

Universidad de las Ciencias Informáticas



Facultad 2

Título: Desarrollo de un Mercado de Datos para el Área de Ingreso.

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

Autor(es): Yoraima Peña Hinojosa

Fernando García Pérez

Tutor: Ing. Yanet del Risco Batista

Co-Tutor: Krysia Llull Céspedes

Ciudad de La Habana, 2012

“Año 54 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 2012.

Yoraima Peña Hinojosa

Firma del Autor

Fernando García Pérez

Firma del Autor

Ing. Yanet del Risco Batista

Firma del Tutor

RESUMEN

El presente trabajo tiene como objetivo principal desarrollar un Mercado de Datos para el Área de Ingreso del Sistema Penitenciario Cubano, el cual permite realizar análisis estadísticos con la información que se encuentra en los centros penitenciarios. Esta solución le facilita a los directivos del área la realización de los cierres trimestrales, además permite tener un amplio conocimiento del estado en que se encuentran los principales indicadores del Área de Ingreso para realizar el proceso de toma de decisiones. La construcción del mercado está basado en el ciclo de vida dimensional del negocio que plantea Ralph Kimball en su metodología, realizando las tareas que propone en cada etapa de desarrollo, siendo la de mayor labor el proceso de Extracción, Transformación y Carga de los datos (ETL por sus siglas en inglés). La arquitectura conceptual de datos está conformada por tres capas, las cuales representan los diferentes entornos por los que pasan los datos en su camino hacia el mercado. Para la elaboración del mismo se utilizó como herramientas, el Sistema Gestor de Base de Datos Oracle 11g, el Oracle Warehouse Builder para realizar el proceso ETL, el Business Intelligence Answer para la creación de los reportes, el Business Intelligence Intercative Dashboard para la configuración de los cuadros de mando y el Embarcadero ER Studio para el modelado.

Palabras Claves: Sistema Penitenciario, Área de Ingreso, Mercado de Datos, Kimball

ÍNDICE

INTRODUCCIÓN.....	1
CAPÍTULO 1: MARCO TEÓRICO DE LA INVESTIGACIÓN	7
Introducción.....	7
1.1 Sistema Penitenciario	7
1.1.1 Tendencias actuales de los Sistemas Penitenciarios.....	7
1.2 Almacén de Datos.....	9
1.2.1 Los principales objetivos de un almacén de datos son:	10
1.2.2 Las principales características de un almacén de datos son	11
1.2.3 Ventajas de los Almacenes de Datos	13
1.2.4 Mercado de Datos	13
1.3 Modelado Multidimensional	14
1.4 Bases de Datos OLTP y OLAP	15
1.4.1 Sistemas OLTP- On-Line Transactional Processing	15
1.4.2 Sistemas OLAP- On-Line Analytical Processing	16
1.5 Cubos multidimensionales.....	16
1.6 Esquemas Dimensionales	17
1.7 Tipo de Organización Física.....	18
1.8 Métodos más usados en la Construcción de Almacenes de Datos	19
1.9. Arquitectura de un Almacén de Datos.....	19
1.9.1 Arquitectura conceptual de los datos del Mercado de Datos	19
1.10 Metodologías de desarrollo de un Almacén de Datos	21
1.10.1 Metodología de Ralph Kimball – Ciclo de Vida dimensional	21
1.11 Herramientas de creación de un Almacén de Datos	23
Conclusiones Parciales	25
CAPÍTULO 2: ANÁLISIS Y DISEÑO DEL MERCADO DE DATOS	26
Introducción.....	26
2.1 Identificación de Temas Analíticos	26
• Técnicas de Identificación de Requisitos	26
• Área de Ingreso.....	27

• Descripción de los Procesos por Tema Analítico:.....	28
2.1.1 Análisis de Requisitos:.....	29
• Requisitos de Información	29
• Especificación de Requisitos	34
• Usuarios del Sistema	39
• Técnicas de Validación de Requisitos	39
2.2 Modelado Dimensional	40
• Granularidad.....	40
• Tablas Dimensiones:.....	40
• Tablas de Hechos	41
• Matriz Dimensional.....	41
• Patrones de Diseño.....	42
• Modelo de Datos	43
2.3 Diseño de las ETL.....	45
Conclusiones Parciales	49
CAPÍTULO 3: IMPLEMENTACIÓN Y PRUEBA.....	50
Introducción.....	50
3.1 Implementación.....	50
3.1.1 Implementación de la Extracción Transformación y Carga (ETL).....	50
3.1.2 Configuración de la herramienta de Administración	62
3.1.3 Construcción de reportes en la herramienta de Respuesta	65
3.1.4 Configuración del Tablero de Mando	68
3.1.5 Tipos de Procesamiento.....	69
3.2 Pruebas	69
• Descripción de los métodos de prueba a realizar	70
• Pruebas de Caja Negra o Funcionales	70
• Lista de Chequeo	71
3.2.1 Resultados de las Pruebas.....	72
CONCLUSIONES	75

INTRODUCCIÓN

En la era actual con el amplio y novedoso desarrollo de las tecnologías y las comunicaciones, se ha incrementado sustancialmente la velocidad del proceso de gestionar la información. Desde un inicio, las organizaciones utilizan las bases de datos (BD) relacionales para almacenar los datos generados en las operaciones transaccionales de su actividad diaria, lo cual contribuyó a que se convirtieran en una herramienta fundamental. En el transcurso de los años las organizaciones han acumulado un gran volumen de información debido a todas las operaciones realizadas en este tiempo, por lo que en la actualidad le han prestado un mayor interés a buscar alternativas para sacar un mejor provecho a la información almacenada, que les permita tomar decisiones a partir de un amplio conocimiento del estado en el cual se encuentra la organización. Con el objetivo de lograr que el proceso de toma de decisiones esté fundamentado sobre un minucioso análisis de la información, es necesario realizar algunas transformaciones para convertirlas en conocimiento, y de esta forma las empresas e instituciones pueden trazar sus metas de acuerdo a los resultados generados.

La aplicación de las soluciones de Inteligencia de Negocios (Business Intelligence, por sus siglas en inglés BI) le brinda a las organizaciones la posibilidad de desarrollarse objetivamente, al contar con la información exacta del estado en que se encuentra hasta ese momento. Con la implementación de BI las empresas pueden traducir sus objetivos en indicadores, al proporcionar información privilegiada para responder a los problemas de negocio, a través de los procesos de recolección y procesamiento de la información que sustentan la toma de decisiones, generando de esta forma una potencial ventaja competitiva. Sus principales productos en la actualidad son los Sistemas de Información Ejecutiva (Executive Information Systems – EIS), los Sistemas para la Toma de Decisiones (Decision Support Systems – DSS), y los Cuadros de Mando Integrales (CMI), que utilizan como principales componentes de orígenes de datos a los almacenes o mercado de datos.

Con el objetivo de lograr un efectivo análisis y procesamiento de la información, se hace necesario utilizar un Almacén de Datos (Data Warehouse, DW por sus siglas en inglés), al constituir uno de los soportes fundamentales para el proceso de toma de decisiones gerenciales. Las aplicaciones para soporte de decisiones basadas en DW pueden hacer más práctica y fácil la explotación de la información, ya que permiten la integración y centralización de los datos generados por una organización en todos los ámbitos de su actividad diaria. Organiza y almacena los datos necesarios para el procesamiento analítico e

informático sobre una amplia perspectiva de tiempo, además de poder adaptarse a los cambios que pueden ocurrir en el entorno del negocio.

El año 1989 marca para Cuba el inicio de la automatización e informatización del Sistema Penitenciario a nivel nacional, con un sistema de control penal, el cual permitía la gestión de los principales datos de los reclusos y algunos aspectos de la gestión penitenciaria. Años después en cumplimiento de la orden 43/99 del Viceministro Primero del Interior se crea el sistema para el control del recluso (SACORE), terminando su desarrollo en el año 2002 y comenzando su explotación en el 1er trimestre del año 2003. En la actualidad el sistema cuenta con tres módulos principales: Control Penal, Tratamiento Educativo y Orden Interior. A pesar de todas las facilidades que brinda todavía existen deficiencias en dichos módulos. (1)

A partir del Plan 20 X 50, para la modernización tecnológica del Ministerio del Interior (MININT), en colaboración con la Universidad de las Ciencias Informáticas (UCI), se comienza con el proyecto Prisiones Cuba. El proyecto tiene como objetivo principal unir los tres sistemas existentes, el sistema para el control del recluso (SACORE), el sistema de incidencias (SAIDEP) y el de capacidades (SACDEP), en la Dirección de Establecimientos Penitenciarios (DEP) y añadir un grupo de funcionalidades necesarias para lograr un buen funcionamiento de los procesos en los Centros Penitenciarios (CP). (1)

Los Centros Penitenciarios de Cuba cuentan con un gran volumen de información de todos los procesos que se manejan en las distintas áreas o especialidades, generada en el trayecto de todos los años de servicios, la cual debe seguir en aumento en años venideros. Esta información resulta ser muy útil si es utilizada con inteligencia y oportunidad para la retroalimentación de los centros, de modo que los directivos puedan tomar decisiones más oportunas. Actualmente los directivos de la DEP tienen dificultad para realizar el análisis de toda la información acumulada, porque no cuentan con una adecuada estructura organizativa que les permita generar reportes y estadísticas sobre los indicadores más relevantes, a partir de los cuales se toman decisiones importantes. Además no tienen la cantidad de especialistas necesarios para generar los reportes de las distintas áreas en los cierres trimestrales.

La realización de estos cierres se toman aproximadamente 15 días, para poder monitorear y generar todos los reportes de los indicadores más importantes de cada especialidad. Esto se realiza a través de consultas a la aplicación web, insertando los datos obtenidos en una plantilla Excel para posteriormente generar los gráficos y tablas asociados a las mismas, este es sin duda un proceso muy engorroso por lo que se pueden pasar por alto algunos detalles que puedan perjudicar el resultado final.

Las Áreas de Ingreso, Observación, Evaluación y Diagnóstico son una unidad organizativa y en su actividad aplican tecnologías, procedimientos de trabajo y especialidades del sistema penitenciario. Es la encargada de estudiar la personalidad del interno teniendo en cuenta el carácter complejo, dinámico, constante y continuo de su evolución (2), además de la relación directa del diagnóstico de personalidad criminal con el juicio del pronóstico inicial. Al final el equipo de trabajo debe fundamentar su propuesta mediante el carácter colegiado en la toma de decisión para la clasificación y ubicación del interno.

Toda la información manejada en el Área de Ingreso es de la gestión operativa de los establecimientos penitenciarios, motivo por el cual la DEP necesita generar las estadísticas lo más exactas posible, con el objetivo de lograr que el proceso de toma de decisiones esté bien fundamentado.

A partir de todo lo anteriormente planteado surge como **problema a resolver** ¿Cómo facilitar la toma de decisiones de los especialistas de dirección sobre el Área de Ingreso del Sistema Penitenciario Cubano?

Para llevar a cabo la investigación se plantea como **objeto de estudio** los procesos para la toma de decisiones del Sistema Penitenciario Cubano y como **objetivo general** Desarrollar un Mercado de Datos para apoyar el proceso de toma de decisiones de los directivos sobre el Área de Ingreso del Sistema Penitenciario Cubano. La investigación tiene como **campo de acción** los procesos para la toma de decisiones en el Área de Ingreso en el Sistema Penitenciario Cubano. La **idea a defender** planteada para la presente investigación es: el desarrollo de un mercado de datos para el análisis de la información referente al Área de Ingreso del Sistema Penitenciario Cubano facilita el proceso de toma de decisiones en dicha área.

El **objetivo general** está dividido en los siguientes **objetivos específicos**:

1. Elaborar el Marco Teórico de la Investigación.
2. Analizar los Requisitos del mercado de datos para el Área de Ingreso.
3. Realizar el Modelado Dimensional del mercado de datos para el Área de Ingreso.
4. Realizar la Extracción, Transformación y Carga de los datos del mercado de datos para el Área de Ingreso.
5. Desarrollar el Tablero de Mando del mercado de datos para el Área de Ingreso.
6. Realizar las pruebas al mercado de datos del Área de Ingreso.

Los objetivos específicos anteriormente planteados abarcan un grupo de tareas para darles cumplimiento, que a continuación se detallan:

1. Elaborar el marco teórico de la investigación
 - Estudio del estado del arte
 - Definición de conceptos fundamentales de los almacenes de datos
 - Descripción de la arquitectura y la metodología de desarrollo
 - Descripción de las herramientas a utilizar en el desarrollo del mercado de datos
2. Analizar los requisitos del mercado de datos para el Área de Ingreso
 - Identificación de temas analíticos del Área de Ingreso
 - Identificación de los procesos de negocio del Área de Ingreso
 - Identificación de los requisitos de información del mercado de datos para el Área de Ingreso
 - Priorización de los procesos del Área de Ingreso
3. Realizar el modelado dimensional del mercado de datos para el Área de Ingreso
 - Identificación de los niveles de granularidad del mercado de datos para el Área de Ingreso
 - Identificación de las dimensiones y medidas del mercado de datos para el Área de Ingreso
 - Identificación de atributos de dimensiones y tablas de hechos
 - Construcción de la Matriz Procesos-Dimensiones
 - Diseño del modelo dimensional detallado
 - Revisión y validación del modelo dimensional
4. Realizar la extracción, transformación y carga (ETL por sus siglas en inglés) de los datos del mercado de datos para el Área de Ingreso
 - Identificación de los datos que serán cargados de la fuente origen
 - Diseño de las ETL del mercado de datos para el Área de Ingreso
 - Implementación de las ETL del mercado de datos para el Área de Ingreso

5. Desarrollar el tablero de mando del mercado de datos para el Área de Ingreso
 - Configuración del repositorio de datos en la herramienta de administración
 - Implementación de los reportes del mercado de datos para el Área de Ingreso
 - Diseño del tablero de mando del mercado de datos para Área de Ingreso
6. Realizar las pruebas al mercado de datos para el Área de Ingreso
 - Selección de las técnicas de pruebas realizadas al mercado de datos para el Área de Ingreso
 - Realización de las pruebas al mercado de datos para el Área de Ingreso
 - Recolección y clasificación de las no conformidades
 - Aplicación de las pruebas de regresión

Posibles resultados:

- Un Mercado de Datos para el Área de Ingreso
- Tableros de Mando, Histórico y Operativo para el Área de Ingreso

Durante la investigación se utilizaron los siguientes **Métodos Científicos:**

Métodos Teóricos:

- Analítico Sintético: Análisis de la documentación existente sobre los procesos que se realizan en el Área de Ingreso, desde las leyes y normativas como metodologías, además de la bibliografía sobre la construcción de los almacenes de datos y sintetizar los elementos necesarios a utilizar en el desarrollo del trabajo.
- Histórico Lógico: Se hace un estudio del estado del arte sobre las soluciones informáticas que son utilizadas en los Sistemas Penitenciarios de otros países, y se analiza cómo realizan el proceso de análisis de la información para la toma de decisiones, con el objetivo de ver cuál puede ser útil para la realización de este trabajo.

