

UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS

FACULTAD 7 - MININT

TÍTULO: *Diseño del Datamart del subsistema de Conducidos.*

TRABAJO DE DIPLOMA PARA OPTAR POR EL TÍTULO DE
INGENIERO EN CIENCIAS INFORMÁTICAS.

AUTOR: *Alfredo Sifontes Mavilio.*

TUTORES: *Cap. Javier Fernández Barrero.*

My. Manuel Cárdenas Cabrera.

ASESOR (ES): *Msc. Maritza Mavilio Ortega.*

... "El vicio tiene tantos cómplices en el mundo, que es necesario que tenga algunos cómplices la virtud. Se puede ser, y se debe ser cómplice de la virtud. Al corazón se le han de poner alas, no anclas." (1)

José Martí

DATOS DE CONTACTO

CPT. Javier Fernández Barrero

Especialista A en Informática. Empresa DATYS. MININT.

e-mail: javier.fernandez@datys.cu

MY. Manuel Cárdenas Cabrerías

Jefe Grupo de Diseño Informático. MININT Sta. Clara.

e-mail:

AGRADECIMIENTOS

A mi familia, por su apoyo, su amor y por estar siempre presente en todos los momentos de mi vida.

A mi novia Karin, por llenar de amor y alegría todos los días de mi vida.

A todos mis amigos, por estar siempre conmigo y formar parte de todo lo bueno que me rodea.

A mi tutor, por toda la ayuda prestada y por haber sabido guiarme con sabiduría durante todo el trayecto del trabajo de diploma.

A todo el equipo de trabajo que apoyó el desarrollo de la tesis.

A todo el que de una forma u otra ha hecho posible la realización de este sueño.

Al Ministerio del Interior MININT por permitirme formarme dentro de sus gloriosas filas.

A la Revolución y en especial al Comandante Fidel por haber apostado y creado este gran proyecto de futuro que hoy es la UCI.

DEDICATORIA

A mi familia, cariñosamente a mi novia y todos mis amigos.

RESUMEN

El presente trabajo tiene como objetivo desarrollar un datamart para tener una reportería eficiente relacionada con el subsistema de los conducidos del sistema integral de la estación PNR. Surge por la necesidad que posee la dirección de la estación PNR de monitorear diversas estadísticas relacionadas con el subsistema de los conducidos para trazar estrategias a seguir en el mejoramiento de la eficiencia y eficacia del trabajo policial a diferentes niveles en nuestro país.

ÍNDICE

Introducción	1
CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA.	8
1.1. Sistema Gestor de Bases de Datos Oracle.....	8
1.1.1. La plataforma.....	8
1.1.2. La Base de Datos.....	9
1.1.3. Developer	9
1.1.4. Comunicaciones.....	10
1.2. Tecnología Datawarehouse.	11
1.2.1. Objetivos fundamentales de un Datawarehouse.....	11
1.2.2. Características del Datawarehouse	13
1.2.3. Los procesos básicos del Datawarehouse (ETL)	14
1.2.4. Base de datos multidimensional	15
1.2.5. Proceso OLAP.....	18
1.2.6. ROLAP	20
1.2.7. MOLAP	21
1.2.8. HOLAP	21
1.2.9. Sistemas existentes a nivel internacional.....	21
1.2.10. Sistemas existentes a nivel nacional	22
1.3. Tecnología Datamart	23
1.4. Fundamentación de la metodología a utilizar.....	25
1.5. Sistema Integral Informativo de la Estación P.N.R.	27
1.6. Herramientas propuestas para el desarrollo del Datamart del subsistema de Conducidos.....	30
CAPÍTULO 2. CARACTERÍSTICAS DEL SISTEMA.....	32
2.1. Análisis del proceso	32
2.2. Análisis de requerimientos.....	33
2.2.1. Indicadores	33

2.3. Modelo conceptual.....	35
2.4. Análisis de los OLTP	35
2.5. Elaboración del modelo lógico del Datamart	37
2.6. Despliegue de la solución.	46
CONCLUSIONES	48
RECOMENDACIONES	49
REFERENCIAS BIBLIOGRÁFICAS	50
BIBLIOGRAFÍA	51

INTRODUCCIÓN

En el mundo contemporáneo existen numerosos desafíos en cuanto al uso adecuado de las tecnologías para mejorar la vida social. El vertiginoso avance de la informática y las comunicaciones, provoca que la informatización de la sociedad constituya un instrumento indispensable para avanzar hacia un mundo de justicia social en el cual no prime la Globalización Neoliberal. Con este fin, en el año 1997 fueron definidos los “Lineamientos estratégicos para la informatización de la Sociedad Cubana” por el Comité Ejecutivo del Consejo de Ministros de Cuba y dos años después comienza un amplio programa de acciones.

En pos de cumplir dichos objetivos, se instrumenta un proceso continuo de transformación y desarrollo de la gestión de la información y el conocimiento en la Dirección de Informática del Ministerio del Interior para lograr la excelencia de las tareas de este ministerio, abarcando una serie de cuestiones que van desde los sistemas de información hasta la capacitación del capital humano, la seguridad informática y las redes.

El reto planteado por el líder de la Revolución Fidel Castro, en el aniversario 40 de la PNR, de llegar a ser la más profesional del mundo, sigue siendo bandera de lucha.

Fundado el 6 de junio de 1961, el Ministerio del Interior ha sido un infranqueable escudo a la incesante y despiadada agresión de los enemigos de la Revolución.

Desde el día del triunfo e incluso desde antes, la Revolución no descuidó "ni tantito así", como diría el Che, el deber primero de defenderse de sus enemigos.

En previsión de la guerra que inevitablemente vendría después de la victoria, desde la Sierra Maestra comenzaron a organizarse futuros pilares de la defensa de la Revolución: las Fuerzas Armadas Revolucionarias, derivadas del propio

Ejército Rebelde; la Policía Nacional Revolucionaria y la Seguridad del Estado, con antecedentes en el Servicio de Inteligencia Rebelde (SIR), el Departamento de Inteligencia del Ejército Rebelde (DIER), y la Policía Rebelde.

Nacidas el cinco de enero de 1959, a unos días del triunfo popular, las fuerzas de la Policía Nacional Revolucionaria han sido garantía de tranquilidad ciudadana y bastión importante en la preservación de las conquistas del país ante las acciones de la contrarrevolución interna, las amenazas y agresiones del imperialismo.

Cuba debía entonces formar una Policía de nuevo tipo, pues no se trataba del cuerpo armado diseñado para defender a los poderosos y reprimir a las mayorías humildes. Convertirse en instrumento del poder del pueblo para defender la Patria en cualquier circunstancia.

Así, los policías estuvieron junto con los milicianos y el Ejército Rebelde en las primeras movilizaciones de los años 60: Playa Girón, la Lucha contra Bandidos, la Crisis de Octubre, así como contra las numerosas acciones terroristas e infiltración de agentes enemigos, y en cada lugar donde se defendía la Patria. En el cumplimiento de esas misiones numerosos miembros de ese Cuerpo ofrendaron su vida.

Al cambiar la esencia del trabajo policial, como consecuencia de las profundas transformaciones que han tenido lugar en el país, aparecieron nuevas manifestaciones delictivas, las cuales exigieron una labor constante de superación técnica y profesional, para situarse a la altura de las exigencias contemporáneas.

La tarea no ha sido fácil, pues sin detenerse recorrieron el largo y complejo camino de la superación como única vía para alcanzar el dominio de los medios modernos puestos a su disposición y dominaron los métodos científicos en las diferentes especialidades que integran la institución.

La Estación PNR es la organización básica del sistema de enfrentamiento a la criminalidad y la más próxima a la comunidad, en ella se materializan las funciones principales de la Policía y constituye el sistema integral para el enfrentamiento a los fenómenos delictivos y la antisocialidad en una demarcación que, como regla general, coincide con el territorio del municipio.

La Estación PNR concreta su labor de Prevención en: la obstaculización de los hechos delictivos y de las ilegalidades, la neutralización de los delincuentes, la influencia positiva sobre las personas con conductas desviadas, la disminución de las causas y condiciones que favorecen las actividades ilegales y en la creación de un clímax de tranquilidad ciudadana.

En relación con el desarrollo de la informática y las comunicaciones y los sistemas automatizados, la dirección de la PNR se ha dado a la tarea de desarrollar y perfeccionar estos avances de la ciencia y la técnica. En la actualidad existe un sistema automatizado que informatiza gran parte de los procesos de trabajo de la estación PNR. En este sentido se desarrollan importantes investigaciones entre las que podemos señalar:

- Administración y desarrollo de aplicaciones en base de datos Oracle del Mayor Manuel Cárdenas.
- Diseño y construcción de datawarehouse y trabajo con las herramientas de inteligencia de negocios de Oracle del Capitán Javier Fernández.

A pesar de lo antes expuesto se evidencia la siguiente **situación problemática**: Necesidad de perfeccionar el actual sistema de gestión de la información acorde con los reajustes, procedimientos y procesos de trabajo actuales, que eleven la eficiencia del trabajo policial y de la dirección de la Estación PNR. Existen limitaciones en la explotación eficiente del actual sistema, en la preparación de la fuerza técnica en las estaciones para administrarlo y mantenerlo, en la agilidad y rapidez del flujo de información a la dirección de la Estación PNR.

Por lo antes expuesto nos planteamos el siguiente **problema a resolver**:
¿Cómo lograr mayor eficiencia en el proceso de gestión de la información en el sistema informático de la Estación PNR?

Se define como **objeto de estudio** el proceso de gestión de la información en la Estación PNR.

El **objetivo general** de esta investigación es diseñar un sistema informático de datos de conducidos que facilite el proceso de gestión de la información de la Estación PNR.

El **campo de acción** es el subsistema informático de los conducidos de la Estación PNR.

Para dar cumplimiento al objetivo general de esta investigación, se plantean las siguientes **tareas de investigación**:

1. Sistematización de los fundamentos teóricos sobre gestión de la información de los sistemas informáticos de la Estación PNR.
2. Caracterización del datamart de hechos denunciados de la Estación PNR en el proceso de gestión de la información.
3. Acercamiento al Escenario de Trabajo:
4. Recibir un ciclo de conferencias sobre la estructura y los sistemas de trabajo de la Estación PNR.
5. Recibir un ciclo de conferencias sobre los sistemas automatizados existentes en la Estación PNR y los que se incluirán en un plazo corto.
6. Recibir un ciclo de conferencias sobre lo realizado en el Datamart de Hechos Denunciados: Sajo y los dashboards construidos.
7. Incorporarse al Grupo de Especialistas de la PNR de las distintas Áreas que trabajan en la identificación de los indicadores de trabajo.
8. Acercamiento a las Tecnologías de Trabajo:
9. Recibir un ciclo de Conferencias sobre las herramientas de Inteligencia de Negocios de Oracle

10.Recepción, organización de la documentación que constituye base material de estudio y elaborar plan individual de asimilación tecnológica incluyendo:

- ✓ La Tecnología MOLAP y ROLAP de Oracle.
- ✓ Los Utilitarios de diseño y ETL
- ✓ La Suite de Inteligencia de Negocios BIEE.

