

Universidad de las Ciencias Informáticas



Facultad 2

Título: Sistema de Apoyo a la Toma de Decisiones del Sistema Penitenciario Cubano.

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

Autor(es): Krysia Llull Céspedes  
Yaimara Méndez Gutiérrez

Tutor: Ing. Yanet del Risco Batista

Ciudad de La Habana, 2011  
“Año 53 de la Revolución”

## 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 2011.

Krysia Llull Céspedes

\_\_\_\_\_

Firma del Autor

Yaimara Méndez Gutiérrez

\_\_\_\_\_

Firma del Autor

Ing. Yanet del Risco Batista

\_\_\_\_\_

Firma del Tutor

## AGRADECIMIENTOS:

De Krysia:

Le agradezco a la Revolución. A Fidel y Raúl por todo su empeño en construir un futuro mejor para los cubanos.

A mis padres por todo su sacrificio y apoyo, a mi hermano por contarme chistes pesaos y contar los días que me faltan para el pase.

A mi abuela por ser mi segunda madre y un ejemplo a seguir.

A mi novio Emilio, mi choc lindo, por todo su cariño y demostrarme que siempre puedo contar con él. A su mama Mirtha por ser la razón de que él sea tal y como es.

A mis primos y tíos por todas las alegrías compartidas.

A nuestra tutora por todo su apoyo y paciencia.

A Leandro por toda su ayuda en la Base de Datos.

De manera general a todos aquellos que de una manera u otra me han ayudado a lo largo de estos 5 años, los profes y a los amigos de la antigua facultad 4 del grupo 2 y a las chicas de mi apartamento actual apto (13 209).

Gracias.

### DEDICATORIA

Dedico este trabajo a las personas más importantes en mi vida

A mi mamá, por su optimismo, por ser mi guía y por su inmenso amor.

A mi papá, por su confianza y por darme fuerzas para seguir adelante.

A mi hermano, por su cariño y por ser un ejemplo a seguir.

Yaimara

Le dedico este trabajo a mis padres, a mi hermano,  
mi abuela y a todos mis primos(as) y tíos(as).

Krysia

### RESUMEN

En el año 1989 comienza en Cuba el sistema de informatización de los centros penitenciarios, con la automatización de los datos principales del recluso y ciertos aspectos de control penal, pero no es hasta algunos años más tarde que a raíz del cumplimiento de la orden 43/99 del Viceministro Primero (VMP) se crea un Sistema Automatizado para el Control del Recluso (SACORE) el cual comenzó a dar respuesta en gran medida a muchos de los problemas existentes en el momento. A partir del Plan 20 x 50 y en colaboración con la Universidad de las Ciencias Informáticas, se decide por parte de la Jefatura del MININT crear el proyecto Humanización Penitenciaria Cuba. Entre los aportes de este proyecto se encuentra dotar a la jefatura de la Dirección de Establecimientos Penitenciarios (DEP) de una herramienta que permita la toma de decisiones.

El presente trabajo tiene como objetivo apoyar el proceso de toma de decisiones de la especialidad Registro Legal del Sistema Penitenciario Cubano, mediante el desarrollo de un Data Mart, basándose en información integrada y global de su negocio, facilitará la aplicación de técnicas estadísticas del análisis, proporcionando así la capacidad de aprender de los datos del pasado y de predecir situaciones futuras en diversos escenarios de la organización. El Data Mart está basado en la metodología Hefesto, su modelamiento, análisis y posterior implementación son consecuentes con esta. El proceso de construcción consta de varias fases siendo las más importantes el proceso de extracción, transformación y carga junto al procesamiento analítico en línea (Configuraciones previas). Se utilizó para su desarrollo el enfoque de Ralf Kimball (Enfoque de abajo hacia arriba), la arquitectura de tres capas, para su implementación como tal, la Suite de Inteligencia de Negocio Oracle Business Intelligence y como gestor de base de datos Oracle 11g.

Palabras Claves: Data Mart, Hefesto.

## ÍNDICE

INTRODUCCIÓN .....	8
CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA .....	13
1.1 Marco conceptual .....	13
1.1.1 Sistema penitenciario .....	13
1.1.2 Almacén de Datos .....	13
1.1.3 Data Mart .....	14
1.1.3 Sistemas valorados: .....	14
1.3.1 Sistema Penitenciario del Gobierno de Panamá.....	14
1.3.2 Sistema Automatizado para el Control del Recluso (SACORE) .....	15
1.3.3 Sistema de Gestión Penitenciaria Ecuador.....	15
1.3.4 Sistema de gestión penitenciaria (SIGEP):.....	16
1.4 Objetivos y Características de un Almacén de Datos. ....	17
1.5 Modelado multidimensional .....	19
1.6 Esquemas dimensionales:.....	20
1.7 Organización física:.....	20
1.8 Métodos más usados en la construcción de Almacén de Datos. ....	21
1.9 Arquitectura conceptual de los datos.....	22
1.10 Diferencias entre OLTP y OLAP .....	23
1.10.1 OLTP.....	23
1.10.2 OLAP.....	23
1.10.3 Comparación entre OLTP Y OLAP .....	24
1.11 Modelado de datos.....	25
1.12 Metodologías evaluadas .....	26
1.12.1 Metodología DWEP .....	26
1.12.2 Metodología Hefesto .....	27
1.13 Herramientas de creación de un Almacén de Datos.....	27
1.13.1 Arquitectura suite de Oracle .....	28
1.13.2 Herramientas de la suite Oracle .....	28
Conclusiones parciales. ....	30

CAPÍTULO 2: ANÁLISIS Y DISEÑO DEL DATA MART .....	31
2.2 Análisis de requerimientos .....	31
2.2.1 Identificar perspectivas .....	35
2.2.2 Indicadores identificados .....	36
2.3 Análisis de los OLTP .....	40
2.3.1 Conformar indicadores .....	41
2.3.2 Establecer correspondencias .....	44
2.3.4 Modelo conceptual ampliado .....	48
Conclusiones .....	50
CAPÍTULO 3: IMPLEMENTACIÓN DEL DATA MART .....	51
3.2 Modelo lógico del Data Mart .....	51
3.2.1. Tipo de modelo lógico del Data Mart .....	51
3.3 Tablas de dimensiones .....	51
3.4 Tablas de hechos .....	52
3.5 Diseño de los procesos ETL.....	53
3.6 Configuración del Data Mart.....	67
Conclusiones .....	71
CONCLUSIONES GENERALES.....	72
RECOMENDACIONES .....	73
REFERENCIAS BIBLIOGRÁFICAS .....	74
BIBLIOGRAFÍA.....	75

## INTRODUCCIÓN

El progreso a nivel mundial se encuentra a una velocidad considerable en cuanto al desarrollo de las tecnologías de la informática y comunicaciones, en la última década se han logrado muchos más avances en este aspecto que en los pasados 50 años, más acentuadamente con la aplicación de la inteligencia de negocio en la rama de la informática. Esto ha dado cabida al surgimiento de una nueva era en donde la información y el conocimiento juegan un papel primordial para toda organización. Los volúmenes de información generados diariamente crecen de forma exponencial por lo que su correcta gestión es fundamental para la toma de decisiones, que es un recurso vital para toda organización. Este tipo de información sólo se convierte en conocimiento cuando los individuos la aplican para la resolución de un problema determinado.

La capacidad para tomar decisiones precisas y de forma rápida, es un factor clave para que una organización pueda desarrollarse objetivamente. Puede parecer que el análisis de datos es un proceso sencillo y fácil de conseguir mediante los sistemas de información tradicionales (como la mayoría de los programas de gestión, las aplicaciones a la medida, e incluso los sistemas de planificación de recursos empresariales (ERP) más sofisticados), estos sistemas suelen disponer de una serie de informes predefinidos en los que presentan la información de manera estática, no permiten profundizar en los datos, navegar entre ellos, manejarlos desde distintas perspectivas y suelen poseer una estructura muy inflexible para este fin. Aunque su diseño se adapta con mayor o menor medida para manejar los datos de la organización, no permiten obtener toda la información disponible de las mismas, y mucho menos extrapolar el conocimiento almacenado en el día a día de las bases de datos. En el año 1958 surge un concepto estrechamente relacionado con lo que se ha venido planteando hasta el momento: la inteligencia empresarial, inteligencia de negocios o BI (del inglés Business Intelligence), no es hasta los últimos 15 años cuando la Informática lo empezó a utilizar en sus soluciones cuando realmente alcanzó la importancia y magnitud que este tiene. La BI no es más que un conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento mediante el análisis de datos existentes en una organización para facilitar la toma de decisiones.

La gestión y aprovechamiento de información no puede realizarse de manera efectiva sin un Almacén de Datos debido a que se han convertido en un componente esencial de los sistemas modernos de toma de



decisiones en la mayoría de las organizaciones. Estos sistemas ofrecen un acceso eficiente a los datos integrados e históricos de las diferentes fuentes de información y constituyen un apoyo indispensable para los directivos en su planificación y toma de decisiones.

En el año 1989 comienza en Cuba el sistema de informatización de los centros penitenciarios, con la automatización de los datos principales del recluso y ciertos aspectos de Control Penal, pero no es hasta algunos años más tarde que a raíz del cumplimiento de la orden 43/99 del Viceministro Primero (VMP) se crea un Sistema Automatizado para el Control del Recluso (SACORE) el cual comenzó a dar respuesta en gran medida a muchos de los problemas descritos anteriormente. Dicho sistema culminó su desarrollo a finales del 2002, poniéndose en marcha a principios del 2003 y cuenta en la actualidad con tres módulos principales: Control Penal, Reeducación Penal y Orden Interior, los cuales aumentan las especificaciones de la automatización de los datos principales del recluso y los aspectos de control penal existentes, además de adicionar un mayor número de facilidades. En los 8 años de explotación del sistema, a pesar de sus facilidades y ayuda brindada a dichos centros aún cuenta con requisitos incompletos o pendientes de dichos módulos, por no contar la institución con el tiempo para su análisis, diseño y programación, y por carecer de las condiciones tecnológicas para su implantación posterior. (Del Risco Batista, 2009)

A partir del Plan 20 x 50, para la modernización tecnológica del MININT, en colaboración con la Universidad de las Ciencias Informáticas, se crea el proyecto Humanización Penitenciaria Cuba. Este 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 el funcionamiento óptimo de los procesos en los centros penitenciarios, aprovechando el desarrollo de las tecnologías de la informática y las comunicaciones que se está llevando a cabo en el país. (Del Risco Batista, 2009)

En el sistema penitenciario cubano se cuenta con un gran cúmulo de información sobre los principales indicadores que definen la situación operativa del mismo. El departamento de Dirección es el encargado de realizar el análisis de la información y generar los reportes sobre el comportamiento de los indicadores en un determinado período. Actualmente el departamento está compuesto por 5 especialistas que son los encargados de generar los reportes y brindar información a la jefatura para la toma de decisiones.

Este proceso se hace muy engorroso, porque la cantidad de especialistas actual no puede generar toda la información que la jefatura necesita para tomar decisiones oportunas y garantizar la situación operativa.

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

Teniendo en cuenta lo anterior se define como **objeto de estudio** de la investigación los procesos de toma de decisiones en sistemas penitenciarios y **como objetivo general** desarrollar un Data Mart para apoyar el proceso de toma de decisiones de la especialidad de Registro Legal del Sistema Penitenciario Cubano. El **campo de acción** está centrado en los procesos de toma de decisiones en el Sistema Penitenciario Cubano.

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

1. Evaluar y seleccionar las metodologías para la creación de un Almacén de Datos.
2. Establecer los requisitos del sistema.
3. Definir la arquitectura del sistema.
4. Diseñar el Data Mart.
5. Implementar el Data Mart.

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

### **Objetivo 1:**

- Investigación de las metodologías existentes para el diseño y construcción de un Almacén de Datos.
- Selección de la metodología de construcción de un Almacén de Datos.

### **Objetivo 2:**

- Obtención y análisis de requisitos del sistema.
- Validación de los requisitos del sistema con el cliente

### **Objetivo 3:**

- Determinar los requisitos no funcionales del sistema.
- Análisis y definición de la arquitectura del sistema.

### **Objetivo 4:**

- Diseño conceptual del Data Mart.
- Diseño lógico específico del Data Mart.
- Diseño físico del Data Mart.

### **Objetivo 5.**

- Implementación del Data Mart.
- Diseño e implementación de los reportes.

### **Organización del documento**

El presente documento se estructura en: resumen, introducción, tres capítulos de contenidos, conclusiones, recomendaciones, referencias bibliográficas y anexos.

### **CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA**

En este capítulo se realiza un estudio de los sistemas de soporte a la decisión existentes en el mundo, en el país y específicamente en nuestra universidad. Definiendo además la metodología y las herramientas a utilizar. Se recogen conceptos, objetivos y características de los almacenes de datos, las tecnologías y software empleados en su diseño, así como un análisis de las diferencias y funcionalidades de los principales gestores de bases de datos que soportan almacenes de datos.

### **CAPÍTULO 2: ANÁLISIS Y DISEÑO DEL DATA MART**

En este capítulo se describe el análisis realizado a la especialidad de Registro Legal de la DEP, el mismo es consecuente con la metodología Hefesto, seleccionada anteriormente por ser la más adecuada para dar cumplimiento a los objetivos propuestos y así diseñar el Data Mart para la especialidad.

### **CAPÍTULO 3: IMPLEMENTACIÓN DEL DATA MART**

En este capítulo se construyen las estructuras físicas del Data Mart, se define el tipo de estructura empleada para el almacenamiento y se crean los correspondientes esquemas lógicos. Se llevan a cabo los procesos ETL, se definen los procesos de carga, se configura el depósito de datos y se implementan los informes solicitados por los clientes, haciendo uso de las herramientas de consulta.

## **CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA**

### **Introducción**

En el presente capítulo se realiza un estudio de las tendencias actuales en el uso de las aplicaciones informáticas para la gestión de la información, para conocer cómo se han venido desarrollando las mismas dentro y fuera de nuestro país. Además se realiza un análisis de los conceptos, objetivos y características de los almacenes de datos. Se evidencian también las diferencias y funcionalidades de los principales gestores de bases de datos, para la exploración de datos y la extracción de conocimiento, además de fundamentar la selección de las tecnologías que se utilizarán para el desarrollo de este trabajo.

### **1.1 Marco conceptual.**

#### **1.1.1 Sistema penitenciario**

Es el encargado de 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. (Del Risco Batista, 2009)

#### **1.1.2 Almacén de Datos**

Este concepto significa el almacenamiento de información homogénea y fiable, en una estructura basada en la consulta y el tratamiento jerarquizado de la misma, y en un entorno diferenciado de los sistemas operacionales. Las dos personalidades más destacadas en este tema sin duda son Bill Inmon, conocido como el padre de los almacenes de datos, y Ralph Kimball, considerado el principal promotor del enfoque dimensional; ambos han desarrollado sus propios enfoques, modelos y arquitecturas. Se puede definir como una colección de datos proveniente de diferentes bases de datos operacionales o relacionales en la cual se encuentra integrada la información de la organización, estando disponible para que los usuarios finales puedan consultarla fácilmente, generar reportes y dar soporte al proceso de toma de decisiones. La definición más formal es la brindada por Inmon en 1992 donde dice: "... es una colección de datos orientado a temas, integrado, no volátil y de tiempo variante, organizados para dar soporte al proceso de

ayuda a la toma de decisiones”, puesto que permite consolidar y administrar datos de variadas fuentes con el propósito de responder preguntas de negocios y tomar decisiones. (Ing. Bernabeu, 2007)

### **1.1.3 Data Mart**

Un Data Mart es la implementación de un Almacén de Datos con alcance restringido a un área funcional, problema en particular, departamento, tema o grupo de necesidades. Conjunto de hechos y datos organizados para soporte decisonal basados en la necesidad de un área o departamento específico. Estos datos son orientados a satisfacer las necesidades particulares de un departamento dado teniendo sólo sentido para el personal de ese departamento y sus datos no tienen porque tener las mismas fuentes que los de otro Data Mart. (Ing. Bernabeu, 2010)

### **1.1.3 Sistemas valorados:**

En nuestros días el desarrollo informático ha propiciado que las organizaciones basen su desempeño en el uso de aplicaciones informáticas para la gestión de la información, toda la información que es generada por estas organizaciones se guardan en bases de datos relacionales. Los datos que son almacenados se acumulan formando así grandes volúmenes de datos que son humanamente difíciles de manejar en un tiempo razonable.

A continuación se mencionan algunas aplicaciones informáticas existentes en los sistemas penitenciarios. Las mismas fueron valoradas como posible solución para realizar este trabajo, teniendo en cuenta las funcionalidades que brindan.