Métodos Empíricos:

- Entrevistas: Se realizan un conjunto de entrevistas cerradas con el cliente para identificar los procesos de negocio que se realizan en el Área de Ingreso, los cuales son posteriormente agrupados por temas analíticos. Además se realiza la captura de los requisitos de información necesarios para darle respuesta a las necesidades del cliente.

La presente investigación, está estructurada en tres capítulos, los cuales son:

Capítulo 1. “Marco Teórico de la Investigación” se realiza un estudio sobre soluciones informáticas similares existentes en Sistemas Penitenciarios de varios países. Además se analizan los principales elementos teóricos que constituyen la base de la investigación, entre los cuales se encuentran la metodología de desarrollo para el mercado de datos y las herramientas utilizadas en la propuesta de solución.

Capítulo 2. “Análisis y Diseño del Mercado de Datos” se definen los temas analíticos existentes y los procesos que los componen. Se realiza la especificación de los requisitos de información y el modelado dimensional para los procesos del caso de estudio. También forma parte del capítulo la matriz dimensional, y el diseño del proceso ETL.

Capítulo 3. “Implementación y Prueba del Mercado de Datos” se describe la implementación del proceso ETL de los procesos descritos en el capítulo anterior. También se detalla la configuración del repositorio de datos en la herramienta de administración, para la realización de los reportes que posteriormente conformarán el tablero de mando. Además se describe cómo se llevan a cabo las pruebas para validar la propuesta de solución.

CAPÍTULO 1: MARCO TEÓRICO DE LA INVESTIGACIÓN

Introducción

En el presente capítulo se realiza un estudio de las diferentes aplicaciones informáticas en sistemas penitenciarios de diferentes países, incluyendo a Cuba. Además se explican conceptos, características, objetivos y ventajas de los almacenes de datos, y su estructura más pequeña el Mercado de Datos. Se hace un análisis de los diferentes enfoques que hay en la actualidad referente a la construcción de los almacenes de datos, y además se fundamenta las tecnologías que se utilizarán para el desarrollo de este trabajo.

1.1 Sistema Penitenciario

El 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, dirigido 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. (3)

Área de Ingreso

Las Áreas de Ingreso, Observación, Evaluación y Diagnóstico son unidades organizativas independientes, subordinadas al Jefe del Centro de Internamiento. Como parte de su actividad aplican tecnologías, procedimientos de trabajo y especialidades del sistema penitenciario, con el objetivo de garantizar los siguientes procesos: Ingreso, Características Socio-Antropológicas, Evaluación Criminológica, Clasificación en Régimen y Ubicación. (2)

La Evaluación Criminológica no se tiene en cuenta en la realización del Mercado de Datos, por no encontrarse modelado este proceso en la fuente de origen, pero se le agrega el proceso Vínculo debido a la importancia que tiene para el Sistema Penitenciario.

1.1.1 Tendencias actuales de los Sistemas Penitenciarios

El desarrollo de las nuevas tecnologías, ha propiciado que diferentes organizaciones basen su desempeño corporativo en el uso de soluciones informáticas, que sean capaces de soportar el proceso

de la toma de decisiones, así como la gestión de la información que se llevan a cabo dentro de la misma.

Para la realización de este trabajo se analizaron un grupo de soluciones informáticas, que son utilizadas en los Sistemas Penitenciarios de varios países, para llevar a cabo las operaciones diarias que se realizan en los mismos. Se tiene en cuenta además, cómo se procesa y analiza la información para la generación de informes, que den soporte al proceso de la toma de decisiones. A continuación se describen estas soluciones, con el correspondiente análisis:

- **Sistema de Gestión Penitenciaria de Panamá**

El sistema fue creado para la gestión penitenciaria, en el año 1997 gracias a un proyecto de colaboración financiado por las Naciones Unidas y el Gobierno de España, con la mediación de la Dirección General de Sistemas Penitenciarios de Panamá (antes Dirección Nacional de Corrección). Este sistema permite básicamente almacenar en una base de datos los registros de los internos que se encuentran detenidos en los centros penitenciarios a nivel de nación y generar estadísticas e informes para la Dirección General de Sistemas Penitenciarios de Panamá (DGSP). El sistema solo almacena la información referente al interno mientras se encuentra en el centro, al obtener la libertad, su expediente es eliminado, afectando los informes y estadísticas históricas que se puedan generar para la toma de decisiones. (4)

- **Sistema de Gestión Penitenciaria de Ecuador (e-SIGPEN)**

El Sistema de Gestión Penitenciaria de Ecuador (e-SIGPEN) facilita el control y manejo de los procesos que se realizan en las áreas de los Centros de Rehabilitación Social. El mismo permite a los directivos del Ministerio de Justicia del país y otras instituciones monitorear las actividades y obtener datos estadísticos reales de la situación en que se encuentran los internos. Todo ello permite llevar un control adecuado de la información de los privados de libertad y tomar decisiones eficaces y oportunas para mejorar el ambiente carcelario del país. (5)

- **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. (6)

El sistema cuenta con una Sala Situacional la cual es la encargada de analizar el comportamiento de los principales indicadores de las áreas Control Penal, Registro y Control de Visitantes, Observación e Historia Clínica del SIGEP. Se analiza la información histórica y operativa para la realización de los reportes, los cuales se muestran a través de gráficos ya que facilitan la comprensión y el análisis de los datos mostrados. La fuente de información para la Sala Situacional es la BD centralizada del SIGEP, lo cual trajo consigo afectaciones en su rendimiento a la hora de mostrar determinados reportes. (8)

Por decisión propia del Proyecto Humanización Penitenciaria se realizaron dos mercados de datos como fuente de información para la Sala Situacional, los cuales muestran los reportes a través de tablas dinámicas y gráficos. Estas soluciones nunca entraron en explotación porque no se encontraban dentro del contrato.

- **Sistema Automatizado para el Control del Recluso (SACORE)**

El sistema entra en explotación en el año 2003, con el objetivo de darle respuestas inmediatas a las solicitudes de información de los diferentes órganos e instituciones sobre el comportamiento de los reclusos en las áreas: Control Penal, Tratamiento Educativo y Orden Interior. El sistema tiene 9 años de explotación, por lo cual actualmente cuenta con un gran cúmulo de información, de los distintos procesos que se llevan a cabo dentro del centro.

La información almacenada es procesada para emitir reportes en los cierres trimestrales, por parte de los especialistas de dirección que atienden las distintas áreas. Para generar los reportes los especialistas sin ser expertos deben de realizar consultas a la aplicación, y los datos obtenidos transportarlos a herramientas de cálculo, como el Excel, para posteriormente generar las tablas y los gráficos que finalmente conforman los reportes.

Este proceso es muy complejo, y consume mucho tiempo para realizar los cierres, los cuales son los que orientan a los directivos para llevar a cabo el proceso de toma de decisiones.

Se deduce de los sistemas analizados que la mejor experiencia para el análisis de la información, por las facilidades que ofrece, es la empleada por el SIGEP con la utilización de los almacenes de datos.

1.2 Almacén de Datos

Las personalidades que han trascendido en la historia por sus grandes aportes en este tema son Ralph Kimball, considerado el principal promotor del enfoque dimensional, y Bill Inmon reconocido como el

Capítulo 1: Marco Teórico de la Investigación

padre del concepto de los Almacenes de Datos, al ser el primero en escribir un libro sobre este tema, y es el que le da la definición:

“Los *almacenes de datos* o Data Warehouse (DW) por su nombre en inglés, es una colección de datos orientados a temas, integrados, no-volátiles y variante en el tiempo, organizados para soportar necesidades empresariales”.

Un almacén de datos se crea al extraer datos desde varias fuentes, que pueden ser una o más bases de datos operacionales. Los datos extraídos son transformados para eliminar inconsistencias y resumir si es necesario, para luego ser cargados en el almacén.

En la actualidad es el centro de atención para las grandes empresas, porque les proveen un ambiente favorable, al permitirles darle un mejor uso a la información que está siendo administrada por diversas aplicaciones operacionales.

Un almacén de datos se puede caracterizar partiendo de una comparación, de cómo los datos de una institución, al almacenarlos en un almacén de datos, difieren de los datos operacionales usados por las aplicaciones de producción, donde las instituciones hacen sus operaciones cotidianas de los procesos que se realizan en la misma. (8)

Base de Datos Operacional	Almacén de datos
Datos Operacionales	Datos del negocio para Información
Orientado a la aplicación	Orientado al sujeto
Actual	Actual + histórico
Detallada	Detallada + más resumida
Cambia continuamente	Estable

Tabla 1: Diferencias en el almacenamiento de datos entre BD Operacionales y Almacén de datos

1.2.1 Los principales objetivos de un almacén de datos son:

1. Proporcionar una vista total e integrada de la empresa.
2. Permitir que la información actual e histórica de la empresa esté disponible de manera práctica para el análisis.

3. Permitir ejecutar las operaciones de soporte a las decisiones sin afectar los sistemas operacionales.
4. Garantizar la consistencia de la información.
5. Constituir una fuente flexible e interactiva de información estratégica.

1.2.2 Las principales características de un almacén de datos son:

- Orientado a temas
- No volátil
- Variante en el tiempo
- Integrado

Orientado a temas: La información es clasificada de acuerdo a los aspectos que son más importantes para la empresa. Existe una gran diferencia en la forma de diseñar la estructura para almacenar los datos en un DW con respecto a los clásicos procesos operacionales que están orientados a las aplicaciones. Los procesos que se llevan a cabo mediante las aplicaciones, responden al manejo de todos los datos de las operaciones realizadas en la empresa, mientras que un almacén de datos sólo tiene en cuenta aquella información de interés para una oportuna toma de decisiones, por ejemplo, para una institución financiera el ambiente operacional se diseña a través de funciones tales como préstamos, ahorros, tarjeta bancaria y depósitos, porque son muy importantes para sus operaciones productivas, sin embargo no se toma en cuenta en un almacén de datos ya que carecen de valor para la toma de decisiones, en este caso un ambiente de almacén de datos se organiza teniendo en cuenta la información que se relacione con el cliente, vendedor, producto y actividad. La principal ventaja que tiene el DW al estar orientado a temas es que los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. (8)

Variante en el tiempo: Esta característica básica de un almacén de datos es muy diferente de la información que se encuentra en un ambiente operacional, en el cual los datos se obtienen en el momento de acceder. En un almacén de datos como la información es solicitada en cualquier momento, los datos encontrados en el depósito se llaman de "tiempo variante". Esto es precisamente una de las principales ventajas de un DW, los datos que se almacenan tienen asociado una variable tiempo, esta cualidad no se encuentra en las bases de datos operacionales, ya que simplemente los

Capítulo 1: Marco Teórico de la Investigación

datos se insertan sin tener, necesariamente, en cuenta una variable tiempo. Además de que el almacén de datos al contar con el sello del tiempo te permite acceder a la data histórica de la información que almacena, y en una base de datos operacional los datos históricos son de poco uso. El almacenamiento de datos históricos, es lo que permite al DW desarrollar pronósticos y análisis de tendencias y patrones, a partir de una base estadística de información. (8)

El tiempo variante se puede ver de diversas maneras:

1° La información representa los datos sobre un horizonte largo de tiempo (de 5-10 años). El horizonte de tiempo representado para el ambiente operacional es mucho más corto (desde valores actuales hasta 60-90 días), debido al diseño de aplicaciones rígidas.

2° 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 o mes. 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.

3° 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. La información del almacén de datos es, para todos los propósitos prácticos, una serie larga de "snapshots" (vistas instantáneas). En caso de que los snapshots de los datos se han tomado incorrectamente, entonces pueden ser cambiados, pero de no ser así, ellos no son alterados una vez hechos.

Integrado: Los datos almacenados en el almacén de datos deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas. 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 consistentes, en atributos físicos de los datos consistentes, fuentes múltiples y otros. Cualquiera que sea la forma del diseño, el resultado es el mismo: la información necesita ser almacenada en el almacén de datos en un modelo globalmente aceptable y singular, aun cuando los sistemas operacionales subyacentes almacenen los datos de manera diferente. (8)

No volátil: Es necesario para la toma de decisiones que la información que se utilice para ello sea estable. Los datos en los procesos operacionales cambian de un momento a otro mientras que en un DW una vez almacenados no se puede cambiar. La actualización (insertar, borrar y modificar), se hace

regularmente en el ambiente operacional sobre una base de registro por registro, en este aspecto también difiere el DW, ya que la manipulación sobre los datos que almacena es mucho más simple porque solamente puede hacer dos tipos de operaciones: la carga inicial de los datos y el acceso a los mismos. La mayoría de los datos que se extraen de la fuente se alteran física y radicalmente cuando se mueven al depósito, no es la misma data que reside en el ambiente operacional desde el punto de vista de integración. (8)

1.2.3 Ventajas de los Almacenes de Datos

Entre las ventajas de los almacenes de datos se encuentran: (8)

- Transforma datos orientados a las aplicaciones en información orientada a la toma de decisiones.
- Integra y consolida diferentes fuentes de datos (internas y/o externas) y departamentos empresariales, que anteriormente formaban islas, en una única plataforma sólida y centralizada.
- Provee la capacidad de analizar y explotar las diferentes áreas de trabajo y de realizar un análisis inmediato de las mismas.
- Permite reaccionar rápidamente a los cambios del mercado.
- Aumenta la competitividad en el mercado.
- Elimina la producción y el procesamiento de datos que no son utilizados ni necesarios, producto de aplicaciones mal diseñadas o ya no utilizadas.
- Mejora la entrega de información, es decir, información completa, correcta, consistente, oportuna y accesible. Información que los usuarios necesitan, en el momento adecuado y en el formato apropiado.

