

**Universidad de las Ciencias Informáticas**

**Facultad 2**



**Título:** Diseño e Implementación de un mercado de datos para el área de Registro Legal.

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

**Autor(es):** Yailin González Ferrer

Yudit Carmenate Acevedo

**Tutores:** Ing. Krysia Llull Céspedes

Ing. Yanet del Risco Batista

**La Habana, junio 2012**

## DECLARACIÓN DE AUTORÍA

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

Para que así conste firmamos la presente a los \_\_\_\_ días del mes de \_\_\_\_\_ del año 2012.

**Yailin González Ferrer**

\_\_\_\_\_

Firma del Autor

**Yudit Carmenate Acevedo**

\_\_\_\_\_

Firma del Autor

**Ing. Krysia Llull Céspedes**

\_\_\_\_\_

Firma del Tutor

**Ing. Yanet del Risco Batista**

\_\_\_\_\_

Firma del Tutor



*“La estadística es una de las disciplinas menos comprendida y conocida, pero de una gran importancia.”*

*El Che.*

*De Yudit:*

*A mis padres por apoyarme en cada momento de mi vida.*

*A mi hermana por su amor y cariño incondicional.*

*A mis abuelos y bisabuelos por todo el amor y siempre estar ahí para mí.*

*A mi novio Eddy por todo su amor y por hacerme feliz con su manera divertida de ver la vida.*

*A mi compañera de tesis por su preocupación y entrega con la tesis.*

*A mis amistades por el apoyo durante estos 5 años.*

*A mis tutoras por la paciencia y el tiempo dedicado a que todo nos quedara con la mejor calidad.*

*A la Revolución por darme la oportunidad de estudiar en sistema educacional gratis.*

*De Yailin:*

*A mis padres que a ellos le debo lo que soy, gracias por todo.*

*A mi hermano, a Nancy, a mi familia por estar siempre ahí.*

*A mi novio Eddy porque siempre me apoyó cuando estaba triste o alegre.*

*A mi compañera de tesis por no volverse loca con mi insistencia.*

*A mis tutoras por ayudarme en la realización de este trabajo.*

*A mis amistades por su apoyo.*

*De Yudit:*

*A mi mamita linda y a mi padre por ser los motores impulsores de todos mis logros.*

*A mi hermana para ser un ejemplo en su vida.*

*A mis abuelos y bisabuelos en especial a mi bisabuelo Pedro que sin su amor no hubiese logrado muchas cosas en mi vida.*

*De Yailin:*

*A mis padres por ser mi razón de ser.*

*A mi hermano, mis abuelos, mis tíos, mis primos, en fin a mi familia, a mi novio, mis amistades a todos por haberme apoyado.*

### RESUMEN

El presente trabajo tiene como objetivo apoyar al proceso de toma de decisiones del sistema penitenciario cubano, mediante la construcción de un mercado de datos para el área de Registro Legal. El desarrollo del mercado de datos brinda fácil acceso a los datos y permite que exista información relevante para la toma de decisiones, la cual puede tener una representación gráfica para su mejor comprensión. El modelamiento, análisis y posterior implementación del mercado de datos está basado en la metodología de Kimball.

En el proceso de construcción se tuvo en cuenta varias fases, siendo las más importantes el proceso de extracción, transformación y carga junto al procesamiento analítico en línea. Se utilizó para su desarrollo la arquitectura de tres capas, para su implementación, la Suite de Inteligencia de Negocio Oracle Business Intelligence y como gestor de base de datos Oracle 11g.

**Palabras claves:** almacén de datos, inteligencia de negocio, mercado de datos, Registro Legal.

**ÍNDICE**

INTRODUCCIÓN ..... 1

CAPÍTULO I: MARCO TEÓRICO DE LA INVESTIGACIÓN..... 7

    1.1.    Sistemas Valorados ..... 7

        1.1.1.    Sistema de Gestión Penitenciaria de Venezuela..... 7

        1.1.2.    Sistema de Gestión Penitenciaria de Ecuador (e-Sigpen)..... 8

        1.1.3.    Sistema Penitenciario del Gobierno de Panamá ..... 8

        1.1.4.    Sistema Automatizado para el Control del Recluso (SACORE) ..... 8

    1.2.    Almacén de datos ..... 9

    1.3.    Mercado de Datos..... 10

    1.4.    Esquemas dimensionales ..... 10

    1.5.    Arquitectura de datos..... 12

    1.6.    Cubos multidimensionales ..... 13

        1.6.1.    Cubos OLAP..... 13

    1.7.    Modelo de datos ..... 14

    1.8.    Herramientas de creación de un Almacén de Datos ..... 15

        1.8.1.    Sistema de Gestor de Base de Datos Oracle 11g..... 15

        1.8.2.    Oracle Warehouse Builder 11g (OWB). ..... 15

        1.8.3.    Oracle Business Intelligence Enterprise Edition 11g ..... 16

        1.8.4.    Herramienta para el modelado de datos ..... 17

    1.9.    Metodología a utilizar ..... 17

    1.10.    Conclusiones parciales..... 19

CAPÍTULO 2: ANÁLISIS Y DISEÑO DEL MERCADO DE DATOS ..... 20

    2.1.    Identificación de los temas analíticos ..... 20

        2.1.1.    Caracterización de los temas analíticos ..... 20

        2.1.2.    Descripción de los procesos ..... 21

    2.2.    Requisitos del negocio ..... 22

    2.3.    Especificación de requisitos de información..... 25

    2.4.    Matriz dimensional ..... 28

    2.5.    Validación de requisitos ..... 29

2.6.	Modelado Dimensional .....	29
2.6.1.	Granularidad.....	29
2.6.2.	Tablas de Dimensiones .....	30
2.6.3.	Tablas de hechos y medidas .....	30
2.6.4.	Modelo dimensional detallado.....	31
2.6.5.	Patrones de diseño para Almacenes de Datos .....	32
2.6.6.	Modelo de datos .....	33
2.7.	Diseño de las ETL.....	34
2.8.	Conclusiones parciales .....	37
CAPÍTULO 3: IMPLEMENTACIÓN Y PRUEBAS .....		38
3.1.	Implementación .....	38
3.1.1.	Implementación de las ETL.....	38
3.1.2.	Especificación y desarrollo de las aplicaciones de Inteligencia de negocio.....	49
3.1.3.	Tipos de procesamiento.....	55
3.2.	Pruebas .....	56
3.3.	Resultados de las pruebas.....	56
3.4.	Conclusiones parciales .....	59
CONCLUSIONES GENERALES.....		60
RECOMENDACIONES .....		61
GLOSARIO DE TÉRMINOS.....		62
REFERENCIAS BIBLIOGRÁFICAS.....		64
BIBLIOGRAFÍA.....		65
ANEXOS.....		67



## INTRODUCCIÓN

En la actualidad las tecnologías de la informática y las comunicaciones han alcanzado un desarrollo considerable a nivel mundial. Los volúmenes de información generados diariamente crecen rápidamente, por lo que se requiere una correcta gestión de la información para la toma de decisiones.

La habilidad de tomar decisiones concretas y de forma rápida, facilita que una organización logre los resultados deseados a partir de la información que maneja. Aunque el proceso de análisis de datos parezca fácil de resolver mediante los sistemas informáticos tradicionales, en estos sistemas generalmente se presenta información de forma estática, datos en los que no se puede profundizar y se hace difícil navegar entre ellos, obteniendo una estructura poco flexible. A partir de esto surge el concepto de Inteligencia de Negocios o BI (del inglés Business Intelligence), que es el conjunto de herramientas y estrategias utilizadas en la administración y creación del conocimiento. Posibilita el análisis de datos existentes en una organización y de esta manera facilita la toma de decisiones.

La inteligencia de negocios le permite a las empresas obtener información privilegiada para responder a los problemas de negocio, a través de los procesos de recolección y procesamiento de la información que sustentan la toma de decisiones. Una de las formas de utilizar la Inteligencia de Negocio para la gestión de la información es mediante los almacenes de datos, brindándole a las organizaciones la posibilidad de desarrollarse objetivamente. Los almacenes de datos ofrecen el acceso a los datos integrados e históricos de las fuentes de información y a su vez proporcionan un apoyo necesario para los directivos en la planificación y la toma de decisiones.

En Cuba, como parte del proceso de informatización, se persigue la utilización ordenada y masiva de las TIC (Tecnologías de la Información y las Comunicaciones) para satisfacer las necesidades de información y conocimiento de todas las personas y esferas de la sociedad. Para ello en el año 1989 se comienza la informatización del sistema penitenciario cubano, con un sistema de Control Penal, que le permitía la gestión de los principales datos del recluso y algunos aspectos del trabajo en el área. En el año 2002 dando cumplimiento a la orden 43/99 del Viceministro Primero del Interior se diseña y programa el Sistema Automatizado para el Control del Recluso (SACORE), utilizando la tecnología existente en ese entonces en el Ministerio del Interior (MININT),

comenzando su utilización en enero del 2003. El sistema cuenta con 9 años de explotación y a pesar de sus facilidades y ayuda brindada a dichos centros, aún tiene requisitos incompletos o pendientes, por no contar la institución con el tiempo requerido para su análisis, diseño y programación. (1)

A partir del Plan 20 x 50 para la modernización tecnológica del MININT y en colaboración con la Universidad de las Ciencias Informáticas, en el año 2009 se crea el proyecto Prisiones Cuba. El cual tiene como objetivo fundamental unir los tres sistemas que existen en la Dirección de Establecimientos Penitenciarios (DEP) y adicionar un conjunto de funcionalidades necesarias para un mejor funcionamiento de los procesos en los centros penitenciarios. (1)

El Sistema Penitenciario Cubano cuenta con una aplicación informática que contiene los datos que la jefatura necesita para generar información, sin embargo se hace engorroso el proceso de crear los reportes para los cierres trimestrales. Aunque los reportes se obtienen en formato digital es necesario hacer el trabajo manualmente. Para elaborar la información los especialistas deben trasladar los datos obtenidos en las tablas predefinidas en formato HTML hacia documentos en formato Excel, para conformar los gráficos que son usados por la jefatura para la toma de decisiones.

Los reportes que tiene el sistema penitenciario se encuentran predefinidos y al surgir una nueva solicitud de información se hace necesaria la creación de un nuevo reporte. Los informáticos encargados de la implementación de los reportes deben responder a una gran cola de peticiones, a la que no se le puede dar solución en el tiempo requerido, dificultando el análisis de la información a los especialistas del área de Registro Legal.

Los pocos especialistas que integran el departamento de dirección no pueden generar en tiempo toda la información que la jefatura requiere para la toma de decisiones. Debido a lo antes expuesto y al gran cúmulo de información que genera la situación operativa de los Centros Penitenciarios Cubanos.

Por tanto surge como **problema a resolver** ¿Cómo facilitar la toma de decisiones sobre el área de Registro Legal del Sistema Penitenciario Cubano?

Teniendo en cuenta todo lo anterior se define como **objeto de estudio** los procesos para la toma de decisiones en sistemas penitenciarios y como **objetivo general** desarrollar un mercado de datos para apoyar el proceso de toma de decisiones para el área de Registro Legal del Sistema Penitenciario Cubano.

**El campo de acción** está enfocado en el proceso de toma de decisiones para el área de Registro Legal del Sistema Penitenciario Cubano.

**Idea a defender:** El desarrollo de un mercado de datos para el análisis de la información asociada al área de Registro Legal del Sistema Penitenciario Cubano, facilita el proceso de toma de decisiones en el área de Registro Legal.

El **objetivo general** se desglosa en los siguientes **objetivos específicos**:

- Elaborar el marco teórico de la investigación.
- Analizar los requisitos del mercado de datos para el área de Registro Legal.
- Realizar el modelado dimensional del mercado de datos para el área de Registro Legal.
- Implementar el proceso de extracción, transformación y carga (por sus siglas en inglés ETL) del mercado de datos para el área de Registro Legal.
- Implementar el mercado de datos para el área de Registro Legal.
- Realizar las pruebas al mercado de datos para el área de Registro Legal.

Para dar cumplimiento a los **objetivos específicos** anteriormente planteados se definen las siguientes tareas por objetivos:

### 1. **Elaboración del diseño teórico de la investigación**

- Análisis de soluciones similares a las que se quiere implementar.
- Descripción de las herramientas y metodología a utilizar para el desarrollo del mercado de datos.

### 2. **Análisis de los requisitos del mercado de datos para el área de Registro Legal**

- Identificación de temas analíticos del mercado de datos para el área de Registro Legal.

- Identificación de procesos de negocio del mercado de datos para el área de Registro Legal.
- Construcción de la matriz procesos-dimensiones.
- Identificación de los requisitos del mercado de datos para el área de Registro Legal.
- Priorización de proceso del mercado de datos para el área de Registro Legal.

### **3. Realización del modelado dimensional del mercado de datos para el área de Registro Legal**

- Identificación de los niveles de granularidad del mercado de datos para el área de Registro Legal.
- Identificación de las dimensiones, tablas de hechos y medidas del mercado de datos para el área de Registro Legal.
- Identificación de atributos de dimensiones y tablas de hechos del mercado de datos para el área de Registro Legal.
- Diseño del modelo dimensional detallado del mercado de datos para el área de Registro Legal.
- Diseño del modelo datos del mercado de datos para el área de Registro Legal.
- Revisión y validación del modelo dimensional del mercado de datos para el área de Registro Legal.

### **4. Implementación del proceso de ETL del mercado de datos para el área de Registro Legal**

- Diseño de las ETL del mercado de datos para el área de Registro Legal.
- Realización del proceso de ETL para cargar las dimensiones y tablas de hechos del mercado de datos para el área de Registro Legal.

### **5. Implementación del mercado de datos para el área de Registro Legal**

- Realización de la configuración de la herramienta de administración del mercado de datos para el área de Registro Legal.
- Implementación de los reportes del mercado de datos para el área de Registro Legal.
- Desarrollo del tablero de mando del mercado de datos para el área de Registro Legal.

### 6. Realización de las pruebas al mercado de datos para el área de Registro Legal

- Selección de las técnicas de pruebas a realizar del mercado de datos para el área de Registro Legal.
- Aplicación de las pruebas para el mercado de datos del área de Registro Legal.
- Recolección y clasificación de las no conformidades.
- Aplicación de las pruebas de regresión.

#### Posibles resultados:

- Diseño e implementación del mercado de datos de Registro Legal.
- Implementación de los reportes de Registro Legal.
- Tablero de Mando de Registro Legal.

#### Métodos de Investigación

##### Métodos Teóricos

- **Análisis Histórico – lógico:** se hace un estudio del estado del arte sobre las aplicaciones informáticas que se utilizan en los sistemas penitenciarios, con el objetivo de conocer como estas aplicaciones manejan la información para la toma de decisiones.
- **Analítico – sintético:** se analiza de la documentación existente sobre los procesos que afectan la toma de decisiones en el área de registro legal, así como sobre los almacenes de datos, sintetizando la información que es importante para el presente trabajo.

- **Modelación:** se logra una relación entre el modelo y el objeto, para ello se definen las dimensiones del mercado de datos, los hechos asociados a las dimensiones definidas, se estructura el modelo dimensional y se transforma del modelo dimensional al diseño físico.

## **Métodos Empíricos**

- **Entrevista:** se entrevista al cliente para el levantamiento de los requisitos y a partir de ellos se conforman los requisitos de información.

**El presente documento cuenta con una estructura de 3 capítulos:**

### **Capítulo 1: “Marco teórico de la investigación”.**

En este capítulo se realiza un estudio de los sistemas informáticos implantados en centros penitenciarios de varios países, además de la metodología y las herramientas a utilizar. Se profundiza en los conceptos, características y principales enfoques de los almacenes de datos, las tecnologías y software empleado en el diseño y desarrollo del mercado de datos de Registro Legal.

### **Capítulo 2: “Análisis y diseño del mercado de datos”.**

En este capítulo se analizan los principales temas analíticos a desarrollar en el área de Registro Legal, identificando los requisitos de los usuarios, indicadores y perspectivas, especificando el modelo conceptual. Se define el modelo lógico del mercado de datos, diseñando las tablas de dimensiones y de hechos, así como el diseño de la extracción, transformación y carga.

### **Capítulo 3: “Implementación y pruebas del mercado de datos”.**

En este capítulo se realiza la implementación de las ETL para cada cubo del mercado de datos; se configura el repositorio de metadatos para el acceso a los datos, se implementan los reportes y se desarrollan los tableros de mando. Además, se definen y aplican las pruebas al mercado de datos con el objetivo de validar la solución propuesta.

## **CAPÍTULO I: MARCO TEÓRICO DE LA INVESTIGACIÓN.**

### **Introducción**

En el desarrollo del capítulo se abordan las tendencias actuales del uso de las aplicaciones informáticas que facilitan la gestión de información para la toma de decisiones en centros penitenciarios. Se mencionan los conceptos, objetivos y características de los almacenes de datos. Se fundamentan las tecnologías y herramientas que se utilizan para el desarrollo del trabajo.

### **1.1. Sistemas Valorados**

Los sistemas penitenciarios permiten garantizar el proceso de ejecución de la sanción de privación de libertad, de la sanción de trabajo correccional con internamiento, la medida de seguridad reeducativa de internamiento y la medida cautelar de prisión provisional. Este sistema, regido por la Dirección de Establecimientos Penitenciarios del Ministerio del Interior, se sustenta en la integración de principios, conceptos, procedimientos, fuerzas y medios que garantizan el funcionamiento de los centros destinados al internamiento y el tratamiento a los internos. (1)

A continuación se mencionan y se hace un estudio de algunas aplicaciones informáticas existentes en los sistemas penitenciarios, las cuales fueron valoradas para realizar este trabajo, teniendo presente las funcionalidades que brindan.

#### **1.1.1. Sistema de Gestión Penitenciaria de Venezuela**

El Sistema de Gestión Penitenciaria (SIGEP) desarrollado en la Universidad de Ciencias Informáticas por el proyecto Prisiones Venezuela, el mismo cuenta con una sala situacional donde se muestran los reportes en tablas y en gráficos en forma de pastel o barra. El SIGEP tiene definido una cantidad exacta de reportes para la sala situacional sobre el Control Penal, lo cual dificulta a los directivos del sistema de justicia venezolano la toma de decisiones; pues cada vez que surja la necesidad de un nuevo informe, solo el equipo de desarrollo será capaz de implementarlo y no es hasta una próxima versión o con la aplicación de un parche que estará disponible.