### **1.3.1 Sistema Penitenciario del Gobierno de Panamá**

A principios de la década del 90 existía una situación crítica en el gobierno de Panamá con respecto al control y organización de su sistema penitenciario, entre las medidas que fueron tomadas para encaminar el sistema fue la de dotar a la Dirección Nacional de Corrección, de los recursos informáticos imprescindibles para abordar con éxito una reforma integral de la administración penitenciaria y una mejora sustancial de los instrumentos de gestión y control de los centros penitenciarios. El sistema fue creado a principios del año 1997, como resultado de un proyecto financiado por las Naciones Unidas y el Gobierno Español en coordinación con la Dirección General de Sistemas Penitenciarios de Panamá (DGSP). Es un sistema de software que permite básicamente 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 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. (Sistema Penitenciario)

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

Fue puesto en práctica a partir del 2003 garantizando respuestas inmediatas a las solicitudes de información de los diferentes órganos e instituciones, es un sistema que recoge prácticamente la totalidad de la información de los reclusos en todas las especialidades. Cuenta con más de 200 reportes impresos y permite el traslado automático de todos los datos del recluso al nivel nacional. A pesar de que este sistema resolvió una gran problemática en el momento de su creación, en la actualidad no resuelve todas las necesidades de información necesarias al departamento de dirección de la DEP. Cuenta con 8 años de explotación, generando un gran cúmulo de información, la misma resulta muy difícil de procesar, los especialistas deben realizar, sin ser expertos consultas a la base de datos y exportar la información a Excel, de esta manera es que pueden generar los gráficos que necesitan para realizar los análisis estadísticos. Estas dificultades provocan que el análisis estadístico se convierta en un proceso engorroso y en muchas ocasiones lento, propiciando la demora en la entrega a tiempo de los informes, afectando de esta forma la toma de decisiones. (Del Risco Batista, 2009)

### **1.3.3 Sistema de Gestión Penitenciaria Ecuador**

El Sistema de Gestión Penitenciaria e-Sigpen facilita el control y manejo de las actividades que realizan los centros de rehabilitación social en cada una de sus áreas, logrando de esta manera tener un control adecuado y oportuno de la información de las personas privadas de libertad.

El Ministerio de Justicia, derechos humanos y cultos conjuntamente con la Dirección Nacional de Rehabilitación Social están en continuo monitoreo de las actividades realizadas en los Centros de Rehabilitación Social por medio del e-Sigpen, el cual permite obtener datos estadísticos reales de la situación en que se encuentran centros, logrando de esta manera tomar decisiones eficaces y oportunas para mejorar el ambiente carcelario en Ecuador. (Justicia)

### 1.3.4 Sistema de gestión penitenciaria (SIGEP):

El Sistema de Gestión Penitenciaria (SIGEP), desplegado en Venezuela y desarrollado en la Universidad de las Ciencias Informáticas tiene como objetivo general “desarrollar e implantar un sistema informático que soporte las decisiones estratégicas del Ministerio del Interior y Justicia y de la Dirección General de Custodia y Rehabilitación del Recluso”. Este es el sistema informático penitenciario estudiado y analizado a fondo en la presente investigación. De él observamos lo siguiente:

En análisis referente a la generación de información de apoyo para la toma de decisiones se observa el uso de herramientas de inteligencia de negocio (Pentaho), sobre un Almacén de Datos con esquema dimensional para la generación de estadísticas y reportes, alternativa esta que fue adoptada luego de que con la implantación del SIGEP (Sistema de Gestión Penitenciaria) en el sistema penitenciario venezolano, se detectara que los volúmenes de datos generados por el sistema crecieron exponencialmente, lo que provocó que este se viera afectando en cuanto a rendimiento, más específicamente cuando se requerían mostrar reportes complejos que involucran la búsqueda de información en varias tablas e incluso diferentes módulos a la vez. A esto se le sumó además la necesidad, por parte de los clientes, de otros reportes no definidos en un principio y que luego con el tiempo se fueron haciendo necesarios para el trabajo en las penitenciarías, aspecto que el sistema no proveía y que dificultaba a los directivos del sistema de justicia venezolano la toma de decisiones, pues cada vez que surgía la necesidad de un nuevo informe, solo el equipo de desarrollo sería capaz de implementarlo y no es hasta una próxima versión o con la aplicación de un parche que estaría disponible.

Los sistemas antes mencionados tienen como principal limitante que la definición de los procesos penitenciarios que los fundamentan y la infraestructura física sobre la cual se implantan son poco generales, es decir, la mayoría de estos sistemas se implementan en términos de la legislación vigente en los países para los cuales fueron definidos, lo que los convierte en sistemas difícilmente portables a otros sistema penitenciario. Aunque es necesario aclarar que a pesar de las limitaciones estos sistemas sirven de apoyo a los procesos de sus sistemas penitenciarios y funcionan como herramientas de trabajo a nivel local e institucional.

Tomando punto de partida las experiencias del SIGEP desplegado en Venezuela, queda justificada para el sistema de apoyo a la toma de decisiones de la Especialidad Registro Legal el uso de herramientas de



Inteligencia de Negocio para la generación de reportes variables haciendo uso de un Almacén de Datos con esquema dimensional, para asegurar los aspectos relacionados con el rendimiento.

## 1.4 Objetivos y Características de un Almacén de Datos.

- Los objetivos más importantes de un Almacén de Datos son:
- La organización cuenta con información accesible.
- La organización cuenta con información consistente y adaptable.
- Es la fundamentación de la toma de decisiones.

Entre sus principales **características** tenemos:

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

**Orientado al tema:** Existe una diferencia entre las formas que se diseña la estructura de datos de las bases datos operacionales y los almacenes de datos, pues en el ambiente operacional la estructura de datos se diseña en torno a las aplicaciones y funciones tales como préstamos, ahorro y tarjetas; sin embargo en el ambiente de los depósitos de datos se diseña en función de los principales temas de la organización, tales como clientes, vendedores, productos y actividades. Una de las diferencias que existen entre el diseño orientado a las aplicaciones y el orientado a los temas es tanto el nivel de detalles de los datos como el contenido de los mismo, es decir en el depósito de datos son excluidos los datos que no serán utilizados por el sistema de soporte a la toma de decisiones, mientras que los sistemas orientados a las aplicaciones contienen datos operacionales para satisfacer requerimientos inmediatos funcionales o de procesamiento los cuales pueden ser utilizados o no por los analistas de los sistemas de soporte a la toma de decisiones. Los almacenes de datos son organizados por temas porque esto facilita su acceso y

entendimiento por parte de los usuarios finales.

## Figura 1

**Integrado:** El aspecto más importante del ambiente de un Almacén de Datos es que la información encontrada en su interior esté siempre integrada. La integración de datos se muestra de muchas maneras: en convenciones de nombres consistentes, en la medida uniforme de variables, en la codificación de estructuras, en atributos físicos de los datos, fuentes múltiples y otros. El contraste de la integración encontrada en el Almacén de Datos con la carencia de integración del ambiente de aplicaciones, se muestran en la figura, con diferencias bien marcadas.

## Figura 2

**De tiempo variante:** Los datos son relativos a un periodo de tiempo y estos deben ser integrados con frecuencia y los mismos son almacenados como fotos. En un Almacén de Datos, a diferencia de una base de datos operacional, los datos son los que corresponden a un instante de tiempo, es decir son como unas series de fotografías del ambiente operacional que se realizan cada cierto periodo de tiempo y una vez que son almacenadas no pueden ser actualizadas. Estas fotografías sólo pueden ser consultadas con el objetivo de ser utilizadas en comparaciones, tendencias y previsiones. Esto no ocurre así en las base datos operacionales, pues en ella los datos son los que corresponden al momento en que se accede y van siendo actualizados según surjan las necesidades. **Figura 3**

El tiempo variante se muestra de varias maneras:

La más simple es que la información representa los datos sobre un horizonte largo de tiempo desde cinco a diez años. El horizonte de tiempo representado para el ambiente operacional es mucho más corto, debido al diseño de aplicaciones rígidas.

La segunda manera en la que se muestra el tiempo variante en el Almacén de Datos está en la estructura clave. Cada estructura clave en el Almacén de Datos contiene, implícita o explícitamente, un elemento de tiempo como día, semana, mes ó año. En ocasiones, el elemento de tiempo existirá implícitamente, como el caso en que un archivo completo se duplica al final del mes, o al cuarto.

La tercera manera en que aparece el tiempo variante es cuando la información del Almacén de Datos, una vez registrada correctamente, no puede ser actualizada. Dicha información es, para todos los propósitos prácticos, una serie larga de vistas instantáneas. Si las vistas de los datos se han tomado incorrectamente, entonces pueden ser cambiados. Asumiendo que los snapshots (vistas instantáneas) se han tomado adecuadamente, ellos no son alterados una vez hechos.

**No volátil:** Esta característica hace referencia a que los datos que son almacenados no sufren ninguna actualización, ellos solo son incrementados. El período cubierto para un Almacén de Datos va de 2 a 10 años, esto se encuentra fundamentado porque la información es útil sólo cuando es estable. Los datos operacionales cambian sobre una base momento a momento. Lo esencial para el análisis y la toma de decisiones, es que se requiere de una base de datos estable. En los ambientes operacionales se pueden realizar operaciones de insertar, eliminar y modificar. Sin embargo en los depósitos de datos sólo se pueden realizar operaciones de carga y acceso de los datos. Entre las ventajas que tiene esta característica es que al no realizarse ningún tipo de actualización sobre el Almacén de Datos, permite que el diseño físico del mismo tenga suficientes libertades como para optimizar el acceso a los datos, más específicamente en el tema de la normalización y desnormalización. **Figura 4**

## 1.5 Modelado multidimensional

El modelo dimensional es una propuesta de Ralph Kimball y es uno de los más utilizados para la construcción de los almacenes de datos. Esta técnica de diseño lógico tiene como objetivo representar los datos dentro de un marco de trabajo estándar e intuitivo, para permitir su acceso con un alto rendimiento. Cada modelo dimensional está compuesto por una llamada tabla de hechos, y un conjunto de tablas más pequeñas llamadas dimensiones. Los elementos de estas tablas se pueden definir como:

**Dimensiones:** definen como están los datos organizados lógicamente y proveen el medio para analizar el contexto del negocio. Representan los ejes del cubo, y los aspectos de interés, mediante los cuales el usuario podrá filtrar y manipular la información almacenada en la tabla de hechos. Ejemplos de dimensiones son: clientes, productos, provincias, tiempo. En una misma dimensión pueden existir varias jerarquías y estas están compuestas por dos o más niveles. Una dimensión es una colección de miembros, unidades o individuos del mismo tipo.

**Hechos:** Las tablas de hechos contienen los hechos, medidas o indicadores que serán utilizados por los analistas de negocio para apoyar el proceso de toma de decisiones. Los hechos son datos instantáneos en el tiempo y posee una clave primaria que está compuesta por las claves primarias de las tablas de dimensiones relacionadas a este. Cada hecho representa una parte del negocio, una transacción o un evento.

**Medidas:** Son atributos numéricos de un hecho que representan el comportamiento del negocio relativo a una dimensión.

## 1.6 Esquemas dimensionales:

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

**Esquema en estrella (Star Scheme):** consta de una tabla de hechos central y de varias tablas de dimensiones relacionadas a esta a través de sus respectivas claves.

### Figura 5:

El modelo debe estar totalmente desnormalizado, es soportado por casi todas las herramientas de consulta y análisis. Posee los mejores tiempos de respuesta, su diseño es fácilmente modificable sin embargo es el menos robusto para la carga y es el más lento de construir.

**Esquema copo de nieve (Snowflake Scheme):** representa una extensión del modelo en estrella con la diferencia de que las dimensiones se organizan en jerarquías de dimensiones. **Figura 6**

Este modelo es más cercano a un modelo de entidad relación, que al modelo en estrella, debido a que sus tablas de dimensiones están normalizadas. Hace una mejor utilización del espacio, es muy útil en tablas de dimensiones de muchas tuplas, al existir muchas uniones el desempeño puede verse reducido.

**Esquema constelación o copo de estrellas (Starflake Scheme):** está compuesto por una serie de esquemas en estrella está formado por una tabla de hechos principal y por una o más tablas de hechos auxiliares las cuales pueden ser resúmenes de la principal. **Figura 7:**

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

## 1.7 Organización física:

**ROLAP (Relational On Line Analytic Processing):** 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.

**MOLAP (Multidimensional On Line Analytic Processing):** almacenar físicamente los datos en estructuras multidimensionales de manera que la representación externa y la interna coinciden. La tecnología está optimizada para consultas y análisis. No es adecuada para muchas dimensiones ni muchos registros.

**HOLAP (Hybrid On Line Analytic Processing):** constituye un sistema híbrido entre MOLAP y ROLAP pues combina estas dos implementaciones para almacenar algunos datos en un motor relacional y otros en una base de datos multidimensional.

## 1.8 Métodos más usados en la construcción de Almacén de Datos.

En la actualidad los modelos propuestos por William H. Inmon y Ralph Kimball para llevar a cabo el diseño de un Almacén de Datos son los más aplicados, aunque sus criterios difieren en cuanto al modelo de datos y a las arquitecturas.

Ralph Kimball resalta que los Data Mart están basados en los datos de la fuente y no en la visión departamental, el Data Mart es sólo una parte de un producto orientado a una organización. Haciendo

referencia a que la idea de construcción de un Almacén de Datos centralizado, basándose en la visión de controlar toda la información antes de construir el Almacén de Datos como lo propone Inmon, no es realista, siendo más real construirlo en un ambiente descentralizado e incremental, porque las organizaciones están en constante cambio, adquiriendo nuevas fuentes de datos y necesitando nuevas perspectivas. Kimball manifiesta que el proceso de construcción de un Almacén de Datos parte de los sistemas operacionales existentes, creando los diferentes Data Mart basados en la información de dichas fuentes, para luego de tenerlos desarrollados y funcionales se comience con la construcción del Almacén de Datos basado en la información que éstos contienen. (Kimball)

Este último es el método más usado, gracias a las diferentes ventajas que proporciona, permitiendo a las organizaciones acometer los proyectos de manera separada y de esta forma reducir los efectos negativos que tendría fracasar en un intento por construir un Almacén de Datos.

### 1.9 Arquitectura conceptual de los datos

Se conoce como arquitectura la estructura que reúne todos los componentes de un Almacén de Datos, es la forma de representar la organización total de datos, comunicación, procesamiento y presentación. Está compuesta más allá de reglas, procedimientos y funciones que permiten al Almacén de Datos trabajar y cumplir los requisitos de la organización. Define las normas, medidas, diseño general, y técnicas de apoyo.

En un Almacén de Datos suelen representarse varias capas a través de las cuales circulan los datos, de modo que estos se obtienen a partir de los de la capa previa, nombrándose de acuerdo al número de capas que abarcan.

**Arquitectura de datos de Una Capa:** la información se guarda sólo una vez en el Almacén de Datos, recogándose únicamente los datos de tiempo real, sobre los cuales actúan sistemas informacionales y operacionales. Como estos sistemas actúan sobre el mismo conjunto de datos, puede suceder que al mismo de tiempo ambos intenten acceder a ellos, provocando que los datos no estén disponibles para los fines operacionales, al estar siendo consultados, mientras esto sucede no es posible realizar actualizaciones. **Figura 8.**

**Arquitectura de datos de Dos Capas:** esta arquitectura cuenta con una capa inferior, donde se contemplan los datos de tiempo real utilizados por las aplicaciones operacionales en modo lectura/escritura, además una superior, para el almacenamiento de los datos derivados utilizados por las aplicaciones informacionales, aumentando de esta forma los requerimientos de almacenamiento debido a la duplicación de información pero garantizando el acceso de los sistemas operacionales en cualquier instante de tiempo. **Figura 9.**

**Arquitectura de datos de Tres Capas:** es la transformación llevada a cabo en los datos de tiempo real a datos derivados requiere de una capa intermedia para solucionar los problemas de inconsistencias, realizando el procesamiento de los distintos conjuntos de datos de tiempo real adecuadamente. **Figura 10:**

Para realizar el diseño de un Almacén de Datos se puede seleccionar cualquiera de las arquitecturas anteriormente mencionadas de acuerdo a las características y necesidades de la organización.

## 1.10 Diferencias entre OLTP y OLAP

### 1.10.1 OLTP

Los sistemas OLTP (On Line Transaction Processing) son bases de datos orientadas al procesamiento de transacciones. Una transacción es un proceso atómico que puede involucrar operaciones de inserción, modificación y borrado de datos. El proceso transaccional es típico de las bases de datos operacionales. (Sinnexus). En los sistemas OLAP el acceso a los datos está optimizado para tareas frecuentes de lectura y escritura. La estructura de los datos está orientada a la aplicación. El historial de datos suele limitarse a los datos actuales o recientes. El procesamiento de transacciones en línea representa toda aquella información transaccional que genera la empresa en su accionar diario, además, de las fuentes externas con las que puede llegar a disponer. (Ing. Bernabéu, 2007). Entre los OLTP más habituales que pueden existir en cualquier organización se encuentran los archivos de textos, hipertextos, hojas de cálculos, bases de datos operativas.