11.Elaboración del datamart de conducidos para la Estación de la PNR.

12.Constatación de viabilidad del datamart de conducidos para la Estación de la PNR.

El sistema propuesto será de gran utilidad para la gestión de la información relacionada con los conducidos hacia la Estación P.N.R y por consiguiente facilitará la toma de decisiones de la dirección de la misma, se lograrán los siguientes **aportes prácticos**:

- Obtención de la relación de Indicadores y sus reglas de cálculo sobre la Eficacia y Eficiencia del accionar policiaco y de los procesos y sistemas de trabajo de la Estación PNR en las circunstancias actuales.
- Obtención del diseño conceptual de los cubos y dimensiones del Datamart de conducidos.
- Obtención de la implementación, (incluyendo la reportería de dashboards), del Datamart de conducidos.
- Alcanzar la asimilación de conocimientos y las habilidades de uso sobre las tecnologías utilizadas en la tesis con vistas a dar continuidad al proyecto.

Para llevar a cabo la investigación se utilizaron los siguientes **Métodos teóricos**:

El **Analítico – Sintético** se utiliza en la revisión de documentos, libros, artículos e informes para la extracción de elementos importantes que están relacionados

con la investigación, para lograr una mayor visión sobre el tema, permitiendo tener un enfoque global del mismo.

El **Inductivo – Deductivo** se utiliza en el estudio de los diferentes temas relacionados con el objeto de estudio.

Como **Métodos Empíricos** se realizarán **entrevistas** y **observaciones** para la recopilación de información. Se extraerán diferentes puntos de vistas sobre el desarrollo y las mejores prácticas adquiridas en diferentes proyectos desarrollados en el MININT. Para conocer la viabilidad del proceso de gestión de la información en la Estación P.N.R.

El presente trabajo está estructurado por **3 capítulos**, distribuidos de la siguiente manera:

Capítulo 1. Fundamentación Teórica.

En el capítulo 1 abordaremos elementos teóricos relacionados con las características y estructuras del Gestor de Bases de Datos Oracle y la tecnología Datawarehouse. Ofrece una descripción de los sistemas informáticos que utilizan esta tecnología a nivel Internacional, Nacional y específicamente en el Ministerio del Interior, así como las herramientas que utilizaremos para dar solución al problema planteado.

Capítulo 2. Características del sistema.

En el capítulo 2 se realizará un análisis de las características del sistema a desarrollar, haciendo hincapié en la situación problemática que da origen al mismo. Se describe el flujo de trabajo de la Estación PNR comprendido dentro del campo de acción de la presente investigación, se detallan las necesidades de los usuarios, describiéndose las funcionalidades que serán objeto de automatización. Finalmente se presentará una propuesta del sistema a desarrollar, especificando detalladamente las dimensiones utilizadas para dar

solución al problema así como también el diseño del modelo conceptual y el esquema estrella utilizado para representar el modelo lógico del datamart del subsistema de los conducidos.

CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA.

Introducción.

En el presente capítulo se abordarán elementos teóricos relacionados con las características y estructuras del Gestor de Bases de Datos Oracle y la tecnología Datawarehouse. Ofrece una descripción de los sistemas informáticos que utilizan esta tecnología a nivel Internacional, Nacional y específicamente en el Ministerio del Interior, así como las herramientas que utilizaremos para dar solución al problema planteado.

1.1. Sistema Gestor de Bases de Datos Oracle.

Oracle es un sistema gestor de base de datos relacional de última generación, lo cual quiere decir que está orientado al acceso remoto y redes (internet). Hoy por hoy Oracle se puede implementar en diferentes plataformas: Familia de Microsoft, Unix, Linux, Vms, entre otras.

Las arquitecturas en las que se asienta pueden ser: Intel, Alpha, Sparc, Risc, entre otros a nivel de procesadores. Oracle es perfectamente configurable en entornos "OLTP", paralelos, Clúster, e incluso resulta una genial solución a nivel de Datawarehouse y CRM.

1.1.1. La plataforma

La plataforma para la que en realidad está desarrollada Oracle, y pensada desde un principio es Unix (sea cual sea el fabricante), ya que es un sistema que soporta la mayoría de la carga de los sistemas a nivel mundial, así como es un sistema abierto y configurable.

Para la gestión, sea del negocio que sea, Oracle y Unix se complementan de una manera casi inalcanzable para otros sistemas, ya que para su administración, Oracle se amolda y apoya en la potencia de Unix, a la vez que

aprovecha al 100% su versatilidad y configurabilidad. Con respecto a Windows, debemos decir que Oracle ha sido capaz de adaptarse de tal manera que hace firme competencia a la base de datos propia de Microsoft "SQL Server". Son muchas las ventajas y pocas las desventajas que presenta Oracle frente a SQL-server, pero eso es un tema que se escapa de nuestro interés en este momento. La instalación de Oracle, administración, configuración e inclusive la implementación de herramientas visuales, en un entorno Windows, se ha conseguido simplificar al máximo para conseguir dar un servicio óptimo a los usuarios de los sistemas Windows.

1.1.2. La Base de Datos

Oracle, como base de datos, se ha comido literalmente el mercado de su sector. Puesto que ofrecen una solución integral para resolver cualquier cuestión empresarial, Oracle se ha convertido en el software "a batir". Las últimas versiones de la B.D 8.i, 9i se han convertido en la gran aventura de Oracle hacia internet, ofreciendo una interfaz de última tecnología basado en Java y XML, así como un servidor adjunto de aplicaciones para internet, email y seguridad de datos. Se ha mejorado el kernel de la B.D implementado constantemente soluciones para la seguridad y rechazo a usuarios "no deseados" en el sistema, así como una estabilidad y escalabilidad a prueba de bombas: "Clúster". Oracle ofrece así mismo una solución "interactiva" para la solución de problemas, que resulta ser una fuente inagotable de sabiduría, comandada por los DBA`s más experimentados del mundo, a través de lo que se denomina como "Metalink"; es una web en la cual se puede exponer cualquier fallo, duda, sugerencia e incluso se pueden buscar soluciones a problemas ya solucionados anteriormente.

1.1.3. Developer

Oracle, con Developer lo que busca es la solución informática respecto del interface gráfico de acceso a datos.

Forms: Es un módulo de programación destinado a la construcción del interface gráfico y del cuerpo del programa de acceso a datos.

Graphics: Es otro módulo de programación destinado a la fabricación de gráficos de estadísticas con respecto a la información contenida o no en una B.D Oracle.

Procedure Builder: Herramienta destinada a la construcción, depuración y compilación de procedimientos y demás objetos de B.D.

Project Builder: Herramienta destinada a la planificación de trabajos con respecto al desarrollo en un entorno de programación orientado a B.D.

Existen más módulos dentro del paquete Developer, pero estos son los más destacados, o los más usados por la comunidad de profesionales de Oracle. Oracle Developer 6i ofrece ventajas con respecto a sus competidores como puede ser: Un estándar de programación plenamente adaptado a las normas de SQL Oracle; un entorno de programación mejorado para facilitar el desarrollo y la posibilidad de desarrollar junto al servidor IAS (Internet Application Server) una solución para internet; todo ello, siguiendo un estándar, y una normalización de la seguridad de acceso a datos.

1.1.4. Comunicaciones

Oracle, se puede implementar en cliente/servidor con muchas arquitecturas de red, pero las más usadas con TCP e IPX/SPX. La razón de TPC es clara: es el estándar de red usado a nivel internacional en internet. En el caso de IPX resulta de la compatibilidad de Oracle para su funcionamiento con Novel/NetWare.

Oracle, posee su propio lenguaje de red, que se asienta de igual manera sobre casi cualquier protocolo; este es Net8 (antiguo SQL-Net). Este protocolo permite la configuración e implementación sobre otros protocolos debido a su versatilidad, es decir, se adapta a los tamaños de tramas de red, y resulta una buena solución de envío y recepción de datos en cualquier red a nivel LAN o incluso WAN.

Oracle, para poder ver la luz por una red WAN o RAS, bien sea internet, o una red privada virtual, requiere de una herramienta denominada IAS. Esta herramienta es configurada en un servidor y sirve de enlace entre el software que ataca la base de datos y los usuarios, a través de un router, configurando los caminos, y las puertas de enlace.

Actualmente para realizar una instalación de Oracle que nos permita acceder a bases de datos Cliente/Servidor, necesitaremos bastante memoria RAM, unos 256 Mb dedicados, junto a unos 10gb de Disco, así como tener instalado en nuestro sistema (Windows), la máquina virtual de java, y ser administradores de la misma. Requiere de mucha memoria RAM, debido a la interfaz gráfica desarrollada en Java.

1.2. Tecnología Datawarehouse.

El Datawarehouse o Almacén de Datos. Surgió con el objetivo de hacer consultable la información que se tiene de una empresa tanto de meses como de años anteriores. Es un almacenamiento simple, completo y consistente de datos obtenidos desde una variedad de fuentes, disponibles para el usuario final de forma tal que puedan entenderlos y utilizarlos. Facilita la integración de sistemas de aplicación no integrados. Un Datawarehouse soporta el procesamiento informático al proveer una plataforma sólida, a partir de los datos históricos para hacer el análisis que se necesita de la información. El objetivo fundamental de un Datawarehouse, es asistir al usuario en el entendimiento del pasado y contar con los elementos para la planeación del futuro de corto, mediano y largo plazo. Es actualmente, el centro de atención de las grandes instituciones, porque provee un ambiente para que las organizaciones hagan un mejor uso de la información que está siendo registrada por las diversas aplicaciones operacionales.

1.2.1. Objetivos fundamentales de un Datawarehouse.

Los objetivos fundamentales de un Datawarehouse son:

- **Hace que la información de la organización sea accesible:**

Los contenidos del Datawarehouse son entendibles y navegables, y el acceso a ellos es caracterizado por el rápido desempeño. Estos requerimientos no tienen fronteras y tampoco límites fijos. Cuando hablamos de entendible significa, que los niveles de la información sean correctos y obvios. Navegables significa el reconocer el destino en la pantalla y llegar a donde queramos con solo un clic. Rápido desempeño significa, cero tiempo de espera.

- **Hacer que la información de la organización sea consistente:**

La información de una parte de la organización puede hacerse coincidir con la información de la otra parte de la organización. Si dos medidas de la organización tienen el mismo nombre, entonces deben significar la misma cosa. Y a la inversa, si dos medidas no significan la misma cosa, entonces son etiquetados diferentes. Información consistente significa, información de alta calidad. Significa que toda la información es contabilizada y completada.

- **Es información adaptable y elástica:**

El Datawarehouse está diseñado para cambios continuos. Cuando se le hacen nuevas preguntas al Datawarehouse, los datos existentes y las tecnologías no cambian ni se corrompen. Cuando se agregan datos nuevos al Datawarehouse, los datos existentes y las tecnologías tampoco cambian ni se corrompen.

- **Es la fundación de la toma de decisiones:**

El Datawarehouse tiene los datos correctos para soportar la toma de decisiones. Solo hay una salida verdadera del Datawarehouse: las decisiones que son hechas después de que el Datawarehouse haya presentado las evidencias. La original etiqueta que preside el Datawarehouse sigue siendo la mejor descripción de lo que queremos construir: un sistema de soporte a las decisiones.