Como propuesta de solución a los problemas que surgen con la creación de nuevos reportes y realizar el análisis de la información para la toma de decisiones, en el proyecto Prisiones

Venezuela se implementaron dos mercados de datos como iniciativa de los desarrolladores. Los mismos no fueron contratados por el cliente por lo que no se desplegaron. (2)

### **1.1.2. Sistema de Gestión Penitenciaria de Ecuador (e-Sigpen)**

El Sistema de Gestión Penitenciaria de Ecuador (e-Sigpen) facilita el control y manejo de los procesos que se realizan en las áreas de los Centros de Rehabilitación Social. El mismo permite a los directivos del Ministerio de Justicia del país y otras instituciones monitorear las actividades y obtener datos estadísticos reales de la situación en que se encuentran los internos. Todo ello permite llevar un control de la información de los privados de libertad y tomar decisiones eficaces y oportunas para mejorar el ambiente carcelario del país. (3)

### **1.1.3. Sistema Penitenciario del Gobierno de Panamá**

Es un sistema informático que permite almacenar en una base de datos Oracle los registros de los internos que están detenidos en los centros penales a nivel nacional y generar estadísticas e informes para el Sistema Penitenciario del Gobierno de Panamá (DGSP). El sistema solo almacena la información referente al interno mientras se encuentra en el centro, después de obtener la libertad su expediente es eliminado, afectando de esta forma los informes históricos que se puedan obtener.(4)

### **1.1.4. Sistema Automatizado para el Control del Recluso (SACORE)**

El sistema penitenciario nacional utiliza el SACORE desde el año 2003, el cual cuenta con tres subsistemas principales: Control Penal, Tratamiento Educativo y Orden Interior. A este sistema se le incluyeron mejoras pero estas no cumplían totalmente con los requisitos funcionales de la institución por lo que se hizo necesario desarrollar otros dos sistemas automatizados que completaran la información del SACORE: el Sistema Automatizado de Incidencias (SAIDEP) y el Sistema Automatizado de capacidades (SACDEP). Los reportes predefinidos con que cuenta el sistema son los mismos desde su despliegue, los cuales no satisfacen las necesidades de análisis de la información. Como solución a este problema desarrollaron un sitio web donde crean los nuevos reportes, la información es mostrada en formato tabla; por lo que la información no es intuitiva y no permite a los usuarios desde un primer momento ver el comportamiento del sistema penitenciario.



Luego de valorados los sistemas se tiene como conclusión que:

Los sistemas penitenciarios estudiados utilizan base de datos operacionales, las cuales no están preparadas para realizar análisis de tendencias. Además, los datos se actualizan constantemente dificultando a los directivos el análisis de la información histórica. En las base de datos operacionales se maneja un gran volumen de información que puede ser relevante o no para la toma de decisiones, que pueden tener varios formatos y la misma no está integrada. Como solución al problema de la gestión de la información para la toma de decisiones, se propone el desarrollo de un almacén de datos.

## 1.2. Almacén de datos

Los almacenes de datos (Data Warehouse en inglés), o depósitos de datos como también se les conoce, proporcionan un ambiente para que las organizaciones hagan un mejor uso de la información que está siendo administrada por diversas aplicaciones operacionales.

El almacén de datos posibilita la extracción de datos de sistemas operacionales y fuentes externas, la integración y homogenización de los datos de toda la empresa, provee información que ha sido transformada y resumida, para que ayude en el proceso de toma de decisiones estratégicas y tácticas. (5)

### Características de un almacén de datos

Como **características** de un almacén de datos se tiene:

- Orientado al tema
- Integrado
- De tiempo variante
- No volátil

**Orientado a temas:** En el almacén de datos, la información se clasifica en base a los aspectos que son de interés para la organización, sin embargo en el ambiente operacional la estructura de los datos se diseña en torno a las aplicaciones. Las diferencias entre las funciones de las aplicaciones y la orientación a temas, radican en el contenido de los datos a nivel detallado. En el almacén de datos se excluye la información que no será usada por el proceso de sistemas de

soporte de decisiones, mientras que las orientadas a las aplicaciones contienen datos para satisfacer de inmediato los requisitos funcionales, que pueden ser usados o no por el analista de soporte de decisiones. (5) **Figura 1. 1** (Ver anexo)

**Integrado:** La integración implica que los datos de diversas fuentes que son generados por distintos departamentos, secciones y aplicaciones, tanto internas como externas se consoliden en una instancia antes de ser agregados al almacén de datos, y deben por lo tanto ser analizados para asegurar su calidad y limpieza. (5) **Figura 1. 2** (Ver anexo)

**Variante en el tiempo:** Los almacenes de datos son de tiempo variante pues toda la información es requerida en algún momento, donde puede ser consultada con el objetivo de ser utilizada en comparaciones, tendencias y previsiones. Debido a que los datos pueden ser consultados en un corto, largo o mediano intervalo de tiempo. Los datos almacenados en estos, no pueden ser actualizados. Toda estructura clave de un almacén contiene implícita o explícitamente un elemento de tiempo, constituyendo una de las principales ventajas del almacén de datos, pues los datos son almacenados junto a sus respectivos históricos. (5) **Figura 1. 3** (Ver anexo)

**No volátil:** La información de un almacén de datos existe para ser leída, pero no modificada. La información es por tanto permanente, considerando como actualización del almacén de datos, la incorporación de los últimos valores que tomaron las distintas variables contenidas en él, sin ningún tipo de acción sobre los que ya existían. (5) **Figura 1. 4** (Ver anexo)

### 1.3. Mercado de Datos

Un mercado de datos, (Data Mart en inglés), 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 mercado de datos puede ser alimentado desde los datos de un almacén de datos, o integrar por sí mismo un compendio de distintas fuentes de información. (6)

### 1.4. Esquemas dimensionales

Para la creación de los mercados de datos, uno de los modelos más utilizados es el modelo multidimensional. Debido a este tipo de modelo es que surgen las bases de datos

multidimensionales. Brindando un acceso rápido a los datos, para analizar sus relaciones y consiguientes resultados y están orientadas a temas. Este modelo garantiza una búsqueda rápida de los datos y la información va a almacenarse a través de las tablas de hechos y las tablas de dimensiones.

Las tablas de hechos contienen los hechos, indicadores o medidas a analizar por los analistas del negocio para apoyar la toma de decisiones. Las tablas de dimensiones son las que permiten manipular la información contenida en las tablas de hechos definiendo cómo están organizados lógicamente los datos.

Existen tres variantes de modelado de las bases de datos dimensionales:

- Esquema en estrella.
- Esquema copo de nieve.
- Esquema constelación.

### **1.4.1. Esquema en estrella.**

Este esquema consta de una tabla de hechos central que se relaciona con varias tablas de dimensiones. Debe estar totalmente desnormalizado, siendo innecesarias las uniones entre las tablas cuando se realiza una consulta, esto favorece a una mayor rapidez de las consultas, aunque tiene la desventaja de generar cierto grado de redundancia.

Entre sus ventajas más significativas está que es el esquema más simple de interpretar, posee los mejores tiempos de respuesta, su diseño es fácil de modificar y simplifica el análisis. Sin embargo, es el menos robusto para la carga y es el más lento de construir. (5)

### **1.4.2. Esquema copo de nieve.**

Constituye una extensión del esquema en estrella cuando las tablas de dimensiones se organizan en jerarquías de dimensiones. Existe una tabla de hechos central que está relacionada con una o más tablas de dimensiones, quienes a su vez pueden relacionarse o no con nuevas dimensiones. Posibilita la segregación de los datos de las tablas de dimensiones, es muy flexible y puede ser implementado luego de haber desarrollado un esquema en estrella, siendo muy útil en las tablas de dimensiones de muchas tuplas.

Tiene como desventaja que de existir muchas tablas de dimensiones, cada una de ellas con varias jerarquías, pueden crearse abundantes tablas llegando a ser inmanejables. Además su desempeño puede verse reducido si existen muchas uniones y relaciones entre tablas. (5)

### **1.4.3. Esquema constelación.**

Este modelo está compuesto por una serie de esquemas en estrella y lo conforman una tabla de hechos principal y una o más tablas de hechos auxiliares. Estas tablas yacen en el centro del modelo y están relacionadas con sus respectivas tablas de dimensiones, vinculándose las tablas de hechos auxiliares con algunas dimensiones asignadas a la tabla de hecho principal y también con nuevas tablas de dimensiones. (5)

Los esquemas anteriores pueden ser implementados basándose en el modelado dimensional, para ello pueden utilizarse los cubos multidimensionales, que son una de las estructuras más completas para representar los datos.

Después de haber analizado todos los esquemas dimensionales, el esquema a utilizar será el esquema constelación por tener más de una tabla de hechos que facilita el análisis de más aspectos importantes del negocio y contribuye a la reutilización de las tablas de dimensiones, debido a que pueden utilizarse para varias tablas de hechos. Además que está definido por el proyecto.

### **1.5. Arquitectura de datos**

En un almacén de datos se representan varias capas a través de las cuales circulan los datos, de modo que estos van obteniéndose desde la capa previa y se nombran de acuerdo al número de capas que abarcan.

*“...la arquitectura incluye todo lo que se necesita para preparar y guardar los datos. Por otro lado, también contiene todos los recursos para distribuir la información desde el almacén de datos. Está compuesta más allá de reglas, procedimientos y funciones que permiten al almacén de datos trabajar y cumplir los requisitos de la empresa... Define las normas, medidas, diseño general, y técnicas de apoyo”. (7)*

Existen tres tipos de arquitectura de datos, las cuales son: arquitectura de una capa, arquitectura de dos capas y arquitectura de tres capas. La arquitectura de datos a utilizar es la de tres capas,

debido a que se prevé que en el futuro haya más de una fuente de origen y es necesaria la capa intermedia para evitar problemas de inconsistencias de datos.

En la **Arquitectura de Datos de Tres Capas** para llevar a cabo la transformación de los datos de tiempo real a datos derivados, se requiere de una capa intermedia, conocida como capa de datos reconciliados, donde se solucionan los problemas de inconsistencias y se realiza el procesamiento de los distintos conjuntos de datos de tiempo real adecuadamente. **Figura 1. 5** (Ver anexo)

### 1.6. Cubos multidimensionales

Una de las estructuras utilizadas para representar los datos del almacén de datos es el cubo multidimensional, el cual convierte los datos que se encuentran en las filas y columnas, en una matriz de N dimensiones. (2)

Entre los objetos importantes a incluir en el cubo están los siguientes:

**Indicadores:** Son sumalizaciones efectuadas sobre hechos o expresiones que posibilitan analizar los datos almacenados en el mercado de datos, pertenecen a las tablas de hechos.

**Atributos:** Constituyen los criterios de análisis que se utilizarán para analizar los indicadores, pertenecen a las tablas de dimensiones.

**Jerarquías:** Representa una relación lógica entre dos o más atributos.

#### 1.6.1. Cubos OLAP

*“Los sistemas OLAP son bases de datos orientadas al procesamiento analítico. Este análisis suele implicar, generalmente, la lectura de grandes cantidades de datos para llegar a extraer algún tipo de información útil: tendencias de ventas, patrones de comportamiento de los consumidores, elaboración de informes complejo, etc. Este sistema es típico de los mercados de datos.”* (8)

Existen tres tipos de modelos de almacenamiento para los datos, los cuales son: Procesamiento Analítico en Línea Relacional (por sus siglas en inglés ROLAP), Procesamiento Analítico en Línea Multidimensional (por sus siglas en inglés MOLAP) y Procesamiento Analítico en línea Híbrido (por sus siglas en inglés HOLAP). El modelo de almacenamiento definido por el grupo de

arquitectura de SIDEPA (Sistema Informativo de la Dirección de Establecimientos Penitenciarios) es ROLAP:

**ROLAP:** Se implementa sobre tecnología relacional, dispone de algunas facilidades para mejorar el rendimiento. Aunque el almacén de datos se organiza a como una base de datos multidimensional es soportado por un SGBD Relacional, las dimensiones y tablas de hechos serán traducidas a tablas.

Otras de sus ventajas consisten en:

- Puede manejar grandes cantidades de datos y no pone limitación en la cantidad existente de los mismos.
- Aprovecha las funciones inherentes de las base de datos relacionales, pues a menudo éstas traen consigo una serie de funcionalidades tecnológicas ROLAP.

## 1.7. Modelo de datos

Tanto en las bases de datos tradicionales como en los almacenes de datos existen tres niveles de modelado de datos: conceptual, lógico y físico. Las diferencias entre los almacenes de datos con las bases de datos operacionales en cuanto al tipo de consultas y rendimiento esperado, hacen que las estrategias de diseño y los modelos de datos utilizados para el almacén de datos sean diferentes.

### Modelo de datos conceptual

El modelo conceptual brinda una información general del negocio, representando las entidades del dominio del problema y sus relaciones. Refleja más el espacio del problema, que el espacio de la solución.

### Modelo de datos lógico

El modelo lógico es un puente intermedio entre el nivel conceptual y el físico. Generalmente incluye todas las entidades y las relaciones entre ellas, atributos, tipos de datos y llaves, sin tener en cuenta cómo serán posteriormente implementados físicamente en la base de datos. Posee un alcance mayor que cubre las áreas y los procesos más importantes.

## **Modelo de datos físico**

Este modelo describe las estructuras de almacenamiento y los métodos usados para tener un acceso efectivo a los datos.

Para el buen diseño y construcción del mercado de datos se deben utilizar herramientas que sean efectivas para la futura implantación del depósito en el área destinada. Las mismas deben dar un óptimo soporte a los datos y permitir llevar a cabo todo lo que el diseño del almacén exige.

### **1.8. Herramientas de creación de un Almacén de Datos**

Para el diseño e implementación del mercado de datos se hace necesario tener en cuenta un conjunto de herramientas que proporcionen al desarrollador facilidades en el trabajo a ser realizado. Todas las herramientas a utilizar son definidas por el equipo de arquitectura SIDEP y por el cliente.

Para llevar a cabo la construcción del almacén de datos es necesario transitar por una serie de actividades. Entre las que se encuentra gestionar datos almacenados en diversas fuentes y formatos, realizar la depuración e integración de los mismos, y almacenarlos en un solo destino, depósito o base de datos para su posterior análisis y exploración. Debido a requisitos del usuario se utiliza la suite de inteligencia de negocio Oracle que es un sistema integrado que permite llevar a cabo el proceso de desarrollo y creación de un almacén de datos así como su posterior explotación.

#### **1.8.1. Sistema de Gestor de Base de Datos Oracle 11g**

Es un sistema gestor de base de datos líder en la industria, utilizado para almacenar todo tipo de datos, incluyendo datos relacionales, documentos, multimedia, XML y datos de localización. Oracle es el gestor de base de datos a utilizar, que ofrece una excepcional disponibilidad, escalabilidad, fiabilidad y seguridad. (9) Permite dividir grandes tablas e índices en componentes más pequeños y manejables, sin cambios requeridos a la aplicación fundamental. (10)

#### **1.8.2. Oracle Warehouse Builder 11g (OWB).**

Oracle Warehouse Builder es una herramienta para los aspectos de integración de datos. Warehouse Builder aprovecha la base de datos Oracle para transformar los datos. Proporciona

auditoría de datos, modelo relacional y dimensional completamente integrado, la gestión de ciclo de vida completo de los datos y metadatos. Warehouse Builder permite crear almacenes de datos, migración de datos desde los sistemas heredados, limpiar y transformar los datos para proporcionar información en buen estado, y gestión de metadatos corporativos. (11) **Figura 1. 6** (Ver anexo)

### 1.8.3. Oracle Business Intelligence Enterprise Edition 11g

Oracle Business Intelligence Enterprise Edition 11g (OBIEE) es una plataforma de inteligencia empresarial que ofrece una gama de capacidades de análisis y presentación de informes. Diseñado para la escalabilidad, fiabilidad y rendimiento, Oracle Business Intelligence Enterprise Edition 11g proporciona una visión contextual, relevante y efectiva a todos los miembros de una organización, lo que mejora la toma de decisiones, las acciones con mejor información, y procesos de negocio más eficientes. (12)

Oracle Business Intelligence Enterprise Edition está constituido por:

**Administration Tool:** Es la herramienta de administración para administrar el servidor Oracle business intelligence. Permite que el administrador del servidor cree y edite repositorios desde modo en línea o desconectado. El objetivo fundamental de los repositorios es relacionar las herramientas de inteligencia de negocio con los datos almacenados en el mercado de datos. Además se definen las jerarquías de las dimensiones y las medidas con sus respectivas reglas de agregaciones. **Figura 1. 7** (Ver anexo)

**Oracle BI Answers:** Es una solución de análisis y consultas dinámicas sobre un cliente liviano, totalmente integrada con los tableros interactivos. Los usuarios finales pueden rápidamente crear sus propias consultas, luego desglosar, analizar, visualizar e incluir los resultados en sus tableros personalizados. **Figura 1. 8** (Ver anexo)

**Oracle Business Intelligence Interactive Dashboards:** (tableros interactivos) ofrecen una interfaz web donde se pueden ver de manera rápida las métricas más importantes, los informes y visualizaciones. Permiten capacidades completas de navegación y desglose hasta los datos transaccionales, brinda informes resumidos que evalúa la empresa, lo cual es perfecto para los informes de administración. **Figura 1. 9** (Ver anexo)



### **1.8.4. Herramienta para el modelado de datos**

**Er/Studio 8.0:** Es una herramienta de modelado de datos fácil de usar, para el diseño y construcción de bases de datos a nivel físico y lógico. Direcciona las necesidades diarias de los administradores de bases de datos, desarrolladores y arquitectos de datos que construyen y mantienen aplicaciones de bases de datos grandes y complejos.

ER/Studio está equipado para crear y manejar diseños de bases de datos funcionales. Ofrece fuertes capacidades de diseño lógico, sincronización bidireccional de los diseños físicos y lógicos, construcción automática de bases de datos.

### **1.9. Metodología a utilizar**

La metodología a utilizar en la realización del mercado de datos, fue definida por el grupo de arquitectura de SIDEPA, a continuación se mencionan sus principales características.