### 1.10.2 OLAP

La tecnología de software para el análisis en línea, administración y ejecución de consultas que permiten predecir información del comportamiento del negocio (On Line Analytic Processing), es un motor de

consultas especializado. Su principal objetivo es brindar respuesta a preguntas complejas, para interpretar el comportamiento del negocio y tomar decisiones.

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, patrones, elaboración de informes complejos, ect. (Sinnexus)

Los sistemas de procesamiento analítico en línea suelen ser accedidos para operaciones de solo lectura, pues la acción más común es la consulta, con muy pocas inserciones, actualizaciones o eliminaciones. La estructura de los datos está orientada al negocio y los formatos de los datos están integrados de manera uniforme en toda la organización. El historial de datos es a largo plazo, normalmente de dos a cinco años.

Las bases de datos OLAP se suelen alimentar de información procedente de los sistemas operacionales.

### **1.10.3 Comparación entre OLTP Y OLAP**

La diferencia más significativa entre ambos sistemas es la forma en que se organizan los datos en la base de datos. Los sistemas OLTP utilizan el modelo entidad – relación para eliminar la redundancia de los datos y con ello posibilitar que las transacciones que alteran el contenido de los mismos sólo toquen a la base de datos en un punto. Basándose en esto se construyen diagramas complejos, debido a que divide los datos en muchas entidades las cuales son convertidas en tablas las consultas que resumen muchos registros y muchas tablas también se tornan difíciles de comprender por los usuarios y complejas de recorrer por los gestores. Motivos por los cuales Ralph Kimball propone el enfoque dimensional. El modelo dimensional es un modelo que permite a los usuarios una fácil comprensión de las bases de datos, además de posibilitar al software un recorrido eficiente de sus estructuras. Está básicamente compuesto por una tabla principal llamada hechos y un conjunto de pequeñas tablas alrededor de esta, nombradas dimensiones.



	OLTP	Data Warehouse
Objetivo	Soportar actividades transaccionales diarias.	Consultar y analizar información estratégica y táctica.
Tipo de datos	Operacionales.	Para la toma de decisiones.
Modelo de datos	Normalizado.	Desnormalizado.
Consulta	SQL.	SQL más extensiones.
Datos consultados	Actuales.	Actuales e históricos.
Horizonte de tiempo	60 - 90 días.	5 - 10 años.
Tipos de consultas	Repetitivas, predefinidas	No previsibles, dinámicas
Nivel de almacenamiento	Nivel de detalle.	Nivel de detalle y diferentes niveles de sumarización.
Acciones disponibles	Alta, baja, modificación y consulta.	Carga y consulta.
Número de transacciones	Elevado	Medio o bajo
Tamaño	Pequeño - Mediano.	Grande.
Tiempo de respuesta	Pequeño (segundos - minutos).	Variable (minutos - horas).
Orientación	Orientado a las aplicaciones.	Orientado al negocio.
Sello de tiempo	La clave puede o no tener un elemento de tiempo.	La clave tiene un elemento de tiempo.
Estructura	Generalmente estable.	Generalmente varía de acuerdo a su propia evolución y utilización.

**Tabla 1:** Comparación entre sistemas. (Ing. Bernabeu, 2007)

En los OLTP, los usuarios típicamente realizan actualizaciones, tales como agregar, modificar, eliminar y consultar algún registro. En cambio en un OLAP, la única operación que pueden realizar es la de consulta.

Las diferencias entre estos dos tipos de usuarios se pueden apreciar mejor en la siguiente tabla comparativa:

	Usuarios de OLTP	Usuarios de Data Warehouse
<i>Acceso concurrente</i>	Muchos.	Pocos.
<i>Tipo de consultas</i>	Predefinidas.	Complejas, no predecibles y no anticipadas.
<i>Registros consultados</i>	Pocos.	Muchos.
<i>Tiempo de respuesta</i>	Crítico.	No crítico.
<i>Acciones permitidas</i>	Agregar, modificar, eliminar y consultar.	Consultar.

**Tabla 2:** Diferencias entre los usuarios de OLTP y los de Data Warehouse. (Ing. Bernabeu, 2007)

Los OLTP son diseñados para soportar el procesamiento de información diaria de las empresas, y el énfasis recae en maximizar la capacidad transaccional de sus datos. Su estructura es altamente normalizada, para brindar mayor eficiencia a sistemas con muchas transacciones que acceden a un pequeño número de registros y están fuertemente condicionadas por los procesos operacionales que deben soportar, para la óptima actualización de sus datos. Esta estructura es ideal para llevar a cabo el

proceso transaccional diario, brindar consultas sobre los datos cargados y tomar decisiones diarias, en cambio los esquemas de OLAP están diseñados para poder llevar a cabo procesos de consulta y análisis para luego tomar decisiones estratégicas de alto nivel.

## 1.11 Modelado 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 captura la información fundamental acerca de las entidades del dominio del problema y sus relaciones. Este modelo es más cercano al espacio del problema que al espacio de la solución.

**Modelo de datos lógico:** El modelo lógico describe los datos en detalle, generalmente incluye todas las entidades y relaciones entre ellas, sus atributos y tipos de datos, así como las llaves primarias y foráneas, sin tener en cuenta cómo ellos se implementarán físicamente en la base de datos. Es un puente entre el nivel conceptual y el físico.

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

## 1.12 Metodologías evaluadas

Las metodologías de desarrollo de un Almacén de Datos tienen la finalidad de guiar el proceso de creación de un Almacén de Datos para que sea capaz de ser explotado por los usuarios finales. En el presente trabajo las principales metodologías evaluadas fueron DWEP y Hefesto.

### 1.12.1 Metodología DWEP

DWEP (Data Warehouse Engineering Process), es una metodología independiente de cualquier implementación específica, ya sea relacional, multidimensional, orientado a objetos y tiene como principal

característica que está basado en el Proceso Unificado de Desarrollo de Software adoptando además el lenguaje de modelado UML para la representación visual de los modelos generados. (Lujan Mora, 2005)

El método **DWEP** propone la estructuración del Almacén de Datos en cinco etapas y tres niveles.

## **Etapas:**

- **Origen:** Define los orígenes de datos del Almacén de Datos, como los sistemas OLTP, fuentes de datos externas, etc.
- **Integración:** Define el mapeo entre los orígenes de datos y el propio Almacén de Datos
- **Almacén de Datos:** Define la estructura del Almacén de Datos.
- **Adaptación:** Define el mapeo entre el Almacén de Datos y las estructuras empleadas por el cliente.
- **Cliente:** Define las estructuras concretas que son empleadas por los clientes para acceder al Almacén de Datos, como Data Mart o aplicaciones OLAP.

## **Niveles:**

- Conceptual.
- Lógico.
- Físico.

Como el DWEP es una instancia de RUP para el desarrollo de almacenes de datos, establece (al igual que este) el ciclo de vida de un proyecto en cuatro fases: Inicio, Elaboración, Construcción y Transición, así como cinco flujos de trabajo fundamentales: Requerimientos, Análisis y Diseño, Implementación y Prueba. Además adiciona dos nuevas actividades: Mantenimiento y Revisión Pos-desarrollo. En cada flujo de trabajo se utilizan diferentes diagramas UML, para modelar y documentar el proceso de desarrollo.

### **1.12.2 Metodología Hefesto**

La metodología Hefesto es independiente del tipo de ciclo de vida que se emplee para contener la metodología, de las herramientas que se utilicen para su implementación y de las estructuras físicas que

contengan el Almacén de Datos y su respectiva distribución. Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar. Se basa en los requerimientos del usuario, por lo cual su estructura es capaz de adaptarse con facilidad y rapidez ante los cambios en el negocio. Puede ser aplicada tanto en almacenes de datos como a Data Mart.

De forma general esta metodología funciona de la siguiente manera: comienza recolectando las necesidades de información de los usuarios y obteniendo las preguntas claves del negocio. Luego se identifican los indicadores resultantes de los interrogativos y sus respectivas perspectivas de análisis, mediante las cuales se construirá el modelo conceptual de datos del Almacén de Datos. Después se analizan los OLTP para señalar las correspondencias con los datos fuentes y seleccionar los campos de estudio de cada perspectiva. Una vez hecho esto, se pasará a la construcción del modelo lógico del depósito, explicitando las jerarquías que intervendrán. Por último, se definirán los procesos de carga, transformación, extracción y limpieza de los datos fuente.

### 1.13 Herramientas de creación de un Almacén de Datos

Debido a que para llevar a cabo la construcción del Almacén de Datos es necesario transitar por una serie de actividades que van desde 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, es de vital importancia contar con una serie de herramientas informáticas que asistan y faciliten el trabajo. Actualmente existen compañías que brindan colecciones de herramientas de Inteligencia de Negocio que permiten llevar a cabo la construcción del Almacén de Datos, que abarcan desde creación, ETL, acondicionamiento, explotación, e incluso mantenimiento.

Debido a requerimientos del usuario se utiliza la suite de inteligencia de negocio Oracle que es un sistema completo e integrado que permite llevar a cabo el proceso de desarrollo y creación de un Almacén de Datos así como su posterior explotación. Creando soluciones BI empresariales desde fuentes de datos heterogéneas.

**Oracle:** Es un sistema gestor de base de datos líder en la industria, utilizable 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 por ser el sistema gestor de bases de datos definido por parte de la

dirección del MININT además de que ofrece una excepcional disponibilidad, escalabilidad, fiabilidad y seguridad y. La principal desventaja de la utilización de Oracle como herramienta son los elevados precios de las licencias de software y del soporte técnico, lo que lo hace un gestor típico de sistemas informáticos de grandes compañías o instituciones gubernamentales. A ello se suma los elevados requisitos de hardware que tiene.

## 1.13.1 Arquitectura suite de Oracle: Figura 11: Arquitectura suite de Oracle

### 1.13.2 Herramientas de la suite Oracle en su versión empresarial

**Oracle WareHouse Builder (OWB):** es la herramienta que permite realizar el proceso de extracción, transformación y carga (ETL). Permite modelar, implementar y mantener un panorama coherente de las operaciones de datos desde múltiples sistemas de origen como archivos relacionales, multidimensionales tipo ASCII en el Almacén de Datos. Oracle Warehouse Builder es una herramienta orientada no solamente a realizar el proceso de ETL, sino también la definición, administración y mantenimiento de un data warehouse. Está concebida para trabajar integrada con la tecnología de Base de Datos Oracle. **Figura 12:** Oracle WareHouse Builder

**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 online u offline. Sólo está disponible para sistemas operativos Windows. **Figura 13:** Administration Tool

**Oracle Business Intelligence Server (OBI):** es el servidor de BI y ofrece un valioso motor de cálculos y puede integrar múltiples fuentes de datos dispares en una única visualización federada con soporte de fuentes Oracle y no Oracle, incluidas Microsoft SQL Server, Microsoft Office Excel, la mayoría de las fuentes ODBC, fuentes multidimensionales, XML, archivos tipo ASCII.

**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 fácil y rápidamente crear sus propias consultas, luego desglosar, analizar, visualizar e incluir los resultados en sus tableros personalizados o entregarlos en una variedad de formatos y canales mediante BI Publisher. **Figura 14:** Oracle BI Answers

**Oracle Business Intelligence Interactive Dashboards:** (paneles interactivos) ofrecen una interface web browser para ver instantáneamente las métricas importantes, los informes y visualizaciones tales como los medidores, cuadros, permiten capacidades completas de navegación y desglose total hasta los datos transaccionales, brinda informes resumidos que evalúa la empresa, lo cual es perfecto para los informes de administración. **Figura 15**

### **Conclusiones parciales.**

En el presente capítulo se ha realizado un estudio teórico sobre los conceptos y herramientas necesarias para la realización del Data Mart. Para la realización de este se seguirá el enfoque de abajo hacia arriba (bottom-up) propuesto por Kimball, con una arquitectura de datos de tres capas, a pesar de que el Almacén de Datos será poblado desde un único origen ella permite recuperarse de un error sin tener que reiniciar un proceso desde el inicio, además con su aplicación se podrán obtener los datos desde el origen de una forma mucho más rápida, se considera la arquitectura más completa. El diseño del esquema del Almacén de Datos será en constelación basado en el enfoque dimensional con una estructura ROLAP. Las políticas del cliente establecen a Oracle como gestor de base de datos, razón por la que el Almacén de Datos estará soportado sobre esta plataforma. Como herramienta de desarrollo para el Data Mart se seleccionó la suite de Inteligencia de Negocio de Oracle para la generación de reportes el análisis de información histórica y operativa, lo cual además de propiciar estar en total correspondencia con la política del cliente, permite un manejo lineal y transparente entre los datos de origen y el Almacén de Datos. El proceso de desarrollo estará guiado por la metodología Hefesto, con ella se realiza un análisis completo que permite abarcar los procesos principales de la organización, examinando e interpretando de forma óptima las necesidades de información del negocio, los resultados y objetivos esperados se distinguen fácilmente siendo los mismos sencillos de comprobar, la metodología se basa en los requerimientos de los usuarios, reduciendo de esta manera la resistencia al cambio porque involucra al cliente en toda la etapa de desarrollo para que tome decisiones respecto al comportamiento y funciones del Almacén de Datos.

### **CAPÍTULO 2: ANÁLISIS Y DISEÑO DEL DATA MART**

#### **2.1 Introducción**

En este capítulo se presenta el análisis realizado en la especialidad de Registro Legal de la DEP, el cual se realiza a través de la metodología Hefesto, seleccionada anteriormente por ser la más adecuada para dar cumplimiento a los objetivos propuestos y así diseñar el Data Mart. 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. Con los datos se identifican los indicadores que se analizan para realizar el diseño del Data Mart correspondiente.

#### **2.2 Análisis de requerimientos**

Para dar cumplimiento a los pasos planteados por la metodología, se analizan los requerimientos 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 a los jefes de la especialidad 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.

En el momento de realizar análisis estadísticos la información se consulta por los niveles de mando establecidos para establecer comparaciones con mayor precisión.

#### **Niveles de mando**

- Dirección de Establecimientos Penitenciarios.
- Jefaturas Provinciales
- Centros Penitenciarios

#### **Procesos**

**Delito:** se considera delito toda acción u omisión socialmente peligrosa prohibida por la ley bajo conminación de una sanción penal. En el momento que un interno ingresa en prisión puede hacerlo en la



condición de sancionado, acusado o asegurado, en cualesquiera de estas condiciones que se produzca su llegada el centro queda registrado entre otras características el delito cometido.

Con la implementación de este proceso cada uno de los niveles de mando tendrá la posibilidad de realizar análisis basándose en los reportes que serán generados. Los directivos podrán evaluar las tendencias apoyándose en la información histórica, entre los posibles análisis se encuentran, el índice delictivo, el tipo de delito con mayor número de ingresos, el delito con mayor índice de ingresos en una provincia, entre otros. Proceso de gran utilidad para los centros penitenciarios, con su utilización se podrán tomar medidas para contrarrestar los incrementos por un mismo tipo de delito.

**Juicios:** se denomina juicio al proceso, voluntario o contencioso, que se lleva a cabo ante los órganos del Poder Judicial para lograr una decisión.

En el centro donde se encuentre ubicado el interno se recibe la citación para la presentación a juicio oral del mismo, asegurando su presentación el día y hora señalada. En caso de no haber sido presentado a juicio oral, algún interno debidamente citado, el Jefe del Órgano Provincial y del MEIJ (municipio especial Isla de la Juventud) de Establecimientos Penitenciarios, realiza de inmediato las indagaciones en el centro penitenciario correspondiente para conocer las causas que dieron lugar a ello. De no existir motivo debidamente justificado, exige su presentación a juicio. El Órgano Provincial y del MEIJ de Establecimientos Penitenciarios, mensualmente evalúa el comportamiento de las presentaciones a juicios; internos citados, presentados, no presentados; juicios celebrados, suspendidos y motivo de las suspensiones. El proceso será diseñado para controlar la cantidad de juicios que se encuentran citados, realizados o que fueron suspendidos, si el mismo fue suspendido su causa. Esta información podrá obtenerse por los niveles de mando establecidos, permitiéndole realizar las evaluaciones y análisis de tendencias necesarios para la toma de decisiones de manera eficiente.

**Situación legal:** se conoce como situación legal al conjunto de características legales asociadas a un interno. La misma se encuentra vinculada al análisis de los procesos que tienen lugar en un centro penitenciario, entre ellos se encuentran:

- Causas.
- Sanciones.

