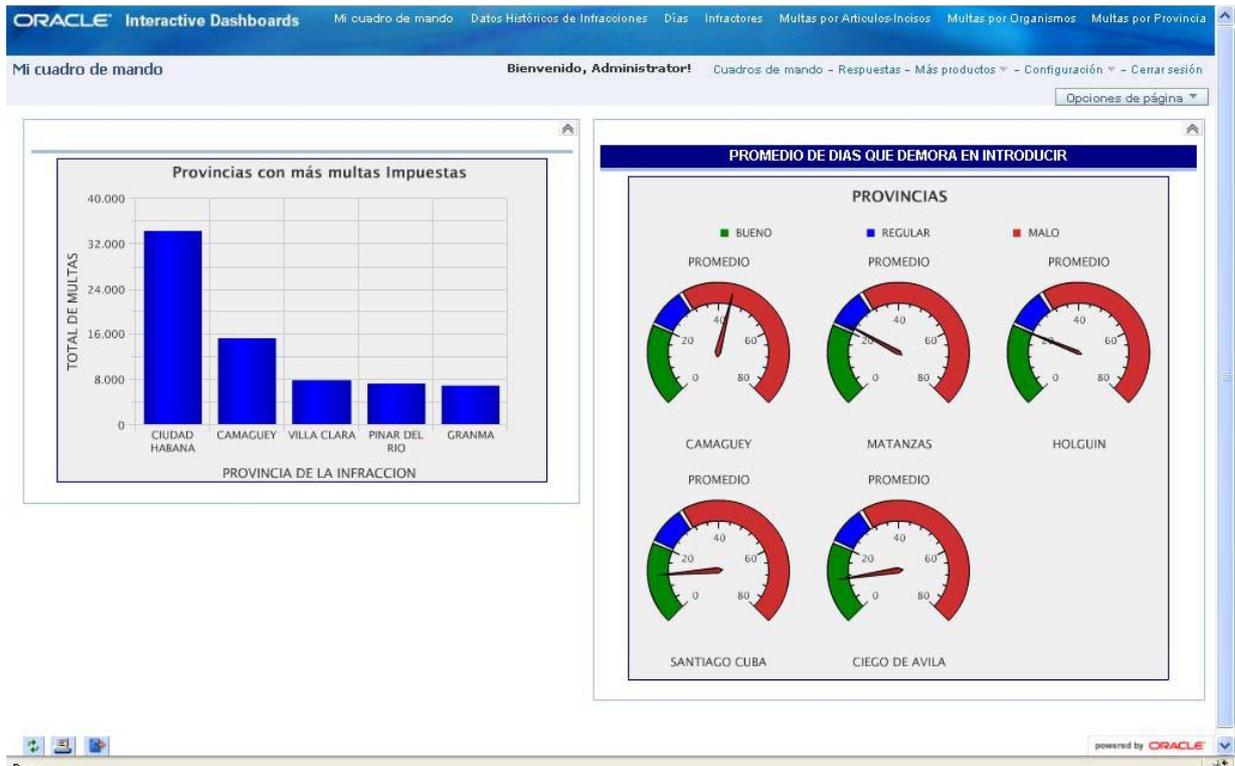


Figura 17: Panorama general

Datos Históricos de Infracciones

Bienvenido, Administrator! Cuadros de mando - Respuestas - Más productos - Configuración - Cerrar sesión