1.2.2. Características del Datawarehouse

Entre sus principales características se encuentran:

- **Orientado al tema**

Una primera característica del Datawarehouse es que la información se clasifica en base a los aspectos que son de interés para la empresa. Siendo así, los datos tomados están en contraste con los clásicos procesos orientados a las aplicaciones.

- **Integrado**

Integra datos recolectados de diferentes sistemas operacionales de la organización y o fuentes externas.

- **De tiempo variante**

Los datos son relativos a un período de tiempo y estos deben ser integrados periódicamente, los mismos son almacenados como fotos que se corresponden a un período de tiempo. Toda la información del Datawarehouse es requerida en algún momento. Esta característica básica de los datos en un depósito, es muy diferente de la información encontrada en el ambiente operacional. En estos, la información se requiere al momento de acceder. En otras palabras, en el ambiente operacional, cuando usted accede a una unidad de información, usted espera que los valores requeridos se obtengan a partir del momento de acceso. Como la información en el Datawarehouse es solicitada en cualquier momento (es decir, no "ahora mismo"), los datos encontrados en el depósito se llaman de "tiempo variante".

- **No volátil**

Los datos que son almacenados no sufren ninguna actualización solo son incrementados. El período cubierto para un DWH va de 2 a 10 años.

1.2.3. Los procesos básicos del Datawarehouse (ETL)

Extracción: Este es el primer paso de obtener la información hacia el ambiente del Datawarehouse. La extracción convierte los datos a un formato preparado para iniciar el proceso de transformación. Una parte del proceso de extracción es la de analizar los datos extraídos, de lo que resulta un chequeo que verifica si los datos cumplen la pauta o estructura que se esperaba, de no ser así, los datos son rechazados. Un aspecto importante que se debe exigir a la tarea de extracción es que ésta produzca un impacto mínimo en el sistema origen. Debido a esto se han tomado diferentes medidas, por ejemplo, si los datos a extraer son muchos, el sistema de origen se podría ralentizar e incluso colapsar, provocando que este no pueda utilizarse con normalidad para su uso cotidiano., por esta razón, en sistemas grandes las operaciones de extracción suelen programarse en horarios o días donde este impacto sea nulo o mínimo.

Transformación: Una vez que la información es extraída hacia el área de tráfico de datos, hay posibles paso de transformación como; limpieza de la información, tirar a la basura lo que no nos sirve, seleccionar únicamente los campos necesarios para el Datawarehouse, combinar fuentes de datos, haciéndolas coincidir por los valores de las llaves, creando nuevas llaves para cada registro de una dimensión. La fase de transformación aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados. Algunas fuentes de datos requerirán alguna pequeña manipulación de los datos.

No obstante en otros casos pueden ser necesarias aplicar algunas de las siguientes transformaciones:

- Seleccionar solo ciertas columnas para su carga (Ej. que las columnas con valores nulos no se carguen).

- Traducir códigos (Ej. Si la fuente almacena una "H" para Hombre y "M" para Mujer pero el destino tiene que guardar "1" para Hombre y "2" para Mujer).
- Codificar valores libres (Ej. convertir "Hombre" en "H" o "Sr" en "1").
- Obtener nuevos valores calculados (Ej. total _ venta = cantidad * precio).

Carga: Al final del proceso de transformación, los datos están en forma para ser cargados. Los Datawarehouse mantienen un historial de los registros de manera que se pueda hacer una auditoria de los mismos y disponer de un rastro de toda la historia de un valor a lo largo del tiempo.

1.2.4. Base de datos multidimensional

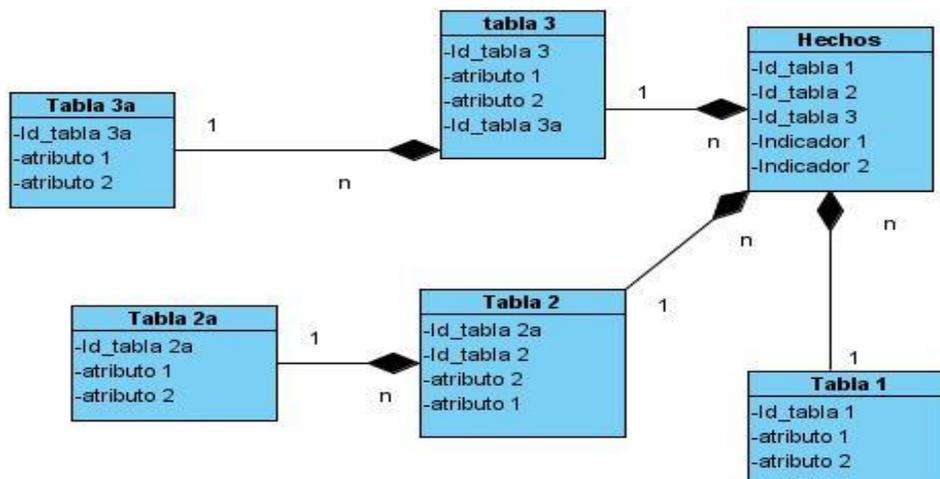
Los Datawarehouse gestionan el depósito de datos y lo organizan en torno a una base de datos multidimensional que tal y como lo indica su nombre almacena los datos en diversas dimensiones, que conforman un cubo multidimensional, en donde el cruce de los valores de los tributos de cada dimensión a lo largo de las abscisas, determinan un hecho específico. Los cálculos que se aplican sobre las dimensiones son matriciales, los cuales se procesan dando como resultado reportes tabulares.

Las bases de datos multidimensionales implican tres variantes posibles de modelamiento, que permiten realizar consultas de soporte de decisión:

- Esquema copo de nieve (Snowflake Scheme).
- Esquema constelación o copo de estrellas (Starflake Scheme).
- Esquema en estrella (Star Scheme).

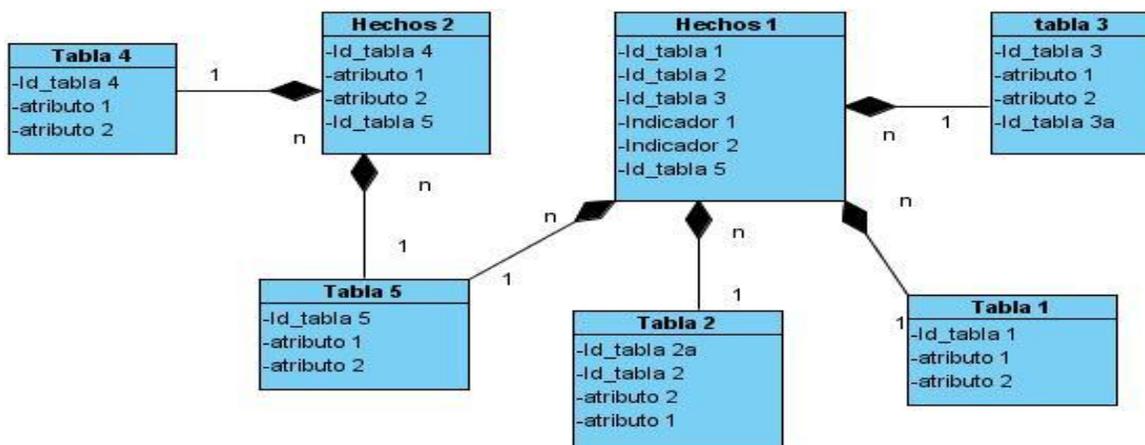
Esquema Copo de Nieve

Este esquema representa una extensión del modelo en estrella cuando las dimensiones se organizan en jerarquías de dimensiones.



Esquema Constelación

Este modelo está compuesto por una serie de esquemas en estrella, y tal como se puede apreciar en la siguiente figura, está formado por una tabla de hechos principal y por una o más tablas de hechos auxiliares las cuales pueden ser sumalizaciones de la principal. Dichas tablas yacen en el centro del modelo y están relacionadas con sus respectivas tablas de dimensiones.



Esquema en Estrella

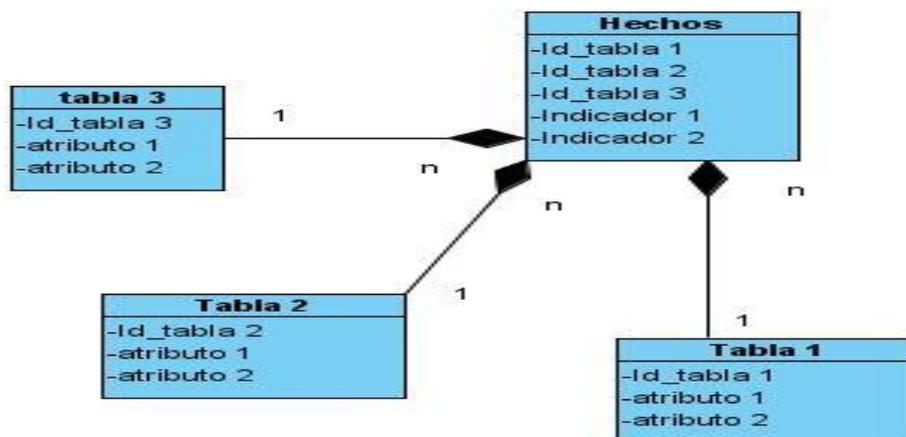
Se definió como variante el **Esquema en Estrella**: Consta de una **tabla de hechos** central y de varias tablas de dimensiones relacionadas a esta, a través de sus respectivas claves.

Tablas de Hechos

Las tablas de hechos contienen los hechos, medidas o indicadores que serán utilizados por los analistas de negocio para apoyar el proceso de toma de decisiones, donde los hechos son datos instantáneos en el tiempo.

El esquema en estrella, consta de una tabla de hechos central y de varias tablas de dimensiones relacionadas a esta, a través de sus respectivas claves. El esquema en estrella es el más simple de interpretar y optimiza los tiempos de respuesta ante las consultas de los usuarios. Este modelo es soportado por casi todas las herramientas de consulta y análisis, y los metadatos son fáciles de documentar y mantener.

En la siguiente figura se puede apreciar un esquema en estrella estándar:



A continuación se destacarán algunas características de este modelo, que ayudarán a comprender mejor el por qué de sus ventajas y de nuestra selección para la confección de la base de datos multidimensional:

- Posee los mejores tiempos de respuesta.

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

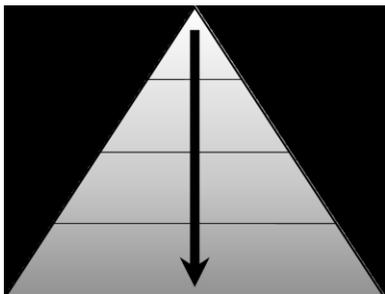
1.2.5. Proceso OLAP

El procesamiento analítico en línea OLAP (On Line Analytic Processing), es la componente más poderosa de los DWH, ya que es el motor de consultas especializado del almacén de datos. Las herramientas OLAP, son una tecnología de software para análisis en línea, administración y ejecución de consultas, que permiten inferir información del comportamiento del negocio. Su principal objetivo es el de brindar rápidas respuestas a complejas preguntas, para interpretar la situación del negocio y tomar decisiones. Cabe destacar que lo que es realmente interesante en OLAP, no es la ejecución de simples consultas tradicionales, sino la posibilidad de utilizar operadores tales como drill-up, drill-down, para explotar profundamente la información. Este tipo de herramientas, se puede analizar el negocio desde diferentes escenarios históricos, y proyectar como se ha venido comportando y evolucionando en un ambiente multidimensional, o sea, mediante la combinación de diferentes perspectivas, temas de interés o dimensiones. Esto permite deducir tendencias, por medio del descubrimiento de relaciones entre las perspectivas que a simple vista no se podrían encontrar sencillamente.