1.2.4 Mercado de Datos

Un Mercado de Datos o Data Mart (DM) por su nombre en inglés, es la implementación de un DW con alcance restringido a un área funcional, problema en particular, departamento, tema o grupo de necesidades en el negocio. (8) Un DM puede ser alimentado desde los datos de un Almacén de datos, o integrar por sí mismo un compendio de distintas fuentes de información. Tiene como principales

ventajas que son simples de implementar, además de que conllevan poco tiempo de construcción y despliegue.

Muchos almacenes de datos comienzan siendo un Mercado de Datos, para, entre otras cosas, minimizar riesgos, pero una vez implementado y explotado con éxito su alcance se va ampliando a medida que pasa el tiempo hasta conformar el DW.

Teniendo en cuenta las operaciones que se desean o requieran desarrollar, los DM pueden adoptar las siguientes arquitecturas:

- **Top-Down:** primero se define el DW y luego se desarrollan, construyen y cargan los DM a partir del mismo. El DW es cargado a través de procesos ETL y luego este alimenta a los diferentes DM, cada uno de los cuales recibirá los datos que correspondan al tema o departamento que traten.

Esta forma de implementación cuenta con la ventaja de no tener que incurrir en complicadas sincronizaciones de hechos, pero requiere una gran inversión y una gran cantidad de tiempo de construcción. (8)

- **Bottom-Up:** en esta arquitectura, se definen previamente los DM y luego se integran en un DW centralizado. Los DM se cargan a través de procesos ETL, los cuales suministrarán la información adecuada a cada uno de ellos. En muchas ocasiones, los DM son implementados sin que exista el DW, ya que tienen sus mismas características pero con la particularidad de que están enfocados en un tema específico. Luego de que hayan sido creados y cargados todos los DM, se procederá a su integración con el depósito. (8)

La ventaja que trae aparejada este modelo es que cada DM se crea y pone en funcionamiento en un corto lapso de tiempo y se puede tener una pequeña solución a un costo no tan elevado. Una vez creados todos los DM, se puede decidir si construir el DW o no. Se decide utilizar este modelo en la solución por todas las ventajas que presenta.

1.3 Modelado Multidimensional

En el modelado de datos es donde se encuentra la principal diferencia entre los sistemas operacionales y el DW: cada uno de ellos es sostenido por un modelo de datos diferente. Los sistemas operacionales se sustentan en el Modelo Entidad Relación (MER) y el DW trabaja con el Modelo Multidimensional. El

modelo multidimensional es una propuesta de Ralph Kimball y es uno de los más utilizados para la construcción de los almacenes de datos. La estructura básica de un DW para el modelo multidimensional está definida por dos elementos: esquemas y tablas. Cada modelo multidimensional se compone de una tabla de hechos central que contienen los valores de las medidas de negocio, y por un conjunto de tablas pequeñas llamadas dimensiones que contienen el detalle de los valores que se encuentran en la tabla de hechos. Los elementos de estas tablas se pueden definir como: (8)

Hechos: Los hechos son datos instantáneos en el tiempo, que son filtrados, agrupados y explorados a través de condiciones definidas en las tablas de dimensiones. 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. Idealmente está compuesta por valores numéricos, continuamente evaluados y aditivos. Su clave primaria está compuesta por las claves primarias de las tablas de dimensiones relacionadas a esta. (8)

Dimensiones: definen cómo están los datos organizados lógicamente y proveen el medio para analizar el contexto del negocio. Contienen datos cualitativos. Representan los aspectos de interés, mediante los cuales los usuarios podrán filtrar y manipular la información almacenada en la tabla de hechos. La tabla de dimensión Tiempo es obligatoria, y la definición de granularidad y estructuración de la misma depende de la dinámica del negocio que se esté analizando. Los atributos dimensionales tienen un rol determinante en un DW, al ser la fuente de todas las necesidades que debieran cubrirse. Esto significa que la base de datos tendrá tanta calidad, como tengan los atributos dimensionales, mientras más descriptivos, y manejables, mejor será el almacén de datos. (8)

1.4 Bases de Datos OLTP y OLAP

Los sistemas de Procesamiento Transaccional y Analítico en Línea conocidos como OLTP y OLAP respectivamente, son conceptos a analizar para un mejor entendimiento y posterior desarrollo de los almacenes de datos, es muy importante reconocer cuando se está en presencia de uno u otro sistema.

1.4.1 Sistemas OLTP- On-Line Transactional Processing

“Son bases de datos orientadas al procesamiento de transacciones, están diseñadas para soportar el procesamiento de información diaria que genera los procesos de una empresa, haciendo un mayor énfasis en maximizar la capacidad transaccional de sus datos. Una transacción genera un proceso atómico que puede involucrar operaciones de inserción, modificación y borrado de datos. El proceso

transaccional es típico en las bases de datos operacionales. En este sistema OLTP el acceso a los datos está optimizado para tareas frecuentes de lectura y escritura, y su historial se limita a los datos actuales o los más recientes. Además de que los datos están estructurados según el nivel de la aplicación”. (9)

Las fuentes de información que manejan los sistemas transaccionales, son de características muy disímiles entre sí, en formato, procedencia, función, etc. Entre los OLTP más habituales que se encuentran en una empresa están los archivos de textos, hipertextos, hojas de cálculos y bases de datos transaccionales.

1.4.2 Sistemas OLAP- On-Line Analytical Processing

En un modelo de datos OLAP, la información es vista como cubos, los cuales consisten de categorías descriptivas (dimensiones) y valores cuantitativos (medidas). El modelo de datos dimensional simplifica a los usuarios formular consultas complejas, arreglar datos en un reporte, cambiar de datos resumidos a datos detallados y filtrar los datos en subconjuntos significativos. (10)

“Los sistemas OLAP son bases de datos orientadas al procesamiento analítico. Este análisis suele implicar, generalmente, la lectura de grandes cantidades de datos para llegar a extraer algún tipo de información útil: tendencias de ventas, patrones de comportamiento de los consumidores, elaboración de informes complejo, etc. Este sistema es típico de los Mercado de Datos.”(9)

En este sistema los datos solamente pueden ser consultados, y para almacenarlos es a través de una carga inicial que se realiza mediante el proceso ETL. El historial de los datos es a largo plazo, pueden ser de 5 – 10 años, y la estructura de sus datos está orientada a las áreas del negocio, además de que los formatos de los datos están integrados de manera uniforme en toda la organización.

1.5 Cubos multidimensionales

El cubo multidimensional es una de las estructura de datos más utilizadas para representar los datos en un DW y cuyo funcionamiento es el más complejo de entender, el cual convierte los datos planos que se encuentran en filas y columnas, en una matriz de N dimensiones.

Los cubos son elementos claves en OLAP (On Line Analytical Processing), una tecnología que provee rápido acceso a datos en un Almacén de Datos. (8)

Los objetos más importantes que se pueden incluir en un cubo multidimensional, son los siguientes:

Indicadores: sumalizaciones que se efectúan sobre algún hecho o expresiones basadas en sumalizaciones, pertenecientes a una tabla de hechos, que serán incluidos en algún cubo multidimensional, con el fin de analizar los datos almacenados en el DW. (8)

Atributos: campos o criterios de análisis, pertenecientes a tablas de dimensiones, que se utilizarán para analizar los indicadores dentro de un cubo multidimensional. (8)

Jerarquías: representa una relación lógica entre dos o más atributos, pertenecientes a un cubo multidimensional; siempre y cuando posean su correspondiente relación “padre-hijo”. Las mismas están compuestas por dos o más niveles, y pueden existir varias en un mismo cubo. (8)

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): deriva su nombre del hecho de que su diagrama forma una estrella. Consta de varias tablas dimensiones, las cuales están relacionadas con una tabla de hechos central a través de sus respectivas llaves. Este modelo debe estar completamente desnormalizado, porque es precisamente esto lo que hace que sea una estrella.

La ventaja que presenta este esquema es que es simple de interpretar y optimiza los tiempos de respuesta ante las consultas de los usuarios. Este modelo es soportado por casi todas las herramientas de consulta y análisis. (8)

Esquema copo de nieve (Snowflake Scheme): se basa esencialmente en la estructura de las tablas dimensiones, ya que en este esquema estas tablas están normalizadas. Este esquema representa una extensión del modelo en estrella cuando las tablas de dimensiones se organizan en jerarquías de dimensiones.

En este modelo existe una tabla de hechos central que está relacionada con una o más tablas de dimensiones, quienes a su vez pueden estar relacionadas o no con una o más tablas de dimensiones. Este esquema hace una mejor utilización del espacio, además de que las tablas al estar normalizadas se requieren menos esfuerzo en el diseño, pero al existir muchas uniones y relaciones entre tablas, el desempeño puede verse reducido. Además si se poseen múltiples tablas de dimensiones, cada una de

ellas con varias jerarquías, se creará un número de tablas bastante considerable, que pueden llegar al punto de ser inmanejables. (8)

Esquema constelación o copo de estrellas (Starflake Scheme): este modelo se compone de varios esquemas de estrella, está formada por una tabla de hechos principal y otras tablas de hechos auxiliares, que están relacionadas con sus respectivas tablas dimensiones, las mismas se encuentran en el centro del esquema.

Su diseño y cualidades son muy similares a las del esquema en estrella, pero posee una serie de diferencias con el mismo. La principal diferencia es que permiten tener más de una tabla de hechos, además de que contribuyen a la reutilización de tablas dimensiones y como desventaja no son soportados por todas las herramientas de consultas y análisis. (8)

El esquema de modelado dimensional obtenido al modelar todos los procesos que engloba el Mercado de Datos, para el Área de Ingreso, es la Constelación. Esto se debe a que un proceso del negocio puede estar compuesto por varios cubos, en dependencia de su complejidad, donde cada cubo forma un modelo de estrella, por lo que al final un mismo proceso puede formar una constelación, y el modelo general del mercado también forma este esquema.

1.7 Tipo de Organización Física

ROLAP (Relational On Line Analytic Processing): es el tipo de organización física para el almacenamiento de los datos que se va a utilizar, el cual se implementa sobre tecnología relacional, pero disponen de algunas facilidades para mejorar el rendimiento. Aunque el Almacén de Datos se organiza como una base de datos multidimensional es soportado por un SGBD Relacional, donde las dimensiones y tablas de hechos serán traducidas a tablas. En los sistemas ROLAP, los cubos multidimensionales se generan dinámicamente al instante de realizar las consultas, haciendo de esta manera el manejo de cubos transparente a los usuarios. Al no tener que intervenir los usuarios en la creación y el mantenimiento explícito de los cubos, ROLAP brinda mucha flexibilidad, posibilitando además la obtención de consultas ad hoc. La principal desventaja de estos sistemas es que no sea muy eficiente en cuanto a la rapidez de respuesta ante las consultas de los usuarios, debido a que los datos de los cubos se deben calcular cada vez que se ejecuta una consulta sobre ellos. (8)

1.8 Métodos más usados en la Construcción de Almacenes de Datos

Los modelos más generalizados para la construcción de almacenes de datos, son los definidos por Ralph Kimball, el principal promotor del modelado multidimensional y el de William H. Inmon, que es considerado el padre de los almacenes de datos.

El paradigma planteado por Kimball expone que el almacén de datos es un conglomerado de todos los mercados de datos dentro de una empresa, siendo una copia de los datos transaccionales estructurados de una forma especial para el análisis, de acuerdo al modelo multidimensional, mientras que Inmon ve la necesidad de transferir la información de los diferentes OLTP (Sistemas Transaccionales) de las organizaciones a un lugar centralizado donde los datos puedan ser utilizados para el análisis, donde los mercados de datos son tratados como subconjuntos del DW corporativo, que son construidos para cubrir las necesidades individuales de análisis de cada departamento, y siempre a partir de este DW central. Desde el punto de vista arquitectónico, la mayor diferencia entre los dos autores es el sentido de la construcción del DW, por un lado se puede comenzar a construir el DW a partir de los mercados de datos o ascendente (Bottom-up, Kimball), o comenzando con todo el DW desde el principio, y después los mercados de datos obtienen la información de él, este enfoque es conocido como descendente (Top-Down, Inmon).

1.9. Arquitectura de un Almacén de Datos

“Una Arquitectura de Almacén de Datos es una forma de representar la estructura total de datos, comunicación, procesamiento y presentación, que existe para los usuarios finales que disponen de una computadora dentro de la empresa.” (11)

Entre los objetivos fundamentales que debe satisfacer la arquitectura de un Almacén de datos se encuentran, soportar diferentes configuraciones de los datos empresariales y al mismo tiempo, brindar facilidades al usuario para la ejecución y la administración de sus tareas en ambientes complejos. Para dar solución a los problemas del mundo real se puede trabajar con diferentes arquitecturas conceptuales que se nombran según la cantidad de capas de datos empresariales que abarcan.