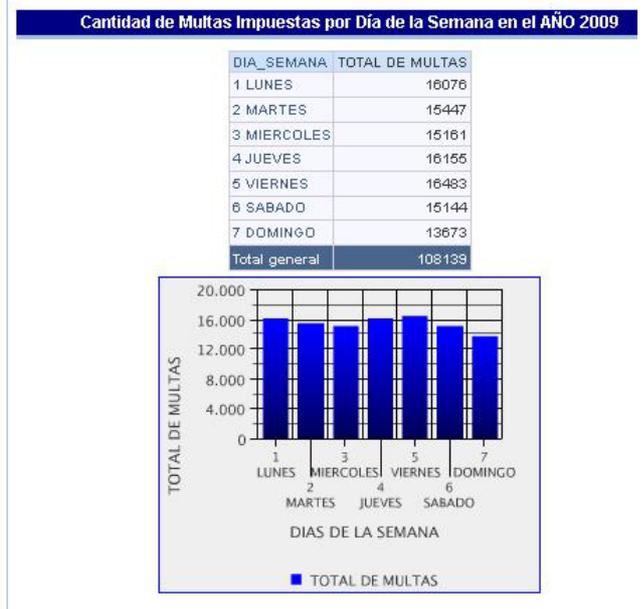
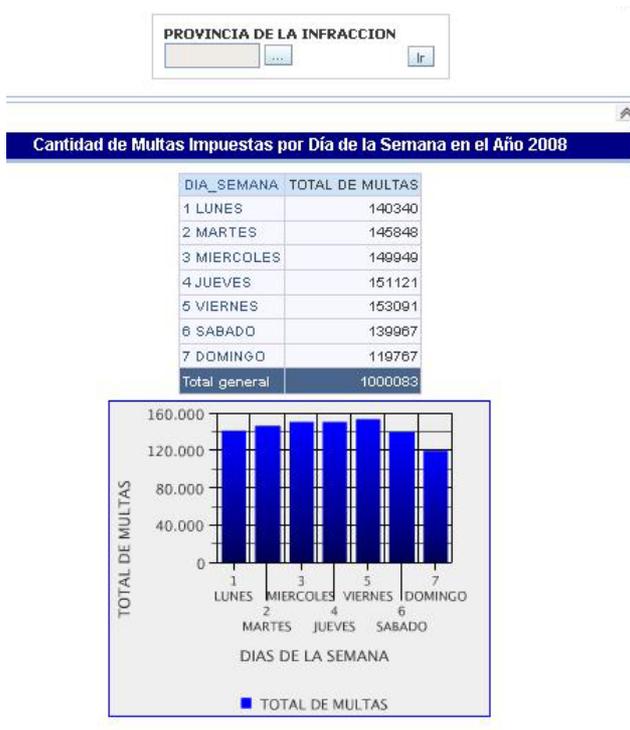
Infracciones | Advertidos | Multas por Dias de la Semana | Multas por Años | Multas por Meses | Multas impuestas en rango de fechas | Multas por meses en años | Total de multas por años, meses y por artículos

Introduzca los datos de una o varias Multas para devolver su Información

DIG: 'L' FOLIO: '875711' FECHA: '05/05/2007'

DIG	FOLIO	FECHA	PROVINCIA DE LA INFRACCION	ARTICULO-INCISO	TRAMO	EXPEDIENTE	UNIDAD QUE IMPONE	VALOR	GESTION	PAGADA	TIPO DE PAGO	ACUMULA PUNTOS	ADVERTIDO	TIEMPO INTRODUCIR	TIEMPO PAGAR	ORGANISMO EMPRESA	USUARIO QUE INSERTA	USUARIO QUE PAGA	USUARIO QUE A...
L	875711	5/05/07	CIUDAD HABANA	ARTICULO 117 INCISO 1	FUERA DE TRAMO	0	NO CONSTA	30	NO SE HA HECHO GESTION	HA SIDO PAGADA	TALON U OFICINA DE COBRO	ACUMULA	NO CONSTA	17	35	NO CONSTA	LOURDEST	ORLAIDA	NI CI

Capítulo 3. Construcción de la solución propuesta



Capítulo 3. Construcción de la solución propuesta

Oracle BI Interactive Dashboards - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://105.34.2.6:9704/analytics/saw.dll?Dashboard&_scid=91W7GPOv8IQ&PortalPath=/shared/Estadisticas/_f

AVG Search Active Surf-Shield Search-Shield AVG Info

Datos Históricos de Infracciones Bienvenido, Administrador! Cuadros de mando - Respuestas - Más productos - Configuración - Cerrar sesión

Infracciones Advertidos Multas por Día de la Semana **Multas por Meses** Multas impuestas en rango de fechas Multas por meses en años Total de multas por años, meses y por artículos Opciones de página