Drill-down:

Permite apreciar los datos en un mayor detalle, bajando por la jerarquía de una dimensión. Esto brinda la posibilidad de introducir un nuevo nivel o criterio de agregación en el análisis, disgregando los grupos actuales.

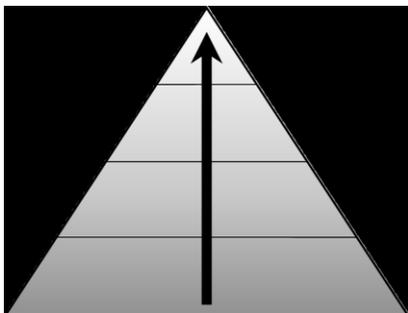
Drill-down es ir de lo general a lo específico. Gráficamente:



Drill-up

Permite apreciar los datos en menor nivel de detalle, subiendo por la jerarquía de una dimensión. Esto brinda la posibilidad de quitar un nivel o criterio de agregación en el análisis, agregando los grupos actuales.

Drill-up es ir de lo específico a lo general. Gráficamente:



Las herramientas OLAP requieren que los datos estén organizados dentro del depósito en forma multidimensional, por lo cual es que utilizan los cubos multidimensionales. Además de las características ya descritas, se pueden enumerar las siguientes:

- Permite recolectar y organizar la información analítica necesaria para los usuarios y disponer de ella en diversos formatos, tales como tablas, gráficos, reportes, etc.

- Soporta análisis complejos de grandes volúmenes de datos.
- Complementa las actividades de otras herramientas que requieran procesamiento analítico en línea.
- Presenta al usuario una visión multidimensional de los datos (matricial).
- Es transparente al tipo de tecnología que soporta el DWH, ya sea ROLAP, MOLAP o HOLAP.
- Permite definir de forma flexible las dimensiones que se quieren analizar, sus restricciones, jerarquías y combinaciones.
- No tiene limitaciones con respecto al número máximo de dimensiones permitidas.
- Permite a los usuarios, analizar la información basándose en más criterios que un análisis de forma tradicional.
- Al contar con muestras grandes, se pueden explorar mejor los datos en busca de respuestas.
- Permiten realizar agregaciones y combinaciones de los datos de maneras complejas y específicas, con el fin de realizar análisis más estratégicos.

1.2.6. ROLAP

Este tipo de organización física se implementa sobre tecnología relacional, pero disponen de algunas facilidades para mejorar el rendimiento. Es decir, ROLAP (Relational On Line Analytic Processing) cuenta con todos los beneficios de un Sistema Gestor de Base de Datos(SGBD) Relacional a los cuales se les provee extensiones y herramientas para poder utilizarlo como un Sistema Gestor de DWH, mucho más lenta que las otras dos estrategias de almacenaje.

1.2.7. MOLAP

Usa unas bases de datos multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos multidimensionalmente. El almacenaje de MOLAP, provee excelente rendimiento y compresión de datos. Como se dice, todo va en el cubo. Tiene el mejor tiempo de respuesta, dependiendo solo en el porcentaje y diseño de las agregaciones del cubo. En general este método, es muy apropiado para cubos con uso frecuente por su rápida respuesta.

1.2.8. HOLAP

Un desarrollo un poco más reciente ha sido la solución OLAP híbrida (HOLAP) Hybrid Online Analytical Process, procesamiento analítico en línea híbrido), la cual combina las arquitecturas ROLAP y MOLAP para brindar una solución con las mejores características de ambas: desempeño superior y gran escalabilidad. Un tipo de HOLAP mantiene los registros de detalle (los volúmenes más grandes) en la base de datos relacional, mientras que mantiene las agregaciones en un almacén MOLAP separado. Los cubos almacenados como HOLAP, son más pequeños que los MOLAP y responden más rápidos que los ROLAP. HOLAP es generalmente usado para cubos que requieren rápida respuesta, para sumalizaciones basadas en una gran cantidad de datos.

1.2.9. Sistemas existentes a nivel internacional

Las empresas del mundo entero están cada vez más interesadas en integrar y centralizar sus datos para reducir costes y conseguir una visión completa de sus clientes, productos, riesgos y finanzas. Algunas de estas compañías son reconocidas internacionalmente por ser grandes corporaciones que manejan enormes cantidades de datos que le son imprescindibles para su correcto funcionamiento. En este caso se encuentran Adidas, Hewlett Packard y Sun, esta última con el Datawarehouse más grande del mundo y todas ellas gigantes corporativos que cuentan con el capital suficiente para realizar una impecable inteligencia de negocio. Estas empresas construyen sus sistemas operacionales

con software propietario, lo cual no proporciona el enfoque más correcto a tener en cuenta, sin embargo brinda una visión muy clara de cómo ordenar la información y el orden de trabajo para la confección del Datawarehouse.

1.2.10. Sistemas existentes a nivel nacional

Las empresas cubanas con el actual avance de tecnologías de la información y las comunicaciones, tiene como premisa fundamental el uso racional de los recursos, tratar de alcanzar la mayor productividad posible y que los productos cada vez tengan una mejor calidad. Por eso se ha pensado desde ya, en el uso de los Datawarehouse. Ejemplos de tal desempeño:

El Datawarehouse implantado por La Corporación CIMEX que se dedica fundamentalmente a la Exportación e Importación de mercancías. El sistema DataWarehouse comercial de la Corporación CIMEX centra su atención en la actividad del comercio, principalmente en la gestión de inventario, permitiendo una gestión de compra-venta eficiente, con una finalidad fundamental: “Disminuir los costos, sin afectar al cliente, permitiendo prestaciones eficientes y con la calidad requerida, aumentando las ganancias o utilidades de las empresas”.

Existencia en CUBAENERGIA de un Datawarehouse para coleccionar toda la documentación concerniente al sistema de información de CUBAENERGIA soportado básicamente en sus sistemas operacionales y aplicaciones en explotación. Los hechos modelados fueron: Ventas, Compras y Gestión de Recursos Humanos por Competencias y por Tiempo.

Estas empresas cubanas brindan un acercamiento mayor a la creación del Datawarehouse, los procesos de ETL y OLAP que son de gran valor para la investigación a pesar de que también trabajan SQL como gestor de base de datos. Estos Datawarehouse de producción nacional ofrecen una visión detallada que contribuirá con el trabajo.

1.3. Tecnología Datamart

Un datamart es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. Un datamart puede ser alimentado desde los datos de un datawarehouse, o integrar por sí mismo un compendio de distintas fuentes de información.



Por tanto, para crear el datamart de un área funcional de la empresa es preciso encontrar la estructura óptima para el análisis de su información, estructura que puede estar montada sobre una base de datos OLTP, como el propio datawarehouse, o sobre una base de datos OLAP. La designación de una u otra dependerá de los datos, los requisitos y las características específicas de cada departamento. De esta forma se pueden plantear dos tipos de datamarts:

Datamart OLAP.

Se basan en los populares cubos OLAP, que se construyen agregando, según los requisitos de cada área o departamento, las dimensiones y los indicadores necesarios de cada cubo relacional. El modo de creación, explotación y mantenimiento de los cubos OLAP es muy heterogéneo, en función de la herramienta final que se utilice.

Datamart OLTP.

Pueden basarse en un simple extracto del datawarehouse, no obstante, lo común es introducir mejoras en su rendimiento (las agregaciones y los filtrados suelen ser las operaciones más usuales) aprovechando las características particulares de cada área de la empresa. Las estructuras más comunes en este sentido son las tablas report, que vienen a ser fact-tables reducidas (que agregan las dimensiones oportunas), y las vistas materializadas, que se construyen con la misma estructura que las anteriores, pero con el objetivo de explotar la reescritura de queries (aunque sólo es posible en algunos sistemas gestores de base de datos avanzados, como Oracle).

Los datamarts que están dotados con estas estructuras óptimas de análisis presentan las siguientes ventajas:

- Poco volumen de datos
- Mayor rapidez de consulta
- Consultas SQL y/o MDX sencillas
- Validación directa de la información
- Facilidad para la historización de los datos



1.4. Fundamentación de la metodología a utilizar.

La metodología a seguir es HEFESTO, que permitirá la construcción del datamart de forma sencilla, ordenada e intuitiva. Hefesto es una metodología bien fundamentada y explícita que permite construir un datawarehouse de manera metódica y sencilla, guiándose por pasos lógicos relacionados sólidamente durante todas las etapas del proceso de confección. Se comienza recolectando las necesidades de información a partir de las preguntas, entrevistas o cuestionarios realizados a los usuarios. Luego, se deben identificar los indicadores resultantes de los interrogativos y sus respectivas perspectivas de análisis, mediante las cuales se construirá el modelo conceptual de datos del datamart, se analizarán los OLTP para señalar las correspondencias con los datos fuentes. Una vez hecho esto, se pasará a la construcción del modelo lógico del depósito, explicitando las jerarquías que intervendrán. Por último, se definirán los procesos de carga, transformación, extracción y limpieza de los datos fuente.

Esta metodología cuenta con las siguientes características:

Los objetivos y resultados esperados en cada fase se distinguen fácilmente y son sencillos de comprender.

- 1 Se basa en los requerimientos del usuario, por lo cual su estructura es capaz de adaptarse con facilidad y rapidez ante los cambios en el negocio.

- 2 Utiliza modelos conceptuales y lógicos, los cuáles son sencillos de interpretar y analizar.
- 3 Es independiente del tipo de ciclo de vida que se emplee para contener la metodología.
- 4 Es independiente de las herramientas que se utilicen para su implementación.
- 5 Es independiente de las estructuras físicas que contengan el DWH y de su respectiva distribución.
- 6 Cuando se culmina con una fase, los resultados obtenidos se convierten en el punto de partida para llevar a cabo el paso siguiente.



1.5. Sistema Integral Informativo de la Estación P.N.R.

En el año 1999, con la creación de la Policía Especializada, se hace necesario implementar un sistema que automatice sus funciones. El perfeccionamiento del trabajo de este órgano implicó un perfeccionamiento del Sistema Automatizado.

La integración de la Policía Especializada a los municipios de la PNR correspondientes conllevó a un rediseño del Sistema, a fin de que se ajustara a los requerimientos de trabajo de una Estación de Policía.

Objetivo del Sistema.

Disponer de un sistema informático que capte, almacene y relacione los datos relativos a la labor de la Estación PNR, en una base a nivel provincial donde se relacionen las informaciones de interés operativo, las personas conducidas, los detenidos, los procesados, las denuncias, el trabajo comunitario y el de la guardia operativa, así como lo relativo a la composición y el desempeño de las Fuerzas de la Estación.