- Antecedentes penales (Primario, Reincidente o Multirreincidente).
- Régimen penitenciario en el que está ubicado.
- Fecha de ingreso y de extinción.

La situación legal del interno está compuesta por aquellas observaciones que son útiles para la comprensión de su situación legal o penitenciaria. Estas valoraciones se consideran de gran importancia para las reglas por las que tendrá que regirse un interno en el centro. La situación legal de un interno es decisiva para la obtención de un permiso, progresión en régimen, libertad anticipada, entre otros beneficios.

El proceso será diseñado para generar los reportes necesarios que permitan los análisis estadísticos para la realización de informes y estudios de las tendencias referentes al mismo. Facilitando además la organización y control de este aspecto en los centros. Esta información podrá obtenerse por los niveles de mando establecidos y será de gran ayuda para una correcta toma de decisiones.

**Trámites legales pendientes:** se considera trámite legal pendiente aquellas situaciones carentes de definición, que presentan los sancionados y asegurados, que precisan de la determinación legal de la autoridad judicial competente o de la solución de otro proceso penal distinto, por el que aparecen sancionados o asegurados los internos.

- Situaciones que se consideran trámites legales pendientes:
- Pendientes de sanción conjunta.
- Pendientes de rectificación de liquidación de sanción.
- Falta de sentencia.
- Falta de liquidación de sanción.
- Denuncias pendientes.
- Expediente de Fase Preparatoria (EFP) pendientes.

- Causas pendientes.
- Rectificación de nombres y apellidos.
- Autos aclaratorios.
- Pendiente de revocación de beneficio de excarcelación anticipada, sanción subsidiaria o de interrupción de medida de seguridad.
- Expedientes pendientes en DNI.

Cuando un interno posee trámites legales pendientes no puede salir en libertad hasta que el mismo sea resuelto, si el mismo llega a la fecha que debe obtener la libertad sin que este sea decidido pasara entonces a prisión provisional. Todos los tipos de trámites tienen un tiempo máximo de resolución.

El proceso de trámites legales pendientes tiene una gran importancia para el sistema penitenciario, su correcto análisis y evaluación de tendencias en el tiempo, permite profundizar en los eventos que provocan en un centro de internamiento el retraso en la solución de los mismos. Esta información podrá ser consultada por los niveles de mando establecidos y potencia la correcta toma de decisiones.

**Comportamiento de la población penal:** es la agrupación de los procesos fundamentales en un centro penitenciario (ingreso, egreso).

**Ingreso:** se considera como ingreso en un centro penitenciario a las personas que arriban al mismo, ya sea en la condición de acusado, asegurado o sancionado.

**Egreso:** se considera egreso a un centro penitenciario cuando el interno obtiene la libertad, es objeto de un traslado hacia otro centro o fallece.

Este proceso se diseña con el objetivo de generar los reportes necesarios para poder evaluar los cambios y tendencias en dichos procesos, además de mejorar el control y organización de los mismos. Esta información podrá obtenerse por los niveles de mando establecidos, apoyando la toma de decisiones.

De los cinco procesos que fueron identificados se seleccionaron los tres críticos y de mayor peso en el sistema penitenciario para detallar el proceso de análisis, diseño e implementación que fue realizado.

### 2.2.1 Identificar perspectivas

Se procede a identificar lo que les interesaba conocer acerca de los procesos, saber cuáles eran las variables o perspectivas<sup>1</sup> que debían tenerse en cuenta para poder tomar decisiones basadas en ello. En cuanto a estas perspectivas se decide utilizar las predefinidas para los cierres con el fin de integrar la totalidad de puntos de vista bajo los que puede observarse la gestión de información en la organización.

A continuación se muestra una tabla que recoge las perspectivas identificadas y una descripción de la función que cumple la misma.

Perspectivas	Es una tabla con sus atributos correspondientes, los cuales nos permitirán filtrar los indicadores propuestos.	
Perspectivas Comunes		
Perspectivas	Descripción	
Tiempo	Tabla que indica especificaciones de la fecha en la que ocurre un evento relacionado con un proceso en específico.	
Interno	Tabla que agrupa las características del interno asociadas a los procesos que se desarrollan en la especialidad.	
Centro	Tabla que recoge las especificaciones de un centro penitenciario que son necesarias para los procesos.	
Perspectivas por los procesos identificados		
Perspectivas	Descripción	

<sup>1</sup> La perspectiva sería en el lenguaje del modelo relacional una tabla de la base de datos, que se identifica por atributos o características. En nuestro caso son tablas que se crean en dependencia del proceso al que pertenece para dar cumplimiento a los indicadores propuestos.

Tipo delito	Tabla que recoge las características específicas de los delitos, está relacionada con el proceso delito.
Comportamiento	Tabla que agrupa los atributos que están relacionados con el ingreso o egreso de los internos en los lugares de internamiento y los motivos correspondientes, se encuentra relacionada al proceso comportamiento de la población penal.
Tipo trámites	Tabla que agrupa los tipos de trámites legales pendientes. Se relaciona con el proceso trámites legales pendientes.
Instancia	Tabla que tiene como atributo la instancia (PNR, Tribunal, Fiscalía) que dio origen al trámite. Se relaciona con el proceso trámites legales pendientes.

Tabla 3: Perspectivas identificadas y su descripción

### 2.2.2 Indicadores identificados

En el levantamiento de requisitos correspondiente se identifican aquellos indicadores<sup>2</sup> en los que estarán basados los reportes finales, en dependencia de las necesidades del cliente por cada uno de los procesos.

A continuación se especifican los indicadores que fueron identificados por cada uno de los procesos

Proceso	Delito
Indicadores	Descripción

<sup>2</sup> Los indicadores son sumalizaciones efectuadas sobre algún hecho o expresiones basadas en sumalizaciones, que serán incluidos en algún cubo multidimensional, con el fin de analizar los datos almacenados en el DW. El valor que estos adopten estará condicionado por los atributos/jerarquías que se utilicen para analizarlos.

Total por tipo de delito	El indicador “Total tipo de delito” representa el total de cada uno de los tipos de delitos existentes.
Total por familia delictiva	El indicador “Total por familia delictiva” representa el total por cada una de las familias delictivas.
Por ciento tipo de delito	El indicador “Por ciento tipo de delito”, representa el por ciento que representa un tipo de delitos de su total.
Por ciento tipo delito_Familia delictiva	El indicador “Por ciento tipo delito_Familia delictiva”, representa el por ciento de un tipo de delitos del total perteneciente a la familia delictiva que este pertenece.
Por ciento familia delictiva _total	El indicador “Por ciento Familia delictiva_total”, representa el por ciento del total de delitos pertenecientes a una familia delictiva del total de delitos existentes.
Proceso	Comportamiento de la población penal
Total de ingresos	El indicador “Total de ingresos” representa el total de internos que ingresan a un centro penitenciario.
Total de egresos	El indicador “Total de egresos” representa el total de internos que provocan un egreso a un centro penitenciario.
Total de ingresos por motivo	El indicador “Total de ingresos por motivos” representa el total de internos que ingresan a un centro penitenciario por cada uno de los motivos existentes.
Total de egresos por motivo	El indicador “Total de egresos por motivos” representa el total de internos que egresan a un

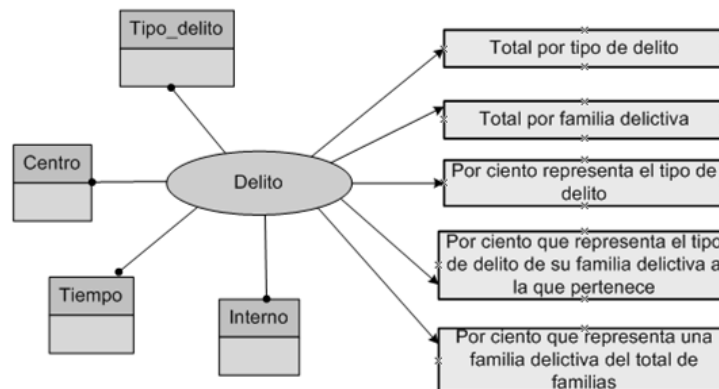
	centro penitenciario por cada uno de los motivos existentes.
Por ciento ingresos	El indicador “Por ciento ingresos”: representa el por ciento de ingresos en el centro con respecto a la población penal del mismo.
Por ciento egresos	El indicador “Por ciento egresos”: representa el por ciento de egresos en el centro con respecto a la población penal del mismo.
Proceso	Trámites Legales Pendientes
Total tipo trámite	El indicador “Total tipo trámites” representa el total por cada uno de los diferentes tipos de trámites, en cada uno de los niveles de mando.
Total de trámites	El indicador “Total de trámites” representa el total de trámites pendientes por los diferentes niveles de mando.
Total tipo trámites por instancia	El indicador “Total tipo trámites por instancia” representa el total de cada uno de los diferentes tipos de trámite que corresponden a una instancia.
Total de trámites resueltos	El indicador “Total de trámites resueltos” representa el total de trámites resueltos por los diferentes niveles de mando.
Por ciento tipo trámites	El indicador “Por ciento tipo trámites”: representa el por ciento que un tipo de trámite representa con respecto de su total.

Por ciento trámites resueltos	El indicador “Por ciento trámites resueltos”: representa el por ciento que representan los trámites resueltos del total de trámites pendientes.
Por ciento tipo de trámites resueltos	El indicador “Por ciento tipo trámites resueltos”: representa el por ciento que un tipo de trámite representa con respecto de su total.

**Tabla 4:** Indicadores identificados por procesos y su explicación.

### 2.2.3 Modelo conceptual

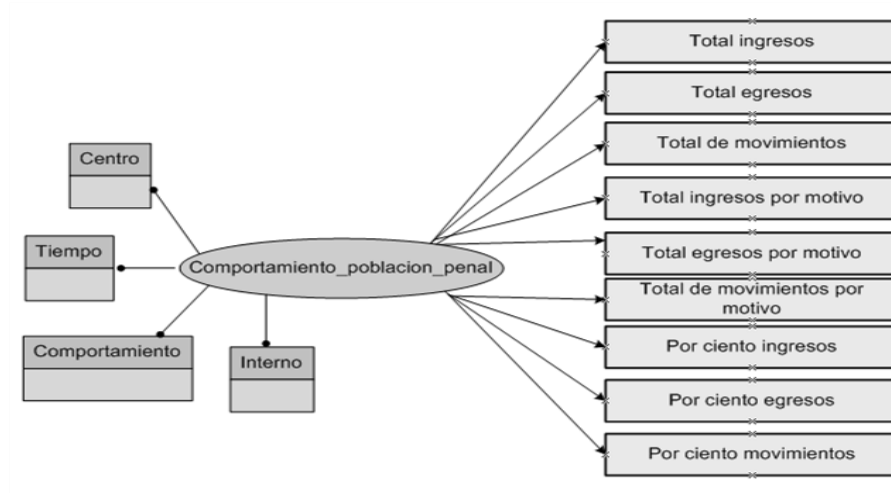
Luego de identificados los principales procesos con las perspectivas y los indicadores, se procede a realizar los modelos conceptuales<sup>3</sup> referentes a cada uno de los procesos identificados con el objetivo de identificar el alcance del proyecto.



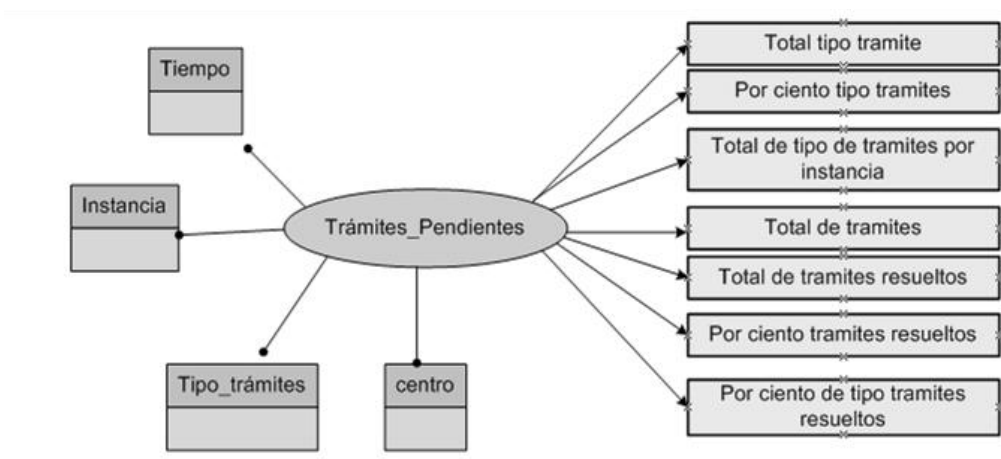
**Figura 16:** Modelo conceptual asociado al proceso Delito

<sup>3</sup> Descripción de alto nivel de la estructura de la base de datos, en la cual la información es representada a través de objetos, relaciones y atributos.





**Figura 17:** Modelo conceptual asociado al proceso Comportamiento de la Población Penal



**Figura 18:** Modelo conceptual asociado al proceso Trámites Legales Pendientes

### 2.3 Análisis de los OLTP

Después del análisis realizado anteriormente se procede a conformar los indicadores por cada uno de los procesos, establecer las respectivas correspondencias entre el modelo conceptual creado en el paso anterior y las fuentes de datos.

### 2.3.1 Conformar indicadores

Por todo el análisis antes realizado se procede a conformar los indicadores que serán necesarios para la implementación de los reportes.

#### **Delito**

##### **Total tipo de delito**

Hechos: tipo de delito

Función de sumarización: COUNT

##### **Total por familia delictiva**

Hechos: familia delictiva

Función de sumarización: COUNT

##### **Por ciento tipo delito**

Ttd: total de un tipo de delito.

Td: total de delitos.

Hechos:  $(Ttd / 100) * 100$

Función de sumarización: Por ciento

##### **Por ciento tipo delito \_Familia delictiva**

Ttd: total de un tipo de delito.

TdFd: total de delitos de su familia delictiva.

Hechos:  $(Ttd / TdFd) * 100$

Función de sumarización: Por ciento

##### **Por ciento familia delictiva\_total**

TdFd: total de delitos que pertenecen a una familia delictiva.

Td: total de delitos.

Hechos:  $(TdFd / Td) * 100$

Función de sumarización: Por ciento

### **Comportamiento de la población penal**

#### **Total de ingresos**

Hechos: ingresos.

Función de sumarización: COUNT

#### **Total de egresos**

Hechos: egresos.

Función de sumarización: COUNT

Total de ingresos por motivo

Hechos: ingresos por motivo.

Función de sumarización: COUNT

#### **Total de egresos por motivo**

Hechos: egresos por motivo.

Función de sumarización: COUNT

#### **Por ciento ingresos**

Ti: total de ingresos.

Pp\_c: población penal del centro.

Hechos:  $(Ti/Pp\_c)*100$

Función de sumarización: Por ciento

### **Por ciento egresos**

Te: total de egresos.

Pp\_c: población penal del centro.

Hechos:  $(Te/Pp\_c)*100$

Función de sumarización: Por ciento

### **Trámites legales pendientes**

Total de trámites

Hechos: trámites

Función de sumarización: COUNT

### **Total de trámites resueltos**

Hechos: trámites resueltos.

Función de sumarización: COUNT

### **Total tipo de trámites**

Hechos: tipo de trámites.

Función de sumarización: COUNT

### **Total tipo de trámites por instancia**

Hechos: trámites por instancia.

Función de sumarización: COUNT

### **Por ciento tipo trámites**

Tt\_t: total tipo de trámites.

Tt: total de trámites pendientes

Hechos:  $(Tt\_t/Tt)*100$

Función de sumarización: Por ciento

### **Por ciento trámites resueltos**

Tt\_r: total trámites resueltos.

Tt: total de trámites pendientes

Hechos:  $(Tt\_r / Tt)*100$

Función de sumarización: Por ciento

### **Por ciento tipo trámites resueltos**

Tt\_tr: total tipo de trámites resueltos.

Tt: total de trámites resueltos

Hechos:  $(Tt\_tr / Tt)*100$

Función de sumarización: Por ciento

### **2.3.2 Establecer correspondencias**

Para poder identificar las correspondencias entre el modelo conceptual y las fuentes de datos se examinan las fuentes de datos disponibles que contenga la información requerida. La fuente de datos examinada fue la base de datos del proyecto “Humanización Penitenciaria”.

En las tablas que serán mostradas a continuación se realizan las correspondencias para cada una de las perspectivas y se ofrece una breve descripción de los datos de análisis que se encuentran presentes.