MES

Cantidad de Multas por Meses del Año 2008

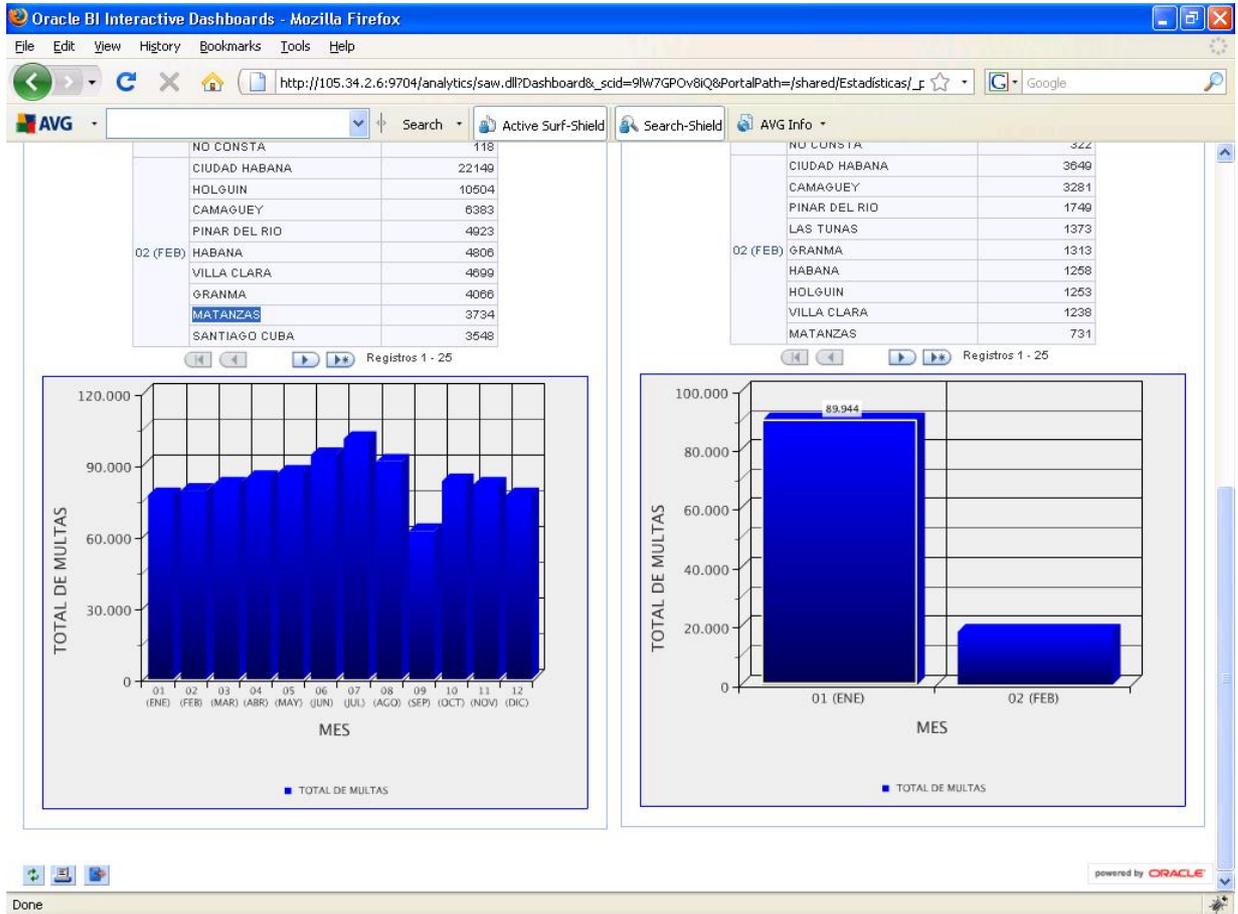
MES	PROVINCIA DE LA INFRACCION	TOTAL DE MULTAS
	CIUDAD HABANA	23476
	HOLGUIN	8254
	PINAR DEL RIO	5823
	VILLA CLARA	5420
	CAMAGUEY	5378
	MATANZAS	4769
	HABANA	4416
01 (ENE)	LAS TUNAS	Resultado
	SANCTI SPIRITUS	3347
	SANTIAGO CUBA	3027
	GRANMA	3026
	CIENFUEGOS	2642
	QUANTANAMO	2079
	CIEGO DE AVILA	1902
	ISLA DE LA JUV.	362
	NO CONSTA	118
02 (FEB)	CIUDAD HABANA	22149
	HOLGUIN	10504
	CAMAGUEY	6383
	PINAR DEL RIO	4923
	HABANA	4806
	VILLA CLARA	4699
	GRANMA	4066
	MATANZAS	3734