#### **Metodología de Ralph Kimball. Características**

La metodología de Kimball apunta a una solución completa que se puede implementar en poco tiempo. Se basa en lo que Kimball denomina Ciclo de Vida Dimensional del Negocio (Business Dimensional Lifecycle), con un enfoque de abajo hacia arriba (Bottom - up). Compuesta por una serie de etapas, las cuales se muestran a continuación. (13)

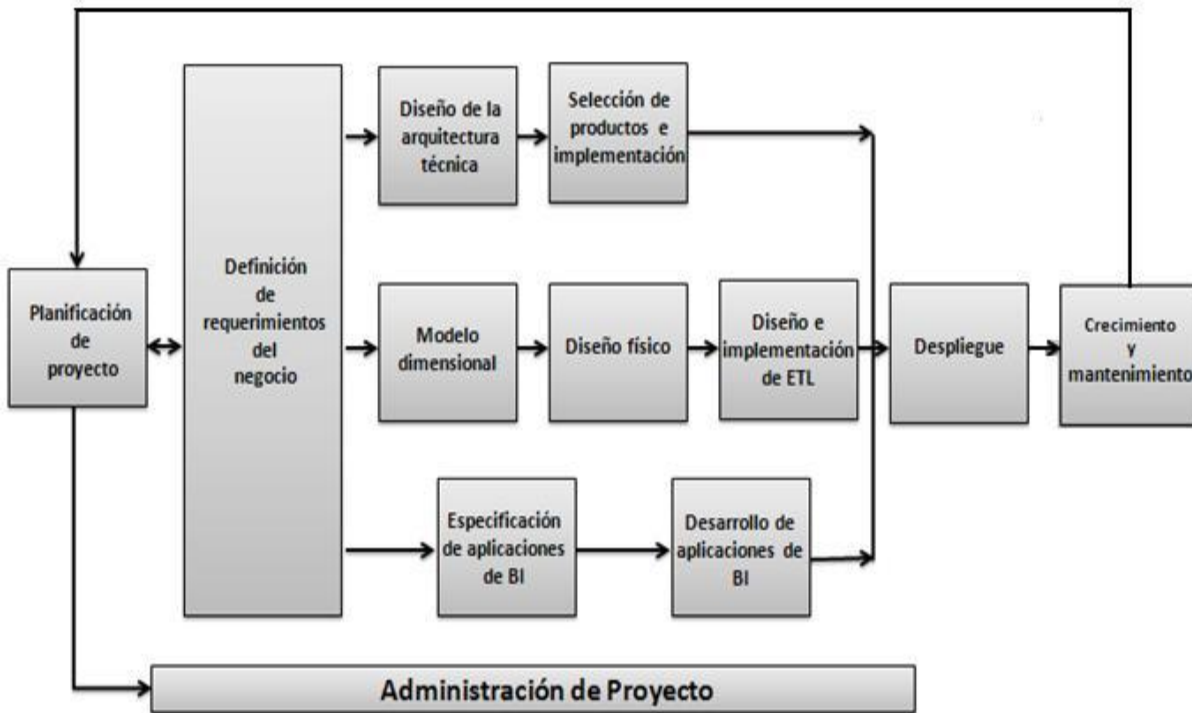


Figura 1. 10: Etapas que componen la metodología de Kimball.

### Planificación

La planificación del proyecto del almacén de datos, es una de las etapas que involucra todo el ciclo de vida de la metodología; desde su inicio hasta su fin. En esta etapa se determinan los objetivos específicos y la meta a alcanzar con el almacén de datos, así como la necesidad de información de la empresa. Además se realiza la captura de los requisitos y se identifican los temas analíticos.

### Análisis de requerimientos

La definición de los requerimientos es el proceso de entrevistar al personal de negocio y técnico, pero siempre conviene tener un poco de preparación previa. Se debe aprender tanto como se pueda sobre el negocio, los competidores, la industria y los clientes del mismo. Hay que leer todos los informes posibles de la organización; rastrear los documentos de estrategia interna; entrevistar

a los empleados, analizar lo que se dice en la prensa acerca de la organización, la competencia y la industria. Se deben conocer los términos y la terminología del negocio.

### **Modelado Dimensional**

En este paso es donde se diseña el modelo lógico del almacén de datos, se establece el nivel de granularidad, se eligen las dimensiones, se identifican las tablas de hechos y se realiza el modelo gráfico de alto nivel.

### **Diseño Físico**

El diseño físico se focaliza sobre la selección de las estructuras necesarias para soportar este diseño. En esta etapa se decide la cantidad de índices a tener en cuenta, el tipo de índice y donde se guardan los mismos. También se especifica el tipo de particionamiento y el rango de este, cuando se utilizan las llaves únicas. Finalmente, se definen los nombres que tendrán los índices, las particiones y las llaves.

### **Diseño del sistema de Extracción, Transformación y Carga.**

El sistema de Extracción, Transformación y Carga es la base sobre la cual se alimenta el almacén de datos. Si el sistema ETL se diseña adecuadamente, da la posibilidad de extraer los datos de los sistemas de origen de datos, aplicar diferentes reglas para aumentar la calidad y consistencia de los mismos, consolidar la información proveniente de distintos sistemas, y finalmente cargar (grabar) la información en el almacén de datos en un formato acorde para ser utilizado con las herramientas de análisis.

### **1.10. Conclusiones parciales**

Se realizó un estudio de algunos sistemas penitenciarios con el objetivo de saber cómo gestionaban la información para la toma de decisiones. A partir de lo estudiado se definió la necesidad de desarrollar un almacén de datos que apoye la toma de decisiones en el Sistema Penitenciario Cubano. El análisis de los conceptos de almacén de datos y mercado de datos identificó las principales características deseables de la propuesta de solución. Para su desarrollo se decidió utilizar las herramientas descritas y la metodología Kimball, con una arquitectura de datos de tres capas. Para el diseño del almacén de datos se definió el esquema constelación, con una estructura de almacenamiento ROLAP.

### CAPÍTULO 2: ANÁLISIS Y DISEÑO DEL MERCADO DE DATOS

#### 2. Introducción

En este capítulo se presenta el análisis realizado en el área de Registro Legal de la Dirección de Establecimientos Penitenciarios (DEP), el cual se realiza a través de la metodología Kimball. Se lleva a cabo un estudio de los procesos que son de mayor interés para el departamento y las perspectivas que son usadas en el análisis. Se determinan los indicadores para realizar el diseño del mercado de datos correspondiente.

##### 2.1. Identificación de los temas analíticos

- **Técnicas de identificación de requisitos**

Para la identificación de requisitos existen diferentes técnicas que propician definir de forma clara y documentada aspectos generales sobre los requisitos del cliente que pueden ser posteriormente refinados con técnicas más específicas. Además de que requieren ser adecuadamente orientadas para cubrir la información que se requiere capturar. A continuación se describen las técnicas que fueron utilizadas para la captura de requisitos del mercado de datos.

**Análisis de documentación:** se realizó un estudio de las leyes, resoluciones y los reglamentos del Sistema Penitenciario Cubano relacionadas con el Área de Registro Legal con el objetivo de lograr una mejor comprensión del negocio.

**Entrevista:** Se realizaron varias entrevistas a los especialistas del área de Registro Legal y a los del Departamento de Dirección de la DEP, la que permite recolectar opiniones, criterios o descripciones sobre los procesos, que aportan una serie de elementos esenciales para el desarrollo del mercado de datos.

Para realizar las entrevistas se utilizaron las preguntas propuestas por la metodología de Kimball, una muestra de ellas se encuentran en la **figura 2. 1**. (Ver anexos)

##### 2.1.1. Caracterización de los temas analíticos

El área de Registro Legal se encarga de la situación legal de los internos y el control del cumplimiento de la sanción del interno durante su tránsito por el sistema penitenciario. (1)

A partir de las entrevistas realizadas se identificaron los siguientes temas analíticos que agrupan requisitos comunes en un tema común, de los procesos del área de Registro Legal.

- **Movimiento:** conformado por los procesos Traslado y Egreso en los que se recoge información relacionada con los traslados de los internos hacia otros centros penitenciarios, así como los egresos por cada proceso legal.
- **Delito:** tiene la información referente a los delitos por los que han sido sancionados los internos.
- **Solicitudes y Decisiones:** integrado por todos los requisitos que recogen toda la información sobre los procesos de solicitudes y decisiones.
- **Salida:** conformado por los conduce y juicios planificados a los centros penitenciarios.

### 2.1.2. Descripción de los procesos

- **Solicitudes:** Son peticiones que realizan los establecimientos penitenciarios a los tribunales o a la Dirección de Establecimientos Penitenciarios. Estas pueden ser: de otorgamiento de beneficios, traslado inter-penal, conducciones, entre otras.
- **Decisiones:** Son las acciones ordenadas por las instancias competentes (autoridades judiciales competentes, los órganos provinciales o la DEP). Las decisiones se comunican a las jefaturas provinciales o centros penitenciarios a través de documentos de carácter legal. Estas determinan la ejecución obligatoria de procesos en los establecimientos penitenciarios.
- **Delito:** Se considera delito toda acción u omisión socialmente peligrosa prohibida por la ley. Este proceso se ocupa de la información relativa a los delitos por los que han sido sancionados los internos.
- **Egreso:** Tiene el control de los internos que salen de los centros penitenciarios por diferentes motivos, ya sea por libertad condicional, cumplimiento de la sanción, fallecimiento, ejecución de la sanción de muerte, entre otras.
- **Conduce:** Los conduce significan las salidas que se realizan al hospital, al tribunal, a las funerarias y domicilio, donde los internos retornan al centro penitenciario. Se encarga de la

planificación y ejecución de los conductos de los internos. De los conductos al tribunal se extrae la información de los juicios planificados y ejecutados por centros penitenciarios.

- **Traslado:** Tiene la responsabilidad de la planificación y ejecución de los traslados inter-unidades o inter-provinciales. Para la planificación se toma como base las decisiones de traslado inter-penal. Los traslados pueden ser individual o colectivos, este último para el caso en que las decisiones de traslado de los internos involucrados, coinciden en fecha y destino, pudiéndose planificar como una única salida.

### 2.2. Requisitos del negocio

Acorde a la metodología, se analizan los requisitos que son de gran importancia para el entendimiento del negocio que se evalúa. Se realiza una entrevista con los usuarios, en este caso el departamento de dirección y con los jefes del área de Registro Legal de la DEP en busca de sus necesidades de información. Se analizan cada uno de los procesos que intervienen en cada área y los indicadores que se tienen en cuenta para la toma de decisiones.

#### Requisitos de información

Los requisitos de información representan la información que debe estar disponible, propiciando que se puedan realizar consultas en todo momento. Constituyen la entrada fundamental para el proceso de inteligencia de negocio y para la creación de los futuros reportes. A continuación se especifican los requisitos de información que fueron identificados.

#### Solicitudes

- R11. Mostrar cantidad de solicitudes por tipo de solicitud por Provincia.
- R12. Mostrar cantidad de solicitudes rechazadas por tipo de solicitud de Provincia.
- R13. Mostrar cantidad de solicitudes por evaluación por tipo de solicitud por Provincia.
- R14. Mostrar cantidad de solicitudes por evaluación por año.

#### Decisiones

- R15. Mostrar cantidad de decisiones por tipo de decisión por Provincia.
- R16. Mostrar cantidad de decisiones por Provincia.

- RI7. Mostrar cantidad de decisiones por año.
- RI8. Mostrar cantidad de decisiones por trimestre.

### **Delito**

- RI9. Mostrar total de delitos por familia delictiva.
- RI10. Mostrar total de delitos por tipo de delito.
- RI11. Mostrar por ciento que representa el tipo de delito.
- RI12. Mostrar por ciento que representa el tipo de delito de su familia delictiva a la que pertenece.
- RI13. Mostrar cantidad de delitos por naturaleza del delito.
- RI14. Mostrar cantidad de delitos que ocasionaron daños a la Revolución por Provincia.
- RI15. Mostrar cantidad de delitos que ocasionaron daños a la Revolución por tipo de delito.
- RI16. Mostrar cantidad de delitos que se produjeron en lugares de internamiento o lugares de detención por Provincia.
- RI17. Mostrar cantidad de delitos donde se emplearon armas o instrumentos para agredir por Provincia.
- RI18. Mostrar cantidad de delitos con connotación social por Provincia.
- RI19. Mostrar cantidad de delitos por connotación social por año.
- RI20. Mostrar cantidad de delitos por naturaleza del delito por año.
- RI21. Mostrar cantidad de delitos hechos en lugar de internamiento por año.
- RI22. Mostrar cantidad de delitos donde se emplearon armas o instrumentos para agredir por año.
- RI23. Mostrar cantidad de delitos por año.

### **Conduce**

- RI24. Mostrar cantidad de juicios celebrados por Provincia.

- RI25. Mostrar cantidad de juicios citados por Provincia.
- RI26. Mostrar cantidad de juicios suspendidos por causa.
- RI27. Mostrar cantidad de juicios por estado por provincia.
- RI28. Mostrar cantidad de conduces por Provincia.
- RI29. Mostrar cantidad de conduces por estado por Provincia.
- RI30. Mostrar cantidad de conduces por urgencia por Provincia.
- RI31. Mostrar cantidad de conduces a juicios cancelados por Provincia.
- RI32. Mostrar cantidad de conduces por motivo por Provincia.

### **Traslado**

- RI33. Mostrar cantidad de traslados por estado por Provincia.
- RI34. Mostrar cantidad de traslados ejecutados que hayan sido prorrogados por Provincia.
- RI35. Mostrar cantidad de traslados por tipo de traslado por Provincia.
- RI36. Mostrar cantidad de traslados por tipo de traslado por Provincia a donde se realiza el traslado.
- RI37. Mostrar cantidad de traslados por tipo de ejecución por Provincia.
- RI38. Mostrar cantidad de traslados por tipo de ejecución por Provincia a donde se realiza el traslado.
- RI39. Mostrar cantidad de traslados interprovinciales a donde se realiza por provincia.
- RI40. Mostrar cantidad de traslados interprovinciales por provincia.
- RI41. Mostrar cantidad de traslados por Provincia.
- RI42. Mostrar cantidad de traslados por Provincia a donde se realiza el traslado.
- RI43. Mostrar cantidad de trasladados con más de 10 internos por Provincia.
- RI44. Mostrar cantidad de trasladados con más de 10 internos por Provincia a donde se realiza el traslado.



- RI45. Mostrar cantidad de traslados por trimestres.
- RI46. Mostrar cantidad de traslados por motivo de traslado por Provincia.
- RI47. Mostrar cantidad de traslados por motivo por Provincia a donde se realiza el traslado.
- RI48. Mostrar cantidad de traslados por tipo de ejecución por año.
- RI49. Mostrar cantidad de traslados por tipo de traslado por año.
- RI50. Mostrar cantidad de traslados cancelados que fueron prorrogados por año.

**Egreso**

- RI51. Mostrar cantidad de egresos por motivo de egreso.
- RI52. Mostrar cantidad de egresos por Provincia.
- RI53. Mostrar cantidad de egresos por cuatrimestre.

**2.3. Especificación de requisitos de información**

Una vez conformados los requisitos de información se realiza la especificación de los mismos para un mejor entendimiento por parte del cliente. A continuación se muestran un fragmento de las especificaciones de requisitos que se encuentran en la planilla de especificación de requisitos de software que está situada en el expediente del proyecto Prisiones Cuba.

Nº	Nombre	Descripción
RI1	Mostrar cantidad de conduces por provincia.	Muestra la cantidad de conduces por provincia que fueron ejecutados. Al seleccionar una provincia muestra el total de conduces ejecutados por centro. Brinda la posibilidad de cambiar la vista del reporte, las cuales pueden ser: gráfica de pastel, barra o área.
	<b>Campos</b>	<b>Tipos de Datos</b>
		<b>Reglas o restricciones</b>

	Provincia	Cadena de caracteres	El nivel de jerarquía definido es provincia y centro, el colectivo no aplica.
	Cantidad de conduces	Numérico	No procede
<b>Nº</b>	<b>Nombre</b>	<b>Descripción</b>	
RI2	Mostrar cantidad de conduces por estado por provincia.	Muestra la cantidad de conduces planificados, ejecutados o los cancelados es decir por el estado del conduce por provincia. Al seleccionar una provincia muestra el total de conduces por estados por centro.	
	<b>Campos</b>	<b>Tipos de Datos</b>	<b>Reglas o restricciones</b>
	Provincia	Cadena de caracteres	El nivel de jerarquía definido es provincia y centro, el colectivo no aplica.
	Cantidad de conduces	Numérico	No procede
	Estado del conduce	Cadena de caracteres	
<b>Nº</b>	<b>Nombre</b>	<b>Descripción</b>	
RI3	Mostrar cantidad de traslados por ejecución por	Muestra la cantidad de traslados ejecutados por tipo de ejecución que puede ser expreso u ordinario por provincia de donde se realiza el	

	Provincia.	traslado. Al seleccionar una provincia muestra el total de traslados ejecutados por centro. Brinda la posibilidad de seleccionar la vista del reporte, las cuales pueden ser: gráfica de pastel, barra o área.	
	<b>Campos</b>	<b>Tipos de Datos</b>	<b>Reglas o restricciones</b>
	Provincia	Cadena de caracteres	El nivel de jerarquía definido es provincia y centro, el colectivo no aplica.
	Cantidad de traslados	Numérico	No procede
	Ejecución del traslado	Cadena de caracteres	
<b>Nº</b>	<b>Nombre</b>	<b>Descripción</b>	
RI4	Mostrar cantidad de traslados por tipo de traslado por Provincia.	Muestra la cantidad de traslados ejecutados según el tipo de traslado que puede ser interpenal o interprovincial por provincia de donde se realiza el traslado. Al seleccionar una provincia muestra el total de traslados ejecutados por tipo de traslado por centro. Brinda la posibilidad de seleccionar la vista del reporte, las cuales pueden ser: gráfica de pastel, barra o área.	
	<b>Campos</b>	<b>Tipos de Datos</b>	<b>Reglas o restricciones</b>
	Provincia	Cadena de	El nivel de jerarquía

		caracteres	definido es provincia y centro, el colectivo no aplica.
	Cantidad de traslados	Numérico	No procede
	Tipo de traslado	Cadena de caracteres	

#### 2.4. Matriz dimensional

La matriz dimensional representa las relaciones que se establecen entre los procesos y las dimensiones. En la matriz se observa la existencia de dimensiones comunes y las que son específicas por cada proceso.

Dimensiones	Procesos				
	Traslado	Solicitudes y decisiones	Egreso	Conduce	Delito
Interno			x		x
Tiempo	x	x	x	x	x
Nivel de Mando	x	x	x	x	x
Traslado	x				
Estado del Traslado	x				
Solicitud		x			
Decisión		x			
Egreso			x		
Estado del Juicio				x	
Estado del conduce.				x	
Conduce				x	
Delito					x

**Figura 2. 2:** Matriz dimensional.

### **2.5. Validación de requisitos**

Los requisitos una vez definidos, necesitan ser validados. La validación de requisitos proporciona un mayor nivel de detalle y permite eliminar las ambigüedades en los datos iniciales. Además de comprobar que los requisitos sean consistentes y precisos.

La metodología propone como técnica de validación de requisitos, la confección de un prototipo funcional. El cual consiste en seleccionar un proceso del negocio, realizando con el mismo una pequeña versión de un mercado de datos. El objetivo del prototipo funcional es mostrarle las facilidades que brinda la herramienta y si cumple con los requisitos definidos. Para realizar el prototipo se utilizó como herramienta de apoyo el ejemplo que trae implementado por defecto la aplicación de Oracle Business Intelligence, usándose los reportes predefinidos con que cuenta dicha herramienta.

### **2.6. Modelado Dimensional**

El modelo dimensional es un proceso iterativo que permite ver los datos desde múltiples perspectivas y en diferentes niveles de detalles. El proceso de diseño comienza con el modelado dimensional de alto nivel que se obtiene a partir de los procesos que se encuentran en la matriz descrita anteriormente. El cual cuenta con los siguientes pasos para su construcción: elegir el proceso de negocio, establecer el nivel de granularidad, elegir las dimensiones e identificar medidas y las tablas de hechos.

#### **2.6.1. Granularidad**

La granularidad representa el nivel de detalle al que se desea almacenar, es la información sobre el negocio que se esté analizando. Mientras mayor sea el nivel de detalle de los datos, se tendrán mayores posibilidades analíticas, porque los mismos podrán ser resumidos. (5)

Para todos los procesos el nivel de granularidad definida para el tiempo es el grano diario. La granularidad especificada para el nivel de mando es centro penitenciario, a excepción de los procesos Egreso y Delito donde su nivel de detalle es colectivo.