Análisis OLTP: Perspectiva Interno		
Atributo Destino	Tablas Origen	Descripción
Sexo	Persona	El atributo representa el sexo del interno.
Raza	Datos _Fisionomicos Nom_colorpiel	Representa el color de piel del interno.
Provincia_residencia	Nom_Provincia	Nombre de la provincia de residencia del interno.
Activo	Expediente_Legal	Indica si el interno está en prisión en ese momento.
Ciudadania_natural	Nom_nacionalidad	Representa la ciudadanía natural del interno.
Ciudadania_adquirida	Interno_ciudadania	En caso de que el interno haya adquirido una ciudadanía, el nombre de la misma
Fecha_nacimiento	Interno	Representa la fecha de nacimiento del interno.
Id_dim_interno	Interno	Llave primaria que representa a la tabla.

**Tabla 5:** Análisis para la perspectiva Interno

Para la perspectiva Tiempo solo es necesario establecer la correspondencia con el campo FECHA\_SQL que existe en la base de datos, pues los restantes son calculados.

Análisis OLTP: Perspectiva Tiempo

Atributo Destino	Descripción
Codigo_Tiempo	Código de una fecha en particular.
Num_Dia	Número del día en que se desarrolla el proceso.
Num_Mes	Número del mes en que se desarrolla el proceso.
Num_Anno	Número del año en que se desarrolla el proceso.
Nomb_Dia	Nombre del día en que se desarrolla el proceso.
Nomb_Mes	Nombre del mes en que se desarrolla el proceso.
Trimestre	Período en que se desarrolla el proceso, en este caso de tres meses.
Cuatrimestre	Período en que se desarrolla el proceso, en este caso de cuatro meses.
Semestre	Período que se desarrolla el proceso, en este caso de seis meses.
Fecha_SQL	Fecha completa del día que se desarrolla el proceso y es mediante este atributo que se establecerá la correspondencia con la fuente de datos.
Num_Dia_Semana	Número de la semana en que se desarrolla el proceso.

**Tabla 6:** Análisis para la perspectiva Tiempo

Análisis OLTP: Perspectiva Centro		
Atributo Destino	Tablas Origen	Descripción
Nombre	Estructura	Nombre del centro penitenciario.

	Nom_tipo_estructura	
Municipio	Nom_municipio	Municipio en el que se encuentra ubicado el centro.
Provincia	Nom_Provincia	Provincia en la que se encuentra ubicada el centro.
Id_dim_centro	Nom_tipo_centro	Clave primaria de la tabla "centro", y representa unívocamente a un centro en particular.

Tabla 7: Análisis para la perspectiva Centro

Análisis OLTP: Perspectiva Comportamiento		
Atributo Destino	Tablas Origen	Descripción
Id_dim_comportamiento	Expediente_legal	Representa la clave primaria de la tabla.
Comportamiento	Ingreso Egreso Reingreso	Representa el tipo de proceso (ingreso, egreso).
Motivo_comportamiento	Nom_Motivo_Ingreso Nom_Motivo_Egreso Nom_Motivo_Reingreso	Hace referencia al motivo por el interno que provoca un (ingreso, egreso) al centro penitenciario.

Tabla 7: Análisis para la perspectiva Comportamiento

Análisis OLTP: Perspectiva Tipo_delito		
Atributo Destino	Tablas Origen	Descripción



Tipo_delito	Nom_delito	Nombre del tipo de delito.
Familia_delictiva	Nom_delito	Nombre de la familia delictiva a la que pertenece el delito.
Id_delito	Nom_delito	Clave primaria del Nom_delito.
Id_dim_tipo_delito	Nom_delito	Clave primaria de la tabla.

**Tabla 8:** Análisis para la perspectiva Tipo\_delito

Análisis OLTP: Perspectiva Tipo_Tramites		
Atributo Destino	Tablas Origen	Descripción
Id_tipo_tramites	Nom_TipoTramite	Representa la clave primaria de la tabla.
Nombre_tipo_tramite	Nom_TipoTramite	Nombre de la familia delictiva a la que pertenece el delito.

**Tabla 9:** Análisis para la perspectiva Tipo\_Tramites

Análisis OLTP: Perspectiva Instancia		
Atributo Destino	Tablas Origen	Descripción
Id_dim_instancia	Nom_Organo	Representa la clave primaria de la tabla.
Instancia	Nom_Organo	Hace referencia al tipo de instancia (tribunal, fiscalía, PNR)

**Tabla 10:** Análisis para la perspectiva Instancia

### 2.3.4 Modelo conceptual ampliado

Se procede a actualizar el modelo conceptual realizado anteriormente incluyendo los campos definidos para cada perspectiva.

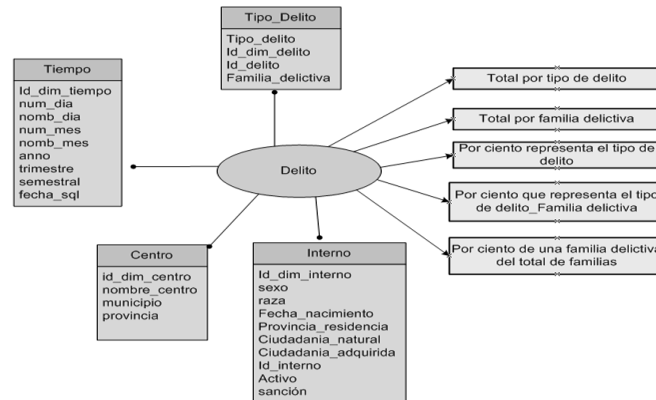


Figura 19: Modelo conceptual ampliado asociado a Delito

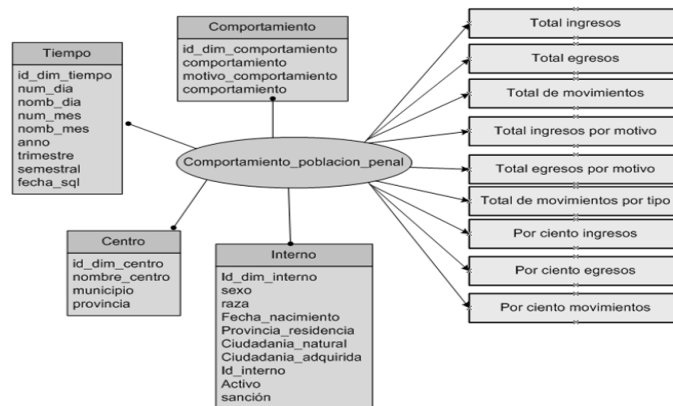


Figura 20: Modelo conceptual ampliado asociado a Comportamiento de la Población Penal

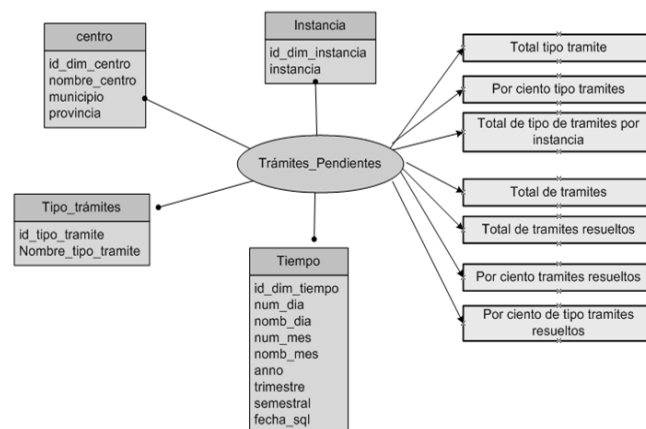


Figura 21: Modelo conceptual ampliado asociado a Trámites Legales Pendientes

### **Conclusiones**

En el presente capítulo se ha realizado un análisis sobre los principales procesos que los especialistas del departamento de la DEP analizan para generar las estadísticas solicitadas. Se identificaron cuatro procesos importantes, los cuales fueron: comportamiento de la población penal, trámites legales pendientes, prisión provisional y juicios. A partir de estos se identificaron las principales perspectivas de análisis y los datos por cada una de ellas que se necesitan. Todo esto es expuesto en los diagramas conceptuales ampliados de cada proceso, los cuales son los artefactos que recogen todo el análisis realizado.

### **CAPÍTULO 3: IMPLEMENTACIÓN DEL DATA MART**

#### **3.1 Introducción**

En este capítulo se construyen las estructuras físicas del Data Mart, se define el tipo de estructura empleada para el almacenamiento y se crean los correspondientes esquemas lógicos. Se llevan a cabo los procesos ETL, se definen los procesos de carga, se configura el depósito de datos y se implementan los informes solicitados por los clientes, haciendo uso de las herramientas de consulta.

#### **3.2 Modelo lógico del Data Mart**

El modelo lógico de la estructura del Data Mart, tiene como base el modelo conceptual que fue creado. Lo primero es definir el tipo de modelo a utilizar y luego se llevan a cabo las acciones propias al caso, para diseñar las tablas de dimensiones y de hechos. Para finalizar se realizan las uniones pertinentes entre las tablas.

##### **3.2.1. Tipo de modelo lógico del Data Mart**

Existen diferentes tipos de esquemas que se utilizan para contener la estructura del depósito de datos. Los mismos se definen como: esquema en estrella, copo de nieve y constelación, siendo este último el seleccionado para contener la estructura del Data Mart.

#### **3.3 Tablas de dimensiones**

En este paso se diseñan las tablas de dimensiones que forman parte del Data Mart. Cada perspectiva definida en el modelo conceptual constituye una tabla de dimensión.

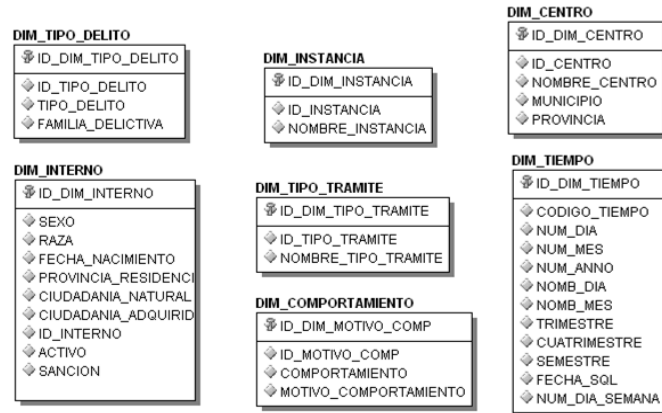


Figura 22: Tablas de Dimensiones

### 3.4 Tablas de hechos

En este paso se definen las tablas de hechos, que son las que contendrán los hechos a través de los cuales se construirán los indicadores de estudio.

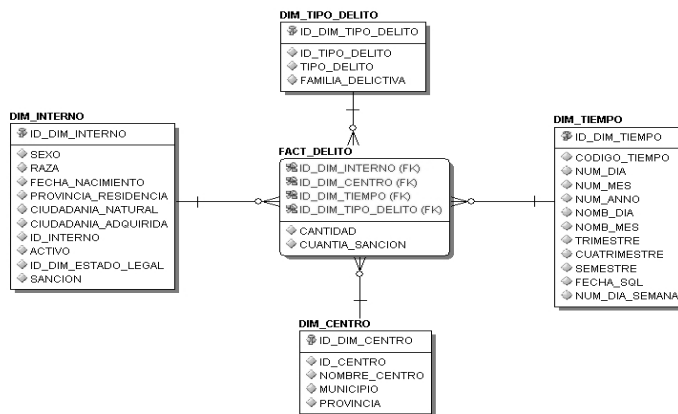


Figura 23: Tabla de hechos correspondiente al proceso Delito.

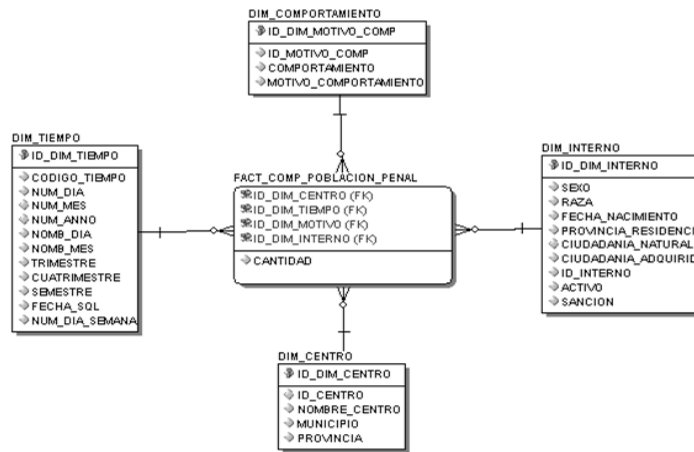


Figura 24: Tabla de hechos Comportamiento de la Población Penal

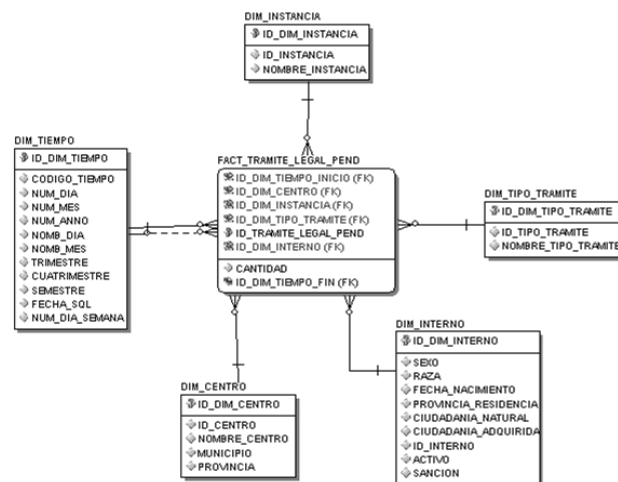


Figura 25: Tabla de hechos Trámites Legales Pendientes

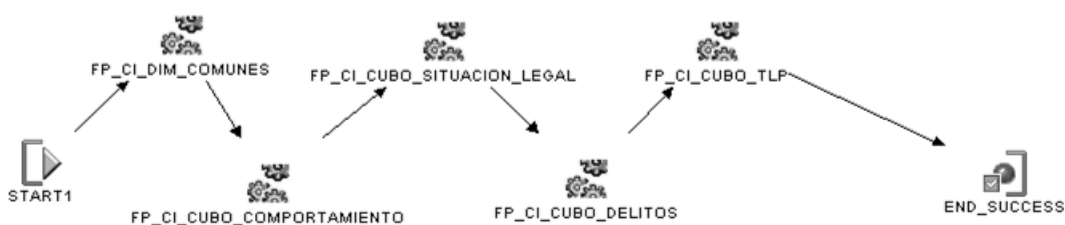
### 3.5 Diseño de los procesos ETL

El Data Mart contiene metadatos sobre los procesos de extracción, transformación y carga de los datos de la base de datos operacional. Los procesos se ejecutarán periódicamente mediante la actualización del almacén de modo que la información continúe siendo relevante y oportuna para la toma de decisiones.

A continuación se muestra el proceso ETL (extracción, transformación y carga), para el Data Mart de Registro Legal, donde primeramente se lleva a cabo la extracción de los datos de las fuentes

correspondientes y a partir de ahí se realiza la transformación, cuya tarea principal es la limpieza de los datos, para evitar redundancia, y otros problemas que pueden provocar que el almacén se vuelva más lento a la hora de consultar los reportes. Después se realiza la carga de los datos y se guardan en los Data Mart creados.

El diseño de las ETL fue realizado con la herramienta Desing Center en su versión 11.1.0.6.0. que no es más que un cliente del OWB.

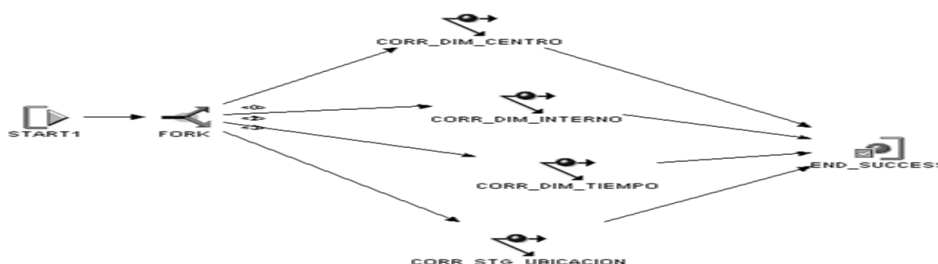


**Figura 26:** Carga inicial.

La carga inicial es el proceso encargado de que se carguen las dimensiones comunes y de poblar los cubos correspondientes.

Se definen como dimensiones comunes aquellas que serán utilizadas por todos los procesos identificados.