Principales Características del Sistema.

- Base de Datos centralizada a nivel provincial.
- Servidor de Aplicaciones en Windows.
- Base de Datos en Unix - Oracle.
- Enlace en red de todos los usuarios de la Estación PNR.
- Acceso por usuario y rol en el Sistema.

Un Sistema Integrado, que permite la interacción de Subsistemas y Sistemas que posibilitan la Automatización de los distintos puestos de trabajo de la Estación PNR. Un Sistema que posibilita la obtención de estadísticas y recuperaciones que se emplean, como herramientas, para lograr mejores resultados en el Enfrentamiento y en la elevación de la efectividad del trabajo de las fuerzas de la PNR.

Sistemas Informáticos que se explotan actualmente en la Estación PNR.

Sistema Automatizado Jurídico Operativo (SAJO).

Es el sistema que recoge las denuncias de los hechos ocurridos, captando todo lo relacionado con los mismos, las víctimas, los acusados, el modo de operar, el lugar del hecho, el delito al que está asociado, la fecha de ocurrencia, la fecha de denuncia. Este sistema permite obtener una gran cantidad de estadísticas relacionadas con todos estos hechos. Se encuentra implementado y consolidado en Ciudad de la Habana e implementado y tratando de consolidarse en el resto de las provincias del país.

Sistema Automatizado de Información a la Población (SAIP).

Es el sistema que controla a todas las personas que se encuentran detenidas, ya sea en la Estación, o en cual otro lugar. Sobre estas personas recoge los datos identificativos, el lugar donde se encuentra, porque delito fue detenida, fecha de detención, si se traslada a otra provincia o municipio. Se encuentra en explotación en Ciudad Habana y en alguna otra provincia.

Sistema de Consultas y Respuestas a los Registros Operativos (SCORE+).

Es el sistema que nos permite obtener todos los antecedentes que se tienen sobre las personas, antecedentes policiacos, migratorios, en prisiones, datos del carnet de identidad. Nos permite además conocer sobre objetos y armas circuladas. Está en explotación en

Sistema de Registro Fotográfico de Personas (SAREFO).

Es el sistema almacena con posibilidad de recuperación, las fichas de las personas con sus fotografías. Está en explotación en todas las Estaciones de Policía.

Sistema de la Guardia Operativa.

Es el sistema que almacena toda la información relacionada con la guardia operativa de una Estación. Solamente se explota en los municipios Habana Vieja y Centro Habana.

Sistema de las Informaciones Operativas.

Es el sistema que se emplea para recoger todas las informaciones de carácter operativo que se generan en la Estación. Solamente se explota en los municipios Habana Vieja y Centro Habana.

Sistema de la Estación.

Es el sistema que capta el enfrentamiento en la vía, los conducidos a la estación, las contravenciones que se aplican en la vía, las contravenciones y advertencias que se aplican en la estación, la definición de la vigilancia y el patrullaje. Solamente se explota en los municipios Habana Vieja y Centro Habana.

Sistema de Atención a la población.

Es el sistema que capta las visitas de las personas a la estación, recogiendo el motivo de la visita y el grado de satisfacción de las personas visitantes. Está en explotación en los municipios Centro Habana, Plaza y Habana del Este.

Sistema de Información de los hechos.

Es el sistema que almacena los hechos ocurridos, este sistema se ejecuta para que lleguen al Puesto de Mando de la Dirección General de la Policía los hechos, y puedan ser representados en el mapa de la criminalidad. Está en explotación en todos los municipios de la Ciudad de la Habana.

En la Unidad Provincial de Patrullas:

SITUP.

Este sistema recoge la información relacionada con el trabajo de esta Unidad: información de los grupos de trabajo, información de los carros, resultados de trabajo, tareas administrativas.

Mapa de la Criminalidad.

Es una aplicación que representa en un mapa los delitos priorizados ocurridos en Ciudad de la Habana y la ubicación de los carros con GPS pertenecientes a varias unidades de la ciudad que pueden apoyar una situación operativa determinada.

Sistema Automatizado del Puesto de Mando de Seguridad Pública.

Es el sistema que permite tener representados en un mapa los carros patrulleros de la unidad provincial que enfrentan inicialmente los hechos delictivos ocurridos en Ciudad de la Habana y que llegan al puesto de mando de la unidad provincial de patrullas por cualquier vía.

1.6. Herramientas propuestas para el desarrollo del Datamart del subsistema de Conducidos.

- Como gestor de base de datos: Oracle.
- Datawarehouse Builder. Herramienta que se utilizará para la construcción física del Datawarehouse.
- Business Intelligence Enterprise Edition. Se utilizará para la construcción de los dashboards o cuadros de mando.

Conclusiones parciales.

Con el estudio de los sistemas de gestión de la información y las búsquedas realizadas con el objetivo de encontrar sistemas similares, se determinaron las características propias del sistema que se va a desarrollar, llegando a la conclusión son escasos y muy eficientes los sistemas que son capaces de

gestionar la información de forma que pueda ser utilizada fácilmente para la toma de decisiones. Esto proporciona la medida de las ventajas que se darán a los usuarios que interactúen con el sistema a implementar. También se realizó un estudio de las tecnologías que actualmente son utilizadas para la implementación de sistemas semejantes al propuesto, seleccionando las que serán utilizadas a lo largo del desarrollo del sistema, fundamentándose las elecciones del sistema gestor de base de datos y la metodología de desarrollo a utilizar.

CAPÍTULO 2. CARACTERÍSTICAS DEL SISTEMA

En el presente capítulo se muestran los resultados y pasos a seguir para el análisis, según los describe la metodología utilizada. Se elabora un modelo conceptual del Datamart a partir de los indicadores obtenidos del negocio. Además, se realiza el modelo lógico del sistema, de forma que quede conformado el diseño del Datamart del subsistema de los conducidos.

2.1. Análisis del proceso

Para una mejor comprensión del proceso que se realiza, se describe a continuación el flujo del proceso de conducción.

- La persona es identificada en la vía y por determinados motivos se decide conducir a la Estación. Puede ser también un autor de un hecho que se detiene en el lugar que esté y se conduce a la Estación.
- El conducido llega a la Estación y pasa por un proceso de recepción: se verifican, se introducen o se completan los datos identificativos.
- El conducido ya fue recepcionado y se completan los datos de la medida que se le va a aplicar, o se deja sin medida y se deja en libertad. Si se le aplica una medida se llenan los datos de la misma, si es una advertencia o contravención, una vez que se le aplica, se deja en libertad, si es una detención, iría hacia el calabozo.
- Dentro del calabozo se decide qué medida aplicarle, que puede ser: traslado a prisión, traslado a otra provincia, entre otras, y culmina el trabajo con el conducido.

2.2. Análisis de requerimientos

El primer paso consiste en obtener los datos que son objeto de análisis en cuanto al tema de los conducidos, lo cual puede llevarse a cabo por diferentes vías, por ejemplo: entrevistas, cuestionarios y observaciones. El objetivo fundamental de este paso es identificar la información clave para el análisis, que será esencial para trazar las metas y estrategias en aras de lograr mayor eficiencia en el trabajo policial.

A partir del análisis realizado se determinó que:

Se desea conocer la cantidad de personas que son conducidas por diferentes motivos hacia una determinada estación de policía, así como su desplazamiento geográfico.

Se necesita tener el control de las actuaciones de las fuerzas tanto internas como externas en las acciones de conducción.

2.2.1. Indicadores

- Analizar los patrones de comportamiento de los motivos de conducción con respecto a: el tiempo ,el espacio según las distintas visiones territoriales , el servicio de vigilancia desplegado ,las fuerzas policiales actuantes y sus unidades organizativas , las medidas tomadas, las unidades PNR y datos asociados a los enfrentados en el momento de las acciones:

Físicas (edad, sexo, color piel).

Lugar residencia (Provincia, municipio).

Con antecedentes o no en los registros en el momento de la acción.

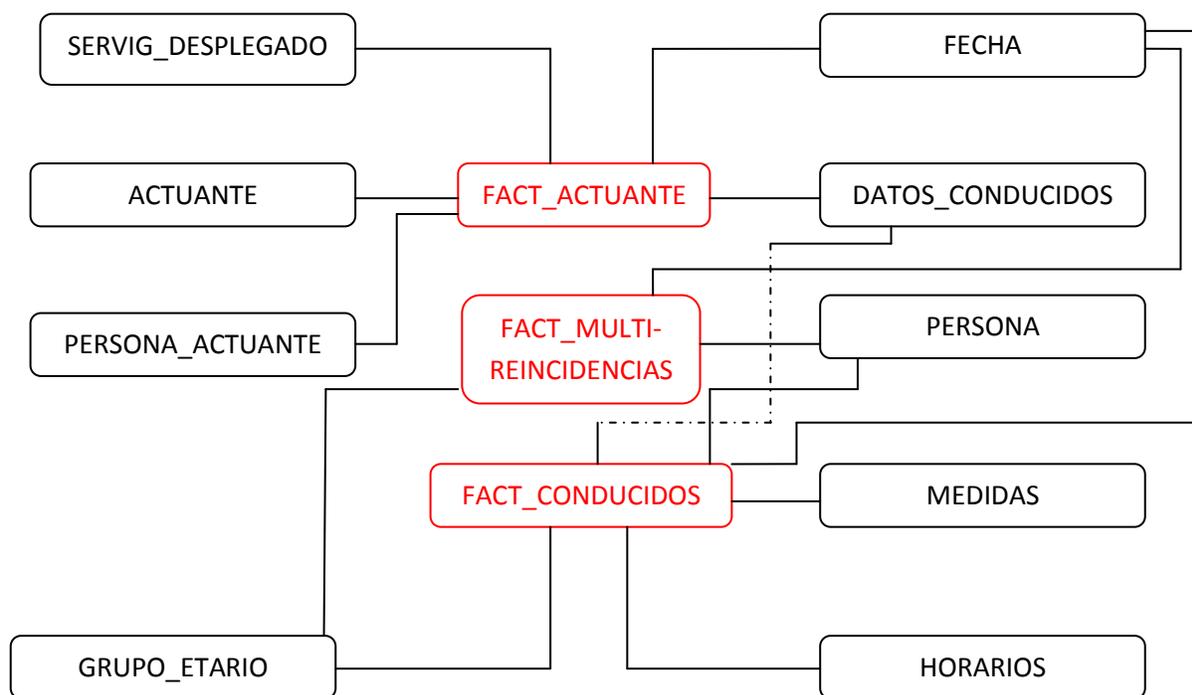
Con respecto a la reincidencia en el momento de la acción: (primario o reincidente) de forma absoluta (desde el comienzo del sistema) o en el año natural en curso.