Cantidad de Multas por Meses del Año 2009

MES	PROVINCIA DE LA INFRACCION	TOTAL DE MULTAS
	CIUDAD HABANA	30523
	CAMAGUEY	12003
	VILLA CLARA	6619
	GRANMA	5590
	PINAR DEL RIO	5548
	HOLGUIN	4912
	LAS TUNAS	4371
01 (ENE)	HABANA	4282
	SANTIAGO CUBA	3292
	SANCTI SPIRITUS	2778
	MATANZAS	2695
	CIENFUEGOS	2526
	QUANTANAMO	2288
	CIEGO DE AVILA	1759
	ISLA DE LA JUV.	466
	NO CONSTA	322
02 (FEB)	CIUDAD HABANA	3649
	CAMAGUEY	3281
	PINAR DEL RIO	1740
	LAS TUNAS	1373
	GRANMA	1313
	HABANA	1258
	HOLGUIN	1253
	VILLA CLARA	1238

Done

Capítulo 3. Construcción de la solución propuesta



Anexo 2 – Modelo Lógico del Esquema Estrella

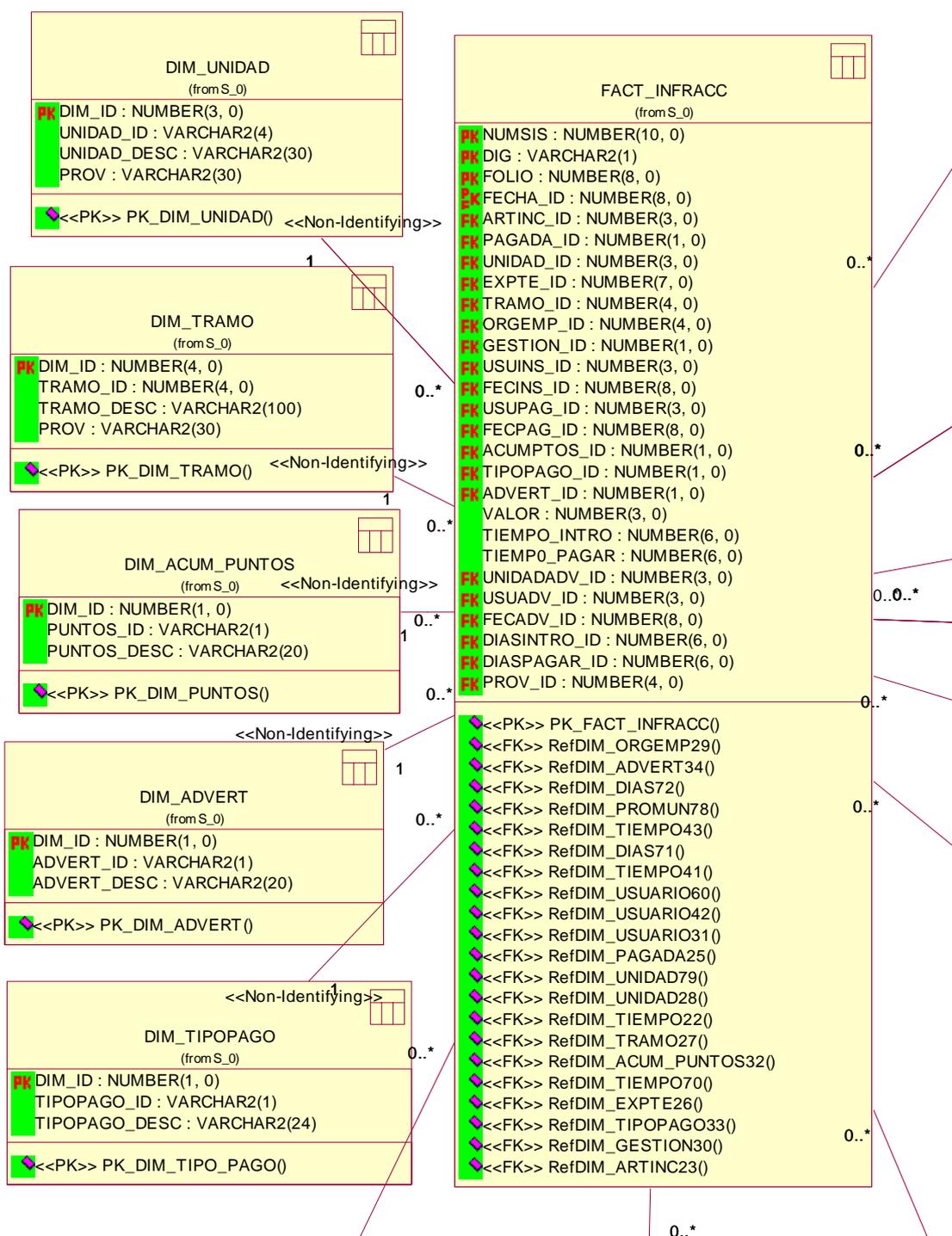


Figura 3. 5 Modelo Lógico del Esquema Estrella. Parte I