1.9.1 Arquitectura conceptual de los datos del Mercado de Datos

La arquitectura de un almacén de datos está formada por diversos elementos que interactúan entre sí, donde es conveniente tener en consideración los diferentes entornos por los que han de pasar los

datos en su camino hacia el DW o DM destino. Dada la cantidad de transformaciones a realizar, y que normalmente el DW, además de cumplir su función de soporte a los requisitos analíticos, realiza una función de integración de los datos que posteriormente conformarán el almacén, se debe de tener en cuenta la cantidad de capas de datos para la representación del mismo. (8)

En la construcción del mercado del Área de Ingreso, se utiliza la arquitectura de tres capas de datos, a continuación se describen las mismas:

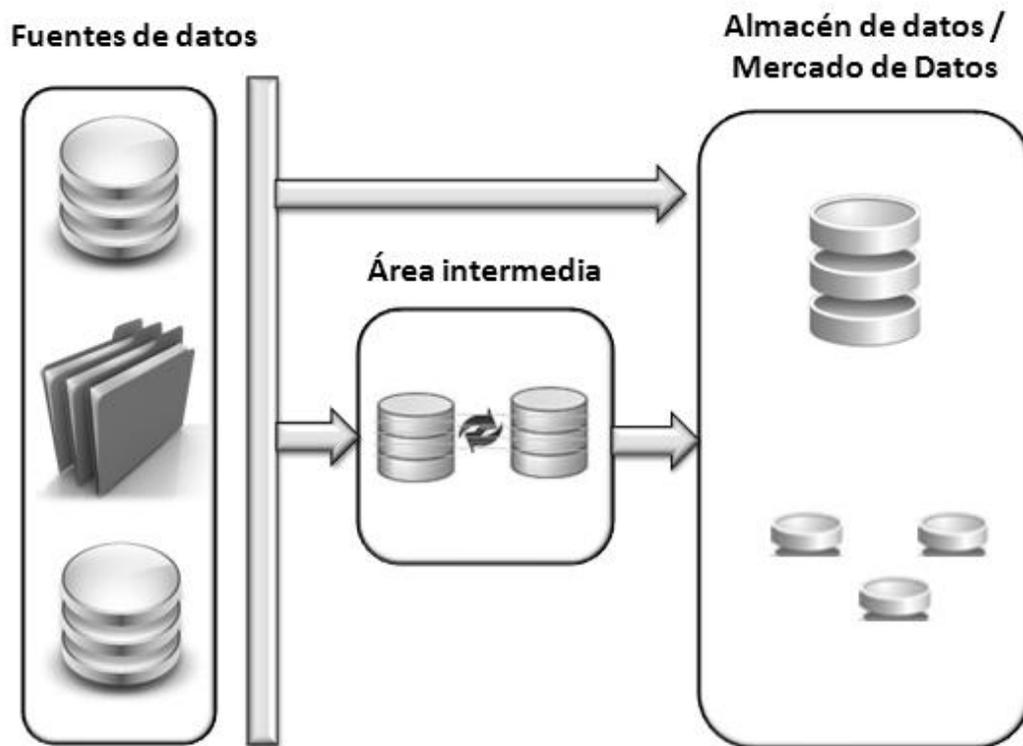


Figura N° 1.10: Arquitectura de un Almacén de Datos

Fuentes de Datos (Data Sources)

En la 1ra capa de datos se encuentran los sistemas operacionales, que representan las diferentes fuentes de datos origen que van a alimentar el mercado. Los datos fuentes pueden estar en cualquier formato, desde archivos de texto, hojas excel y base de datos relacionales.

Área Temporal (Staging Area)

Es un área temporal donde se recogen los datos que se necesitan de los sistemas origen. En esta capa solo se almacenan los datos estrictamente necesarios para poblar el almacén, y se aplica el mínimo de transformaciones posibles en su carga, para evitar un alto valor de afectación en los sistemas operacionales del origen. (8)

Los datos una vez cargados en este área, se le aplican una serie de transformaciones necesarias para almacenarlos posteriormente en el mercado, con el objetivo de eliminar las inconsistencias que puedan traer del sistema origen, además de integrarlos al formato establecido.

Mercado de Datos o Almacén de Datos (Data Mart/ Data Warehouse)

La última capa es el DM o DW, y sus datos provienen del staging area. En esta capa residen los datos almacenados en una estructura dimensional, a partir de los cuales se realiza todo el análisis para la creación de los reportes con las herramientas de consultas.

1.10 Metodologías de desarrollo de un Almacén de Datos

Las metodologías de desarrollo de un almacén de datos, tienen como principal objetivo guiar todo el ciclo de vida para su construcción. En la actualidad existen varios modelos a seguir, pero de manera general pueden ser englobadas en dos grandes bloques: top-down y bottom-up que se corresponden con las metodologías propuestas por Bill Inmon y Ralph Kimball respectivamente. (12)

En la implementación del Mercado de Datos, se va a utilizar la definida por Kimball por decisión del grupo de arquitectura de SIDEPA (Sistema de Información para la Dirección de Establecimiento Penitenciario). A continuación se describe la misma:

1.10.1 Metodología de Ralph Kimball – Ciclo de Vida dimensional

La metodología propuesta se basa en lo que Kimball denomina Ciclo de Vida Dimensional del Negocio, esta metodología se ha convertido en el estándar de facto en el área de apoyo a las decisiones empresariales.

La construcción de una solución de almacenes de datos con inteligencia de negocio es muy compleja, con esta metodología propuesta por Kimball se simplifica esta complejidad, y como respuesta divide el

ciclo de vida de su desarrollo en una serie de etapas, el flujo central de dichas etapas se describe a continuación: (12)

Planificación

Esta es la primera tarea del ciclo de vida de la metodología, donde se determina el propósito, objetivos específicos y el alcance del proyecto, los principales riesgos y una aproximación inicial a las necesidades de información. Además incluye un conjunto de acciones típicas de un plan de proyecto. (12)

Análisis de Requisitos

En el análisis de los requisitos se realizan entrevistas, se hacen encuestas y todo tipo de métodos para conocer a profundidad sobre el negocio en cuestión, los competidores, la industria y los clientes del mismo. Los requisitos del negocio deben determinar el alcance del Almacén de Datos. (12)

Modelado Dimensional

En esta tarea es donde se diseña el modelo lógico del Almacén de Datos, se identifican las tablas de hechos, se eligen las dimensiones, se establece el nivel de granularidad y se realiza el modelo gráfico de alto nivel. (12)

Diseño Físico

El diseño físico se focaliza sobre la selección de las estructuras necesarias para soportar el diseño físico, responde a cuál es la estructura a seguir, cómo se va a configurar, etc. (12)

Diseño e Implementación del Sistema de Extracción, Transformación y Carga

El sistema de Extracción, Transformación y Carga es la base sobre la cual se alimenta el almacén de datos. En este paso se extraen los datos de los sistemas de origen de datos, se aplican diferentes reglas para aumentar la calidad y consistencia de los mismos, y finalmente se realiza la carga de los datos para poblar el almacén de datos. (12)

Despliegue

El despliegue representa la convergencia de la tecnología, los datos y las aplicaciones del usuario final accesibles desde el escritorio de los usuarios de negocios. Una amplia planificación es necesaria para garantizar el buen funcionamiento de todos los elementos que intervienen. La educación de los

usuarios de negocio integrando todos los aspectos de la convergencia, debe ser desarrollada y entregada. Además, el apoyo a los usuarios y la comunicación o las estrategias de retroalimentación debe quedar establecido antes de que los usuarios de negocios puedan tener acceso al almacén de datos. (13)

Mantenimiento y Crecimiento

Queda mucho trabajo a raíz de la implementación inicial del almacén de datos, por lo que se hace necesario seguir centrándose en los usuarios del negocio, proporcionándoles apoyo y educación. También es necesario asegurar que los procesos y procedimientos están en su lugar para el funcionamiento continuo y eficaz del almacén. La aceptación y desempeño del almacén de datos debe ser medido a través del tiempo, y finalmente el plan de mantenimiento debe incluir una estrategia de comunicación de amplio alcance. (13)

1.11 Herramientas de creación de un Almacén de Datos

Para la construcción de un almacén de datos se necesita utilizar una serie de herramientas informáticas capaces de asistir y facilitar el trabajo. En la actualidad existen compañías que brindan colecciones de herramientas de inteligencia de negocio para construir los mismos, un ejemplo de esta es la Suite de Inteligencia de Negocio de Oracle en su Edición Empresarial (OBIEE), que es la utilizada para construir el mercado de datos para el Área de Ingreso, la cual fue definida por parte del grupo de arquitectura del proyecto.

OBIEE es una plataforma de inteligencia empresarial que ofrece una gama completa de capacidades de análisis y presentación de informes. Diseñado para la escalabilidad, fiabilidad y rendimiento, que proporciona una visión contextual, relevante y efectiva a todos los miembros de una organización, lo cual posibilita mejorar el proceso de toma de decisiones. La suite está compuesta por el siguiente conjunto de herramientas:

- **Sistema Gestor de Base de Datos Oracle 11g**

Es el sistema gestor de base de datos (SGBD) líder en la industria, también es considerado uno de los más completos, destacando su soporte de transacciones, estabilidad, disponibilidad, escalabilidad, fiabilidad y seguridad. Oracle puede ejecutarse en todas las plataformas y cuenta con un lenguaje de diseño de base de datos muy completo (PL\SQL). Provee funciones completas para gestionar fácilmente el procesamiento de transacciones, inteligencia de negocio y aplicaciones de gestión de

contenidos. Posee un alto rendimiento para los almacenes de datos, procesamiento analítico en línea, y minería de datos. (14)

- **Oracle Warehouse Builder (OWB)**

Es una herramienta que permite realizar el diseño completo de un almacén de datos y luego cargarlo en un motor de base de datos Oracle mediante la implementación del proceso ETL. La aplicación incluye una potente interfaz gráfica que permite crear dimensiones, cubos, correspondencias y procesos sin tener que escribir directamente el código. Esta herramienta permite tomar los datos desde diferentes fuentes y sistemas e integrarlos para proporcionar información de alta calidad que luego será analizada. Está concebida para trabajar con la tecnología de Base de Datos Oracle.

- **Administration Tool**

Es la herramienta utilizada para administrar el servidor Oracle Business Intelligence. Permite al administrador configurar los repositorios de datos, desde modo en línea o desconectado. Además de crear y editar los usuarios que interactúan con la herramienta.

- **Oracle Business Intelligence Answers**

BI Answer es una poderosa herramienta de visualización con una interfaz web. Esta es capaz de conectarse con el servidor de BI y permitirle al usuario realizar consultas ad-hoc de manera sencilla. Además, se pueden crear fácilmente gráficos, tablas dinámicas e informes visualmente atractivos, lo cual es completamente interactivo y se puede explorar, guardar, compartir, modificar el formato, o personalizarlos. (15)

- **Oracle Business Intelligence Interactive Dashboards**

Cuadros de mando interactivos. Esta solución ofrece una colección de cuadros de mando totalmente interactivo e informes con una rica variedad de visualizaciones. Los paneles de control proporcionan a los usuarios la información filtrada y personalizada para su identidad, función o rol basado en las reglas de seguridad predefinidas. La interfaz de usuario rica e interactiva hace la presentación de los datos intuitivos, relevante y fácil de entender. (15)

ER Estudio como Herramienta de Modelado

El ER Estudio es una herramienta CASE muy fácil de usar, con soporte de múltiples niveles de análisis y que permite realizar tanto el modelado lógico de los datos como la implantación física para el

Capítulo 1: Marco Teórico de la Investigación

Mercado de Datos. Es una de las herramientas más comunes y soporta más de 30 de las bases de datos más importantes de la industria. Además permite el modelado relacional y multidimensional, la integración con productos de “Business Intelligence” y la transformación automática de un diseño lógico a un diseño físico para una plataforma específica de base de datos. (16)

Conclusiones Parciales

En el presente capítulo se realizó un análisis de los sistemas existentes en varios países evidenciando que no utilizan almacenes de datos como parte de su solución, solo el Sistema de Gestión Penitenciaria de Venezuela con la construcción de dos mercados de datos. Se definieron conceptos fundamentales de los almacenes de datos, se describió la arquitectura y metodología a utilizar. Además se describieron las herramientas a utilizar en el desarrollo del mercado de datos.

CAPÍTULO 2: ANÁLISIS Y DISEÑO DEL MERCADO DE DATOS

Introducción

En este capítulo se presenta el análisis en el Área de Ingreso, Observación, Evaluación y Diagnóstico del Sistema Penitenciario Cubano, además del proceso Vínculo¹, para realizar el diseño del mercado de datos siguiendo los pasos que plantea la metodología Kimball. Se define cómo se realiza la captura de requisitos, y se analizan los diferentes procesos que abarca dicha área más otros que son de interés para el centro; para ello se hace una breve descripción de los mismos. Además se identifican las dimensiones y los hechos necesarios para diseñar el mercado de datos, donde se describen detalladamente solo los procesos del caso de estudio, que son el de Ingreso y Características Socio-Antropológicas.

2.1 Identificación de Temas Analíticos

- **Técnicas de Identificación de Requisitos**

En la etapa del análisis de la construcción de los almacenes de datos, se realiza la tarea identificación de los requisitos, donde se aplican un conjunto de técnicas para la captura de los mismos. Siguiendo la metodología de Kimball, el propone fundamentalmente las entrevistas cerradas con el cliente, para la cual define una serie de preguntas enfocadas al rol que ejecuta el que se va a entrevistar.

Entrevistas: Las entrevistas se llevan a cabo individualmente o en grupos muy pequeños. Las mismas necesitan ser programadas en un tiempo limitado para no perjudicar el calendario de los entrevistados. Además promueven un alto grado de participación del cliente, lo cual posibilita generar una mayor cantidad de información, así como un mayor intercambio entre las partes involucradas.

Este tipo de técnica se utilizó con el propósito de conocer cuáles son las metas y objetivos del negocio para el Área de Ingreso, sus prioridades, y los procesos que abarca. Otro motivo de la entrevista es saber cuál es el personal que trabaja en dicha área y las responsabilidades que le competen a cada uno, así como identificar sus necesidades de información.

¹ Vínculo: proceso del negocio encargado de recoger las relaciones que tienen los internos dentro y fuera del centro penitenciario, conocidos como vínculos internos y externos respectivamente.

Otra técnica a utilizar para la identificación de los requisitos, es el análisis de la documentación del negocio, específicamente la correspondiente con el Área de Ingreso, con el objetivo de obtener mayor conocimiento sobre los procesos que se realizan en la misma. Entre los documentos a analizar se encuentran las leyes y normativas a seguir en los centros penitenciarios, para los distintos procesos que se tienen en cuenta en esta área. El documento de mayor importancia es sobre la metodología rectora para el Área de Ingreso, Observación, Evaluación y Diagnóstico dirigida al equipo de trabajo, donde se describe en detalle todos los pasos a seguir una vez que el interno es acogido en una unidad receptora, hasta que es ubicado en un centro penitenciario.

A partir de las entrevistas y del análisis de la documentación, se identificaron los temas analíticos y procesos del negocio. Los temas analíticos agrupan requisitos comunes en un tema común.

- **Área de Ingreso**

El Área de Ingreso, Observación y Diagnóstico se encarga de realizar un estudio amplio sobre el comportamiento y la conducta del interno, partiendo de su ingreso al centro, donde se recogen todos los datos necesarios para su correcta identificación. Posterior al ingreso se realiza una propuesta de clasificación en régimen, la cual es analizada por la comisión de clasificación para ubicarlo oficialmente. Otro proceso a tener en cuenta es el de Vínculo, el cual recoge las relaciones del interno, y aunque no tiene una vinculación directa con el Área de Ingreso, es muy importante por la influencia que pueden ejercer sobre el interno.

Partiendo de todo lo explicado anteriormente se identificaron los siguientes temas analíticos para esta área:

Registro de Ingreso: Este tema va dirigido a la actividad administrativa de registro, por lo que dentro de ella se encuentran los procesos Ingreso y Características Socio-Antropológicas.