- Aportar la información estadística para monitorear en tiempo real el estado del proceso de conducción hasta su conclusión en la toma o no de medidas con las personas involucradas. Incluir el monitoreo del subproceso de órdenes de traslado de los conducidos hacia la estación.
- Analizar las acciones de conducción ineficaces que no terminan con medidas policiales, o terminan con medidas débiles aplicadas (profilaxis o advertencia verbal) y que por consiguiente no avalan el gasto de tiempo y recursos involucrados con las mismas.
 - Identificar en tiempo y espacio al personal policial y el servicio de vigilancia y patrullaje desplegado y su tipo de fuerza actuante involucrados con las acciones de conducción ineficaces que traen como consecuencia: gastos de recursos, pérdida de tiempo y molestias e irritación de la población.
 - Identificar el grado de repetitividad con que ocurren estas acciones con respecto a las mismas personas enfrentadas involucrando el mismo personal policial actuante.
- Analizar el nivel de repetitividad de las personas conducidas de forma absoluta, en el año, trimestre y mes con la combinación de:
 - Los diferentes motivos de conducción.
 - Espacio territorial donde ocurrieron las acciones.
 - Medidas tomadas.
 - Fuerzas policiales actuantes.
 - Servicios de vigilancia actuantes.
 - Edad del enfrentado en el momento de la acción.

Los elementos señalados en **ROJO** constituyen los principales indicadores del datamart del subsistema de los conducidos.

2.3. Modelo conceptual

En esta etapa se construye el modelo conceptual a partir de los indicadores obtenidos, lo que permitirá observar con claridad el alcance del Datamart. En la figura se muestra dicho modelo.

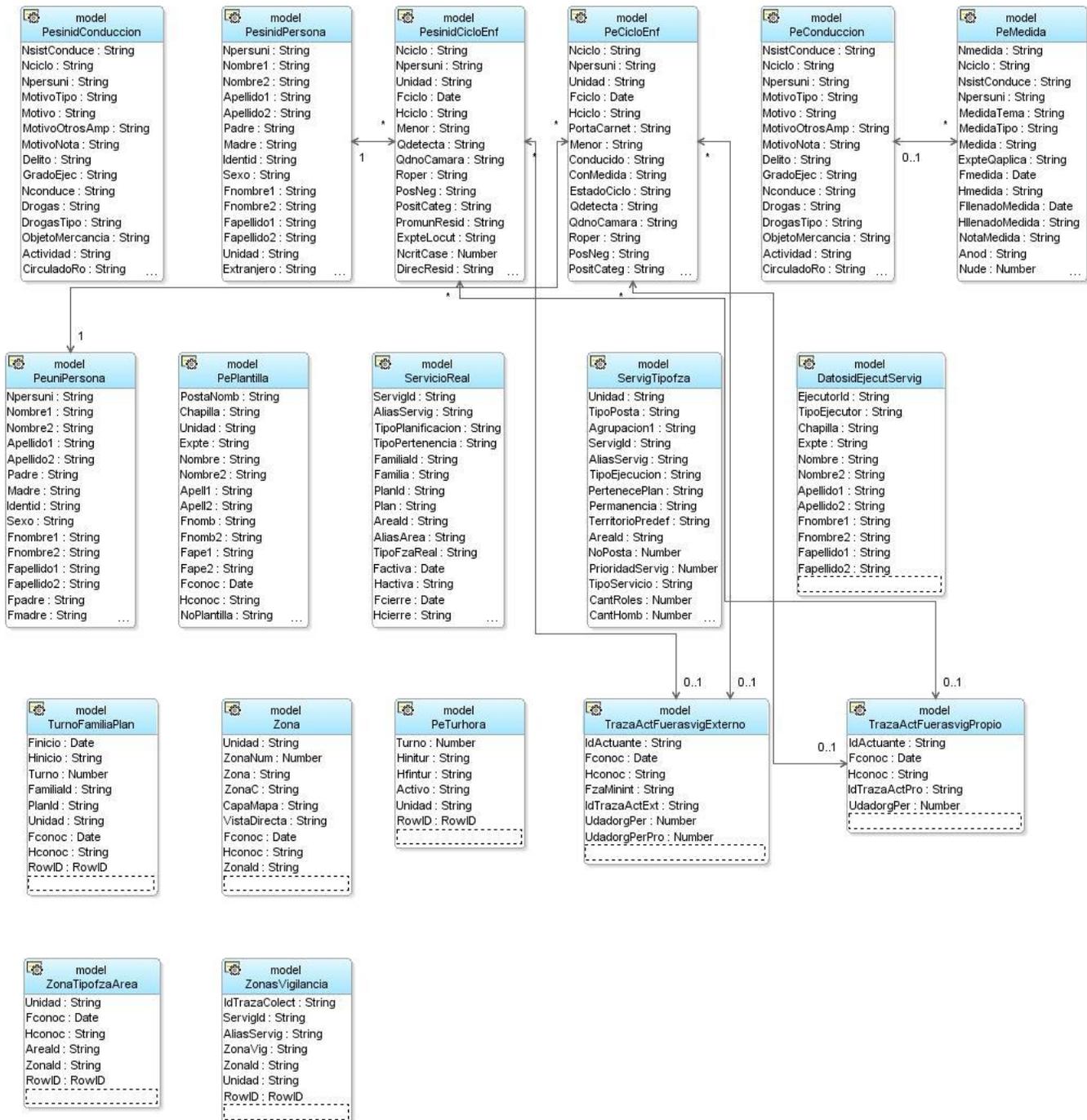


Modelo conceptual del datamart del subsistema de los conducidos.

2.4. Análisis de los OLTP

El segundo aspecto a tener en cuenta es analizar y examinar los OLTP (Sistema Estación (esquema Bespecial y esquema Codif) disponibles que contengan la información requerida. Las siguientes figuras muestran los atributos

pertencientes a las tablas de las bases de datos de Bespecial y Codif, respectivamente.



Fragmento del esquema BESPECIAL.



Fragmento del esquema CODIF.

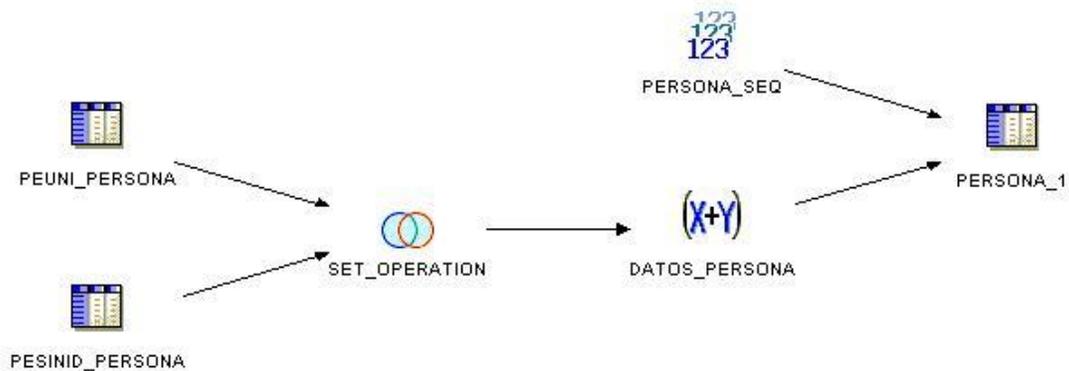
2.5. Elaboración del modelo lógico del Datamart

En este paso se confecciona el modelo lógico, que no es más que la representación de una estructura de datos, que puede procesarse y almacenarse en algún sistema gestor de base de datos, en este caso Oracle. Es muy importante definir en este punto el esquema que se empleará, el definido fue el esquema constelación. Antes de confeccionar este diagrama fue preciso elaborar y mapear las dimensiones asociadas a las tablas de hechos, a continuación mostramos una selección de las más importantes.

Dimensión Persona.

The screenshot shows a window titled 'PERSONA_1' with a tree view of columns and keys. The columns section lists: DIMENSION_KEY (70g), PROVINCIA_MUN_ID (70g), PROVINCIA_MUN_COD (abc), NOMBRES (abc), DOCUMENTADO (abc), EXTRANJERO (abc), PERSONA_ID (70g), COMBATIENTE_FAR (abc), PASAPORTE (abc), PROVINCIA_MUN (abc), PERSONA_COD (abc), CARNET_IDENTIDAD (abc), SEXO (abc), and CARNET_FAR (abc). The keys section shows a primary key 'PERSONA_1_DIMENSION_KEY_PK' on the 'DIMENSION_KEY' column (70g).

Tabla de la dimensión Persona con todos sus atributos.



Mapeo de la dimensión Persona.

The screenshot shows a software interface for defining hierarchies for a dimension named 'PERSONA'. The interface is organized into three main sections:

- Atributos:** A list of attributes with their data types. The attributes listed are: ID (7dg), COD (abc), NOMBRES (abc), DOCUMENTADO (abc), CARNET_IDENTIDAD (abc), PROVINCIA_MUN (abc), SEXO (abc), EXTRANJERO (abc), PASAPORTE (abc), COMBATIENTE_FAR (abc), and CARNET_FAR (abc).
- Niveles:** A list of levels for the hierarchy. It includes 'PROVINCIA_MUN' and 'PERSONA'. Each level has a list of its own attributes. For 'PROVINCIA_MUN', the attributes are ID and COD. For 'PERSONA', the attributes are DOCUMENTADO, NOMBRES, EXTRANJERO, COMBATIENTE_FAR, ID, PASAPORTE, COD, PROVINCIA_MUN, CARNET_IDENTIDAD, CARNET_FAR, and SEXO.
- Jerarquías:** A list of hierarchy types. It includes 'STANDARD' and 'PERSONA'.

Two callout boxes are present:

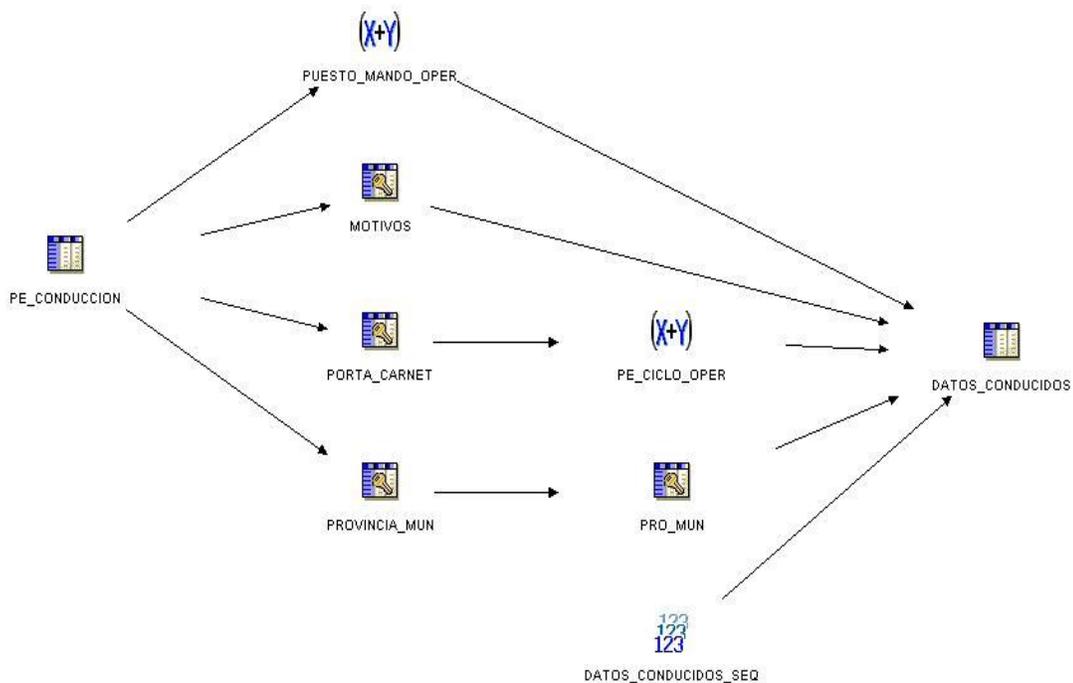
- 'Niveles de Jerarquía' points to the 'PROVINCIA_MUN' and 'PERSONA' levels in the 'Niveles' section.
- 'Tipo de Jerarquía' points to the 'STANDARD' hierarchy type in the 'Jerarquías' section.