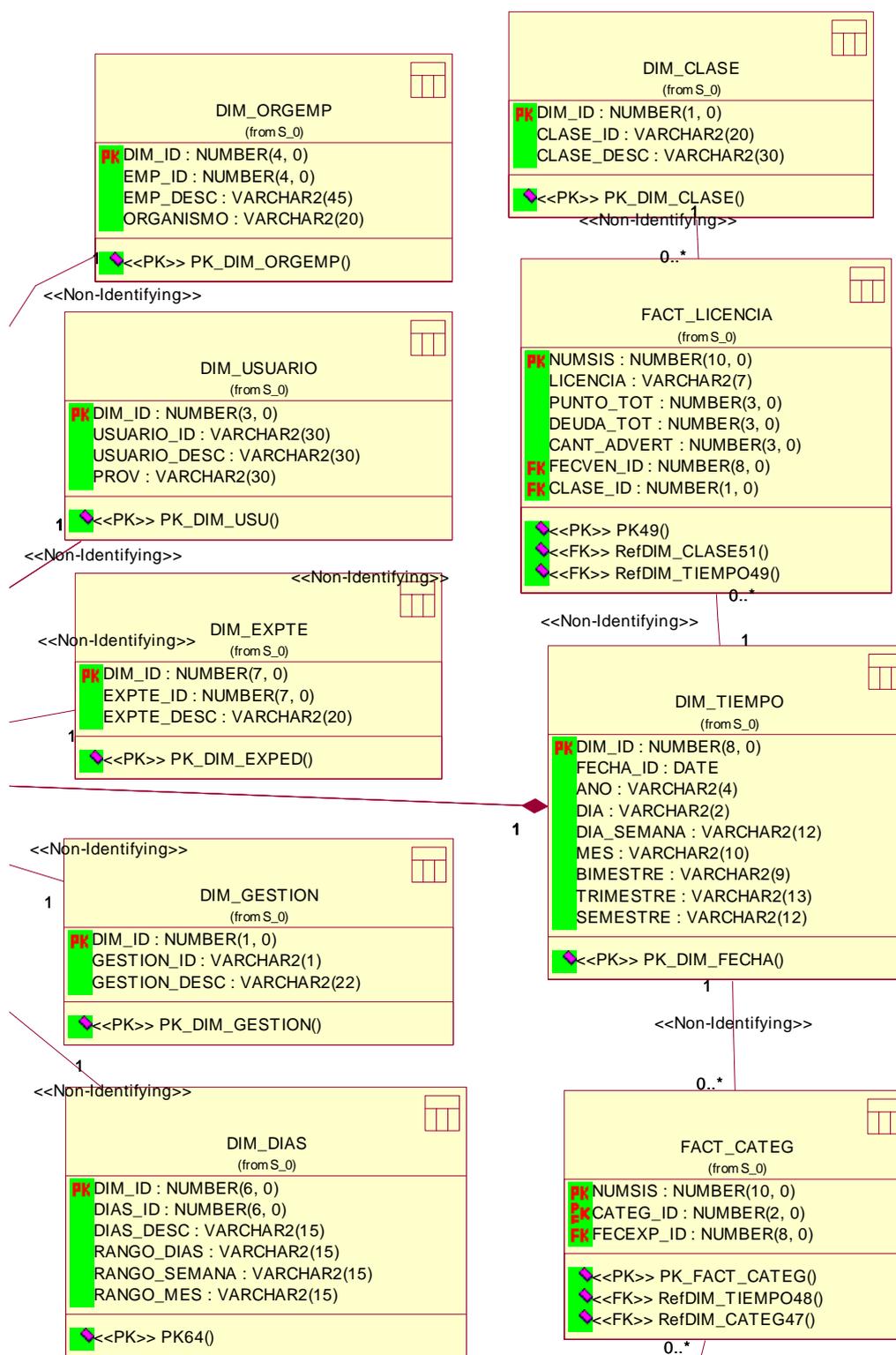


Figura 3. 6 Modelo Lógico del Esquema Estrella. Parte II

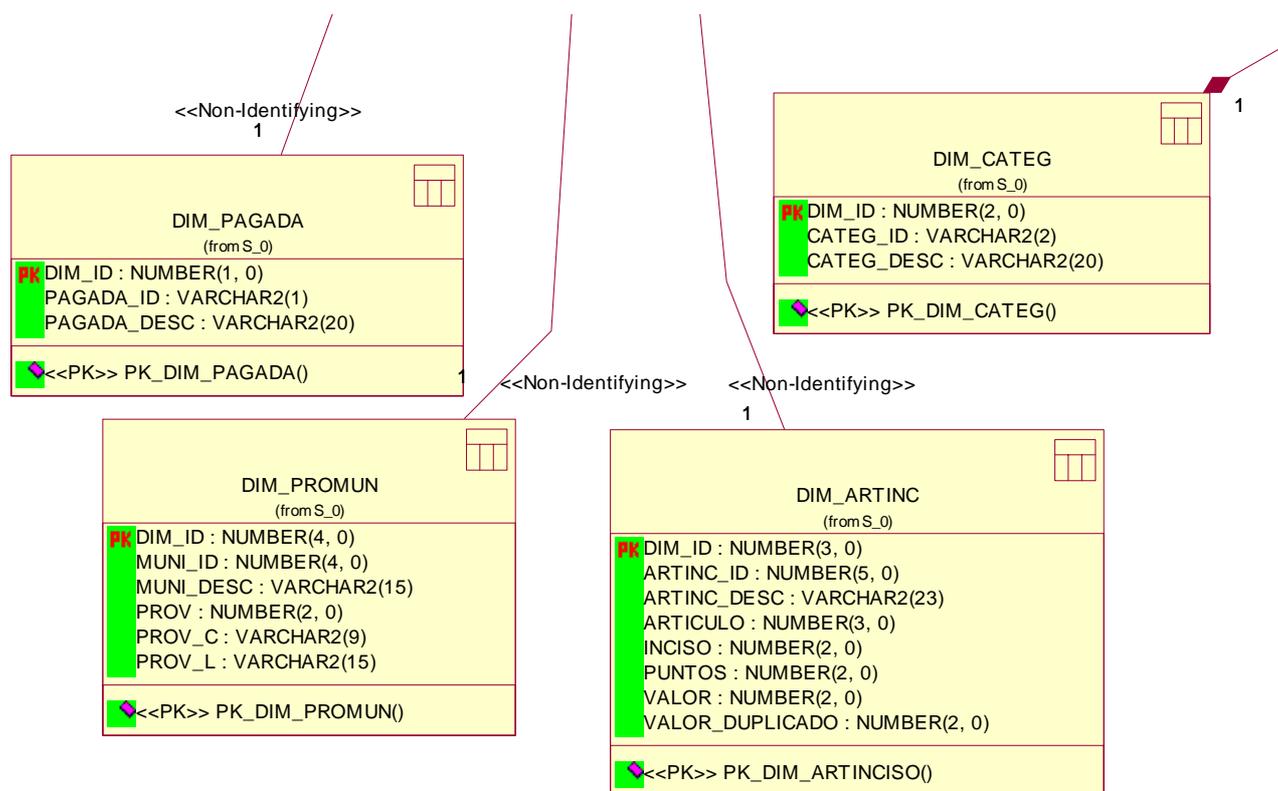


Figura 3. 7 Modelo Lógico del Esquema Estrella. Parte III

Anexo 3: Validación

Encuesta

Lee detenidamente cada afirmación y evalúala en la escala de muy bajo a muy alto (1 a 5).