Clasificación: Abarca lo relacionado con la Clasificación en Régimen del interno, y en dependencia de esta entonces se definirá la Ubicación del mismo en el centro.

Vínculo: Este tema va orientado a conocer las relaciones del interno, para saber las influencias que tienen los internos, y cómo inciden sobre ellos, para ello se analiza el proceso Vínculo.

- **Descripción de los Procesos por Tema Analítico:**

- **Registro de Ingreso**

Ingreso

Se considera como ingreso a un centro penitenciario receptor a las personas que arriban al mismo provenientes de los órganos de justicia, ya sea en la condición de acusado, asegurado o sancionado.

El diseño de este proceso posibilita analizar el nivel de ingresos en los centros penitenciarios receptores, la cantidad de internos que reingresan a los centros. Gracias a este proceso se conoce cuantos internos son reincidentes, y de esta manera se les puede dar un mejor seguimiento a las personas luego de salir de un centro de reclutamiento.

Características Socio-Antropológicas

Cada interno, como un individuo en particular, tiene sus propios rasgos físicos, características personales y conductuales. Si se analizan las particularidades de varios sancionados a la vez, se podrán establecer patrones comunes entre todos, lo que beneficiaría a los directivos de cada instancia al tener un conocimiento más preciso de la población que albergan.

Conocer por rango de edades cuántos internos ingresa a los centros penitenciarios en un período de tiempo determinado, posibilita analizar los grupos de edades que más incurren en indisciplinas y tomar decisiones encaminadas a erradicar esta situación.

Analizar el nivel cultural de la población que está sancionada, el sexo, el oficio que realizaban, el estado civil y otras características más, indican también cuáles son los grupos de personas mayoritarios en los centros penitenciarios y qué trabajo preventivo ha de realizarse desde la sociedad para disminuir la cifra de sancionados.

- **Clasificación**

Clasificación en Régimen

Es la clasificación de severidad que tiene un área de reclusión dentro de un establecimiento penitenciario.

Con este proceso se quiere conocer la cantidad de internos por clasificación de régimen teniendo en cuenta el tipo de clasificación, así como la cantidad de clasificaciones dadas por los niveles de mando en el tiempo.

Ubicación

Registra la ubicación física de los privados de libertad en el área de reclusión de los establecimientos penitenciarios.

El diseño de este proceso posibilita analizar la cantidad de internos existente por local y colectivo, además de saber la cantidad asociado al tipo de régimen de colectivo. Con estos datos se pueden realizar estudios sobre las tendencias referentes al mismo. Facilitando la toma de decisiones.

➤ Vínculo

El tema vínculo solo analiza al proceso vínculo, el cual recoge las relaciones de los internos con otras personas. Estas relaciones pueden ser tanto dentro del centro penitenciario como fuera, los cuales son conocidos como vínculos internos y externos respectivamente.

El proceso será diseñado para generar reportes necesarios que permitan los análisis estadísticos para la realización de informes y estudios, sobre la cantidad de internos que tienen vínculos en los centros. Además posibilita conocer cuál es el motivo de vínculo más común, así como el tipo de influencia que predomina en los internos.

2.1.1 Análisis de Requisitos:

El análisis de los requisitos recoge todas las necesidades del usuario final, y la misma está compuesta por los requisitos de información y la especificación de requisitos, los cuales deben ser descriptivos y orientados al usuario final. Además se describen los usuarios que van a interactuar con el sistema.

- **Requisitos de Información**

Describen qué información debe almacenar el sistema para satisfacer las necesidades de clientes y usuarios. Identifican los conceptos relevantes sobre los que se debe almacenar información y los datos específicos que son de interés. (17)

Los requisitos de información representa toda la información necesaria para los especialistas de la Dirección de Establecimientos Penitenciarios (DEP), con el propósito de realizar los análisis pertinentes que serán tomados en cuenta en la proyección estratégica del centro. Para darle respuesta a las necesidades del cliente, se identificaron un conjunto de tablas dimensiones y medidas, por los diferentes procesos que abarca el Área de Ingreso. Los requisitos de información identificados se muestran a continuación desglosados por cada proceso, para un mejor entendimiento y organización, su enumeración es continua:

➤ **Proceso Ingreso**

RI1_Mostrar total de ingresos y reingresos

RI2_Mostrar total de ingresos y reingresos por año

RI3_Mostrar total de ingresos y reingresos por provincia

RI4_Mostrar total de ingresos por trimestre y por provincia

RI5_Mostrar total de ingresos por provincia

RI6_Mostrar total de ingresos por año y por provincia

RI7_Mostrar total de ingresos por delito y por año

RI8_Mostrar total de ingresos por año y por trimestres

RI9_Mostrar total de ingresos por edad de ingreso y por año

RI10_Mostrar la relación entre edad de ingreso y provincia

RI11_Mostrar la relación entre edad de ingreso y delito

RI12_Mostrar total de delitos por provincia

RI13_Mostrar total de delitos por motivo de ingreso

RI14_Mostrar la comparación de edad por ingresos y reingresos

RI15_Mostrar total de reingresos por provincia

RI16_Mostrar total de reingresos por trimestre y por provincia

RI17_Mostrar total de reingresos por año y por provincia

RI18_Mostrar total de reingresos por edad de reingreso y por año

RI19_Mostrar total de reingresos por año y por trimestre

RI20_Mostrar total de reingresos por delito y por año

RI21_Mostrar total de reingresos por edad de reingreso y por provincia

RI22_Mostrar la relación entre la edad de reingreso y el delito cometido

RI23_Mostrar total de reingresos por delito y por provincia

RI24_Mostrar total de delitos por motivo de reingreso

➤ **Proceso Vínculo**

RI25_Mostrar total de vínculos internos por año y por trimestre

RI26_Mostrar total de vínculos internos por provincia y por año

RI27_Mostrar total de vínculos internos por tipo de influencia, por motivo del vínculo, por tipo de vínculo y por provincia

RI28_Mostrar total de vínculos internos por provincia

RI29_Mostrar total de vínculos internos por tipo de documento

RI30_Mostrar total de vínculos internos por tipo de vínculo y por motivo del vínculo

RI31_Mostrar total de vínculos internos por tipo de influencia y por motivo del vínculo

RI32_Mostrar total de vínculos internos por provincia y por año

RI33_Mostrar total de vínculos internos por tipo de documento y por tipo de vínculo

RI34_Mostrar total de vínculos internos por tipo de documento y por motivo del vínculo

RI35_Mostrar total de internos sin vínculos

RI36_Mostrar total de vínculos externos por año y por trimestre

RI37_Mostrar total de vínculos externos por provincia y por año

RI38_Mostrar total de internos con vínculos por provincia

RI39_Mostrar por ciento de internos con vínculos

RI40_Mostrar total de vínculos externos por tipo de vínculo y por motivo del vínculo

RI41_Mostrar total de vínculos externos por tipo de influencia, por motivo del vínculo, por tipo de vínculo y por provincia

RI42_Mostrar total de vínculos externos por tipo de influencia y por visita

RI43_Mostrar total de vínculos externos por tipo de vínculo externo, influencia y motivo del vínculo

➤ **Proceso Características Socio-Antropológicas**

RI44_Mostrar total de internos por creencias religiosas y por provincia de nacimiento

RI45_Mostrar total de internos por creencias religiosas y por provincia

RI46_Mostrar total de internos por antecedentes penales y por provincia

RI47_Mostrar total de internos por color de piel y por provincia

RI48_Mostrar total de internos por color de piel y por provincia de nacimiento

RI49_Mostrar total de internos por creencias religiosas, por color de piel y por provincia

RI50_Mostrar total de internos por estado civil y por provincia

RI51_Mostrar total de internos por nacionalidad y por provincia

RI52_Mostrar total de internos por provincia

RI53_Mostrar total de internos por número de hijos, por estado civil y por provincia

RI54_Mostrar total de internos por número de hijos

RI55_Mostrar total de internos por oficio y por provincia

RI56_Mostrar total de internos por oficio, por nivel escolar y por provincia

RI57_Mostrar total de internos por provincia de procedencia

RI58_Mostrar total de internos por región

RI59_Mostrar total de internos por rango de edad y por provincia

RI60_Mostrar total de internos por rango de edad, por antecedentes y por provincia

RI61_Mostrar relación entre antecedentes, color de piel y provincia

RI62_Mostrar relación entre antecedentes, nivel escolar y color de piel

RI63_Mostrar relación entre rango de edad, color de piel y provincia

RI64_Mostrar relación entre rango de edad, región y color de piel

RI65_Mostrar relación entre provincia de nacimiento y rango de edad

RI66_Mostrar relación entre región de procedencia y rango de edad

RI67_Mostrar relación entre nivel escolar, color de piel y provincia

➤ **Proceso Ubicación**

RI68_Mostrar total de internos ubicados por año y por trimestre

RI69_Mostrar total de internos ubicados por colectivo y por provincia

RI70_Mostrar total de internos ubicados por régimen y por año

RI71_Mostrar relación entre régimen de colectivo y local

RI72_Mostrar total de internos ubicados por régimen de colectivo

RI73_Mostrar relación entre centros y locales

RI74_Mostrar total de internos ubicados por colectivo, por régimen y provincia

RI75_Mostrar total de internos ubicados por año y por provincia

➤ **Proceso Clasificación en Régimen**

RI76_Mostrar total de internos por clasificación en régimen y por provincia

RI77_Mostrar total de internos por tipo de régimen

RI78_Mostrar total de internos por tipo de régimen y por clasificación

RI79_Mostrar total de internos por tipo de clasificación y por año

Capítulo 2: Análisis y Diseño del Mercado de Datos

RI80_Mostrar total de internos clasificados por año y por tipo de régimen

RI81_Mostrar clasificación histórica

RI82_Mostrar total de internos por clasificación en régimen y por trimestre

RI83_Mostrar total de internos clasificados por tipo de clasificación, por tipo de régimen y por provincia

RI84_Mostrar total de internos por clasificación en régimen, por provincia y por año

- **Especificación de Requisitos**

Para realizar la descripción de la especificación de requisitos, y analizar los restantes temas del capítulo que van orientado a un mayor nivel de detalle, se toman los procesos del caso de estudio por ser los más importantes del área:

Proceso Ingreso

Nº	Nombre	Descripción	Complejidad	Prioridad para el cliente
RI 3.	Total de Ingresos y Reingresos por Provincia	Visualiza el total de ingresos y reingresos por provincia en un gráfico de barra, donde cada barra representa los ingresos, reingresos y sus respectivos totales por provincia. Al seleccionar una provincia, muestra los ingresos y reingresos por centro, si selecciona un centro muestra los ingresos y reingresos por colectivo.	Alta (ISEC-SIDEP-0132_EvalReq)	Alta
	Campos	Tipos de Datos	Reglas o Restricciones	
	Provincia	Cadena de Caracteres	No procede	
	Ingresos y Reingresos	Cadena de Caracteres		

Capítulo 2: Análisis y Diseño del Mercado de Datos

	Cantidad de Ingresos y Reingresos	Numérico		
	Observaciones			
RI 6.	Total de Ingresos por Año y por Provincia.	Visualiza el total de ingresos por provincia y por año en una gráfica de barra, donde cada barra representa una provincia y cada color de la barra representa un año. Al seleccionar una barra con su respectivo color, muestra los ingresos por centro y los meses del año seleccionado, si selecciona un año muestra los ingresos por provincia y los meses del año seleccionado.	Alta (ISEC-SIDEP-0132_EvalReq)	Alta
	Campos	Tipos de Datos	Reglas o Restricciones	
	Provincia	Cadena de caracteres	No procede	
	Año	Fecha		
	Cantidad de Ingresos	Numérico		
	Observaciones			
RI 12.	Total de Delitos por Provincia.	Visualiza el total de ingresos por delito y por provincia en una gráfica de área, donde cada área representa las provincia y cada color del área representa un conjunto de delitos al seleccionar	Alta (ISEC-SIDEP-0132_EvalReq)	Alta

Capítulo 2: Análisis y Diseño del Mercado de Datos

		un área muestra las provincias y los delitos pertenecientes al conjunto de delitos.		
	Campos	Tipos de Datos	Reglas o Restricciones	
	Delito	Cadena de Caracteres	No procede	
	Provincia	Cadena de Caracteres		
	Cantidad de Ingresos	Numérico		
	Observaciones			
RI 13.	Total de Delitos por Motivo de Ingreso	Visualiza el total de ingresos por delito y por motivo en un gráfico de barras, donde cada barra representa el motivo del ingreso variando entre las familias de delitos. Al seleccionar una familia de delitos, muestra los delitos pertenecientes a dicha familia por los motivos de ingresos.	Alta (ISEC-SIDEP-0132_EvalReq)	Alta
	Campos	Tipos de Datos	Reglas o Restricciones	
	Delito	Cadena de Caracteres	No procede	
	Motivo	Cadena de Caracteres		
	Cantidad de Ingresos	Numérico		
	Observaciones			

Proceso Características Socio-Antropológicas

Nº	Nombre	Descripción	Complejidad	Prioridad para el cliente	
RI 46.	Total de Internos por Antecedentes Penales y por Provincia	Visualiza el total de internos por antecedentes penales y por provincia en un gráfico de barras, donde cada barra representa el tipo de antecedente penal en una provincia. Al seleccionar una provincia, muestra los internos por antecedentes penales y por centro, si selecciona un centro, muestra los internos por antecedentes penales y por colectivo.	Alta (ISEC-SIDEP-0132_EvalReq)	Alta	
	Campos	Tipos de Datos			Reglas o Restricciones
	Antecedente Penal	Cadena de Caracteres			
	Provincia	Cadena de Caracteres			
	Cantidad de internos	Numérico			
	Observaciones				
RI 47.	Total de Internos por Color de Piel y por Provincia	Visualiza el total de internos por color de piel y por provincia en una gráfica de barra, donde cada barra representa el color de piel y esta se encuentra en una provincia determinada. Al seleccionar una barra muestra los internos por color de piel y por centro. Al	Alta (ISEC-SIDEP-0132_EvalReq)	Alta	

Capítulo 2: Análisis y Diseño del Mercado de Datos

		seleccionar nuevamente una barra muestra los internos por color de piel y por colectivo.		
	Campos	Tipos de Datos	Reglas o Restricciones	
	Color de piel	Cadena de Caracteres		
	Provincia	Cadena de Caracteres		
	Cantidad de internos	Numérico		
	Observaciones			
RI 52.	Total de Internos por Provincia	Visualiza el total de internos por provincia en un gráfico de pastel, donde cada cuña representa el Por ciento de internos por provincia. Al seleccionar una provincia, muestra el por ciento de internos por centro, si selecciona un centro, muestra el por ciento de internos por colectivo.	Alta (ISEC-SIDEP-0132_EvalReq)	Alta
	Campos	Tipos de Datos	Reglas o Restricciones	
	Provincia	Cadena de Caracteres		
	Por ciento de internos	Numérico		
	Observaciones			
RI 55.	Total de Internos por Oficio y por Provincia	Visualiza el total de internos por oficio y por provincia en un gráfico de barras, donde cada barra representa un oficio variando por las provincias. Al selecciona una	Alta (ISEC-SIDEP-0132_EvalReq)	Alta

Capítulo 2: Análisis y Diseño del Mercado de Datos

		provincia, muestra los internos por oficio y por centro, si selecciona un centro, muestra los internos por oficio y por colectivo.		
	Campos	Tipos de Datos	Reglas o Restricciones	
	Oficio	Cadena de Caracteres		
	Provincia	Cadena de Caracteres		
	Cantidad de internos	Numérico		
	Observaciones			

- **Usuarios del Sistema**

A partir de las entrevistas realizadas, dentro del proceso de captura de requisitos, se identificaron 3 usuarios principales que van a interactuar con el sistema, los cuales se describen a continuación:

1. Oficial de Observación y Diagnóstico: Se autentica, consulta y analiza los reportes, exporta los reportes y configura la visualización del reporte
2. Especialista del Área de Ingreso, Observación, Evaluación y Diagnostico: Encargado de gestionar los reportes.
3. Administrador del Sistema: Realiza el proceso de extracción, transformación y carga de los datos además de gestionar los usuarios y los roles.