### 2.6.2. Tablas de Dimensiones

Las tablas de dimensiones son elementos para ubicar datos que participan en el análisis, son datos que deben de tenerse en cuenta para poder tomar decisiones basadas en ellos.

Entre las dimensiones que se definieron se encuentran las siguientes:

**Dimensión Tiempo:** Contiene los datos fundamentales relacionados con el tiempo. Integrada por los siguientes atributos: num\_día, nom\_día, num\_mes, nom\_mes, año, trimestre, cuatrimestre, semestre, nonestre y fecha\_sql.

**Dimensión Nivel de Mando:** Guarda la información referente a los distintos niveles de mando relacionados con los centros penitenciarios. Integrada por los siguientes atributos: nombre, provincia, municipio, colectivo, tipo\_estructura y tipo\_centro.

**Dimensión Traslado:** Contiene información relacionada al traslado. Constituida por los siguientes atributos: tipo\_ejecucion, tipo\_traslado y motivo\_traslado.

**Dimensión Estado del Traslado:** Contiene información referente a los diferentes estados que puede tomar un traslado y como atributo que lo integra se tiene el siguiente: estado\_traslado.

**Dimensión Conduce:** Contiene información relacionada con los conduce. Constituida por los siguientes atributos: urgencia y motivo.

**Dimensión Estado del Conduce:** Contiene información referente a los diferentes estados que puede tomar un conduce y como atributo que lo integra se tiene el siguiente: estado\_conduce.

### 2.6.3. Tablas de hechos y medidas

Una tabla de hechos es una representación de un proceso de negocio. Esta tabla se encuentra rodeada de tablas de dimensiones que determinan cómo están distribuidos y relacionados los datos lógicamente en el ambiente del negocio.

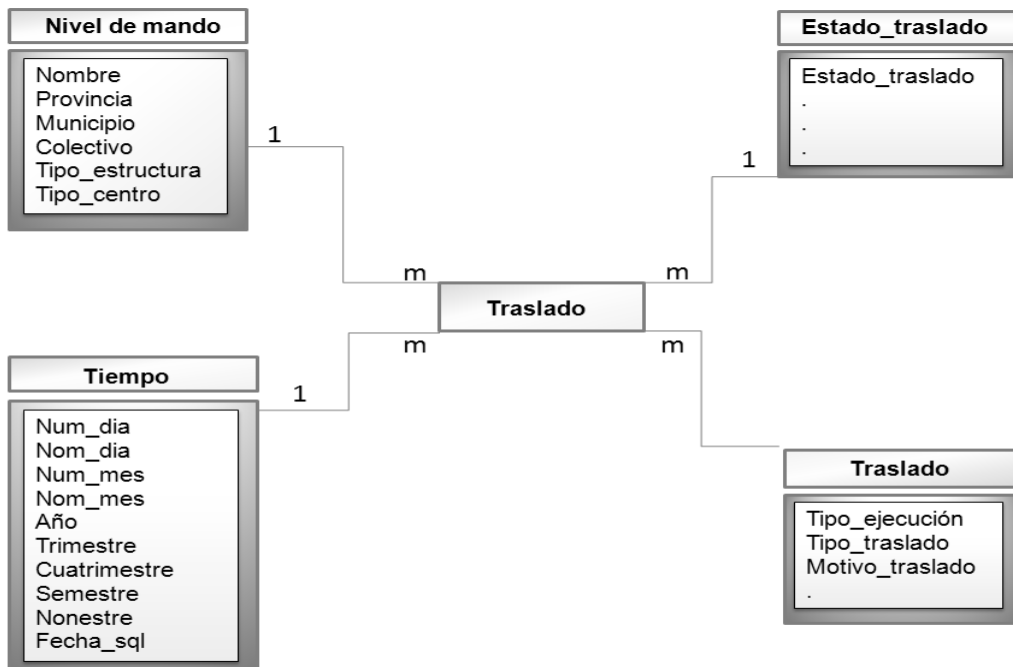
Entre las tablas de hechos que se definieron se encuentran las siguientes:

**Tabla de hecho Traslado:** está conformada por las llaves primarias de las dimensiones que caracterizan al traslado y la cantidad de internos que se trasladan.

**Tabla de hecho Conduce:** contiene las llaves primarias de las dimensiones que caracterizan al hecho de los conduces que se le hacen a los internos y la cantidad de internos que se conducen..

**2.6.4. Modelo dimensional detallado**

Al ser identificados los procesos y las dimensiones, se procede a conformar el modelo dimensional del mercado de datos. Este se realiza a partir de la construcción del modelo conceptual detallado, donde la información se representa a través de relaciones y atributos. De esta manera, quedan relacionadas las tablas de hechos con sus dimensiones correspondientes.



**Figura 2. 3:** Modelo dimensional detallado del proceso traslado.

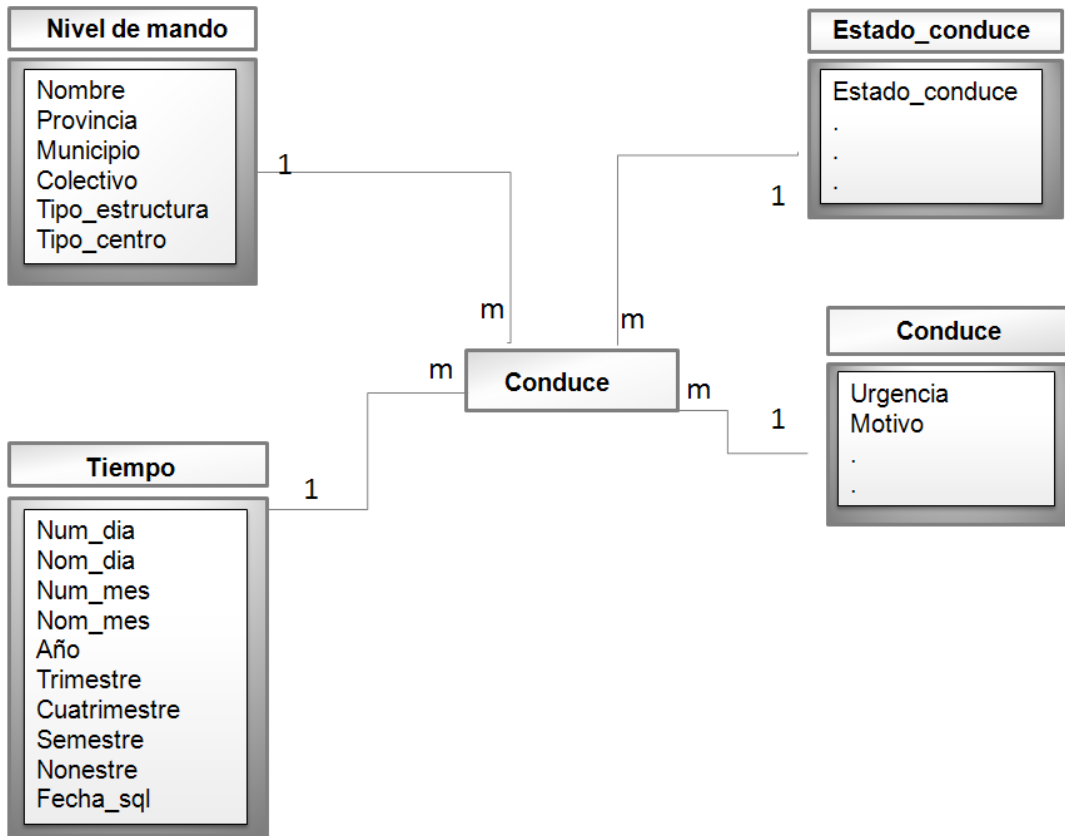


Figura 2. 4: Modelo dimensional detallado del proceso conduce.

### 2.6.5. Patrones de diseño para Almacenes de Datos

Para diseñar y construir un almacén de datos se requiere de mucho esfuerzo y análisis. En la actualidad los almacenes de datos suelen ser muy grandes y el uso de los patrones de diseño hacen que el trabajo sea más fácil y permite la solución de un problema de manera simple.

Para el diseño del almacén de datos se utilizaron los siguientes patrones:

**Llaves subrogadas:** se aplica para darle solución al problema de que los identificadores de las tablas de dimensiones no se repitan. Esto se resuelve mediante la creación de secuencias,

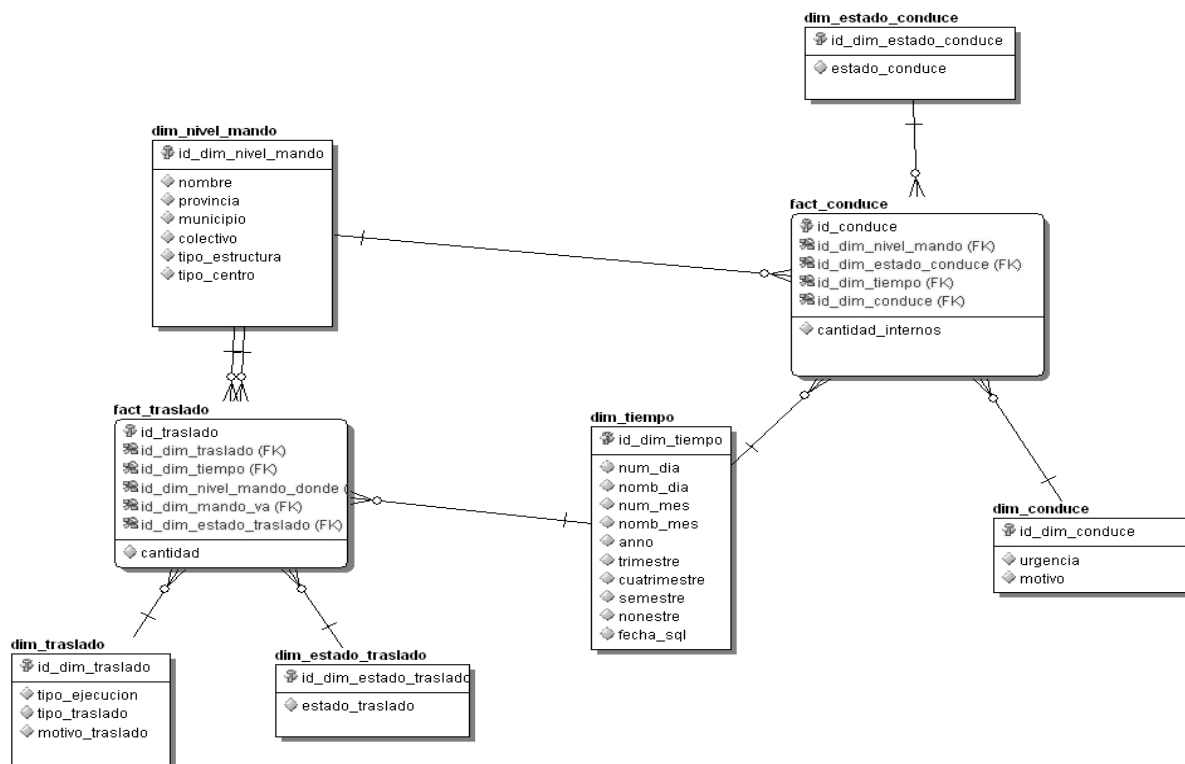


generándose una llave primaria para cada entidad. Esta solución garantiza que el identificador de la dimensión sea único.

**Dimensiones con varios roles:** cuando una dimensión interviene en una tabla de hechos varias veces, pero con diferentes roles, ocasiona que las dimensiones se almacenen en múltiples ocasiones. Por lo que es necesario crear un alias para cada referencia a la dimensión, tratándolas como dimensiones diferentes. De esta forma solo se almacenan una vez, permitiendo que la misma dimensión desempeñe varias funciones según el lugar donde se utilice.

### 2.6.6. Modelo de datos

Una vez obtenido los requisitos de información se confecciona el modelo de datos. Para ello primero se define el tipo de modelo que se utiliza, eligiéndose constelación por las características mencionadas en el capítulo anterior. Posteriormente se diseñan las tablas y columnas físicas del mercado de datos y finalmente se establecen las relaciones pertinentes entre ellas. En la figura 2.5 muestra una parte del modelo lógico para el modelo de datos.

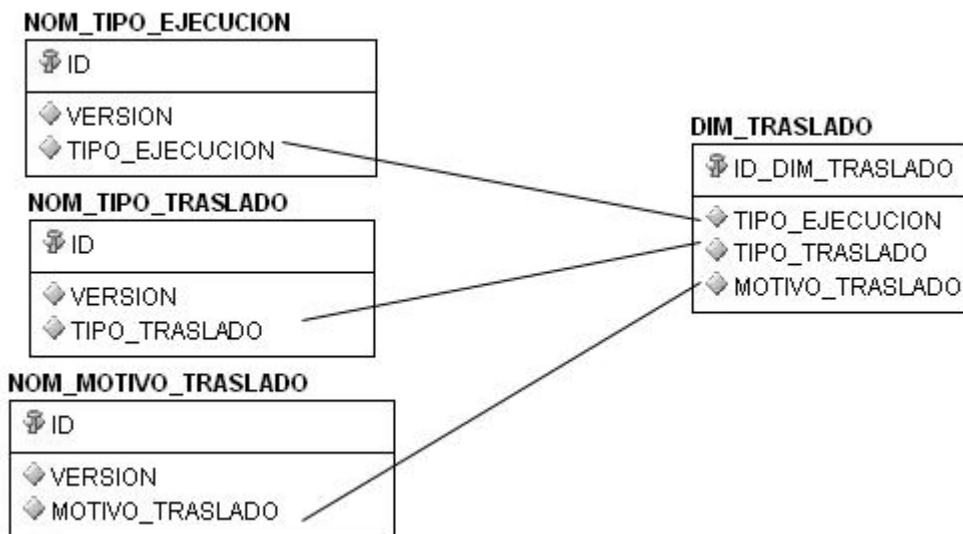


**Figura 2. 5:** Modelo de datos del mercado de datos.

**2.7. Diseño de las ETL**

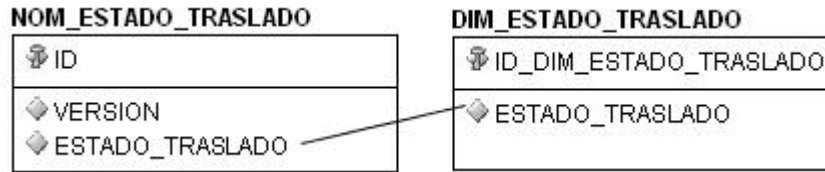
Luego de realizarse el modelo dimensional se inicia el proceso de extracción, transformación y carga, el cual tiene como resultado final la carga de los datos en el mercado de datos. Mediante la extracción se obtiene la información almacenada en las fuentes de datos origen, en la transformación se realizan operaciones de limpieza sobre los datos y en la carga se almacenan los datos en el mercado de datos.

Para el diseño de las ETL se identifican las tablas de la base de datos origen que permiten obtener los valores necesarios para el mercado de datos. Se utiliza como fuente de datos la base de datos de la solución SIDEPA del proyecto “Prisiones Cuba”. A continuación se muestran los siguientes diseños de las ETL:



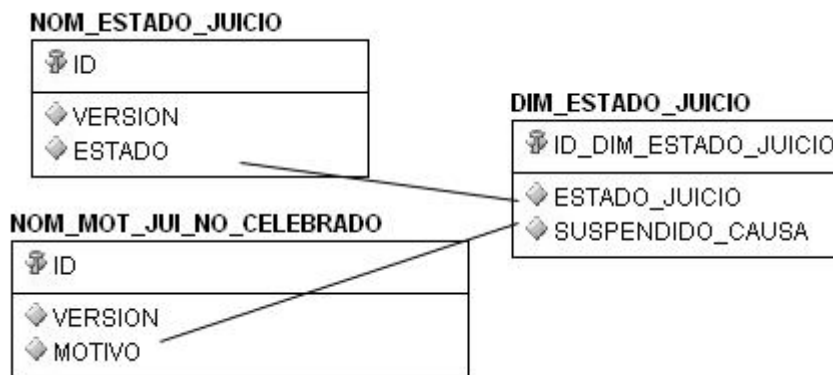
**Figura 2. 6:** Diseño de la ETL para la dimensión dim\_traslado.

Para poblar la dimensión dim\_traslado se identifican los siguientes atributos: tipo\_ejecucion de la tabla nom\_tipo\_ejecucion, tipo\_traslado de la tabla nom\_tipo\_traslado y motivo\_traslado de la tabla nom\_motivo\_traslado, como se muestra en la figura 2.6.



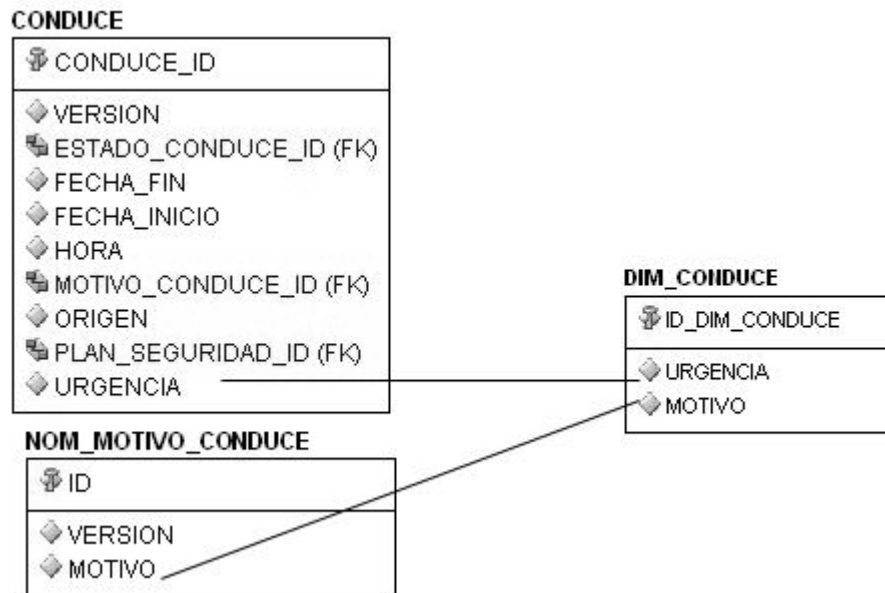
**Figura 2. 7:** Diseño de la ETL para la dimensión dim\_estado\_traslado.

Para cargar la dimensión dim\_estado\_traslado se identifica el atributo estado\_traslado de la tabla nom\_estado\_traslado, como se muestra en la figura 2.7.



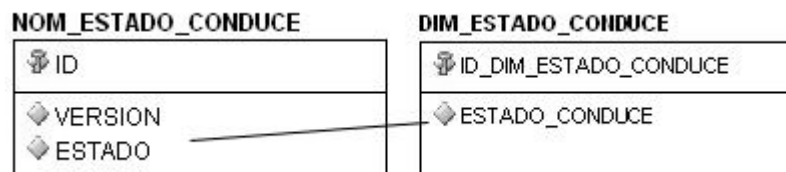
**Figura 2. 8:** Diseño de la ETL para la dimensión dim\_estado\_juicio.