Definición de jerarquías en la dimensión Persona.

Dimensión Datos_Conducidos.

DATOS_CONDUCIDOS	
Columnas	
DIMENSION_KEY	78g
PROVINCIA_MUN_ID	78g
PROVINCIA_MUN_COD	abc
DATOS_COND_ID	78g
PORTA_CARNET	abc
MOTIVO	abc
PROVINCIA_MUN	abc
PUESTO_MANDO	abc
TIPO_MOTIVO	abc
ESTADO_CICLO	abc
DATOS_COND_COD	abc
CON_MEDIDA	abc
Claves	
DATOS_COND_DIMENSION_KEY_PK	
DIMENSION_KEY	78g

Tabla de la dimensión Datos_Conducidos con todos sus atributos.



Mapeo de la dimensión Datos_Conducidos.

DATOS_CONDUCIDOS

- Atributos
 - ID 78g
 - COD abc
 - TIPO_MOTIVO abc
 - MOTIVO abc
 - CON_MEDIDA abc
 - PUESTO_MANDO abc
 - ESTADO_CICLO abc
 - PROVINCIA_MUN abc
 - PORTA_CARNET abc
- Niveles
 - PROVINCIA_MUN
 - ID
 - COD
 - DATOS_COND
 - CON_MEDIDA
 - MOTIVO
 - COD
 - PORTA_CARNET
 - TIPO_MOTIVO
 - ESTADO_CICLO
- Jerarquías
 - STANDARD
 - PROVINCIA_MUN
 - DATOS_COND

Niveles de Jerarquía

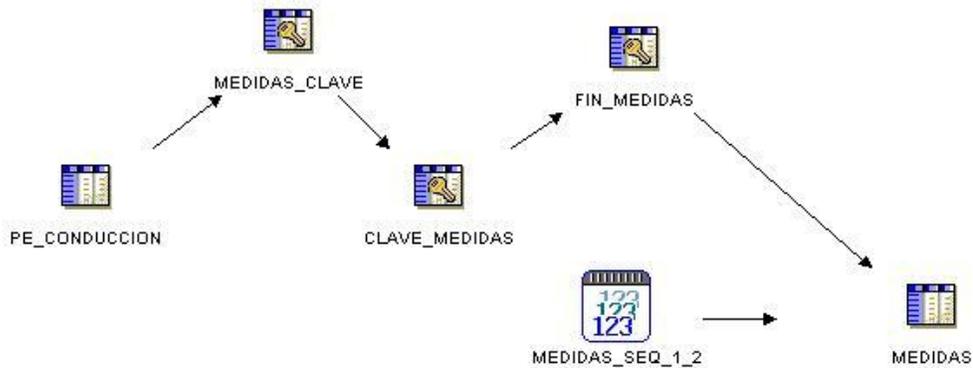
Tipo de Jerarquía

Definición de jerarquías en la dimensión Persona.

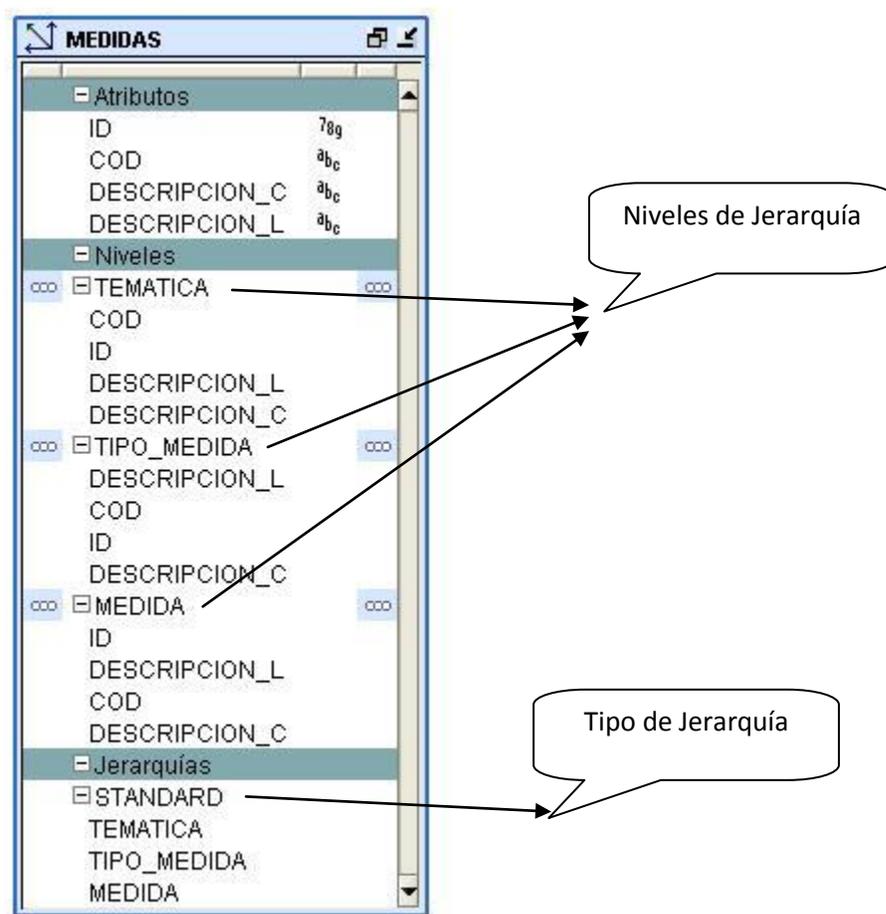
Dimensión Medidas.

MEDIDAS	
= Columnas	
DIMENSION_KEY	78g
TEMATICA_DESCRIPCION_C	abc
TEMATICA_DESCRIPCION_L	abc
TEMATICA_COD	abc
TEMATICA_ID	78g
TIPO_MEDIDA_ID	78g
TIPO_MEDIDA_DESCRIPCION_C	abc
TIPO_MEDIDA_DESCRIPCION_L	abc
TIPO_MEDIDA_COD	abc
MEDIDA_COD	abc
MEDIDA_DESCRIPCION_L	abc
MEDIDA_DESCRIPCION_C	abc
MEDIDA_ID	78g
= Claves	
MEDIDAS_DIMENSION_KEY_PK	
DIMENSION_KEY	78g

Tabla de la dimensión Medidas con todos sus atributos.

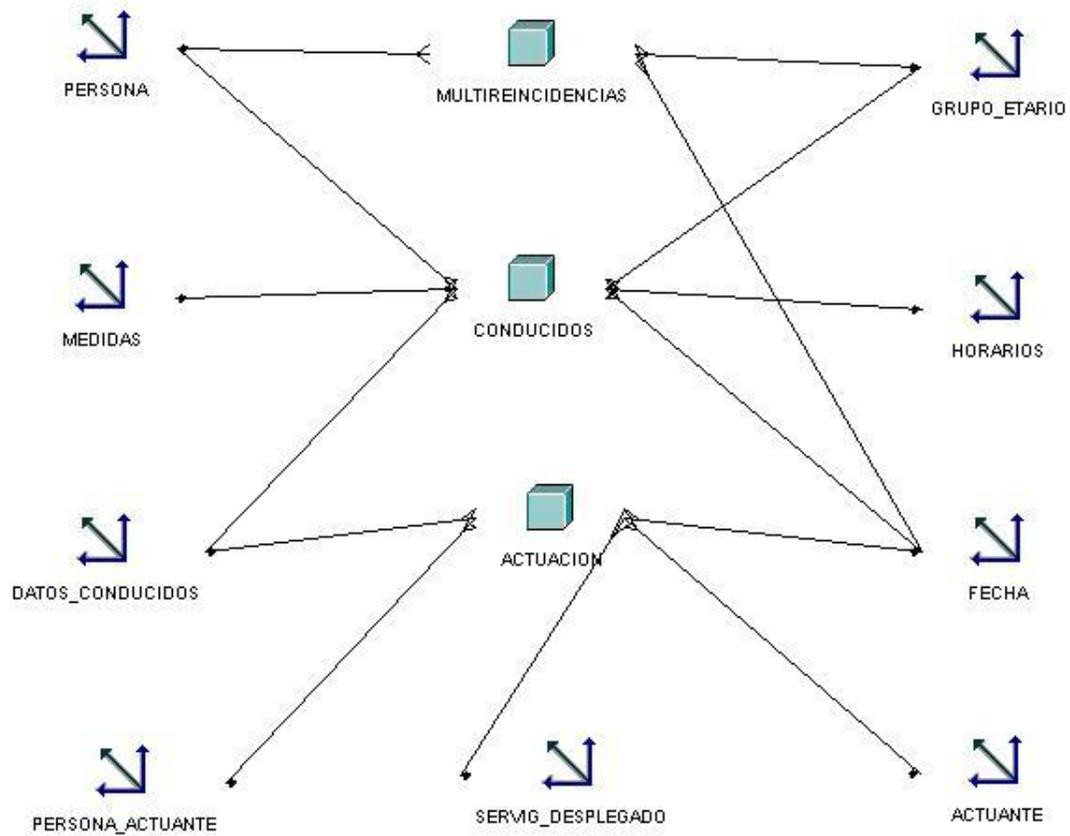


Mapeo de la dimensión Medidas.