- **Técnicas de Validación de Requisitos**

Posteriormente al proceso de captura de requisitos, se aplican técnicas para la validación de los mismos. Es necesario asegurar que el análisis realizado y los resultados obtenidos en la etapa de identificación de requisitos sean correctos. Para validar los requisitos se va a utilizar la técnica prototipo funcional, la cual se describe a continuación:

Técnica Prototipo: Los prototipos son simulaciones del posible producto, con el objetivo de mostrarle al cliente una versión reducida del producto final, permitiendo conseguir una importante retroalimentación en cuanto a saber si el producto diseñado con base a los requisitos recolectados, le permite al usuario

realizar su trabajo de manera eficiente. Para realizar el prototipo se utilizó como herramienta de apoyo los ejemplos que trae por defecto la Suite de Inteligencia de Negocios de Oracle.

2.2 Modelado Dimensional

Dentro del Modelo de Datos se encuentra el “*modelo dimensional*”, que tiene como elementos principales las dimensiones y los hechos, los cuales recogen de los procesos que abarca el subsistema, los atributos necesarios para dar respuesta a los requisitos de información. Las dimensiones describen los diferentes puntos de vista del hecho, y el hecho es el punto de intersección entre las diferentes dimensiones. Es importante conocer la composición de ambos y la relación que existe entre ellos para lograr un buen funcionamiento del mercado de datos.

- **Granularidad**

La granularidad define el máximo nivel de detalle que va a tener los datos en las tablas dimensiones. En el Área de Ingreso, se llega hasta el nivel de detalle del interno, donde se definen un conjunto de atributos muy particulares del mismo los cuales deben ser analizados para responder a determinados reportes, en el caso de la dimensión tiempo se llega a un grano diario, y en el nivel de mando su más fino detalle es el colectivo.

- **Tablas Dimensiones:**

Las tablas de dimensiones contienen la descripción textual del negocio. En el modelo dimensional poseen varios atributos que en su conjunto definen una fila en la tabla de dimensión. Los atributos de las dimensiones sirven como fuente primaria de las restricciones de las consultas, agrupaciones y las etiquetas de los reportes.

Para el diseño del Mercado de Datos se identificaron 11 tablas dimensiones, y 8 tablas de hechos de los 5 procesos que componen el Área de Ingreso. A continuación se describen brevemente las dimensiones de los procesos del caso de estudio:

dim_interno: Contiene toda la información que se relaciona con las características principales del interno.

dim_nivel_mando: Guarda la información referente a los distintos niveles de mando existentes en los centros penitenciarios.

dim_tiempo: Contiene todos los datos relacionados con el tiempo, hasta el nivel más bajo de la granularidad, que en este caso es hasta el día del mes.

dim_mot_ingreso: Recoge el valor de los motivos por los cuales se ingresa.

dim_mot_reingreso: Recoge el valor de los motivos por los cuales se reingresa.

dim_tipodelito: Guarda los tipos de delitos que se pueden cometer, y la familia a la cual pertenece el mismo.

dim_oficio: Recoge el valor de diferentes oficios que un interno puede tener.

- **Tablas de Hechos**

Las tablas de hechos son las fuentes de almacenamiento de las medidas numéricas y constituyen la fuente principal de información, donde se integran los distintos puntos de vistas de los datos que se muestran en las tablas dimensiones.

Para el modelo del proceso Ingreso y Características Socio-Antropológicas se definieron 4 tablas de hechos, las cuales se describen a continuación:

fact_ingreso_delito: Guarda el hecho correspondiente a los ingresos que se han realizado en cada centro penitenciario y los delitos asociados a los mismos.

fact_reingreso_delito: Guarda el hecho correspondiente a los reingresos que se han realizado en cada centro penitenciario y los delitos asociados a los mismos.

fact_caract_socioatplg_ofc: Guarda el hecho correspondiente a los oficios que tienen ya verificados los internos.

fact_caract_socioatplg: Guarda la interrelación del interno con el nivel de mando al cual pertenece, cuando se le crea el expediente legal.

- **Matriz Dimensional**

En la matriz dimensional o Bus Matrix por sus siglas en inglés, se recoge la relación existente entre las dimensiones y los procesos que componen el Área de Ingreso. En los modelos propuestos existen dimensiones comunes y específicas de cada proceso. En la Tabla 2.1 que se muestra a continuación se observa esta matriz:

Dimensiones	Procesos				
	Ingreso	Caract. Socio.- Ant.	Ubicación	Vínculo	Clasificac. Régimen
dim_interno	x	x	x	x	x
dim_tiempo	x	x	x	x	x
dim_nivel_mando	x	x	x	x	x
dim_oficio		x			
dim_mot_ingreso	x				
dim_mot_reingreso	x				
dim_tipodelito	x				
dim_ubicación			x		
dim_vinculo_interno				x	
dim_vinculo_externo				x	
dim_clasificacion					x

Tabla 2.1: Matriz Dimensional

- **Patrones de Diseño**

El Modelo Dimensional de los procesos que se analizan, cuentan con un total de 7 tablas dimensiones y 4 tablas de hechos. Para el diseño del mismo se utilizaron los siguientes patrones de almacenes de datos:

1. Claves subrogadas: Se utilizó para resolver el problema de que las llaves de todas las dimensiones sean comunes en cuanto al mismo tipo de datos, que sea un único valor y que tenga una probabilidad muy baja de repetición, lo que posibilita mayor facilidad al realizar las consultas, al ser llaves artificiales secuenciales numéricas, sin ninguna relación con el negocio. Además permite que la construcción y mantenimiento de índices sea una tarea sencilla.
2. Dimensiones Lentamente Cambiantes (SCD): se utiliza este patrón debido a que hay determinados atributos en las dimensiones que pueden cambiar su valor, para un mismo registro, este patrón brinda 3 posibles soluciones, de ellas se va a utilizar la SCD Añadir Fila, porque permite el análisis histórico de los datos, al adicionar un nuevo registro con el valor

modificado pero sin borrar el valor que tenía anteriormente. Es otra evidencia de por qué es necesario utilizar las llaves subrogadas.

3. Productos Heterogéneos: Se utiliza para resolver el problema de la herencia que viene de los datos operacionales, lo cual permite consultar más en profundidad y facilidad a cada tipo de dimensión.

- **Modelo de Datos**

En consecuencia de los requisitos de información definidos para los procesos de Ingreso y Características Socio-Antropológicas, se elaboró su modelo de datos. Cada modelo está compuesto por las tres dimensiones comunes, que son `dim_interno`, `dim_tiempo` y `dim_nivel_mando`, y las dimensiones propias de su proceso, en el caso del Ingreso las `dim_mot_ingreso`, `dim_mot_reingreso` y `dim_tipodelito`, y para Características Socio-Antropológicas solamente la `dim_oficios`. Además de las tablas de hechos que contienen las medidas, formando cada esquema una constelación, al tener por cada proceso más de un cubo dimensional.