- Tiempo
- Interno
- Centro

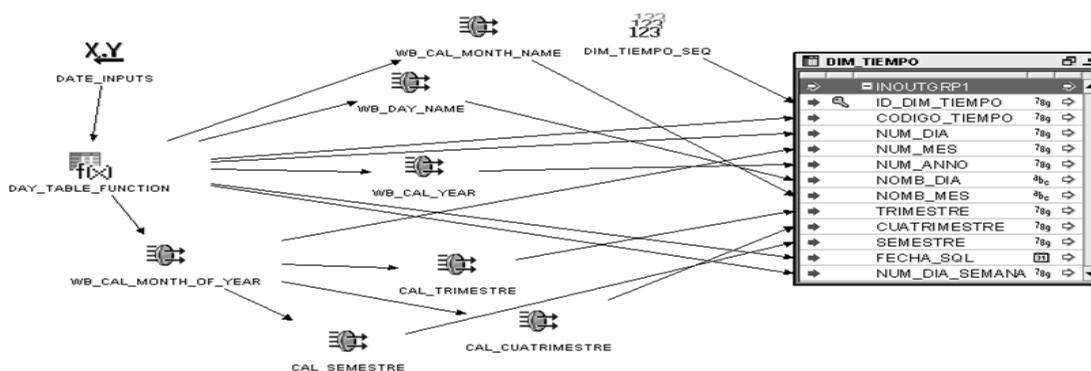


**Figura 27:** Carga inicial de las dimensiones comunes

Las tareas que se llevan a cabo en este proceso son las siguientes:

- START1: actividad de inicio, encargada de que comience el flujo de actividades que componen el proceso.
- FORK: actividad encargada de iniciar actividades múltiples de forma simultánea.
- Se ejecutan los pasos necesarios para cargar las dimensiones comunes, más adelante se especifican los pasos para realizar este proceso.
- END\_SUCCESS: actividad final, indica que el proceso ha concluido con éxito.

A continuación se especifican los pasos para cargar la DIM\_TIEMPO.



**Figura 28:** Carga de la DIM\_TIEMPO

Este procedimiento realiza lo siguiente:

- Recibe como parámetros los valores de "Fecha\_Inicio" y "Cantidad\_años".
- Recorre una a una las fechas que se encuentran dentro de este intervalo.
- Analiza cada fecha y realiza una serie de operaciones para crear los valores de los campos de la tabla de la dimensión FECHA:



- Se utilizaron algunas funciones propias del OWB y otras que fueron necesarias de implementar para solventar los requerimientos de la fecha solicitados por el cliente. Las mismas son:

**CAL\_SEMESTRE**

```

trimestre number;
mes number;
BEGIN
trimestre := ((num_mes-1)/6);
trimestre := floor(trimestre);
RETURN trimestre + 1;
END;
```

• **CAL\_CUATRIMESTRE**

```

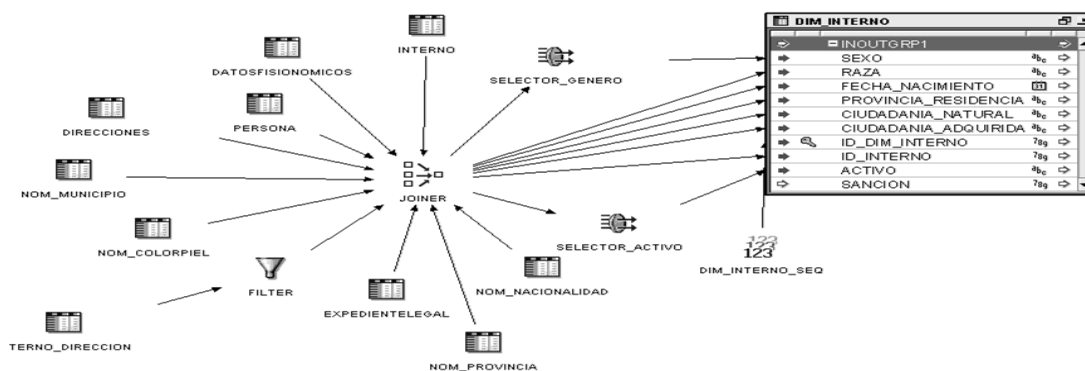
trimestre number;
mes number;
BEGIN
trimestre := ((num_mes-1)/4);
trimestre := floor(trimestre);
RETURN trimestre + 1;
END;
```

**CAL\_TRIMESTRE**

```

trimestre number;
mes number;
BEGIN
trimestre := ((num_mes-1)/3);
trimestre := floor(trimestre);
RETURN trimestre + 1;
END;
```

A continuación se especifican los pasos para cargar la DIM\_INTERNO.

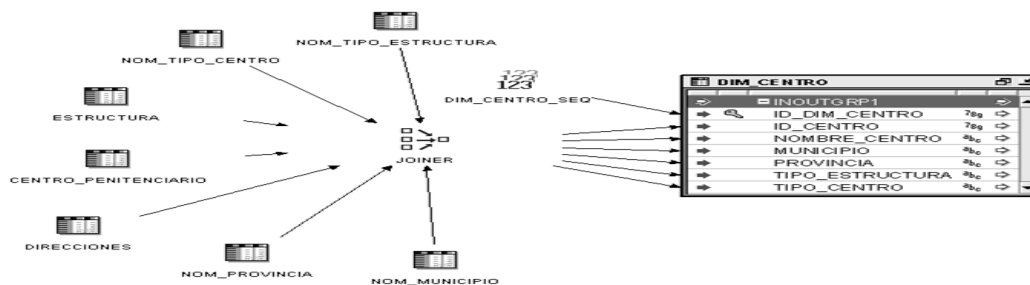


**Figura 29:** Carga de la DIM\_INTERNO

- Se toma como fuente de entrada los datos necesarios que se encuentran en las tablas:
- Interno, DatosFisionomicos, Persona, Direcciones, Nom\_Municipio, Nom\_colorpiel, Interno\_direccion, ExpedienteLegal, Nom\_Provincia, Nom\_nacionalidad.
- Se utiliza el operador FILTER, el objetivo del mismo es ser conectado a otro operador para aplicar una condición de filtro definida y enviar un subconjunto de filas al siguiente operador, su salida es el resultado del filtro aplicado, en este caso el FILTER es aplicado a la tabla Interno\_direccion.

- Se procede a utilizar el operador JOINER, es el encargado de unir varios conjuntos de filas de diferentes fuentes con diferentes cardinalidades y produce una salida. En este caso es el encargado de unir las tablas antes mencionadas, para producir la salida encargada de poblar la DIM\_INTERNO.
- El operador Selector\_Genero es una función implementada que tiene como entrada un número entero (0,1) que identifica el sexo del interno y tiene como salida la descripción asociada al valor (femenino, masculino).
- Finalmente se le agrega un id secuencial, utilizando la operador de secuencia DIM\_INTERNO\_SEQ.

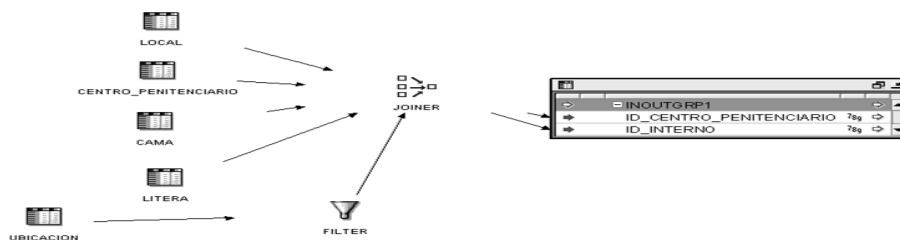
A continuación se especifican los pasos para cargar la DIM\_CENTRO.



**Figura 30:** Carga de la DIM\_CENTRO

- Se toma como fuente de entrada los datos necesarios que se encuentren en la tablas:
- Centro\_Penitenciario, Nom\_TipoCentro, Nom\_TipoEstructura, Nom\_Provincia, Nom\_Municipio, Direcciones, Estructura.
- Se procede a utilizar el operador JOINER, es el encargado de unir varios conjuntos de filas de diferentes fuentes con diferentes cardinalidades y produce una salida. En este caso es el encargado de unir las tablas antes mencionadas, para producir la salida encargada de poblar la Dim\_centro.
- Finalmente se le agrega un id secuencial, utilizando el operador de secuencia DIM\_Centro\_SEQ.

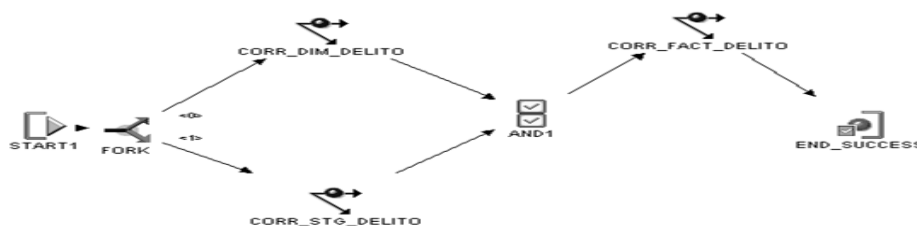
A continuación se especifican los pasos para cargar el STG\_Ubicacion.



**Figura 31:** Carga de el STG\_UBICACION

- Se toma como fuente de entrada los datos necesarios que se encuentren en la tablas: Litera, Cama, Centro\_Penitenciario, Local, Ubicación.
- Se utiliza el operador FILTER, el objetivo del mismo es ser conectado a otro operador para aplicar una condición de filtro definida y enviar un subconjunto de filas al siguiente operador, su salida es el resultado del filtro aplicado, en este caso el FILTER es aplicado a la tabla ubicación.
- Se procede a utilizar el operador JOINER, es el encargado de unir varios conjuntos de filas de diferentes fuentes con diferentes cardinalidades y produce una salida. En este caso es el encargado de unir las tablas antes mencionadas, para producir la salida encargada de poblar el STG\_UBICACION.

**Proceso: Delito**

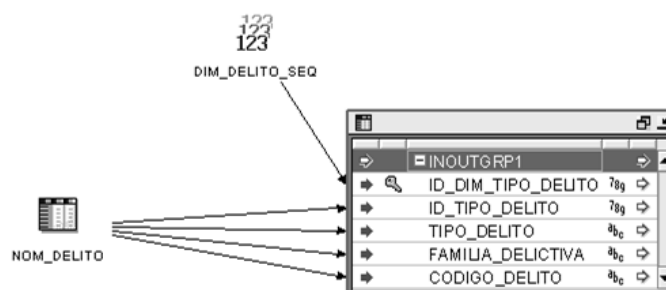


**Figura 32:** Carga inicial del proceso Delito

- Las tareas que se llevan a cabo en este proceso son las siguientes:

- START1: actividad de inicio, encargada de que comience el flujo de actividades que componen el proceso.
- FORK: actividad encargada de iniciar actividades múltiples de forma simultánea.
- CORR\_DIM\_DELITO: ejecuta el contenedor de pasos que cargará la DIM\_DELITO, más adelante se detallará el mismo.
- CORR\_STG\_DELITO: ejecuta el contenedor de pasos necesarios para cargar el STG\_DELITO, más adelante se detalla el mismo.
- AND1: actividad encargada de controlar la correcta ejecución de la actividad siguiente, no permite que comience la ejecución de la misma sin que hayan terminado completamente las anteriores.
- CORR\_FACT\_DELITO: ejecuta el contenedor de pasos para cargar la FACT\_DELITO, más adelante se detalla el mismo.
- END\_SUCCESS: actividad final, indica que el proceso ha concluido con éxito.

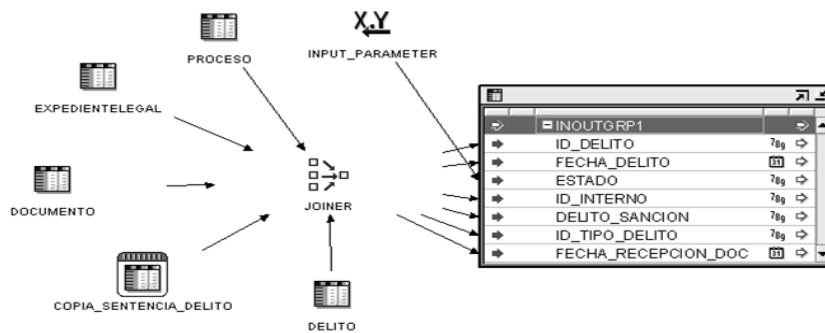
A continuación se especifican los pasos para cargar la DIM\_DELITO.



**Figura 33:** carga de la DIM\_DELITO

- Se tomará como fuente de entrada los datos contenidos en la tabla NOM\_DELITO, además se utiliza una actividad para su id secuencial.

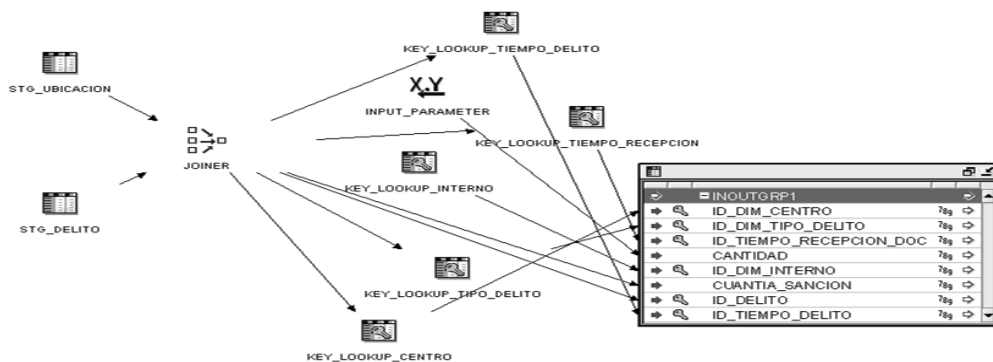
A continuación se especifican los pasos para cargar el STG\_DELITO.



**Figura 34:** Carga del STG\_DELITO

- Se toma como fuente de entrada los datos necesarios que se encuentran en las tablas:
- ExpedienteLegal, Proceso, Documento, Copia\_Sentencia\_Delito, Delito.
- Se procede a utilizar el operador JOINER, es el encargado de unir varios conjuntos de filas de diferentes fuentes con diferentes cardinalidades y produce una salida. En este caso es el encargado de unir las tablas antes mencionadas, para producir la salida encargada de poblar el STG\_DELITO.
- Se utiliza el operador INPUT\_PARAMETER, para cargar datos que no se encuentran en la base de datos, como pueden ser las constantes.

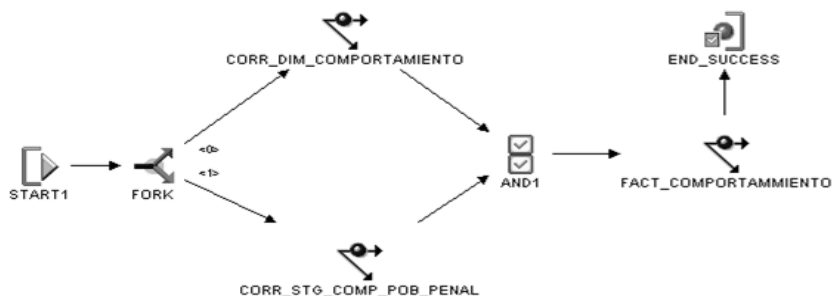
A continuación se especifican los pasos para cargar la FACT\_DELITO.



**Figura 35:** Carga de la FACT\_DELITO

- Se utiliza como datos de entrada el STG\_DELITO y el STG\_UBICACION.
- Se procede a utilizar el operador JOINER, es el encargado de unir varios conjuntos de filas de diferentes fuentes con diferentes cardinalidades y produce una salida.
- El operador Key\_Lookup se utiliza para leer los datos de las dimensiones correspondientes a través de su clave o llave. En esta caso se obtienen como salida los datos de las dimensiones interno, tiempo, tipo delito, centro, cargándose de esta manera los hechos en la FACT\_DELITO.
- Se utiliza el operador INPUT\_PARAMETER, para cargar datos que no se encuentran en la base de datos y que son necesarios en la FACT\_DELITO.

### Proceso: Comportamiento de la Población Penal



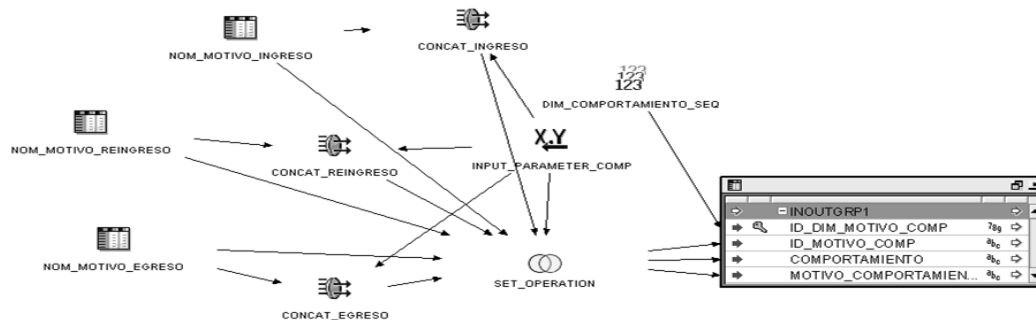
**Figura 36:** carga inicial del proceso Comportamiento de la Población Penal

Las tareas que se llevan a cabo en este proceso son las siguientes:

- START1: actividad de inicio, encargada de que comience el flujo de actividades que componen el proceso.
- FORK: actividad encargada de iniciar actividades múltiples de forma simultánea.
- CORR\_DIM\_COMPORTAMIENTO: ejecuta el contenedor de pasos que cargará la DIM\_COMPORTAMIENTO, más adelante se detallará el mismo.