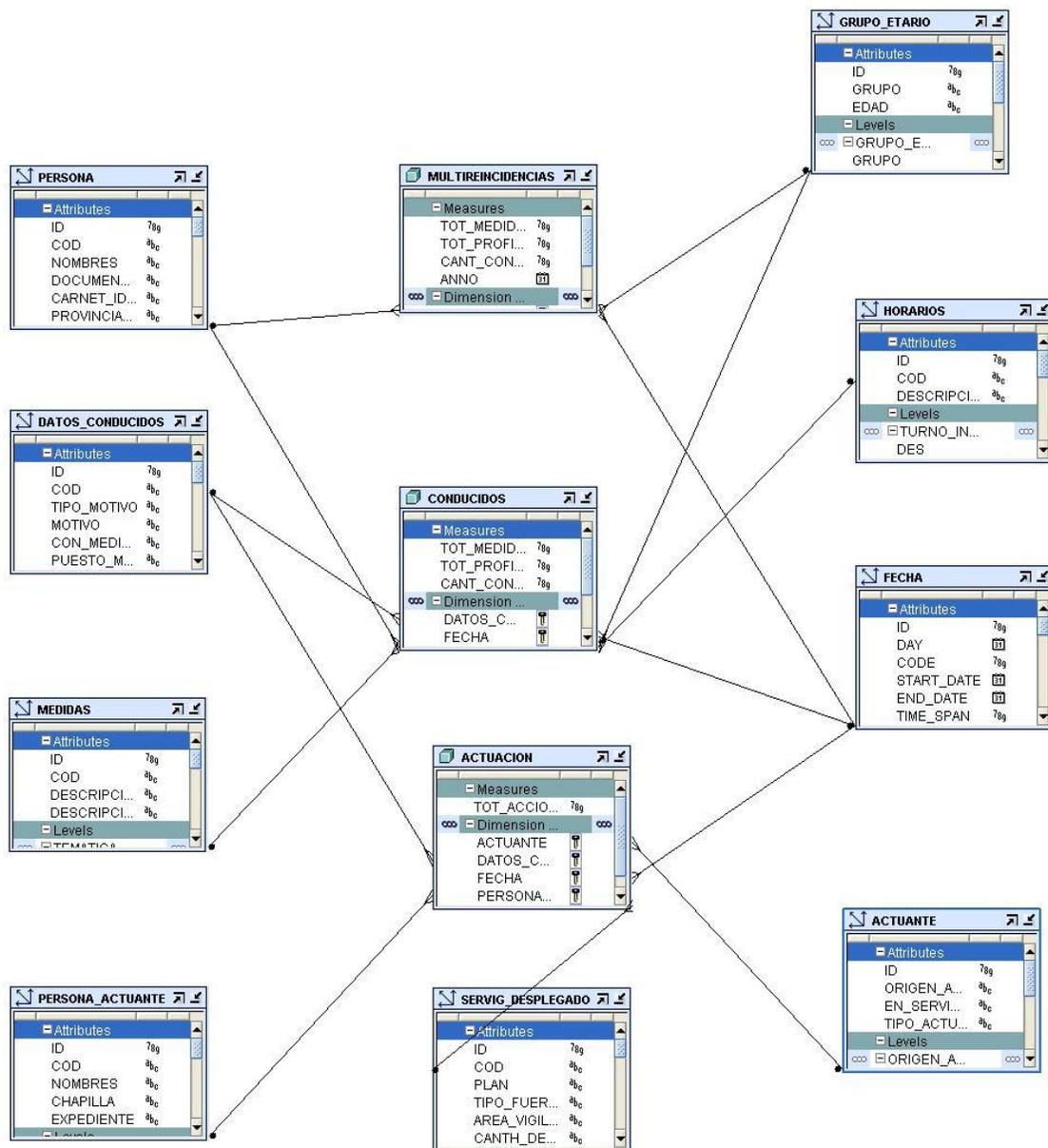


Definición de jerarquías en la dimensión Medidas.

Luego de haber mostrado el mapeo y diseño de algunas de las dimensiones más importantes, se procede a la elaboración del diseño del modelo lógico del datamart del subsistema de los conducidos. En las figuras que se muestra a continuación podemos apreciar el modelo lógico del datamart construido con la herramienta de diseño Warehouse Builder del sistema gestor de bases de datos Oracle teniendo en cuenta el diagrama definido anteriormente, el diagrama constelación.

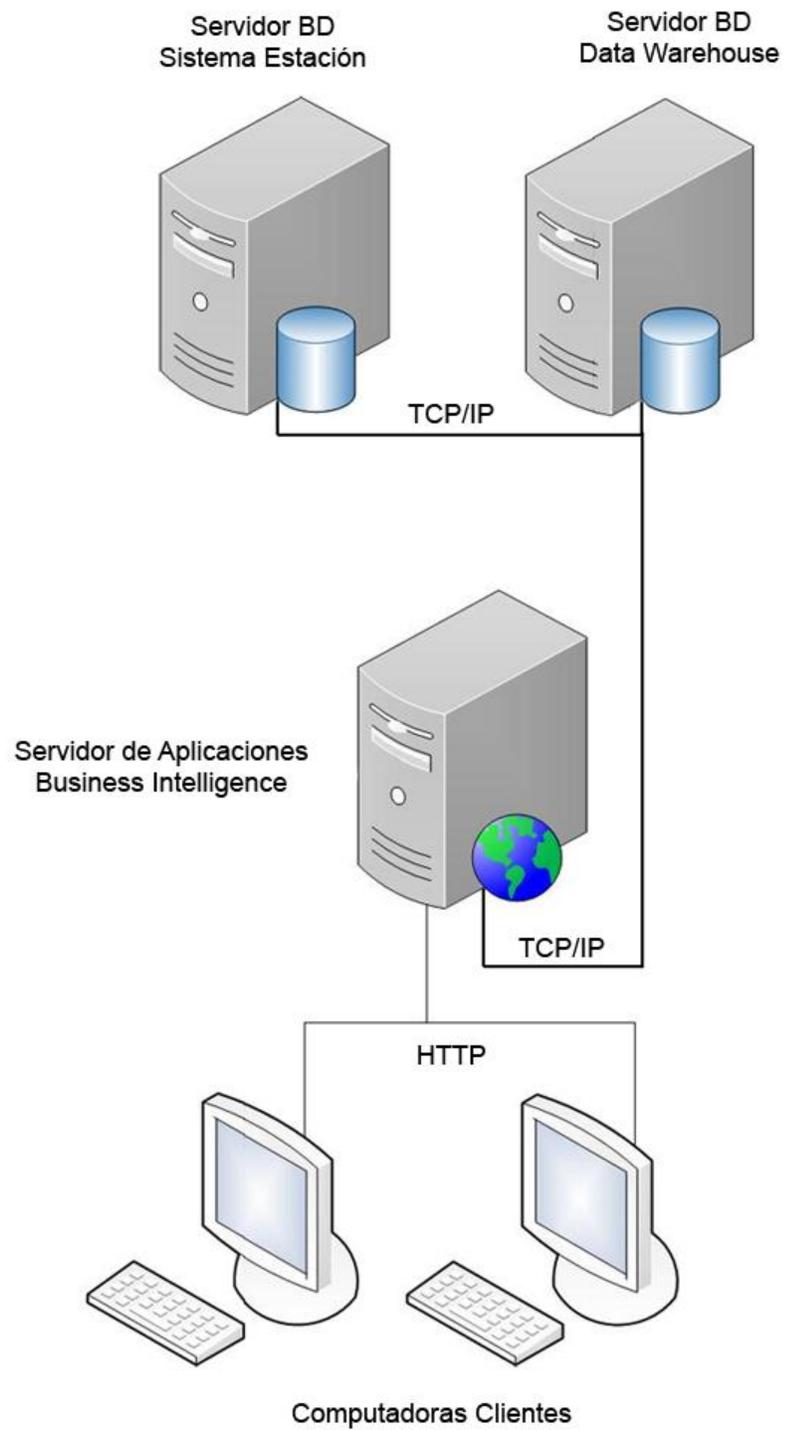


Diseño sencillo de las dimensiones y tablas de hechos.



Diseño de las dimensiones y tabla de hechos con sus atributos asociados.

2.6. Despliegue de la solución.



Conclusiones parciales.

En el transcurso de éste capítulo se expusieron los resultados obtenidos en el diseño del datamart del subsistema de los conducidos, siguiendo los pasos definidos en la metodología escogida. Este segundo capítulo, constituye la base fundamental para la realización exitosa del procesamiento ETL.

CONCLUSIONES

Se utilizó como base la metodología Hefesto que encaminó los pasos para un efectivo desenvolvimiento de los procesos que componen la construcción del Datamart de los conducidos y constituye un paso impulsor para la implementación del datawarehouse integral de la estación.

Al culminar el diseño del datamart del subsistema de los conducidos se dio cumplimiento al objetivo general del trabajo de diploma.

El análisis del sistema operacional de la Estación PNR, junto a la guía de la metodología contribuyó a la selección apropiada de la información a modelar para el diseño del Datamart y a un mejor entendimiento de los datos que se obtienen de este sistema.

Con el resultado obtenido en el análisis de los datos de las fuentes de datos y la guía de la metodología Hefesto se diseñó el datamart del subsistema de los conducidos utilizando Oracle 11g y Windows Vista Business/XP.

RECOMENDACIONES

Culminar la construcción del datamart del subsistema de los conducidos y ponerlo en explotación en las estaciones de la PNR.

Que este diseño sirva como paso impulsor para terminar el datawarehouse integral de la estación PNR.

REFERENCIAS BIBLIOGRÁFICAS

Martí Pérez, José J. Obras completas Tomo VII. Editorial Ciencias Sociales. La Habana, 1975.

Rosabal, Heriberto. Artículo de la Revista Bohemia. Junio 2006, Año 98. Escudo y Leyenda. Edición Digital.

Rodríguez Álvarez, Ángel. Artículo publicado en la Revista Adelante Digital. Nace la PNR al servicio del pueblo cubano.

Pedrito 12. Características de Oracle,
<http://www.todoexpertos.com/categorias/tecnologia-e-internet/bases-de-datos/oracle/expertos/pedrito12>.

Laudon Keneth C. Laudon Jane P. Sistemas de información gerencial 8va cd. México. Pearson Educación, 2004.

Pyle, Dorian. Business Modeling and Data Mining. Morgan Kaufmann, 2003. ISBN 155860653x.

Características de Datawarehouse, <http://www.sqlmax.com/dataw1.asp>.

http://www.sinnexus.com/business_intelligence/datamart.aspx. Business Intelligence + Informática Estratégica.

BIBLIOGRAFÍA

Martí Pérez, José J. Obras completas Tomo VII. Editorial Ciencias Sociales. La Habana, 1975.

Rosabal, Heriberto. Artículo de la Revista Bohemia. Junio 2006, Año 98. Escudo y Leyenda. Edición Digital.

Rodríguez Álvarez, Ángel. Artículo publicado en la Revista Adelante Digital. Nace la PNR al servicio del pueblo cubano.

Pedrito 12. Características de Oracle,
<http://www.todoexpertos.com/categorias/tecnologia-e-internet/bases-de-datos/oracle/expertos/pedrito12>.

Laudon Keneth C. Laudon Jane P. Sistemas de información gerencial 8va cd. México. Pearson Educación, 2004.

Pyle, Dorian. Business Modeling and Data Mining. Morgan Kaufmann, 2003. ISBN 155860653x.

Características de Datawarehouse, <http://www.sqlmax.com/dataw1.asp>.

http://www.sinnexus.com/business_intelligence/datamart.aspx. Business Intelligence + Informática Estratégica.

Griñán García, Amparo y Ávila Domínguez, Yurena. Almacén de datos Julio 2007.

Bahamón L, José Hernando. Construcción de indicadores de gestión bajo el enfoque de sistemas.

Bahamón L, José Hernando. Propuesta de un método para el diseño y modelado de una bodega de datos. Abril de 2003.

Kimbal, Ralph y Ross, Margy. The data warehouse toolkit. Second edition. The complete guide to dimensional modeling.

Bibliografía

Stakowiak, Robert. Rayman, Joseph. Greenwald, Rick. Data warehousing and business intelligence solutions.

Wolf, Carmen. Implementando un data warehouse. Agosto de 2002
<http://www.inf.udec.cl/revista/edicion5/cwolff.htm>.