Para cargar la dimensión dim\_estado\_juicio se identifican los siguientes atributos: estado de la tabla nom\_estado\_juicio y motivo de la tabla nom\_mot\_jui\_no\_celebrado, como se muestra en la figura 2.8.



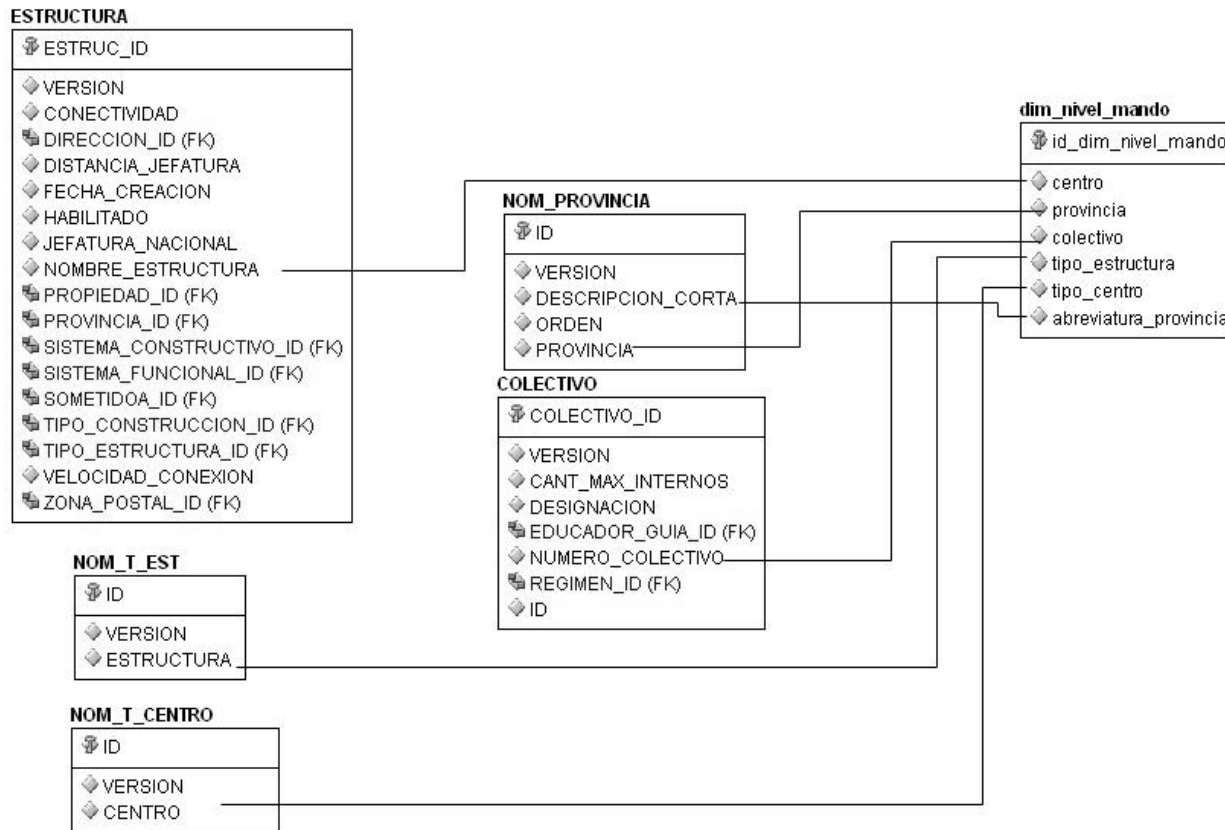
**Figura 2. 9:** Diseño de la ETL para la dimensión `dim_conduce`.

Para cargar la dimensión `dim_conduce` se identifican los siguientes atributos: `urgencia`, de la tabla `conduce` y `motivo`, de la tabla `nom_motivo_conduce`, como se muestra en la figura 2.9.



**Figura 2. 10:** Diseño de la ETL para la dimensión `dim_estado_conduce`.

Para poblar la dimensión `dim_estado_conduce` se identifica el atributo `estado` de la tabla `nom_estado_conduce`, como se muestra en la figura 2.10.



**Figura 2. 11:** Diseño de la ETL para la dimensión `dim_nivel_mando`.

Para cargar la dimensión `dim_nivel_mando` se identifican los siguientes atributos: `nombre_estructura`, de la tabla `estructura`, `provincia` de la tabla `nom_provincia`, `numero_colectivo` de la tabla `colectivo`, `estructura` de la tabla `nom_t_est` y `centro` de la tabla `nom_t_centro`, como se muestra en la figura 2.11.

## 2.8. Conclusiones parciales

En el presente capítulo se realizó un análisis de los principales temas analíticos del área de Registro Legal, donde se determinaron 55 requisitos de información. Además, se diseñó el modelo de datos donde se identificaron las tablas de hechos y sus dimensiones, así como el diseño de las ETL.

### CAPÍTULO 3: IMPLEMENTACIÓN Y PRUEBAS

#### Introducción

Este capítulo aborda el tema de la implementación del mercado de datos y las pruebas realizadas. En la implementación están presentes la extracción, transformación y carga de los datos, la configuración de la Herramienta de Administración, la creación de los reportes y por último la confección del tablero de mando. Además de todo el proceso de implementación se realizan las pruebas para detectar no conformidades al mercado de datos.

#### 3.1. Implementación

Para la implementación es necesario realizar la extracción, transformación y carga de los datos. Una vez cargado el mercado de datos se configura el repositorio de metadatos, el cual le da acceso a los datos a las herramientas de consulta. Por último se crean los reportes y se elabora el tablero de mando. A continuación se describen las actividades antes mencionada.

##### 3.1.1. Implementación de las ETL

En la implementación de las ETL se lleva a cabo la extracción de los datos de las fuentes correspondientes, estos datos se copian a un área temporal para su transformación. Esta tiene como tarea principal la limpieza de los datos, donde se realiza el tratamiento de valores nulos, codificación, entre otras. Por último los datos transformados se cargan en el mercado de datos.

La implementación de las ETL fue realizado con la herramienta Oracle Warehouse Builder.

El proceso de las ETL es realizado a cada cubo<sup>1</sup>. Primero se cargan los datos de las dimensiones. Posteriormente se extraen los datos en las tablas temporales, los cuales son transformados y cargados en la tabla de hechos. Finalmente estos datos son cargados en las tablas de hecho. Los procesos ETL de los cubos Traslado, Conduce y Juicio utilizan las dimensiones comunes: dim\_nivel\_mando y dim\_tiempo.

A continuación se especifica el proceso ETL del cubo Traslado.

- **ETL DIM\_ESTADO\_TRASLADO.**

---

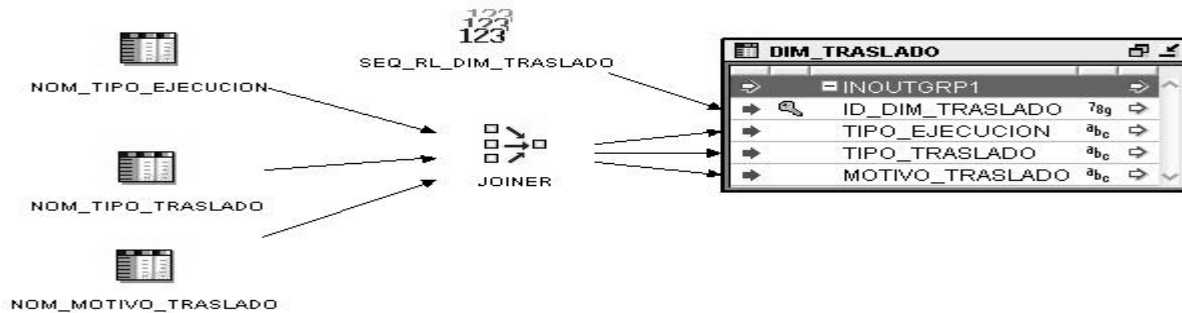
<sup>1</sup> Representación o visualización de una tabla de hecho y sus dimensiones correspondientes.



**Figura 3. 1: ETL DIM\_ESTADO\_TRASLADO.**

Para la carga de la dimensión dim\_estado\_traslado se toma como fuente de entrada los datos necesarios que se encuentra en la tabla nom\_estado\_traslado y como identificador de la dimensión, el valor generado por la secuencia SEQ\_RL\_DIM\_ESTADO\_TRASLADO.

- **ETL DIM\_TRASLADO.**



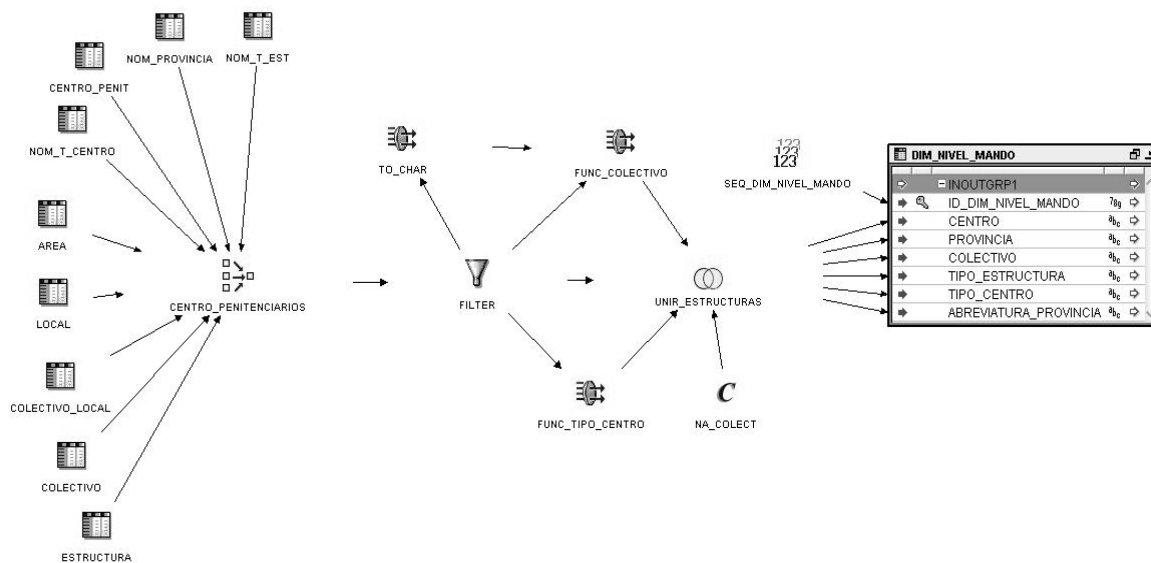
**Figura 3. 2: ETL DIM\_TRASLADO.**

Para la carga de la dimensión dim\_traslado se toma como fuente de entrada los datos necesarios que se encuentran en las tablas: nom\_tipo\_ejecucion, nom\_tipo\_traslado y nom\_motivo\_traslado.

Posteriormente se utiliza el operador JOINER para establecer relaciones entre las tablas (similar al JOIN en SQL) En este caso es el encargado de unir las tablas antes mencionadas.

El identificador de la dimensión es el valor generado por la secuencia SEQ\_RL\_DIM\_TRASLADO.

- **ETL DIM\_NIVEL\_MANDO.**



**Figura 3. 3: ETL DIM\_NIVEL\_MANDO.**

Para la carga de la dimensión dim\_nivel\_mando se toma como fuente de entrada los datos necesarios que se encuentran en las tablas: estructura, colectivo, colectivo\_local, local, area, nom\_t\_centro, centro\_penit, nom\_provincia y nom\_t\_est.

Después se procede a utilizar el operador JOINER “centro\_penitenciaros”, encargado de unir las tablas antes mencionadas.

Luego se utiliza el operador FILTER, que tiene como objetivo descartar los centros penitenciarios que no tienen definido colectivos.

Se utilizan algunas funciones propias del OWB y otras que son necesarias para implementar las transformaciones. Las funciones implementadas son:

- **Func\_tipo\_centro:** recibe como parámetro el tipo de estructura y el tipo de centro. Si el tipo de estructura es ‘Jefatura’ devuelve ‘No Aplica’, en caso contrario devuelve el tipo de centro correspondiente.
- **Func\_colectivo:** recibe como parámetro el tipo de estructura y el número de colectivo. Si el tipo de estructura es ‘Jefatura’ devuelve ‘No Aplica’, en caso contrario devuelve el número de colectivo correspondiente.



Posteriormente se emplea el operador SET\_OPERATION “unir\_estructuras”, su función es combinar los resultados de las consultas en un único resultado. Son unidas dos consultas; la primera devuelve los centros penitenciarios que contienen colectivos asociados y la segunda retorna los mismos centros penitenciarios pero sustituyendo el valor del atributo colectivo por ‘No Aplica’.

Se utiliza el operador CONSTANT “na\_colect”, en el cual se define la constante 'No Aplica' que actúa como parámetro de entrada en la dimensión dim\_nivel\_mando.

El identificador de la dimensión es el valor generado por la secuencia SEQ\_DIM\_NIVEL\_MANDO.

- ETL DIM\_TIEMPO



**Figura 3. 4: ETL DIM\_TIEMPO.**

La carga de la dimensión dim\_tiempo recibe como parámetros la fecha de inicio y la cantidad de años que se quiere generar.

Se utiliza el operador de función de tabla DAY\_TABLE\_FUNCTION para crear cada una de las fechas que se encuentran dentro del intervalo. Función desarrollada por OWB.

Se utilizan algunas funciones propias del OWB y otras que son necesarias para implementar las transformaciones. Las funciones implementadas son:

**cal\_trimestre:** dado el número del mes devuelve el número del trimestre correspondiente.

**cal\_cuatrimestre:** dado el número del mes devuelve el número del cuatrimestre correspondiente.

**cal\_semestre:** dado el número del mes devuelve el número del semestre correspondiente.

- ETL STG\_UBICACION.

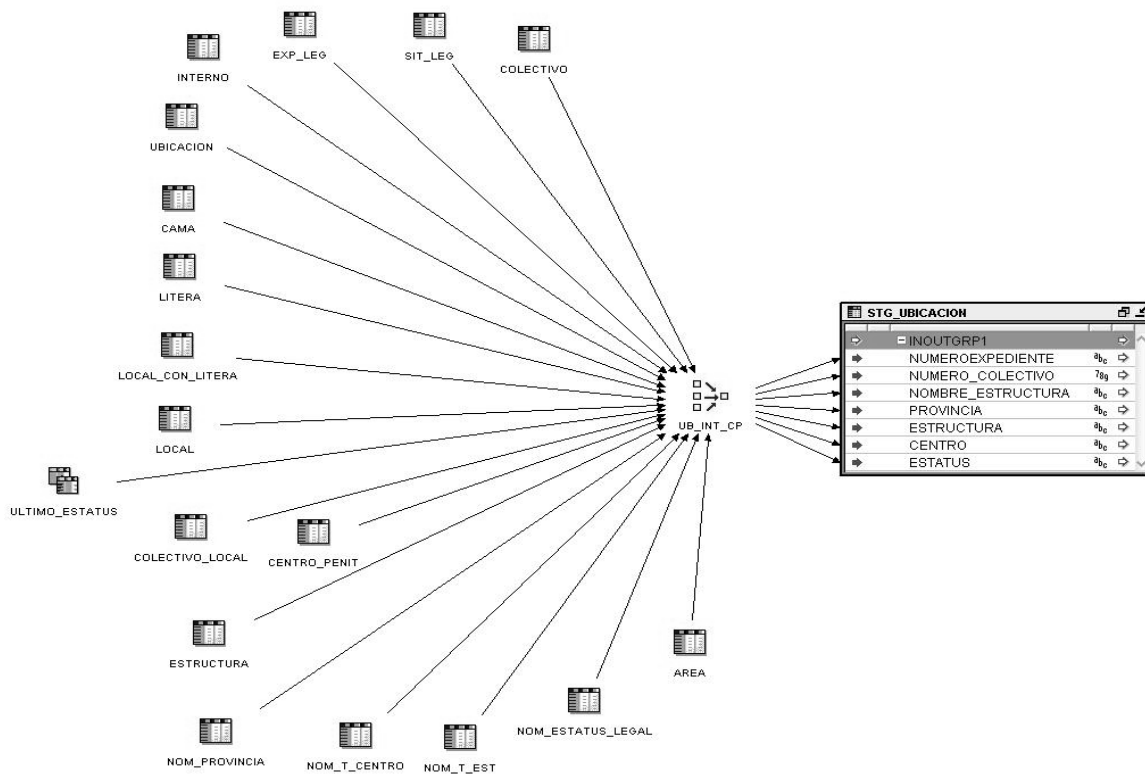


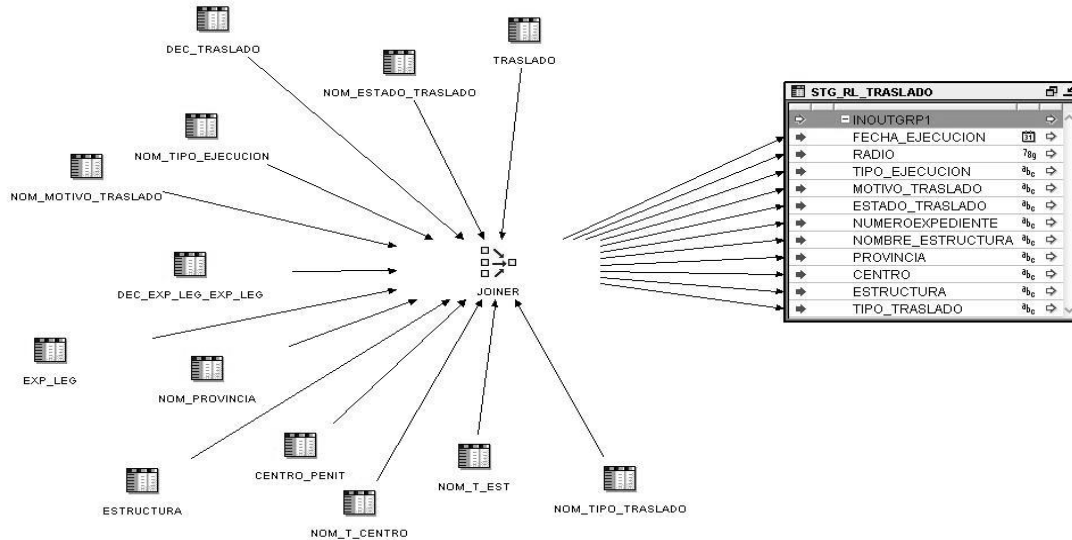
Figura 3. 5: ETL STG\_UBICACION.

Para la carga del stg\_ubicacion se toma como fuente de entrada los datos necesarios que se encuentran en las tablas: colectivo, sit\_leg, exp\_leg, interno, ubicación, cama, litera,

local\_con\_litera, local, colectivo\_local, centro\_penit, estructura, nom\_provincia, nom\_t\_centro, nom\_t\_est, nom\_estatus\_legal y area.

Por último se utiliza el operador JOINER, encargado de unir las tablas antes mencionadas.