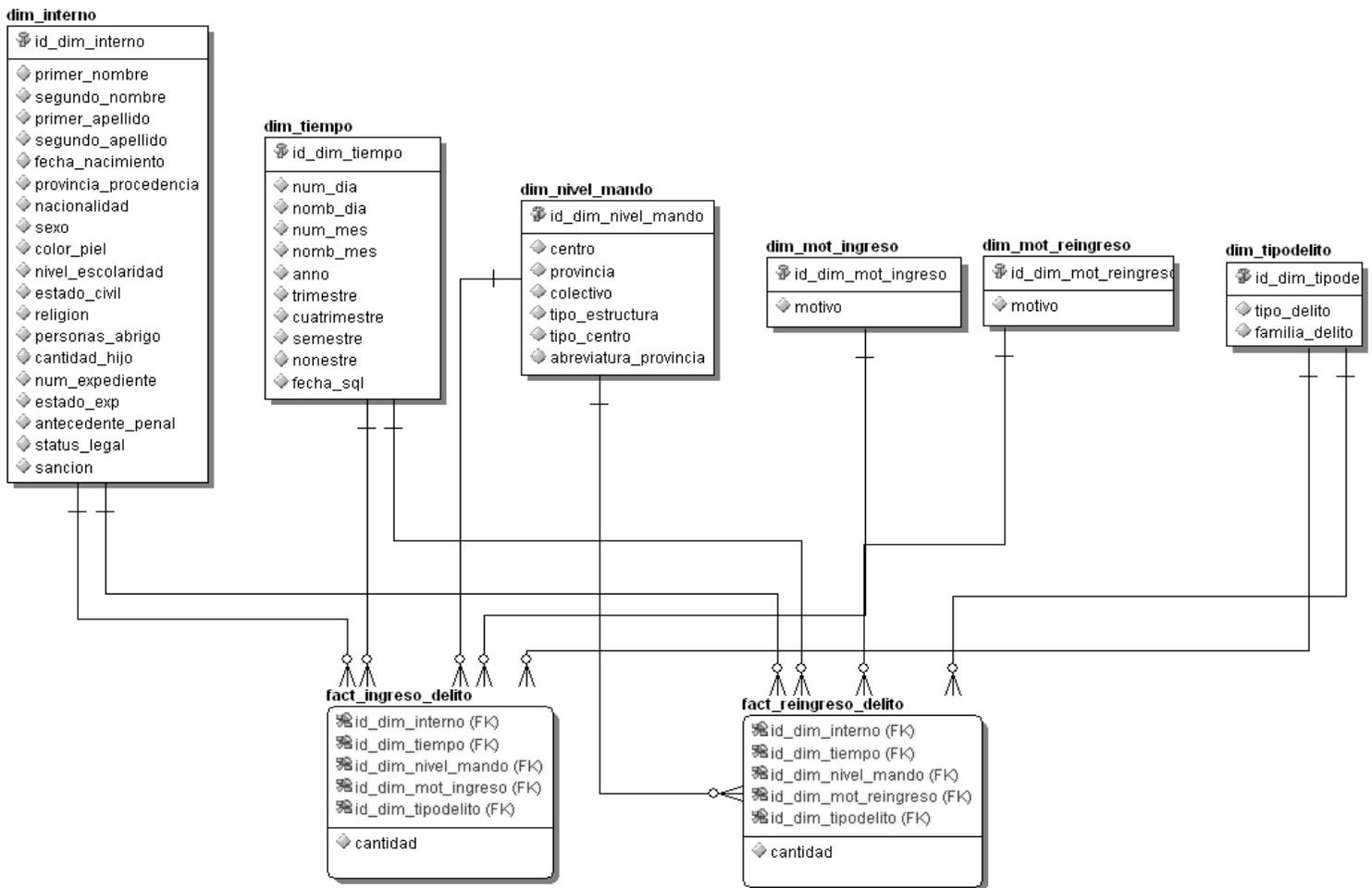


Figura 2.2: Modelo de Datos del proceso Ingreso

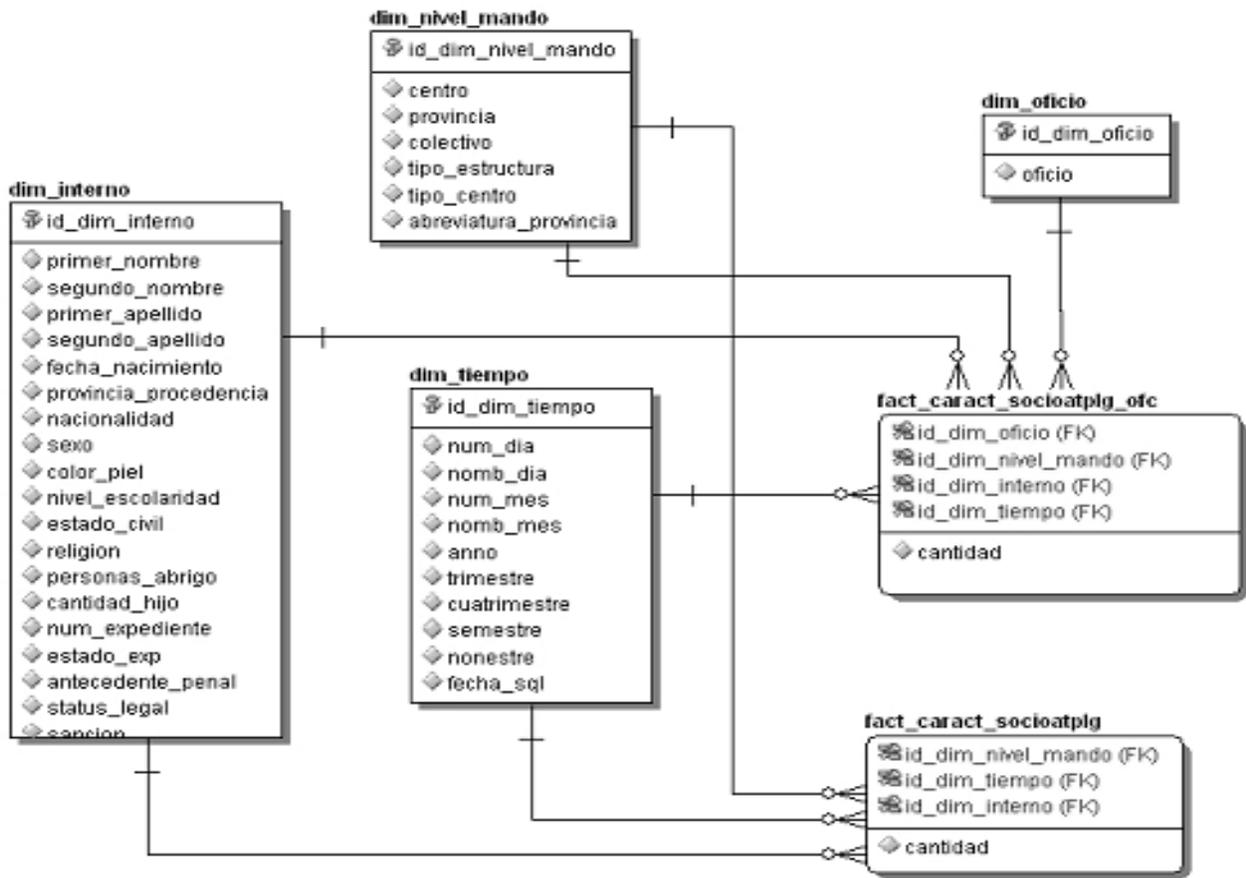


Figura 2.3: Modelo de Datos del proceso Características Socio - Antropológicas

2.3 Diseño de las ETL

En la elaboración del diseño de las ETL se analizó la fuente de datos origen para identificar qué datos son estrictamente necesarios y relevantes para poblar el mercado de datos, en este caso dicha fuente es la base de datos operacional del Sistema Informativo de la Dirección de Establecimientos Penitenciarios (SIDEPE). A continuación se muestran el diseño correspondiente a las dimensiones de los procesos del caso de estudio.

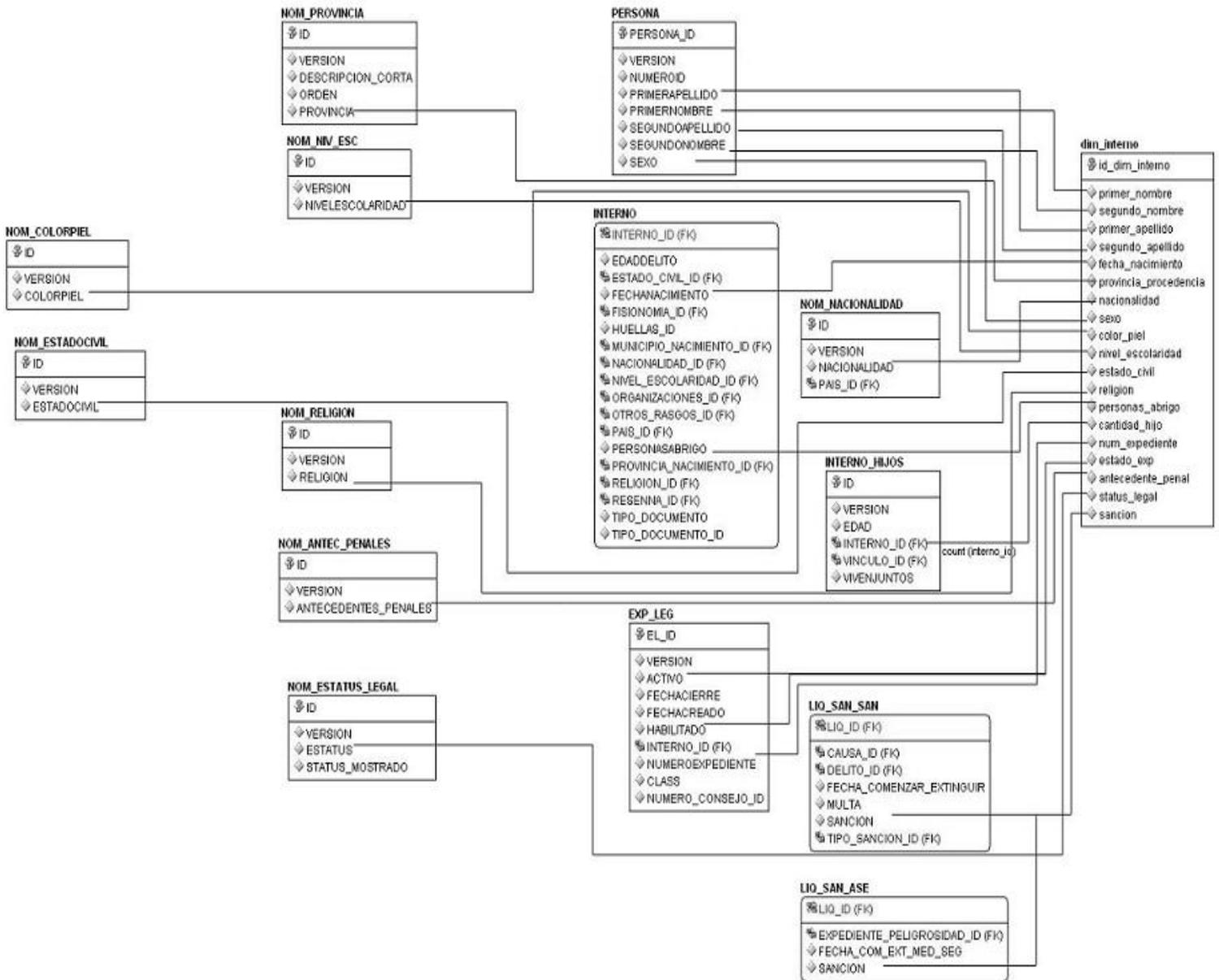


Figura 2.4: Diseño de la ETL de la dim_interno

Los atributos necesarios para poblar la dimensión interno, se obtienen a partir de las tablas PERSONA, INTERNO, INTERNO_HIJOS, EXP_LEG, LIQ_SAN_SAN, LIQ_SAN_ASE, NOM_COLOR_PIEL, NOM_PROVINCIA, NOM_ESTADOCIVIL, NOM_NIV_ESC, NOM_ANTEC_PENALES, NOM_RELIGION, NOM_ESTATUS_LEGAL, y NOM_NACIONALIDAD.

Capítulo 2: Análisis y Diseño del Mercado de Datos

De la tabla PERSONA se toman los atributos primerapellido, primernombre, segundoapellido, segundonombre, sexo.

De la tabla INTERNO se toman los atributos fechanacimiento, personasbrigo, y de la tabla INTERNO_HIJOS el id_interno para hacerle un count, con el objetivo de obtener la cantidad de hijos asociados a cada interno.

De la tabla EXP_LEG se toman los atributos activo, habilitado, numeroexpediente, y de las tablas LIQ_SAN_SAN, LIQ_SAN_ASE el atributo sancion.

De las tablas que son nomencladores como NOM_COLOR_PIEL, NOM_PROVINCIA, NOM_ESTADOCIVIL, NOM_NIV_ESC, NOM_ANTEC_PENALES, NOM_RELIGION, NOM_ESTATUS_LEGAL, y NOM_NACIONALIDAD se toman los atributos correspondientes al nombre de la tabla.

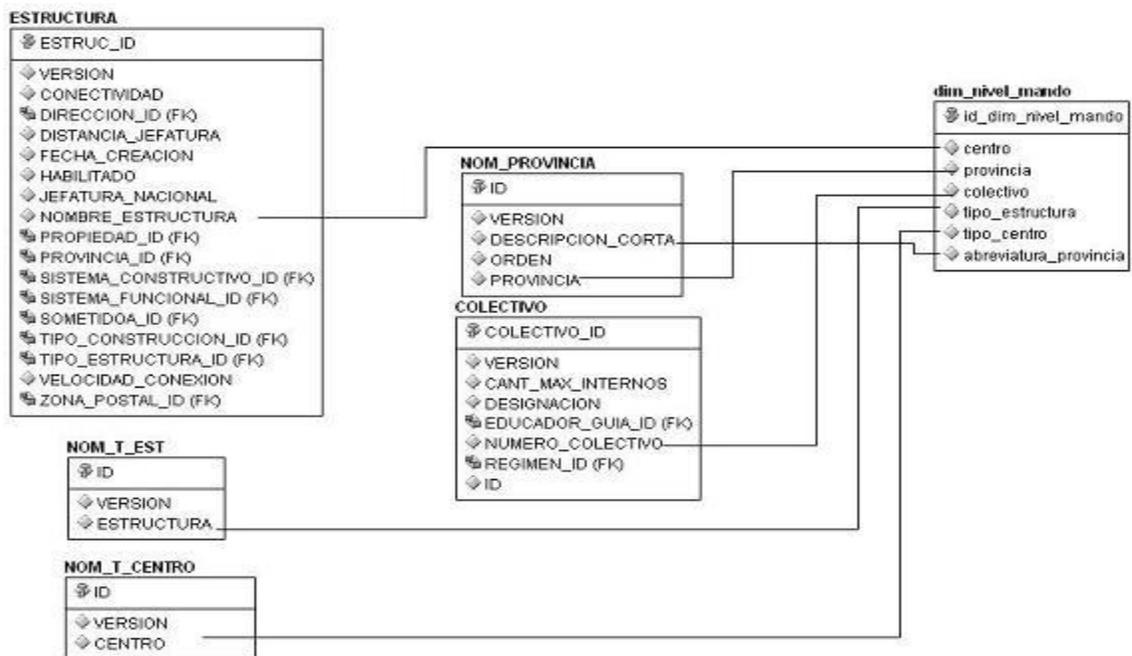


Figura 2.5: Diseño de la ETL de la dim_nivel_mando

Los atributos necesarios para cargar la dimensión nivel de mando, se obtienen a partir de las tablas ESTRUCTURA, COLECTIVO, NOM_T_EST, NOM_T_CENTRO y NOM_PROVINCIA.

De la tabla ESTRUCTURA se toma el atributo nombre_estructura, de COLECTIVO el numero_colectivo.

De las tablas que son nomencladores como NOM_T_EST, se toma estructura, en NOM_T_CENTRO se toma el centro y de NOM_PROVINCIA, provincia y descripción_corta.



Figura 2.6: Diseño de la ETL de la dim_mot_ingreso

Para poblar la dimensión motivo de ingreso, se necesita el atributo motivo_ingreso perteneciente a la tabla NOM_MOT_ING.



Figura 2.7: Diseño de la ETL de la dim_mot_reingreso

Para poblar la dimensión motivo de reingreso, se necesita el atributo motivo_reingreso perteneciente a la tabla NOM_MOT_REI.

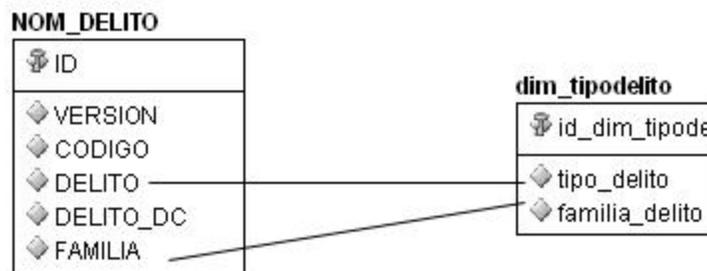


Figura 2.8: Diseño de la ETL de la dim_tipodelito

Los atributos necesarios para cargar la dimensión tipo delito se obtienen a partir de la tabla NOM_DELITO, con los atributos delito y familia.

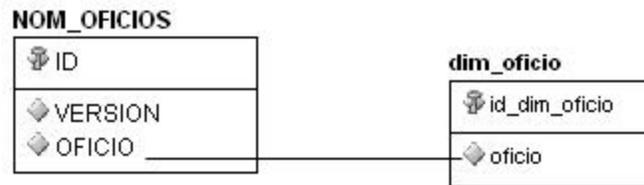


Figura 2.9: Diseño de la ETL de la dim_oficio

Para cargar la dimensión oficio solo es necesario el atributo oficio de la tabla NOM_OFICIO.

Conclusiones Parciales

En este capítulo se describieron los principales procesos que se llevan a cabo en el desarrollo del análisis y diseño del mercado de datos para el Área de Ingreso, los cuales fueron agrupados en 3 temas analíticos, y se identificaron un total de 84 requisitos de información. Se realizó la especificación de los requisitos, de las tablas de dimensiones y los hechos para los procesos del caso de estudio así como su modelo de datos correspondiente. También se diseñó el proceso de extracción, transformación y carga de los datos necesarios que poblarán el mercado.

CAPÍTULO 3: IMPLEMENTACIÓN Y PRUEBA

Introducción

En este capítulo se presenta la implementación del Mercado de Datos, y las pruebas realizadas. Como parte de la implementación se muestra: la implementación del proceso ETL de los datos pertenecientes a los procesos analizados en el capítulo anterior, otro elemento es la configuración del repositorio de datos que será necesario para la construcción de los reportes y por último la edición de los Tableros de Mando. En consecuencia de lo que describe la metodología Kimball para probar el mercado de datos, se realizan las pruebas y analiza el resultado de las mismas.

3.1 Implementación

El proceso de implementación engloba la realización de las ETL de los datos, la configuración del repositorio a través de la herramienta de Administración, la creación de los reportes en la herramienta de consulta y por último la confección de los tableros de mando.

3.1.1 Implementación de la Extracción Transformación y Carga (ETL)

La implementación de las ETL fue realizado con la herramienta Desing Center, que no es más que un cliente del OWB.

El proceso ETL se realiza para cada cubo, con el objetivo de poblar sus dimensiones y la tabla de hechos. Los procesos que se analizan están compuestos por 2 cubos cada uno, donde todos tienen relaciones con las dimensiones comunes, por lo que la explicación de la implementación de las ETL correspondientes a estas se especifica solamente para el primer cubo que se analiza.