1. La profundidad del contenido se adecua al usuario del producto.
1__ 2__ 3__ 4__
2. Logra motivar por su manejabilidad.
1__ 2__ 3__ 4__
3. El sistema cumple con el objetivo para el cual fue diseñado.
1__ 2__ 3__ 4__
4. Los colores empleados en el diseño de la aplicación son los adecuados.
1__ 2__ 3__ 4__
5. Con el sistema implementado se permite visualizar con mayor facilidad y rapidez la información referente a las infracciones del tránsito.
1__ 2__ 3__ 4__
6. El sistema facilita la comparación de los resultados obtenidos en diferentes períodos.
1__ 2__ 3__ 4__
7. A partir de la aplicación del sistema existe una mayor facilidad para el proceso de toma de decisiones sobre las infracciones del tránsito.
1__ 2__ 3__ 4__
8. Existe claridad en la explicación de los temas que se ofrecen en la ayuda.
1__ 2__ 3__ 4__
9. Se observa coherencia entre los objetivos del software y el contenido.
1__ 2__ 3__ 4__
10. La interactividad es apropiada para el usuario.
1__ 2__ 3__ 4__
11. Las instrucciones que se presentan al usuario son claras.
1__ 2__ 3__ 4__

12. Las vistas presentan una cantidad apropiada de información.

1__ 2__ 3__ 4__

13. La estructura de la presentación permite acceder sin dificultad a sus diferentes componentes.

1__ 2__ 3__ 4__

14. El tamaño y el tipo de letra empleados son apropiados.

1__ 2__ 3__ 4__

Observaciones:

Profundidad del contenido

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Muy Alto	4	100,0	100,0	100,0

Manejabilidad

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Alto	1	25,0	25,0	25,0
Muy Alto	3	75,0	75,0	100,0
Total	4	100,0	100,0	

Objetivo del sistema

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Muy Alto	4	100,0	100,0	100,0

Colores empleados

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Alto	2	50,0	50,0	50,0
Muy Alto	2	50,0	50,0	100,0
Total	4	100,0	100,0	

Facilidad y Rapidez para visualizar Información

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Alto	2	50,0	50,0	50,0
Muy Alto	2	50,0	50,0	100,0

Total	4	100,0	100,0
-------	---	-------	-------

Comparación de resultados

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Alto	1	25,0	25,0	25,0
Muy Alto	3	75,0	75,0	100,0
Total	4	100,0	100,0	

Facilidad en la toma de decisiones

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Alto	1	25,0	25,0	25,0
Muy Alto	3	75,0	75,0	100,0
Total	4	100,0	100,0	

Claridad de la ayuda

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Alto	1	25,0	25,0	25,0
Muy Alto	3	75,0	75,0	100,0
Total	4	100,0	100,0	

Coherencia entre objetivos y contenido del sistema

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Muy Alto	4	100,0	100,0	100,0

Interactividad con el usuario

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Alto	2	50,0	50,0	50,0
Muy Alto	2	50,0	50,0	100,0
Total	4	100,0	100,0	

Claridad de instrucciones

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Muy Alto	4	100,0	100,0	100,0

Cantidad de información que tienen las vistas

	Frequency	Percent	Valid Percent	Cumulative Percent
--	-----------	---------	---------------	--------------------

Valid	Muy Alto	4	100,0	100,0	100,0
-------	----------	---	-------	-------	-------

Accesibilidad a diferentes componentes

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Muy Alto	4	100,0	100,0	100,0

Tamaño y tipo de letra

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Muy Alto	4	100,0	100,0	100,0

Anexo 3.1: Encuesta