- **ETL STG\_RL TRASLADO.**

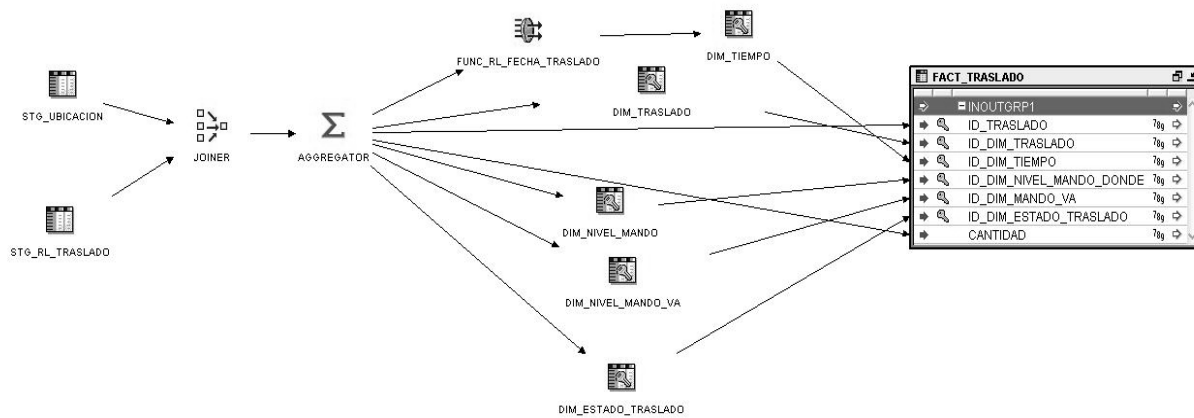


**Figura 3. 6: ETL STG\_RL TRASLADO.**

Para la carga del stg\_traslado se toma como fuente de entrada los datos necesarios que se encuentran en las tablas: traslado, nom\_estado\_traslado, dec\_traslado, nom\_tipo\_ejecucion, nom\_motivo\_traslado, dec\_exp\_leg\_exp\_leg, exp\_leg, estructura, nom\_provincia, centro\_penit, nom\_t\_centro, nom\_t\_est y nom\_tipo\_traslado.

Para terminar se utiliza el operador JOINER, encargado de unir las tablas antes citadas.

- **ETL FACT\_TRASLADO.**



**Figura 3. 7: ETL FACT\_TRASLADO.**

Para la carga de la fact\_traslado se utiliza como datos de entrada las tablas: stg\_ubicacion y stg\_rl\_traslado. Posteriormente se aplica el operador JOINER, su objetivo es establecer relaciones entre las tablas; en este caso es el encargado de unir las tablas antes mencionadas. Se procede a emplear el operador AGREGATOR, el cual es similar a las funciones de agregación en SQL; la función de agregación puesta en práctica es COUNT.

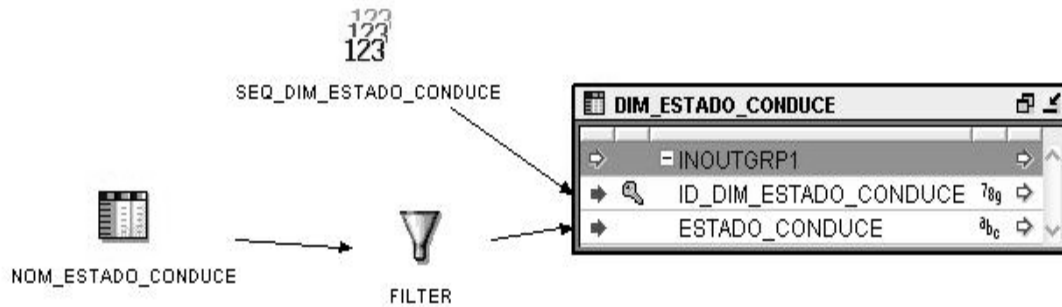
El operador Key\_Lookup u operador de consulta de llaves se utiliza para leer los datos de las dimensiones correspondientes a través de su llave. En este caso se obtienen como salida los datos de las dimensiones: dim\_tiempo, dim\_nivel\_mando, dim\_estado\_traslado y dim\_traslado. La salida de este operador le proporciona la entrada a la tabla de hecho.

Finalizado la ETL FACT\_TRASLADO queda cargado el cubo Traslado en el mercado de datos.

### Proceso ETL del cubo Conduce

La carga del cubo Conduce está integrada por los procesos ETL: DIM\_NIVEL\_MANDO y DIM\_TIEMPO explicados anteriormente, por lo que a continuación solo se aborda el proceso ETL: DIM\_ESTADO\_CONDUCE, DIM\_CONDUCE, STG\_CONDUCE y FACT\_CONDUCE.

- **ETL DIM\_ESTADO\_CONDUCE**



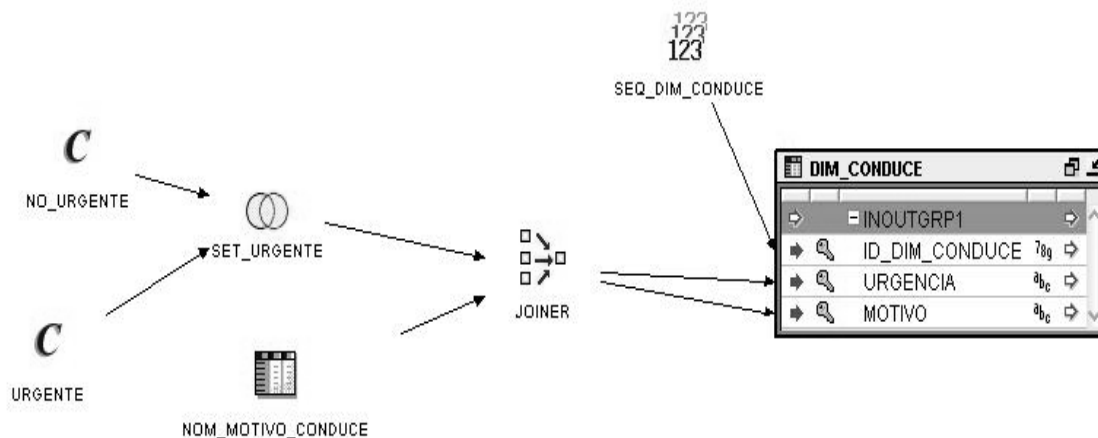
**Figura 3. 8: ETL DIM\_ESTADO\_CONDUCE.**

Para la carga de la dimensión dim\_estado\_conduce se escoge como fuente de entrada los datos necesarios que se encuentran en la tabla: nom\_estado\_conduce.

Luego se utiliza el operador FILTER para seleccionar los valores que sean diferentes de 'En ejecución'.

El identificador de la dimensión es el valor generado por la secuencia SEQ\_DIM\_ESTADO\_CONDUCE.

- **ETL DIM\_CONDUCE**



**Figura 3. 9: ETL DIM\_CONDUCE.**

Para la carga de la dimensión dim\_conduce se elige como fuente de entrada los datos necesarios que se encuentran en la tabla: nom\_motivo\_conduce.

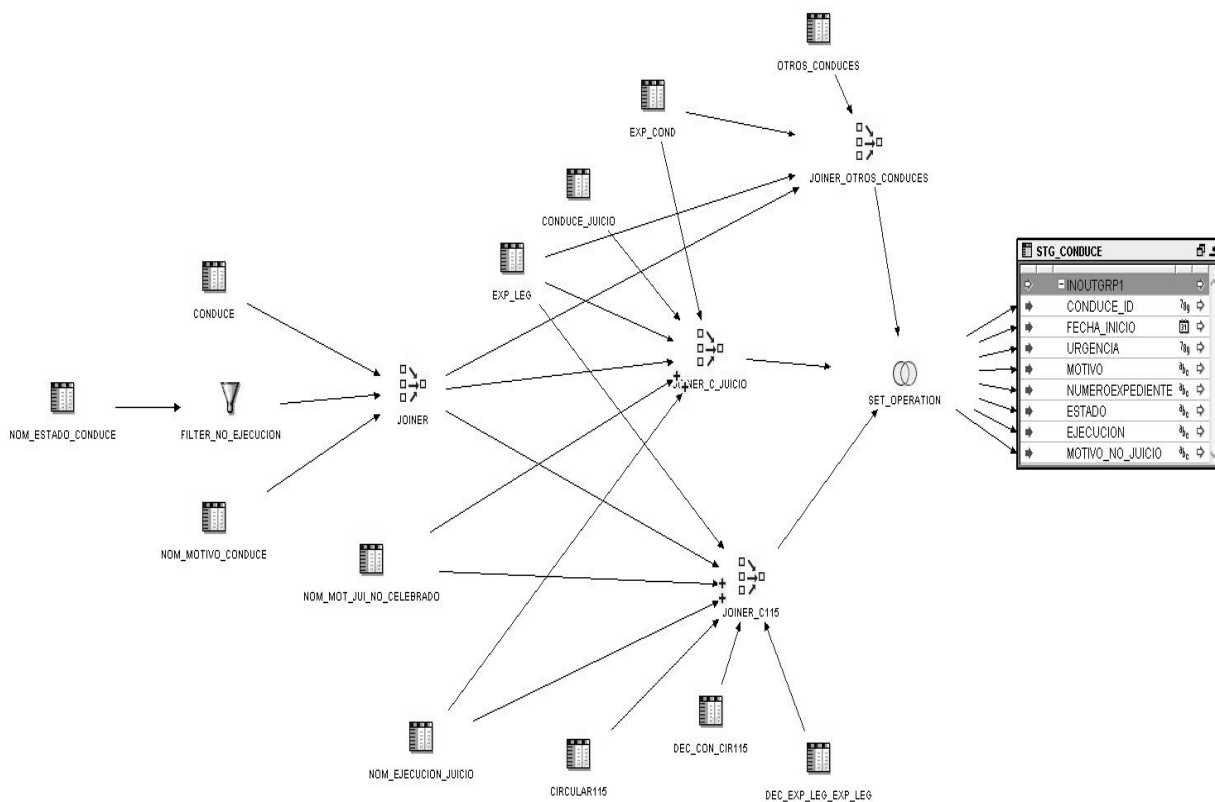
Se emplean los operadores CONSTANT: “no\_urgente” y “urgente”. Su funcionalidad es cargar datos que no se encuentran en la base de datos y que son necesarios en la dimensión dim\_conduce.

Posteriormente se emplea el operador SET\_OPERATION “set\_urgente”, su función en este caso es la de combinar las constantes en un único resultado.

Posteriormente se utiliza el operador JOINER, su función es establecer relaciones entre las tablas. En este caso es el encargado de unir la tabla antes mencionada con la salida del operador SET\_OPERATION.

El identificador de la dimensión es el valor generado por la secuencia SEQ\_DIM\_CONDUCE.

- **ETL STG\_CONDUCE**



**Figura 3. 10: ETL STG\_CONDUCE.**

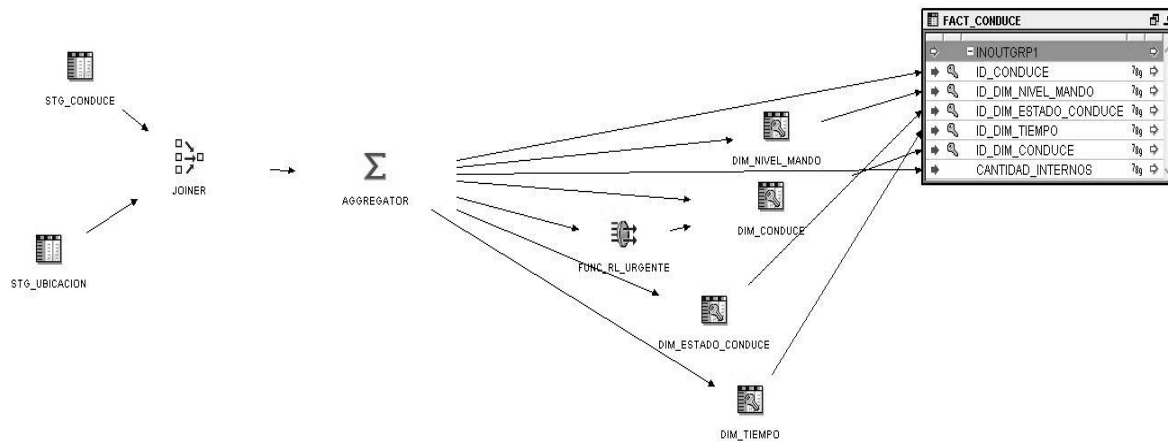
Para la carga del stg\_conduce se toma como fuente de entrada los datos necesarios que se encuentran en las tablas: conduce, nom\_estado\_conduce, nom\_motivo\_conduce, nom\_ejecucion\_juicio, nom\_mot\_jui\_no\_celebrado, circular115, exp\_cond, exp\_leg, conduce\_juicio, dec\_con\_cir115, dec\_exp\_leg\_exp\_leg, otros\_conduce.

Se emplea el operador FILTER “filter\_no\_ejecucion”, su objetivo es aplicar una condición de filtro definida, su salida es el resultado del filtro aplicado.

Posteriormente se utilizan los operadores JOINER: “joiner”, “joiner\_otros\_conduces”, “joiner\_c\_juicio” y “joiner\_c\_conduce”. Su objetivo es establecer relaciones entre las tablas, en este caso es el encargado de unir las tablas citadas anteriormente.

Por último se emplea el operador SET\_OPERATION su función en este caso es la de combinar las salidas de los joiner en un único resultado.

- **ETL FACT\_CONDUCE**



**Figura 3. 11: ETL FACT\_CONDUCE.**

Para la carga de la fact\_conduce se utiliza como datos de entrada las tablas: stg\_ubicacion y stg\_conduce.

Posteriormente se utiliza el operador JOINER, su objetivo es unir varios conjuntos de filas de diferentes fuentes con diferentes cardinalidades. En este caso es el encargado de unir las tablas citadas anteriormente.

Se procede a utilizar el operador AGREGADOR el cual es similar a las funciones de agregación en SQL. La función de agregación utilizada es COUNT.

Para la transformación se aplica la función func\_rl\_urgente. La misma recibe como parámetro de entrada un valor numérico y devuelve “urgente” o “no urgente”.

El operador Key\_Lookup u operador de consulta de llaves se utiliza para leer los datos de las dimensiones correspondientes a través de su llave. En este caso se obtienen como salida los datos de las dimensiones: dim\_tiempo, dim\_nivel\_mando, dim\_estado\_conduce y dim\_conduce. La salida de este operador le proporciona la entrada a la tabla de hecho.



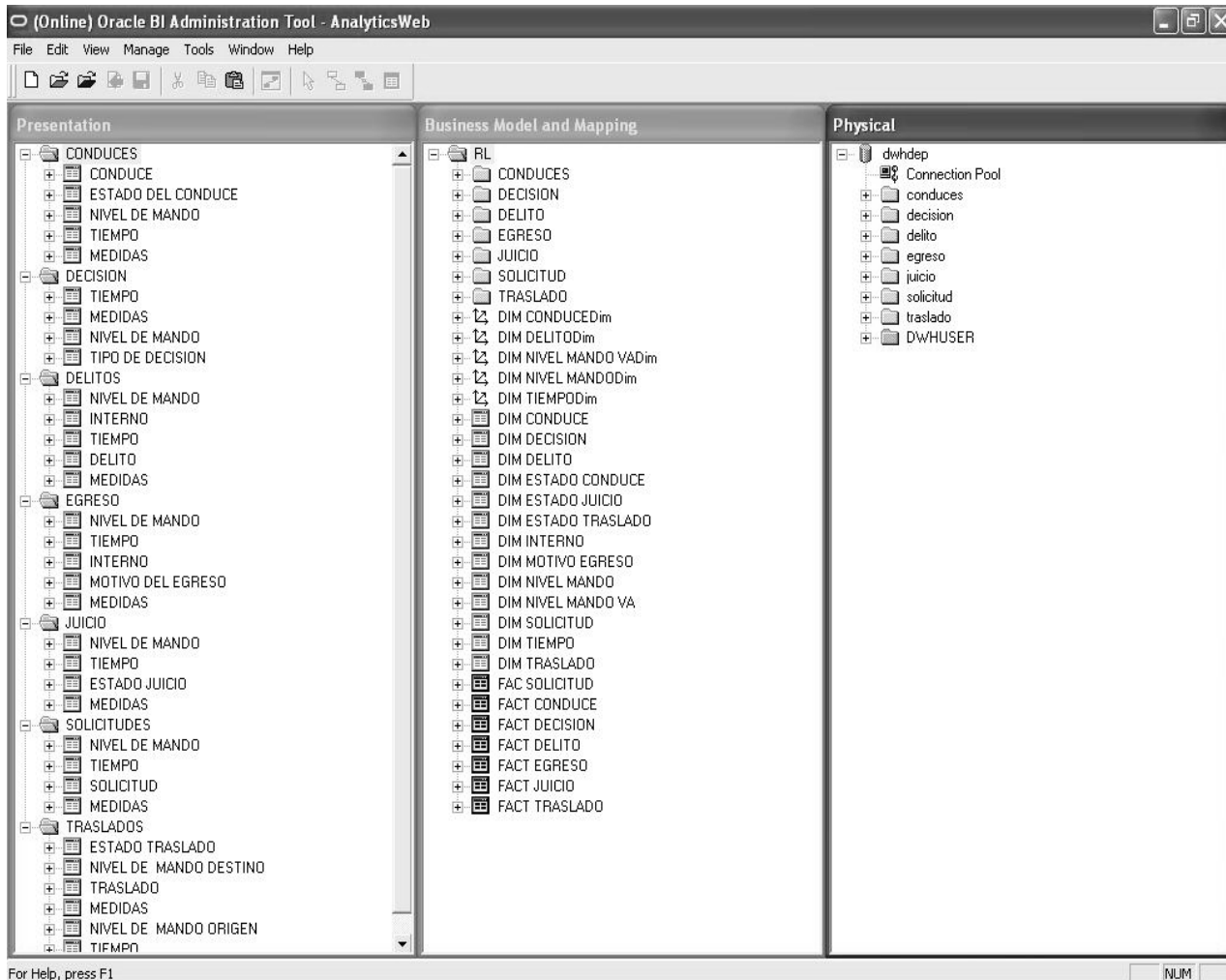
Terminado la ETL FACT\_CONDUCE queda cargado el cubo Conduce en el mercado de datos.

### **3.1.2. Especificación y desarrollo de las aplicaciones de Inteligencia de negocio**

Un elemento fundamental en la explotación de los mercados de datos son las aplicaciones de Inteligencia de Negocio. Estas aplicaciones son concebidas para consultar, analizar y presentar información dirigida a apoyar las necesidades del negocio. Las aplicaciones de inteligencia de negocio proporcionan al usuario el acceso a la información guardada en el mercado de datos. Estas aplicaciones se dividen en las herramientas de construcción de reportes y el tablero de mando.