- CORR\_STG\_COMPORTEAMIENTO: ejecuta el contenedor de pasos necesarios para cargar el STG\_COMPORTEAMIENTO, más adelante se detalla el mismo.
- AND1: actividad encargada de controlar la correcta ejecución de la actividad siguiente, no permite que comience la ejecución de la misma sin que hayan terminado completamente las anteriores.
- CORR\_FACT\_COMPORTEAMIENTO: ejecuta el contenedor de pasos para cargar la FACT\_COMPORTEAMIENTO, más adelante se detalla el mismo.
- END\_SUCCESS: actividad final, indica que el proceso ha concluido con éxito.

A continuación se especifican los pasos para cargar la DIM\_COMPORTEAMIENTO.

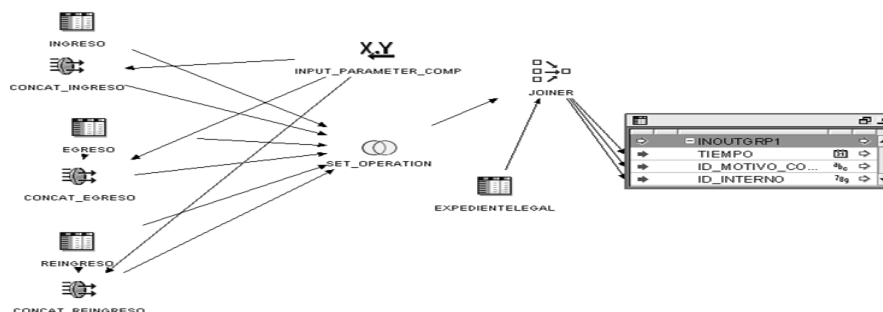


**Figura 37:** Carga de la DIM\_COMPORTEAMIENTO

- Se tomaran como fuente de entrada los datos necesarios que se encuentran en las tablas: Nom\_Motivo\_Reingreso, Nom\_Motivo\_Ingreso, Nom\_Motivo\_Egreso.
- Se utiliza la función CONCAT para cada una de las tablas, la función es equivalente al operador de concatenación, su funcionamiento principal se basa en concatenar atributos del tipo de dato CHAR o VARCHAR, devuelve una cadena de tipo VARCHAR.
- Se utiliza el operador INPUT\_PARAMETER\_COMP, para cargar datos que no se encuentran en la base de datos y que le son necesarios a la función CONCAT.
- Se utiliza el operador SET\_OPERATION, su función en este caso es la de combinar los resultados de las consultas y devolver el resultado de esto.

- Se carga la DIM\_COMPORTEAMIENTO recibiendo como entrada lo que devuelve el operador SET\_OPERATION.
- Finalmente se le agrega un id secuencial, utilizando el operador de secuencia DIM\_COMPORTEAMIENTO\_SEQ.

A continuación se especifican los pasos para cargar el STG\_COMPORTEAMIENTO.



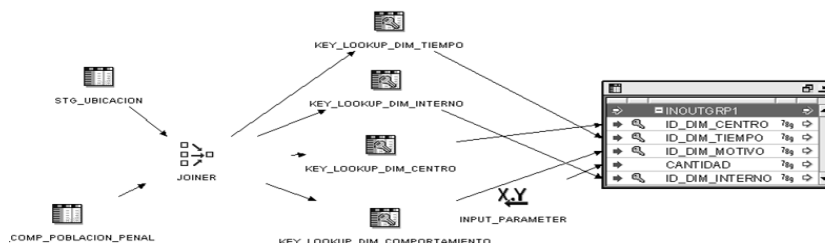
**Figura 38:** Carga del STG\_COMPORTEAMIENTO

- Se toma como fuente de entrada los datos necesarios que se encuentran en las tablas: Ingreso, Egreso, Reingreso, ExpedienteLegal.
- Se procede a utilizar el operador INPUT\_PARAMETER\_COMP.
- Se procede a utilizar la función CONCAT para cada una de las tablas. En este caso se encarga de concatenar las tablas Ingreso, Egreso y Reingreso con los datos que provee el operador INPUT\_PARAMETER\_COMP.
- Se procede a utilizar el operador SET\_OPERATION, en este proceso el operador se encarga de unir el resultado de las consultas y su resultado es la entrada del operador JOINER.
- Se procede a utilizar el operador JOINER, encargado de unir varios conjuntos de filas de diferentes fuentes con diferentes cardinalidades y produce una salida. En este caso específico es el encargado de unir la tabla ExpedienteLegal con la salida que produce el operador SET\_OPERATION.



- Finalmente se carga el STG\_COMPORTEAMIENTO con la salida que produce el operador JOINER.

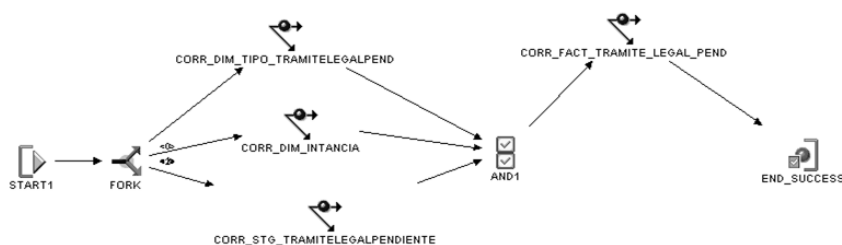
A continuación se especifican los pasos para cargar la FACT\_COMPORTEAMIENTO.



**Figura 39:** Carga de la FACT\_COMPORTEAMIENTO

- Se recibe como entrada los datos necesarios que se encuentran en las tablas:
- STG\_UBICACION, STG\_COMPORTEAMIENTO.
- Se procede a utilizar el operador JOINER, encargado de unir las diferentes filas de las tablas antes especificadas y producir una salida.
- El operador Key\_Lookup se utiliza para leer los datos de las dimensiones correspondientes a través de su clave o llave y produce de esta forma una salida. En este caso se obtienen los datos de las dimensiones interno, tiempo, comportamiento, centro, cargándose de esta manera los hechos en la FACT\_COMPORTEAMIENTO
- Para la carga de la FACT\_COMPORTEAMIENTO también se utilizan los datos que provee el operador INPUT\_PARAMETER.

## Proceso: Trámites Legales Pendientes

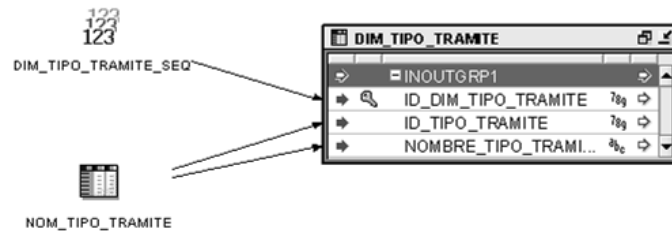


**Figura 40:** carga inicial al proceso Trámites Legales Pendientes

Las tareas que se llevan a cabo en este proceso son las siguientes:

- START1: actividad de inicio, encargada de que comience el flujo de actividades que componen el proceso.
- FORK: actividad encargada de iniciar actividades múltiples de forma simultánea.
- CORR\_DIM\_TIPO\_TRAMITELEGALPEND: ejecuta el contenedor de pasos que cargará la DIM\_TIPO\_TRAMITELEGALPEND, más adelante se detallará el mismo.
- CORR\_DIM\_INSTANCIA: ejecuta el contenedor de pasos necesarios para cargar la DIM\_INSTANCIA, más adelante se detalla el mismo.
- CORR\_STG\_ TRAMITELEGALPEND: ejecuta el contenedor de pasos necesarios para cargar el CORR\_STG\_ TRAMITELEGALPEND, más adelante se detalla el mismo.
- AND1: actividad encargada de controlar la correcta ejecución de la actividad siguiente, no permite que comience la ejecución de la misma sin que hayan terminado completamente las anteriores.
- CORR\_FACT\_TRAMITE\_LEGAL\_PEND: ejecuta el contenedor de pasos para cagar la FACT\_TRAMITE\_LEGAL\_PEND, más adelante se detalla el mismo.
- END\_SUCCESS: actividad final, indica que el proceso ha concluido con éxito.

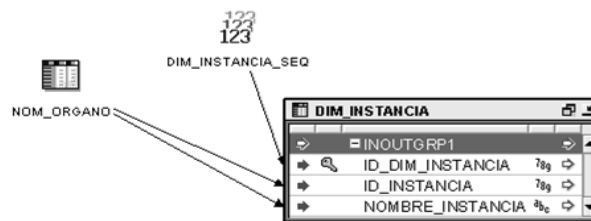
A continuación se especifican los pasos para cargar la DIM\_ TIPO\_ TRAMITELEGALPEND



**Figura 41:** Carga de la DIM\_TRAMITE

- En este proceso se toman como fuente de entrada para carga la dimensión los datos provenientes de la tabla Nom\_Tipo\_Tramite y se adiciona un id secuencial con el operador Dim\_Tipo\_Tramite\_SEQ.

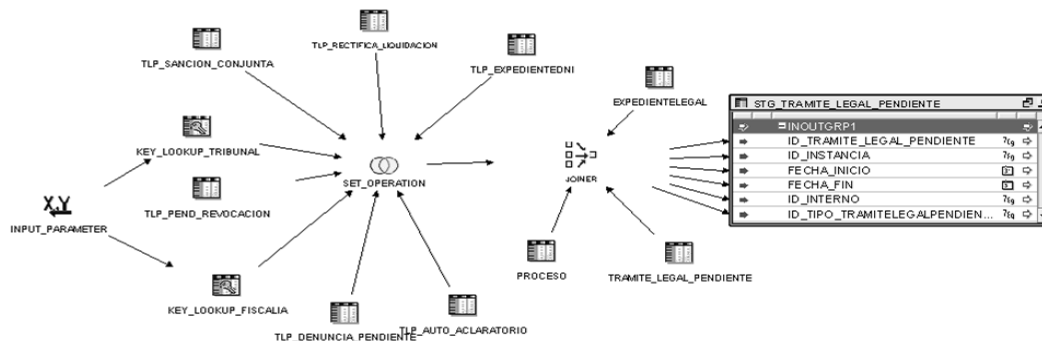
A continuación se especifican los pasos para cargar la DIM\_INSTANCIA



**Figura 42:** Carga de la DIM\_INSTANCIA

- En este proceso se toman como fuente de entrada para carga la dimensión los datos provenientes de la tabla Nom\_Organo y se adiciona un id secuencial con el operador Dim\_Instancea\_SEQ.

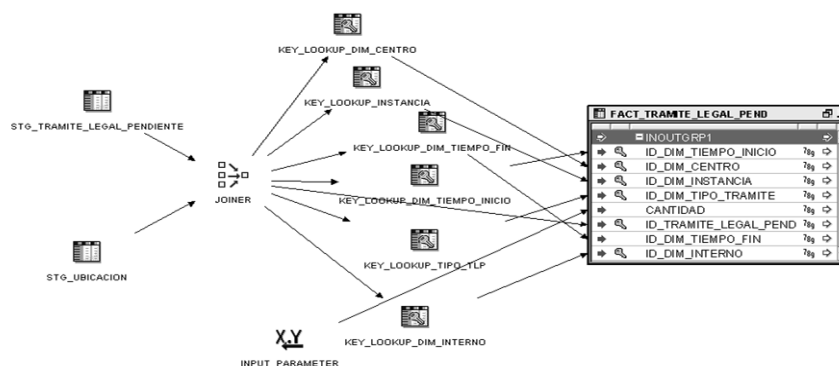
A continuación se especifican los pasos para cargar el STG\_TRAMITELEGALPEND



**Figura 43:** Carga del STG\_TRAMITE

- En el proceso de carga del STG\_ TRAMITELEGALPEND se procede a utilizar los operadores SET\_OPERATION, JOINER, INPUT\_PARAMETER y los KEY\_LOOKUP\_TRIBUNAL, KEY\_LOOKUP\_FISCALIA.
- Como tablas de entrada se utilizaran ExpedienteLegal, Tramite\_Legal\_Pendiente, Proceso, TLP\_ExpedienteNI, TLP\_Rectifica\_Liquidacion, TLP\_Sancion\_conjunta, TLP\_Pendiente\_Revocacion, TLP\_Denuncia\_Revocacion, Auto\_Aclaratorio.

A continuación se especifican los pasos para cargar la FACT\_TRAMITE\_LEGAL\_PEND

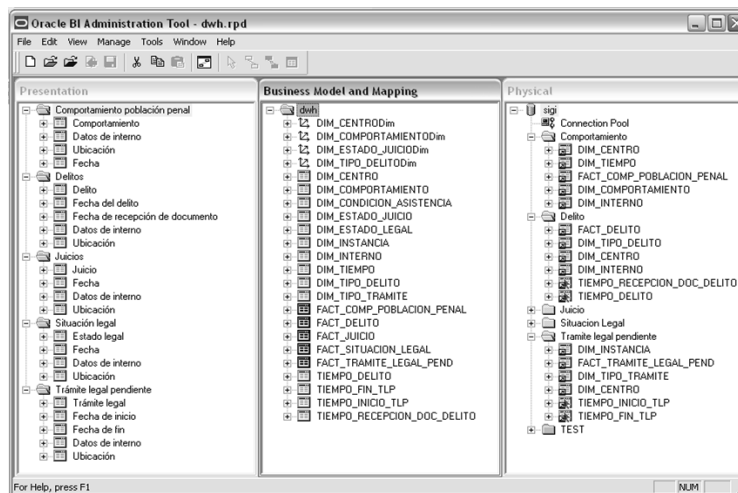


**Figura 44:** Carga de la FACT\_TRAMITE

- En el proceso de carga de la FACT se proceden a utilizar los operadores JOINER, INPUT\_PARAMETER y los KEY\_LOOKUP correspondientes (centro, instancia, tiempo\_fin, tiempo\_inicio, tipo\_tlp, interno).
- Como tablas de entradas se utilizan el STG\_TRAMITE\_LEGAL\_PENDIENTE y el STG\_UBICACION.

### 3.6 Configuración del Data Mart

Después de conformado el Almacén de Datos se procede a su configuración, para que pueda ser explotado por las herramientas de consulta. La herramienta utilizada para esta configuración es la herramienta de administración (Administration Tool).



**Figura 45:** Pantalla principal del Administration Tool

De manera general se importan las tablas del cubo correspondiente existente en el Almacén de Datos hacia la capa física de la herramienta de administración, donde se definen las restricciones de llave primaria y llaves foráneas que definen las relaciones del cubo. En la capa intermedia de mapeo y modelado del negocio es donde se define el modelo de los datos del negocio y se especifican las correspondencias entre el modelo de negocio y la capa física. Además se definen las jerarquías para la dimensiones; se especifican las medidas contenidas en las tablas de hechos con su respectiva regla de agregación y se eliminan los atributos que no se vayan a utilizar en el análisis. Como último paso se construye la capa de presentación, en ella se definen como serán vistos los datos por los usuarios finales a través de las herramientas de consulta. Esta capa permite organizar y renombrar las columnas en las carpetas, para un mejor entendimiento del usuario final.

Para el último paso de la etapa de implementación se utiliza el OBI, como herramienta cliente para la consulta y análisis de la información por parte de los usuarios. Primeramente se procede a la realización de los reportes más importantes para la DEP, así como la configuración de navegabilidad, implementación de los filtros predeterminados, formato requerido, esto se realiza para cada uno de los procesos.

A continuación se muestran los reportes asociados al proceso trámites legales pendientes, que no son más que la implementación de los indicadores definidos en el análisis.



**Figura 46:** Pantalla del Oracle Answers, asociada a los reportes del proceso Trámites Legales Pendientes

No obstante el usuario tiene la posibilidad de configurar nuevos reportes de forma dinámica cuando la necesidad de información lo requiera, también al seleccionar un reporte configurar su navegabilidad y crear filtros.

A continuación se muestra una pantalla de cómo se vería el ambiente de trabajo para el usuario que este configurando un reporte



**Figura 47:** Pantalla del Oracle Answers, ambiente de trabajo para configurar un reporte

En la figura 48 que se muestra a continuación se puede apreciar la navegabilidad por los reportes, en la pestaña del proceso correspondiente aparecen los que han sido predefinidos en formas de pequeñas tablas y gráficos



Figura 48: Oracle Interactive Dashboards, proceso Trámites Pegales Pendientes

### **Conclusiones**

En este capítulo se mostraron como quedaron las estructuras físicas del Almacén de Datos, se definió el tipo de estructura empleado y se crean los correspondientes esquemas lógicos. Además se llevo a cabo el modelado de los procesos ETL con la herramienta Desing Center. Se utilizo la herramienta de administración de Oracle para configurar el Almacén de Datos para que pudiera ser consultado por las herramientas Oracle Interactive Dashboard y Oracle Answers, en las cuales se implementaron los informes solicitados por los clientes.



### **CONCLUSIONES GENERALES**

- Se cumplió en el tiempo estimado, con los objetivos generales y específicos del mismo así como con las tareas investigativas.
- Se llegó a la conclusión de que el proceso de desarrollo de un Almacén de Datos resulta bastante complejo y trabajoso, ya que involucra grandes cantidades de información sobre la organización, así como de las herramientas con la que se trabaja.
- Se comprobó que la metodología Hefesto permite la construcción de un Almacén de Datos de forma sencilla, ordenada e intuitiva.
- Se establecieron los cinco procesos fundamentales en la especialidad Registro Legal del Sistema Penitenciario Cubano.
- Mediante el diseño e implementación de este Data Mart, se podrá proveer al Sistema Penitenciario Cubano, de una herramienta de Inteligencia de Negocio eficiente y dinámica que ayude al proceso de toma de decisiones.
- Al tratarse el desarrollo de este Data Mart como un proceso iterativo e incremental, se podrán incluir en el futuro nuevas necesidades, procesos de negocios y cambios que surjan en el Sistema Penitenciario Cubano.

### **RECOMENDACIONES**

- Se recomienda realizar los Data Mart asociados a las demás especialidades de la Dirección de establecimientos Penitenciarios.
- Implementar un Almacén de Datos central con los Data Mart creados.
- Valorar las nuevas necesidades de información que puedan presentar los usuarios.
- Motivar el estudio y desarrollo de los almacenes de datos en los proyectos de la Universidad.

### REFERENCIAS BIBLIOGRÁFICAS

1. **Del Risco Batista, Yanet. 2009.** *Proyecto Técnico.* 2009.
2. **Ing. Bernabeu, Ricardo Dario. 2007.** *Data wareHouse. Hefesto.* Córdoba, Argentina : s.n., 2007.
3. —. **2010.** *HEFESTO: Metodología para la Construcción de un Data WareHouse.* Córdoba, Argentina : s.n., 2010.
4. **Justicia, Mnisterio de.** Ministerio de Justicia. *Mnisterio de Justicia.* [En línea] [Citado el: 15 de enero de 2011.] <http://www.minjusticia-ddhh.gov.ec:8000/mjdh/admin/login.jsp>.
5. **Kimball, Ralph.** *The Data WareHouse ETL Toolkit, Practical Techniques for Extracting, Cleaning, Conforming and Delivering Data”.*
6. **Lujan Mora, Sergio. 2005.** *Data Warehouse Design with UML.* Universidad de Alicante, España : s.n., 2005.
7. **Sistema Penitenciario, Direccion General del.** Direccion General del sistema penitenciario. *Direccion General del sistema penitenciario.* [En línea] [Citado el: 18 de noviembre de 2010.] <http://www.sistemapenitenciario.gob.pa/>.

## BIBLIOGRAFÍA

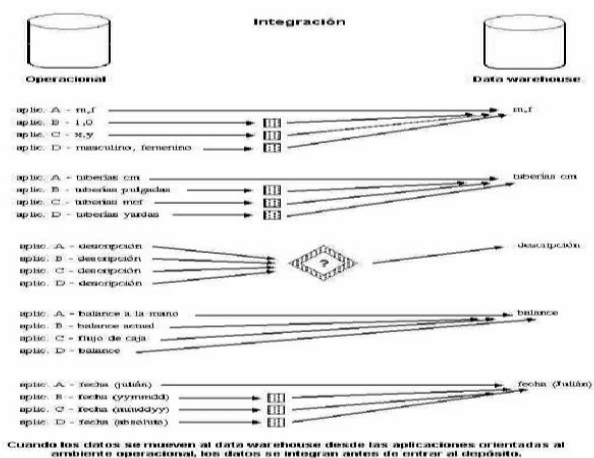
1. **Bravet Mertinez, René and Pedraja Lazo, David. 2010.** *Repositorio de Datos de la Sala Situacional del SIGEP, desde el punto de vista del Control Penal.* Ciudad Habana, Cuba : s.n., 2010.
2. **Cabrera Gutiérrez, Jose Manuel. 2010.** *Ingreso y Datos Generales.* 2010.
3. **Castelán García Leopoldo, Ocharán Hernández Jorge Octavio. 2010.** *Diseño de un Almacén de datos basado en Data Warehouse.* Veracruzana : s.n., 2010.
4. **Dario, Ing. Bernabeu Ricardo. 2010.** *HEFESTO: Metodología para la Construcción de un Data WareHouse.* Córdoba, Argentina : s.n., 2010.
5. **Del Risco Batista, Yanet. 2009.** *Proyecto Técnico.* 2009.
6. **Electrónica, Consejo superior de Administración.** Consejo superior de administración electrónica. *Consejo superior de administración electrónica.* [Online] [Cited: noviembre 20, 2010.] <http://www.csae.map.es>.
7. **Fernández Pedraza, Libisdey and Ramos Hernández, Alejandro. 2010.** *Diseño de un Almacén de Datos para apoyar la toma de decisiones en el Centro de Informática Médica.* Habana, Cuba : s.n., 2010.
8. **Gutiérrez Sánchez, Emilio José. 2009.** *Una propuesta para el desarrollo de Almacenes de Datos XML.* 2009.
9. **Ing. Bernabeu, Ricardo Dario. 2007.** *Data wareHouse. Hefesto.* Córdoba, Argentina : s.n., 2007.
10. **Inmon H, William. 2005.** *Building the Data Warehouse.* Indianapolis y Canadá : Wiley Publishing, 2005. Vol. 4ta Edición. Inc. 543.9780764599446..
11. **INTELLIGENCE, ORACLE BUSINESS.** *Creating a Repository Using the Oracle Business Intelligence Administration Tool.*
- 12.—. *Creating Interactive Dashboards and.*
13. **Justicia, Mnisterio de.** Ministerio de Justicia. *Mnisterio de Justicia.* [Online] [Cited: enero 15, 2011.] <http://www.minjusticia-ddhh.gov.ec:8000/mjdh/admin/login.jsp>.
14. **Kimball, Ralph. 1996.** *El Juego de Herramientas del Almacén de Datos.* New York : John Wiley & Sons, 1996. Inc. 388 p.0471153370.
- 15.—. *The Data WareHouse ETL Toolkit, Practical Techniques for Extracting, Cleaning, Conforming and Delivering Data”.*

16. **Lu Hernández, Jorge Raú. 2010.** *Manual Técnico*. Guatemala : s.n., 2010.
17. **Lujan Mora, Sergio. 2005.** *Data Warehouse Design with UML*. Universidad de Alicante, España : s.n., 2005.
18. **Oracle.** [Online] <http://www.oracle.com/index.html>.
- 19.—. **2004.** *Analytic Workspace Manager and Oracle OLAP 10g*. 2004.
- 20.—. **2006.** *Oracle® Business Intelligence Server Administration Guide*. 2006.
- 21.—. **2009.** *Oracle® Warehouse Builder*. 2009.
22. **Pedraza Fernández, Libisdey and Hernández Ramos, Alejandro. 2010.** *Diseño de un Almacén de Datos para apoyar la toma de decisiones en el Centro de Informática Médica*. Ciudad Habana, Cuba : s.n., 2010.
23. **Ricardo Dario, Bernabeu. 2010.** *DATA WAREHOUSING: Investigación y Sistematización de Conceptos - HEFESTO: Metodología propia para la Construcción de un Data Warehouse. V2.0*. Córdoba, Argentina : s.n., 2010.
24. **Sistema Penitenciario, Direccion General del.** Direccion General del sistema penitenciario. *Direccion General del sistema penitenciario*. [Online] [Cited: noviembre 18, 2010.] <http://www.sistemapenitenciario.gob.pa/>.

**ANEXOS**



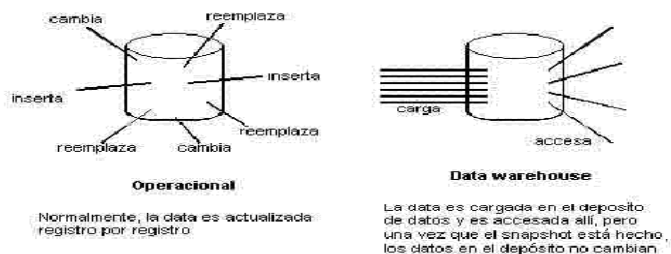
**Figura 1:** Contraste entre la orientación a las aplicaciones y la orientación a temas



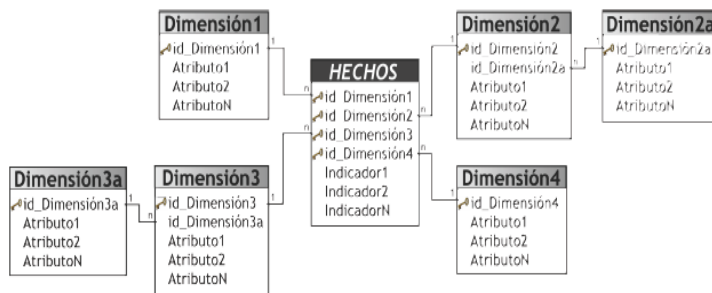
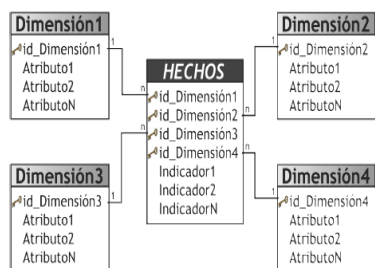
**Figura 2:** Contraste de integración entre el ambiente operacional y el ambiente del depósito de datos.



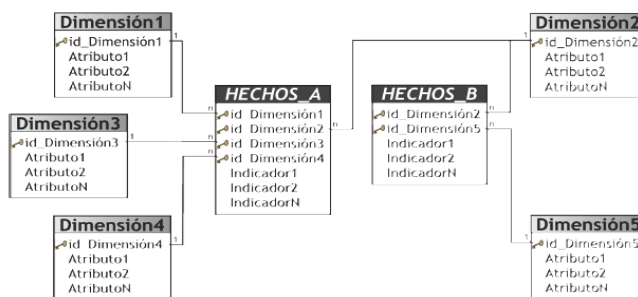
**Figura 3:** Contraste de la característica de variación de tiempo entre el ambiente operacional y el ambiente de depósito de datos.



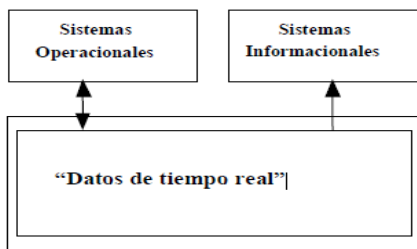
**Figura 4:** Contraste de la característica no volátil entre el ambiente operacional y el ambiente de depósito de datos.



**Figura 5:** Esquema en estrella (Star Scheme) **Figura 6:** Esquema copo de nieve (Snowflake Scheme)



**Figura 7:** Esquema constelación o copo de estrellas (Starflake Scheme):



**Figura 8:** Arquitectura de datos de una sola capa.



**Figura 9:** Arquitectura de datos de dos capas.

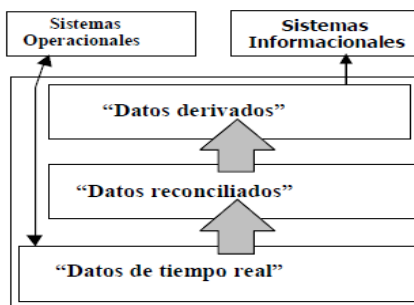


Figura 10: Arquitectura de datos de 3 capas

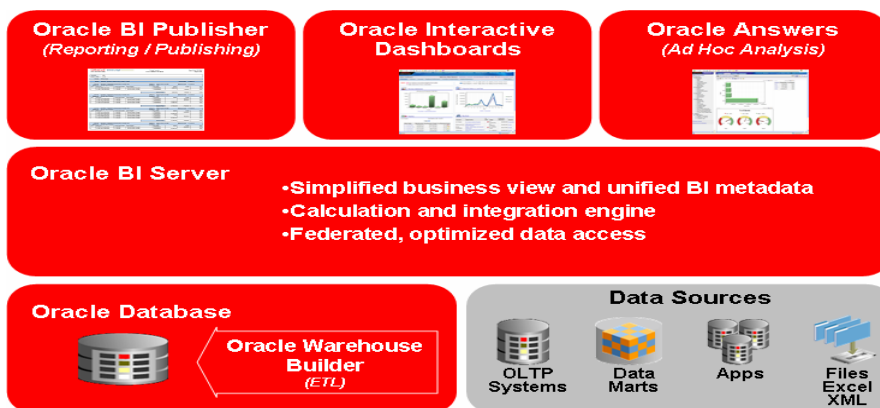


Figura 11: Arquitectura suite de Oracle

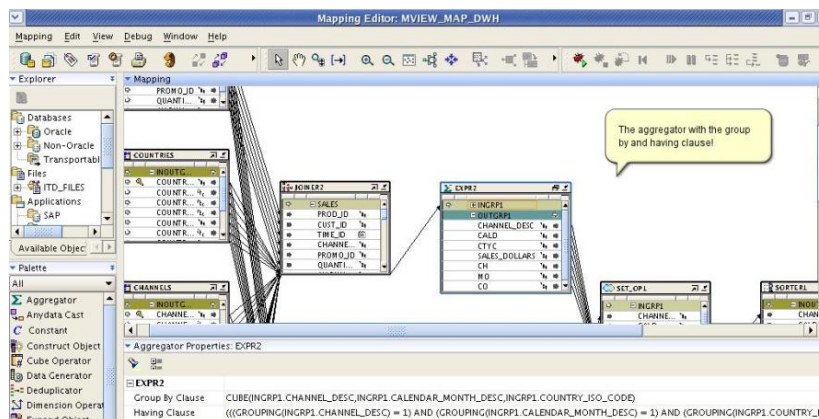


Figura 12: Oracle WareHouse Builder



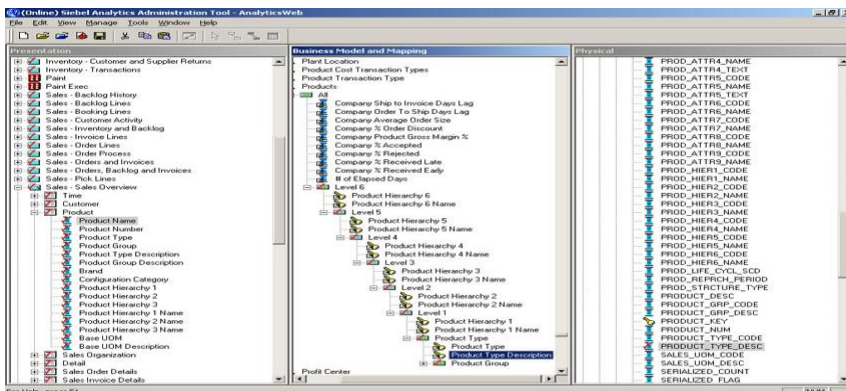


Figura 13: Administration Tool



Figura 14: Oracle BI Answers

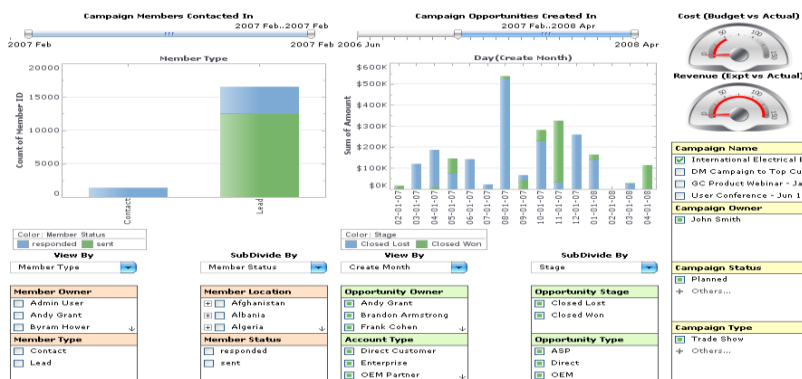


Figura 15: Oracle Business Intelligence Interactive Dashboards