#### **3.1.2.1. Creación del repositorio de metadatos**

Al terminar de ser cargados los cubos pertenecientes al mercado de datos se procede a configurar el repositorio de metadatos con el objetivo de permitirles a las herramientas de inteligencia de negocio interactuar con los datos del mercado. Para realizar dicha configuración, se emplea la herramienta de administración (Administration Tool).



**Figura 3. 12:** Repositorio de metadatos.

El repositorio creado está conformado por tres capas: la capa física, modelo de negocio y la capa de presentación, como se puede apreciar en la figura 3.12. En la creación del repositorio primeramente se importan las tablas de los cubos pertenecientes al mercado de datos en la capa física, definiéndose las restricciones de llaves primarias y llaves foráneas que permiten las relaciones entre las tablas. En el modelo de negocio o capa intermedia se especifican las correspondencias entre el modelo de negocio y los esquemas de la capa física. Se define el modelado de los datos del negocio donde se le asigna a las fuentes físicas el vocabulario de negocio de los usuarios. Además se definen las jerarquías de las dimensiones y las medidas con

sus respectivas reglas de agregaciones. Por último, se construye la capa de presentación, en la cual se propone como serán vistos los datos por los usuarios finales a través de las herramientas de consulta. La cual permite la organización de las tablas por áreas temáticas y reemplazar los nombres de las tablas y columnas por valores representativos del negocio.

### **3.1.2.2. Herramienta de construcción de reportes**

Después de configurada la herramienta de administración, se procede a la creación de los reportes asociados a cada uno de los procesos, en el cual se utiliza como herramienta de trabajo el Oracle BI Answers para realizar consultas Ad-Hoc<sup>2</sup>. El mismo permite a los usuarios hacer sus propias consultas, las que puedan ser visualizadas a través de tablas, gráficas interactivas, tablas dinámicas e informes. A continuación se muestran reportes relacionados al proceso Traslado:

---

<sup>2</sup> Consultas en bases de datos, el sistema le permite al usuario personalizar una consulta en tiempo real, en vez de estar atado a las consultas prediseñadas para informes.

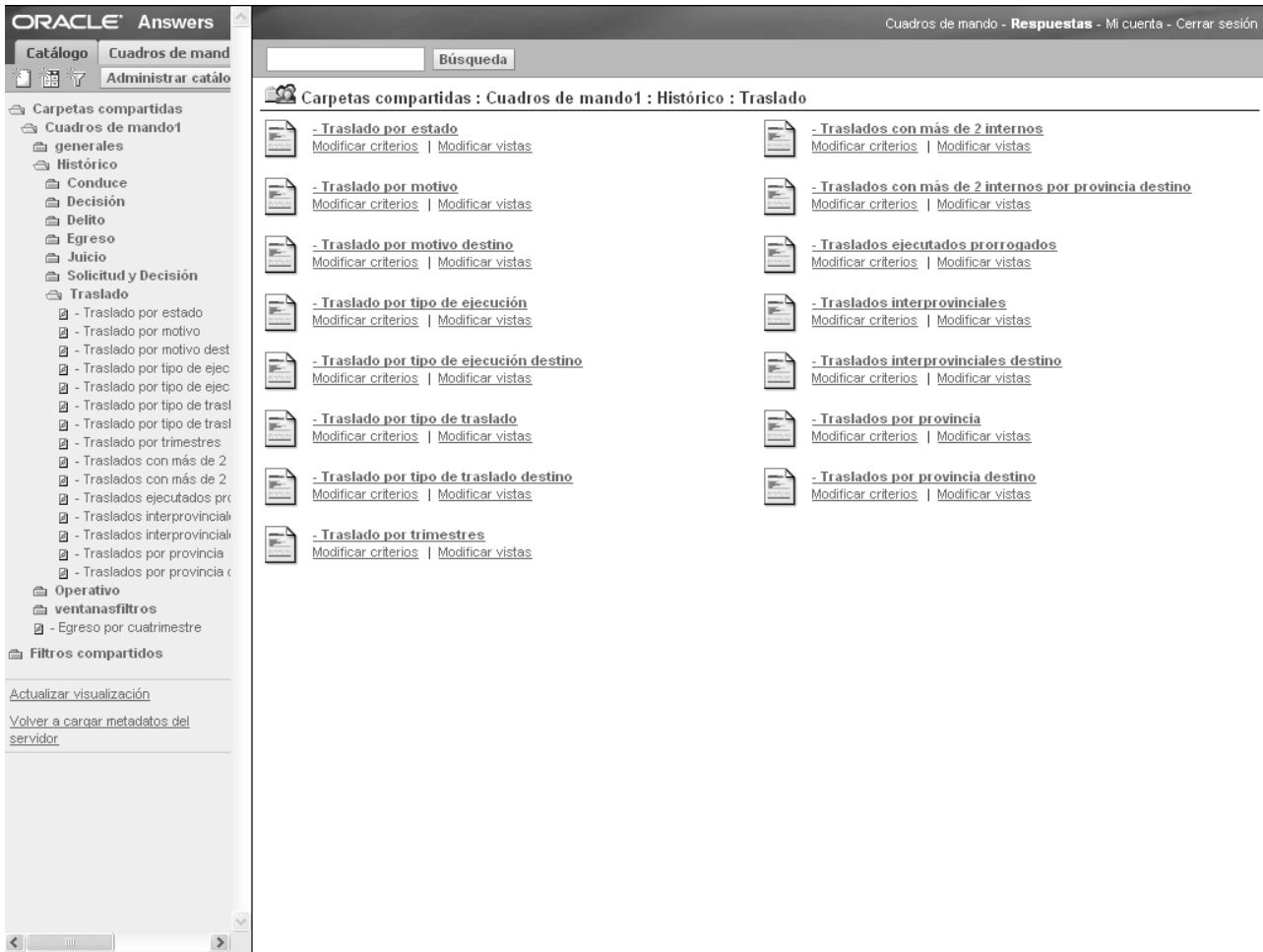


Figura 3. 13: Reportes guardados del proceso Traslado.

Los usuarios tienen la posibilidad de ver y modificar las columnas y filtros de los reportes. Pueden especificar el orden en que deben devolverse los resultados, el formato y las fórmulas de las columnas. A continuación se muestra el espacio de trabajo de creación y configuración de reportes:

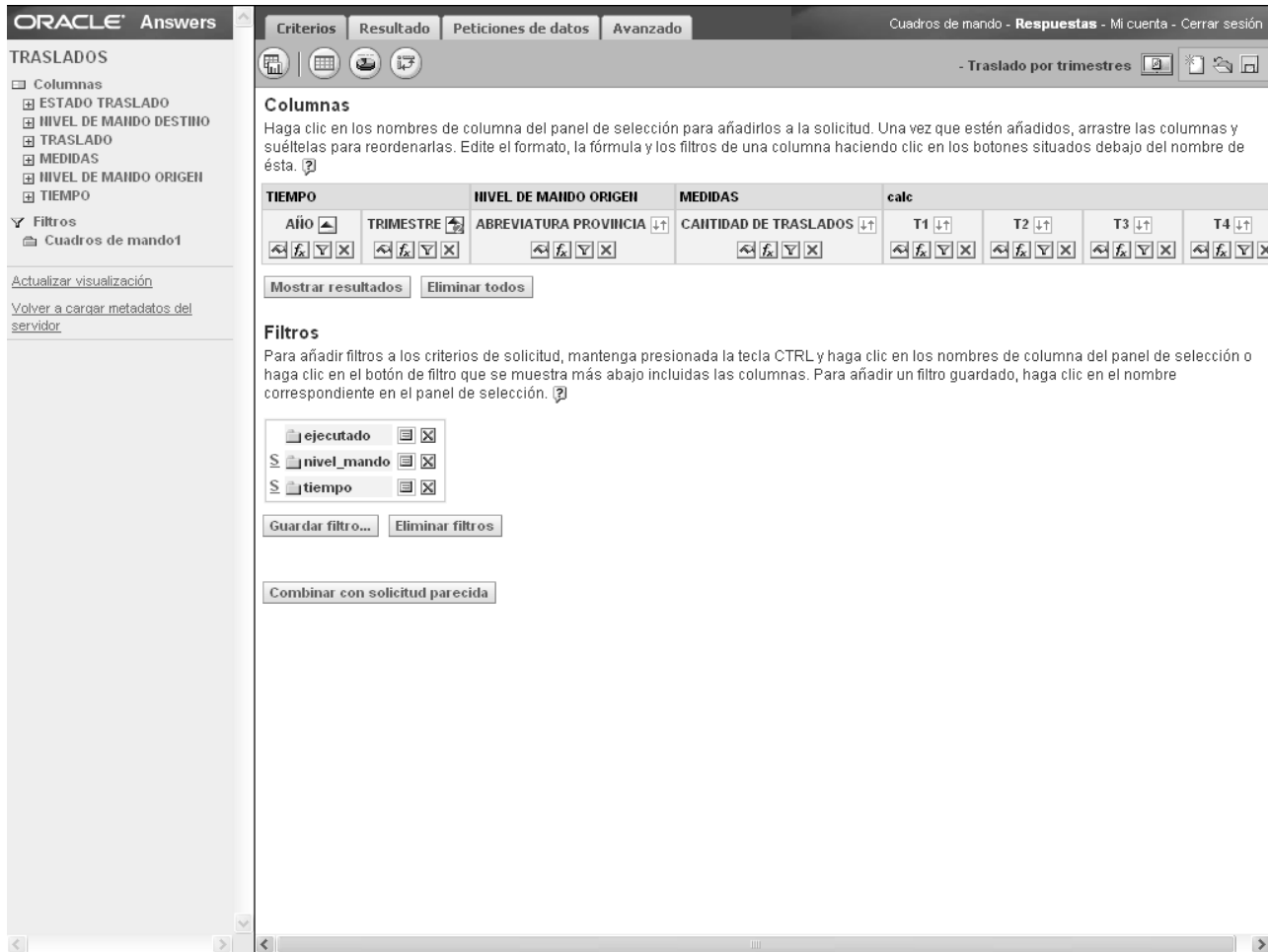


Figura 3. 14: Configuración de reportes en el espacio de trabajo Oracle BI Answer.

### 3.1.2.3. Tableros de Mando

El tablero de mando (Dashboard) les proporciona a los usuarios un acceso interactivo a la información y personalización dinámica. El usuario tiene la posibilidad de navegar por la aplicación, modificar los reportes e interactuar con los resultados.

La solución está integrada por los tableros de mando: Histórico y Operativo. El tablero de mando Histórico contribuye al análisis de toda la información recopilada o la filtrada por un rango de fecha. Por otro lado el tablero de mando Operativo muestra los resultados actuales, concebida para consultar los datos que emplean los usuarios en el trabajo diario. Las siguientes figuras muestran los tableros de mando:



Figura 3. 15: Tablero de mando Histórico.



Figura 3. 16: Tablero de mando Operativo.

### 3.1.3. Tipos de procesamiento

En un mercado de datos los tipos de procesamientos proporcionan un análisis interactivo de los datos mediante las dimensiones y sus jerarquías. A continuación se describen los tipos de procesamiento que tienen lugar en el mercado de datos:

**Drill-Down:** Permite cambiar el nivel de agregación de los datos consultados, navegando a través de las jerarquías de los datos, desde los de mayor nivel de agregación hasta los más detallados. Un ejemplo de jerarquía es: provincia, centro y colectivo en la dimensión nivel de mando.

**Slice y Dice:** Consiste en restringir los valores de una consulta activa. El slice conforma un subconjunto de los datos, aplicándole restricciones a una sola dimensión mediante la elección de un valor o rango de valores. Los dice son la combinación de varios slice, donde se aplican más de una restricción e intervienen dos o más dimensiones.

**Pivot:** permite seleccionar el orden de visualización de las dimensiones, con el objetivo de examinar la información desde diferentes perspectivas.

### 3.2. Pruebas

Las pruebas tienen como objetivo verificar que el sistema cumpla con los requisitos que se describieron para su desarrollo. Son empleadas para encontrar errores en la implementación, que no se hayan encontrado durante su construcción.

Al finalizar el diseño e implementación del mercado de datos del área de Registro Legal, se realizan las pruebas utilizando como herramienta la lista de chequeo, con el objetivo de verificar el buen funcionamiento del mercado de datos.

Las pruebas basadas en lista de chequeo (check list, como se conoce internacionalmente en inglés) realiza una serie de preguntas en forma de cuestionario con el objetivo de comprobar el cumplimiento de los requisitos. Permite identificar problemas existentes en la aplicación.

La metodología de Kimball propone utilizar las listas de chequeo para evaluar la aplicación desarrollada. Realizándose dos iteraciones de pruebas por el grupo de desarrollo, la primera para encontrar las no conformidades y una segunda donde se realizaron las pruebas de regresión para comprobar que no se introdujeron nuevos errores cuando se arreglaron las no conformidades detectadas.

Para medir la calidad de las pruebas que se realizan se define un criterio de evaluación en un intervalo de 1 a 10; donde de 1 a 5 representa mala calidad, de 6 a 7 baja calidad, de 8 a 9 aceptable y 10 calidad superior.

### 3.3. Resultados de las pruebas

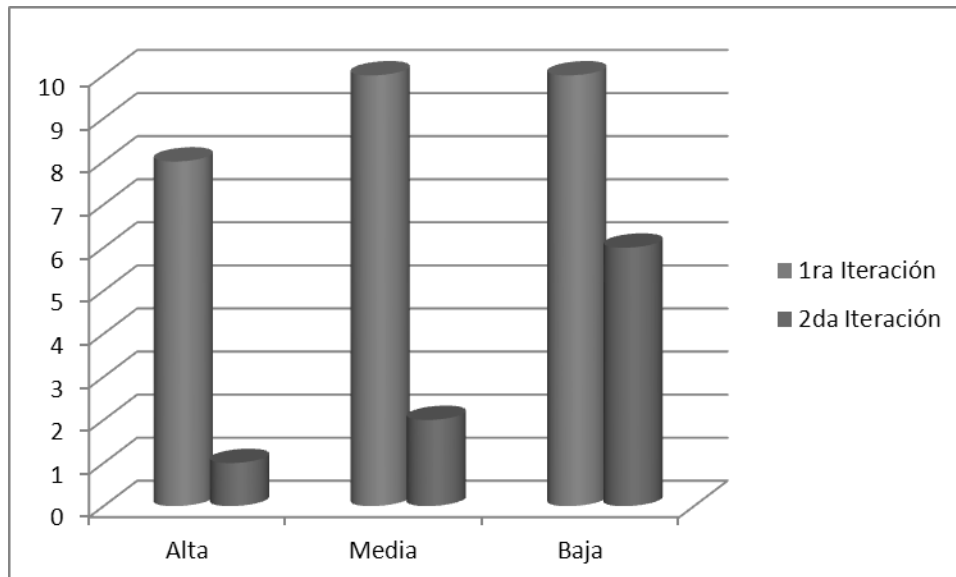
Luego de aplicadas las listas de chequeo a la aplicación se detectó durante la primera iteración de las pruebas 28 no conformidades, a las que se le dieron solución. Fue realizada una segunda para comprobar que no hayan surgido nuevas no conformidades y corregido las anteriores,



detectándose 9 no conformidades, que fueron resueltas. A continuación se muestra la lista de chequeo resumen resultado de las aplicadas por los desarrolladores de la aplicación y el resultado de las pruebas en la siguiente gráfica:

Nivel de importancia	Criterio de evaluación	Eval	NP	Cantidad de elementos afectados	Comentarios
Alta	¿Los subtítulos de las filas y las columnas corresponden a la descripción de sus atributos?	7		2	
Alta	¿Los reportes son configurables a través de la interfaz del sistema?	8		2	
Baja	¿La interfaz está orientada a facilitar el uso de las funciones del sistema por parte de los usuarios?	9		3	
Media	¿Los usuarios son capaces de manipular los resultados de manera que se ajusten a sus necesidades, conformando nuevos reportes?	8		2	
Alta	¿El sistema muestra las funcionalidades	7		3	

	diferentes de acuerdo al rol de los usuarios?				
Alta	¿Se puede navegar a través de los niveles de granularidad de los reportes?	9		1	
Media	¿Los procesos pueden ser analizados desde todas sus dimensiones?	8		3	
Media	¿El título de la gráfica se corresponde con los datos mostrados?	8		2	
Baja	¿Los elementos de la tabla siguen el mismo formato de letra?	8		3	
Media	¿Los parámetros de los requisitos se muestran en los reportes?	8		3	
Baja	¿La interfaz tiene errores ortográficos?	8		2	
Baja	¿El tipo de gráfica es el más representativo para mostrar la información?	8		2	



**Figura 3. 17:** Resultado de las pruebas.

### 3.4. Conclusiones parciales

Con la implementación de las ETL se pobló el mercado de datos para el área de Registro Legal. A las aplicaciones de usuario final se les permitió el acceso a los datos con la configuración del repositorio de metadatos. Se crearon los reportes predefinidos dados los requisitos de información y se configuró el tablero de mando permitiendo visualizarlos. También fueron corregidos los errores que se detectaron con la aplicación de las pruebas al mercado de datos.

### **CONCLUSIONES GENERALES**

Se diseñó e implementó el proceso de extracción, transformación y carga de los datos, que permitió poblar el mercado de datos para el área de Registro Legal.

Se relacionaron las aplicaciones de usuario final con el mercado de datos a través de la configuración del repositorio de metadatos.

Se desarrolló la herramienta de creación de reportes dando la posibilidad a los usuarios de la construcción de nuevas solicitudes o reportes, así como modificar los ya existentes.

Se elaboraron los reportes predefinidos que dan respuesta a los requisitos de información del área de Registro Legal.

Se creó el tablero de mando que permite visualizar los indicadores del área de Registro Legal.

Se desarrolló el mercado de datos para facilitar el proceso de toma de decisiones en el área de Registro legal, permitiendo al usuario de crear nuevos reportes, consultarlos y visualizarlos.

### **RECOMENDACIONES**

Implementar un almacén de datos a partir de los mercados de datos creados de cada área del Sistema Penitenciario Cubano.

Valorar las nuevas necesidades de información que puedan presentar los usuarios.

Desplegar el mercado de datos para el Área de registro legal en los establecimientos penitenciarios del país.

Realizar los mercados de datos a las restantes áreas de la Dirección de Establecimientos Penitenciarios.

### GLOSARIO DE TÉRMINOS

**Almacén de datos:** Es una colección de datos para el soporte del proceso de toma de decisiones, orientado a temas, no volátil, integrado y de tiempo variante.

**Cubo:** Representación o visualización de una tabla de hecho y sus dimensiones correspondientes.

**DEP:** Dirección de Establecimientos Penitenciario.

**Delito:** Acción u omisión socialmente peligrosa prohibida por la ley bajo conminación de una sanción penal.

**Dimensiones:** Son elementos para ubicar datos que participan en el análisis, son datos que deben de tenerse en cuenta para poder tomar decisiones basadas en ellos.

**ETL:** Proceso de Extracción, Limpieza, Transformación y Carga de los datos.

**Inteligencia de negocio:** Es una arquitectura y colección de herramientas que buscan mejorar a las organizaciones, proporcionando vistas de aspectos de negocio a todos los empleados (estratégico, táctico, operacional) para que tomen mejores y más relevantes decisiones en menos tiempo y con la mayor información posible.

**Mercado de datos:** Un mercado de datos es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica.

**MININT:** Ministerio del Interior.

**Registro legal:** Área del Sistema Penitenciario Cubano que se encarga de la situación legal de los internos y el control del cumplimiento de la sanción del interno durante su tránsito por el sistema penitenciario.

**SGBD:** Sistema Gestor de Bases de Datos, tipo de software muy específico, dedicado a servir de interfaz entre la base de datos, el usuario y las aplicaciones que la utilizan.

**SIDEP:** Sistema Informativo de la Dirección de Establecimientos Penitenciarios.

**Sistema penitenciario cubano:** Garantiza el funcionamiento de los centros destinados al internamiento y el tratamiento a los internos.

**Tabla de hechos:** Es una representación de un proceso de negocio. Esta tabla se encuentra rodeada de tablas de dimensiones que determinan cómo están distribuidos y relacionados los datos lógicamente en el ambiente del negocio.

### REFERENCIAS BIBLIOGRÁFICAS

1. **Del Risco Batista, Yanet.** *Proyecto Técnico.* 2009.
2. **Martínez Bravet, René.** *"Repositorio de Datos de la Sala Situacional del SIGEP, desde el punto de vista del Control Pena".* Ciudad de la Habana, Cuba : Universidad de las Ciencias Informáticas, 2010.
3. Portal Iberoamericano de la Justicia Electrónica. [En línea] [Citado el: 30 de Marzo de 2012.] <http://www.piaje.org/ES/Services/Paginas/statesservices.aspx?Country=Ecuador>.
4. **Guerrero, José Iván.** *Evolución histórica del Sistema Penitenciario.* Panamá : s.n.
5. **Bernabeu, Ricardo Dario.** *DATA WAREHOUSING: Investigación y Sistematización de Conceptos - HEFESTO Metodología propia para la Construcción de un Data Warehouse. V2.0.* 2010.
6. **Zimányi, Esteban y Malinowski, Elzbieta.** *Advanced Data Warehouse Design.* 2008.
7. **Ponniah, Paulraj.** *Data Warehousing Fundamentals:A Comprehensive Guide for IT Professionals.* New York, Chichester, Weinheim, Brisbane, Singapore y Toronto : s.n., 2001.
8. Sinneuxs. Bases de datos OLTP y OLAP. [En línea] [Citado el: 30 de noviembre de 2011.] [http://www.sinnexus.com/business\\_intelligence/olap\\_vs\\_oltp.aspx](http://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx).
9. **Oracle® Database.** *2 Day DBA 11g Release.* 2008.
10. Sitio Oficial de Oracle. [En línea] [Citado el: 10 de marzo de 2012.] <http://www.oracle.com/us/products/database/product-editions-066501.html>.
11. **Oracle® Warehouse Builder.** *User's Guide 11g Release.* s.l. : 2007.
12. Oracle Business Intelligence Enterprise Edition 11g. [En línea] [Citado el: 15 de enero de 2012.] <http://www.oracle.com/us/solutions/ent-performance-bi/enterprise-edition-066546.html>.

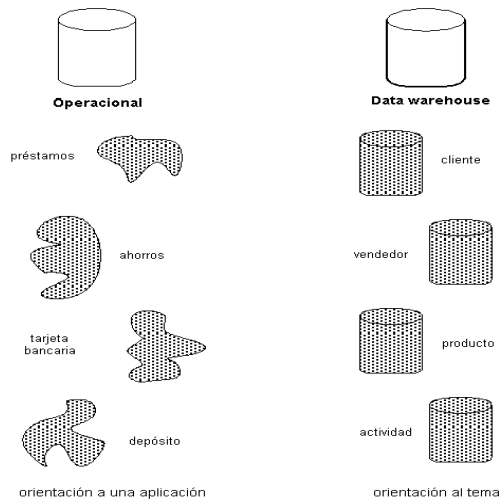


### BIBLIOGRAFÍA

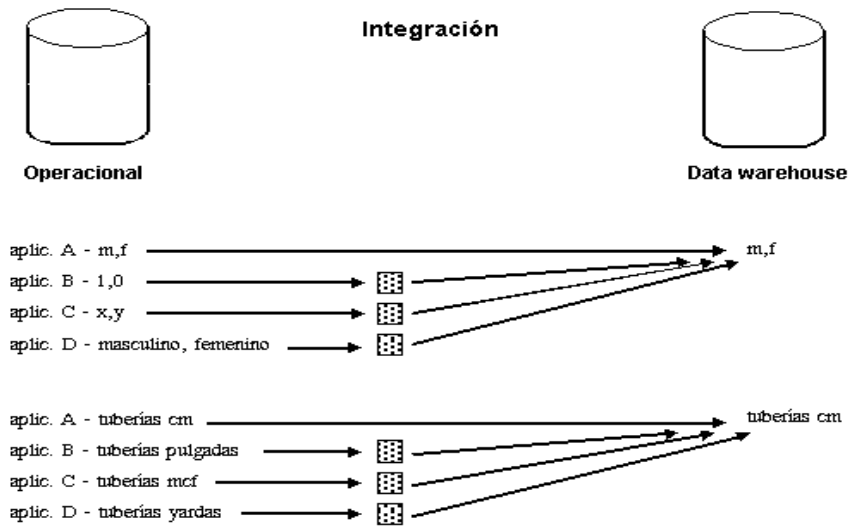
1. **Del Risco Batista, Yanet.** *Proyecto Técnico.* 2009.
2. **Martínez Bravet, René.** *"Repositorio de Datos de la Sala Situacional del SIGEP, desde el punto de vista del Control Pena".* Ciudad de la Habana, Cuba : Universidad de las Ciencias Informáticas, 2010.
3. Portal Iberoamericano de la Justicia Electrónica. [En línea] [Citado el: 30 de Marzo de 2012.] <http://www.piaje.org/ES/Services/Paginas/statesservices.aspx?Country=Ecuador>.
4. **Guerrero, José Iván.** *Evolución histórica del Sistema Penitenciario.* Panamá : s.n.
5. **Bernabeu, Ricardo Dario.** *DATA WAREHOUSING: Investigación y Sistematización de Conceptos - HEFESTO Metodología propia para la Construcción de un Data Warehouse. V2.0.* 2010.
6. **Zimányi, Esteban y Malinowski, Elzbieta.** *Advanced Data Warehouse Design.* 2008.
7. **Ponniah, Paulraj.** *Data Warehousing Fundamentals:A Comprehensive Guide for IT Professionals.* New York, Chichester, Weinheim, Brisbane, Singapore y Toronto : s.n., 2001.
8. Sinneuxs. Bases de datos OLTP y OLAP. [En línea] [Citado el: 30 de noviembre de 2011.] [http://www.sinnexus.com/business\\_intelligence/olap\\_vs\\_oltp.aspx](http://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx).
9. **Oracle® Database.** *2 Day DBA 11g Release.* 2008.
10. Sitio Oficial de Oracle. [En línea] [Citado el: 10 de marzo de 2012.] <http://www.oracle.com/us/products/database/product-editions-066501.html>.
11. **Oracle® Warehouse Builder.** *User's Guide 11g Release.* s.l. : 2007.
12. Oracle Business Intelligence Enterprise Edition 11g. [En línea] [Citado el: 15 de enero de 2012.] <http://www.oracle.com/us/solutions/ent-performance-bi/enterprise-edition-066546.html>.
13. **Kimball, Ralph.** *The Data Warehouse Lifecycle Toolkit.*
14. **Kimball, Ralph y Ross, Margy.** *The Kimball Group Reader;Relentlessly Practical Tools for Data Warehousing and Business Intelligence.* 2010.

15. **Imhoff, Claudia, Glemmo, Nicholas y Geiger, Jonathan G.** *Warehouse Design Relational and Dimensional Techniques*. 2003.
16. **Kimball, Ralph y Caserta, Joe.** *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. 2004.
17. **Oracle.** *Oracle® Business Intelligence Server Administration Guide*. 2006.
18. **Oracle.** *Oracle® Business Intelligence Presentation Services Administration Guide*. 2006.

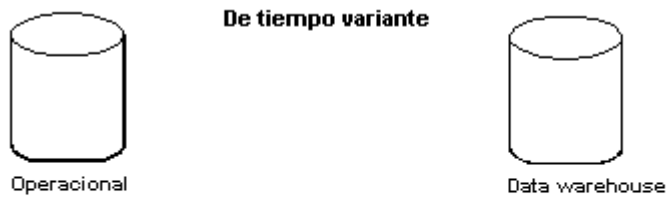
**ANEXOS**



**Figura 1.1:** Almacén de datos orientado a temas.



**Figura 1.2:** Almacén de datos integrado.



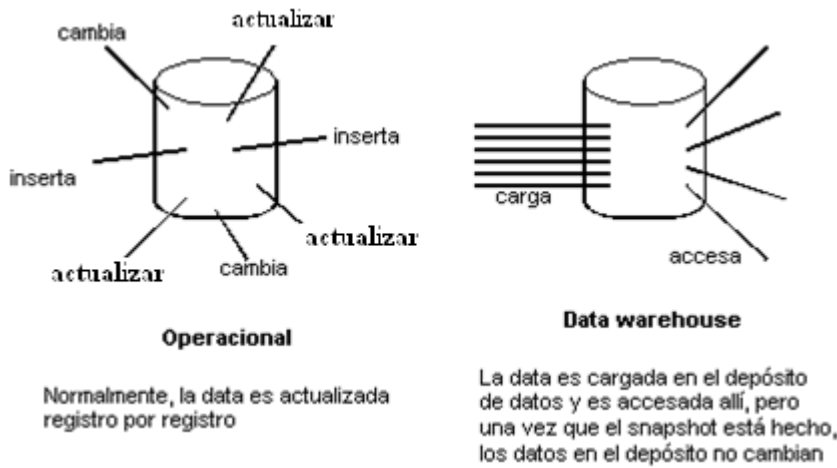
Valor actual de los datos:

- Horizonte de tiempo: 60-90 días
- La clave puede, como no, tener un elemento de tiempo
- Los datos pueden ser actualizados

Datos instantáneos:

- Horizonte de tiempo: 5-10 años
- La clave contiene un elemento de tiempo
- Una vez que el snapshot se realice, el registro no puede ser actualizado

**Figura 1.3:** Almacén de datos variante en el tiempo.



**Figura 1.4:** Almacén de datos no volátil.

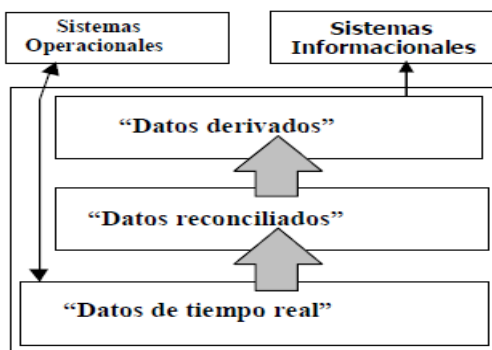


Figura 1.6: Oracle Warehouse Builder (OWB).

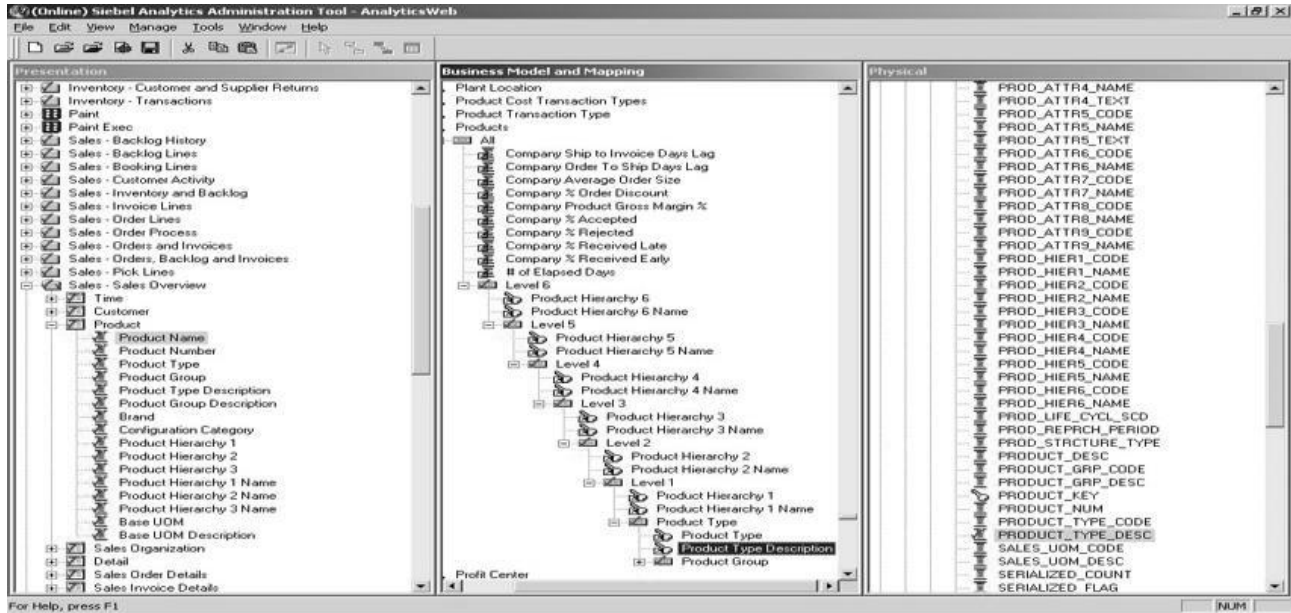


Figura 1.7: Administration Tool.



Figura 1.8: Oracle BI Answers.

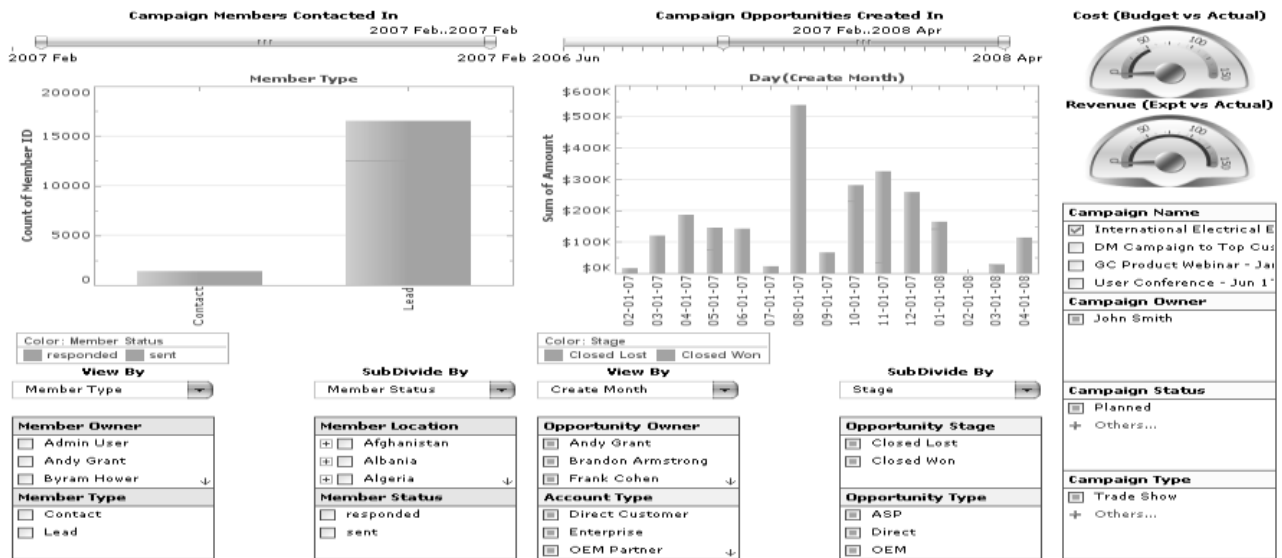


Figura 1.9: Oracle Business Intelligence Interactive Dashboards.

**TEMA: RESPONSABILIDADES**

1. Describe tu organización y sus relaciones internas con el resto de la compañía
2. ¿Cuáles son tus responsabilidades primarias?

**TEMA: ASUNTOS Y OBJETIVOS DE NEGOCIO**

1. ¿Cuáles son los objetivos de tu organización? ¿Qué usted está tratando de lograr?
2. ¿Cuáles son tus metas prioritarias sobresalientes de negocio?
3. ¿Cuál es tu métrica de éxito? ¿Cómo sabe usted que está bien?
4. ¿Cuál es el impacto en la organización?
5. ¿Cómo identifica usted problemas /excepciones o sabe que va encaminado al problema?

**TEMA: ANALISIS DE LOS REQUERIMIENTOS**

1. ¿Qué tipo de análisis de rutina usted hace actualmente?
2. ¿Cómo obtiene usted actualmente los datos?
3. ¿Qué datos son usados?
4. ¿Qué hace con la información una vez que la obtiene?
5. ¿Qué análisis le gustaría realizar?
6. ¿Hay mejoras potenciales para su método /proceso actual?
7. ¿Quién pide análisis ad hoc?
8. ¿Qué hacen con el análisis?
9. ¿Cuáles reportes usa usted actualmente?
10. ¿Qué dato en el reporte es importante?
11. ¿Cómo utiliza usted la información?
12. ¿Qué capacidades analíticas le gustaría a usted tener?
13. ¿Cuánta información histórica es requerida?
14. ¿Cuál es el impacto financiero?

**Figura 2.1:** Ejemplo de preguntas de la entrevista para los especialistas del Área de Registro Legal.

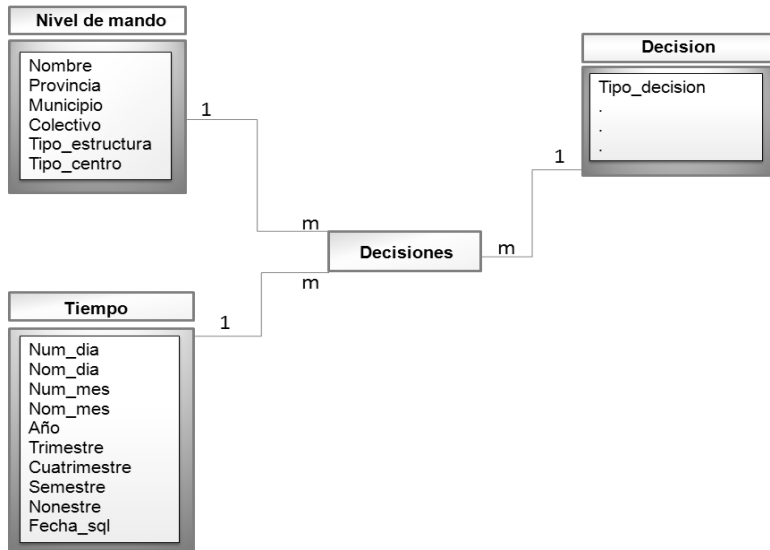


Figura A. 1: Modelo dimensional detallado del proceso Decisión.