Para realizar el proceso de las ETL para las tablas fact, los datos deben ser almacenados primeramente en un área temporal. Esta área, es justamente el paso intermedio entre los sistemas operacionales/transaccionales fuente y el área de presentación de datos. Aquí se obtienen los datos operacionales de la fuente origen, que luego son transformados para que puedan ser utilizados por los usuarios finales del mercado de datos. La extracción de los datos para esta área, se traduce en leer y copiar los datos fuentes que sean requeridos para el mercado de datos. Posteriormente, pueden realizarse transformaciones que deben ser sobre los datos originales ya sean correcciones, integración de múltiples fuentes de datos, conversión a formatos estándar, entre otras.

A continuación se especifica el proceso ETL de un cubo por cada proceso, de Ingreso, y Características Socio - Antropológicas:

- ETL para el cubo Ingreso
 - ETL para la DIM_MOT_INGRESO

La dimensión DIM_MOT_INGRESO, tiene como atributos el id subrogado id_dim_mot_ingreso, y motivo que guarda el motivo del ingreso, para su carga se toma como fuente de entrada el motivo_ingreso que se encuentran en la tabla NOM_MOT_ING y como identificador de la dimensión el valor generado por la secuencia SEC_MOTINGRESO.

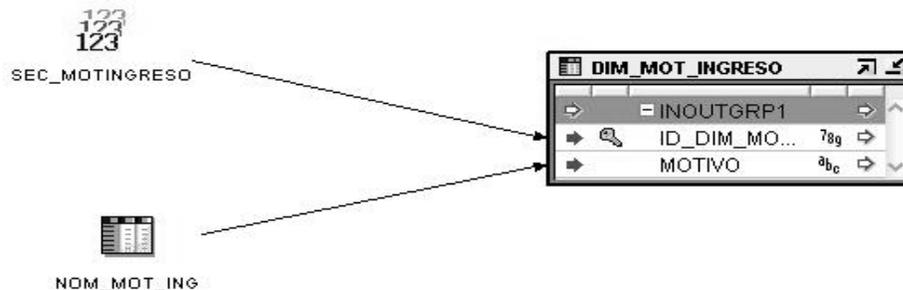


Figura 3.1: ETL DIM_MOT_INGRESO

- ETL para la DIM_TIPODELITO

La dimensión DIM_TIPODELITO, tiene como atributos el id subrogado id_dim_tipodelito, tipo_delito y familia_delito que guarda el tipo de delito y la familia delictiva a la cual pertenece. Para la carga de los datos, se toma como fuente de entrada el atributo delito y familia que se encuentran en la tabla NOM_DELITO. Debido a que los internos que ingresen al centro como asegurados, no tienen delitos asociados, se crea una constante delito que toma como valor 'No aplica delito por ser asegurado', para los atributos tipo_delito y familia_delito, la cual se une con los otros valores que toman de la fuente, a través del operador SET. El identificador de la dimensión toma el valor generado por la secuencia SEC_DELITO.

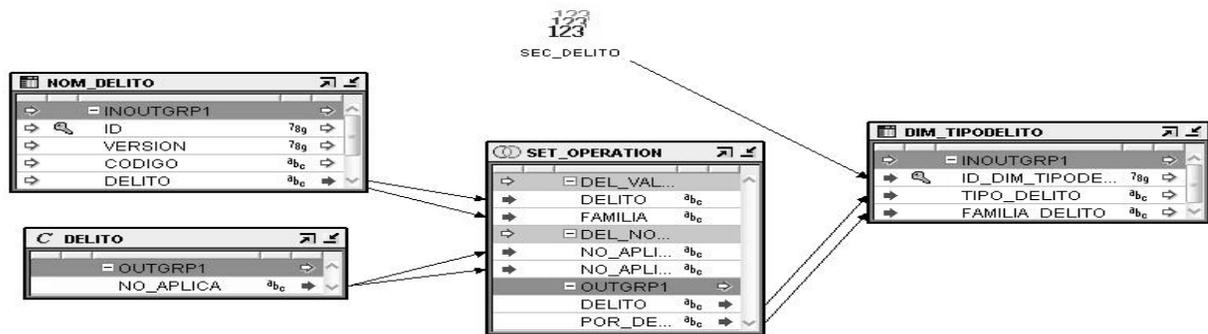


Figura 3.2: ETL DIM_TIPODELITO

- ETL para la dimensión común NIVEL_MANDO

Para la carga de la dimensión DIM_NIVEL_MANDO se toma como fuente de entrada los datos necesarios que se encuentran en las tablas: ESTRUCTURA, COLECTIVO, COLECTIVO_LOCAL, LOCAL, ÁREA, NOM_T_CENTRO, CENTRO_PENIT, NOM_PROVINCIA Y NOM_T_EST.

Se procede a utilizar el operador JOINER titulado CENTRO_PENITENCIARIOS, encargado de relacionar las tablas antes mencionadas.

Luego se utiliza el operador FILTER, el objetivo del mismo es obtener los datos de las estructuras de tipo Jefatura, o los Centros Penitenciarios que tienen asociados colectivos.

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

- **TO_CHAR:** dado un dato de entrada de tipo number o date, lo convierte a tipo de dato varchar o char, con el formato que se le especifique.
- **FUNC_TIPO_CENTRO:** dado el tipo de estructura y el tipo de centro devuelve el tipo de centro correspondiente o 'No Aplica'.
- **FUNC_COLECTIVO:** dado el tipo de estructura y el número de colectivo devuelve el número de colectivo correspondiente o 'No Aplica'.

Posteriormente se emplea el operador SET_OPERATION titulado UNIR_ESTRUCTURAS, su función en este caso es la de combinar los resultados de las consultas para definir el valor que toma colectivo y

tipo de centro para las estructuras de tipo Jefatura y Centro Penitenciario, y los demás datos que debe de almacenar el nivel de mando.

Se utiliza el operador CONSTANT, titulado NA_COLECT para cargar datos que no se encuentran en la base de datos y que son necesarios en la dimensión dim_nivel_mando.

El valor que toma el identificador de la dimensión es el generado por la secuencia SEQ_DIM_NIVEL_MANDO.

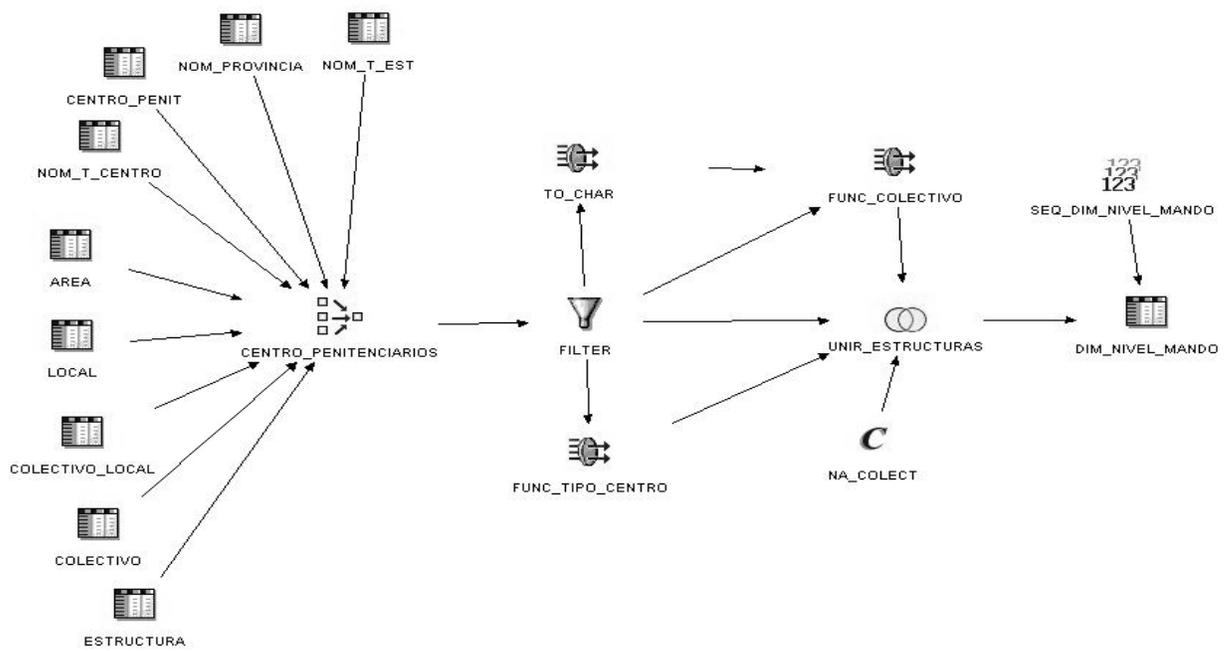


Figura 3.3: ETL DIM_NIVEL_MANDO

- ETL para la dimensión común TIEMPO

La carga de la dimensión DIM_TIEMPO recibe como parámetros, la fecha de inicio y la cantidad de años que se quiere generar.

Posteriormente se utiliza el operador de función de tabla DAY_TABLE_FUNCTION para crear cada una las fechas que se encuentran dentro del intervalo. Función desarrollada por OWB.

Capítulo 3: Implementación y Prueba del Mercado de Datos

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

CAL_TRIMESTRE: dado el número del mes devuelve el número del trimestre correspondiente.

CAL_CUATRIMESTRE: dado el número del mes devuelve el número del cuatrimestre correspondiente.

CAL_SEMESTRE: dado el número del mes devuelve el número del semestre correspondiente.

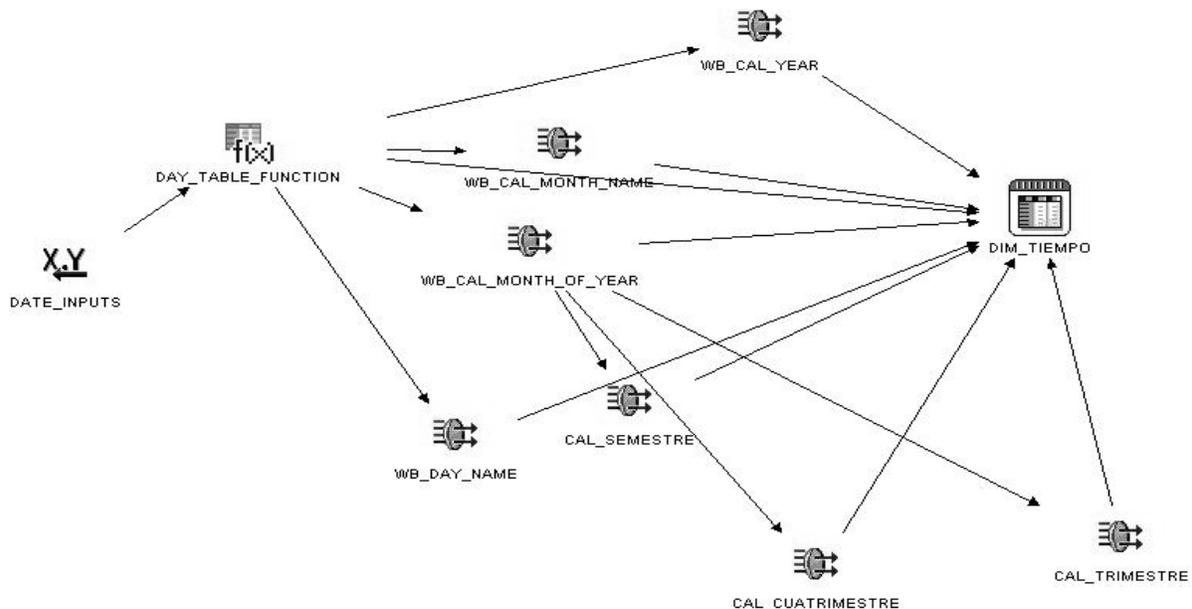


Figura 3.4: ETL DIM_ TIEMPO

- ETL para la dimensión común INTERNO

Para la carga de la DIM_INTERNO se toma como fuente de entrada, los datos necesarios que se encuentran en las tablas: PERSONA, INTERNO, DATOS_FIS, EXP_LEGAL, SIT_LEGAL, PROCESO, INTERNO-HIJOS, DOCUMENTO, LIQ_SAN_SAN, LIQ_SAN_ASE, NOM_COLORPIEL, NOM_ESTADOCIVIL, NOM_NIV_ESC, NOM_RELIGION, NOM_PROVINCIA, NOM_NACIONALIDAD, NOM_ANTEC_PENALES, NOM_ESTADO_PROCESO, NOM_T_DOC, NOM_ESTATUS_LEGAL y la vista ULTIMO_STATUS.

Capítulo 3: Implementación y Prueba del Mercado de Datos

Con el objetivo de obtener el valor del atributo, cantidad de hijos para cada interno, primeramente se utiliza un operador AGGREGATOR nombrado CANTIDAD-HIJOS-INTERNO, que devuelve los internos que tienen hijos, y la cantidad asociada al mismo, posteriormente se utiliza el operador SET nombrado INTERNO-SIN-HIJOS, para obtener los internos que no tienen hijos, y por último un SET nombrado REL-INT-HIJOS que unifica todos los valores del atributo cantidad de hijos del interno.

Para obtener la sanción que le imponen al interno, primero se utiliza un operador SET, nombrado LIQ_SANCION, que unifica en un mismo atributo los valores que puede tomar la sanción, ya que provienen de diferentes tablas del origen. Posterior a este paso se relacionan las tablas que unen al interno con la sanción, a través del operador JOINER nombrado OBTENER_SANCION. Luego se utiliza el operador FILTER, para obtener los documentos de tipo 'Liquidación de Sanción', asociado a los procesos que tiene el interno, con estado 'Cumpliendo' o 'Total Conjunta', y por último el operador AGGREGATOR para escoger el documento que tiene la sanción que está actualmente cumpliendo.

Para obtener la relación del interno con todos los datos generales que se almacenan en la dimensión, se utiliza el operador JOINER. Debido a que muchos datos que devuelve este operador tienen valor nulo o vacío, y que vienen del origen con una nomenclatura distinta a la que debe de almacenar en el mercado, se utilizan funciones de transformación, algunas de estas son:

- **FUNC_NO_DEFINIDO:** dado un dato de entrada, verifica si está vacío o nulo, en dicho caso devuelve 'Por Definir', y en caso contrario, el mismo valor de entrada. Debido a que esta función la utilizan varios atributos, se utilizó el operador de transformación con el nombre del mismo atributo, ejemplo: PROVINCIA_NAC, NIVEL_ESC, etc.
- **FUNC_GENERO:** dado el valor del sexo, que viene del origen con tipo de dato numérico, lo transforma en 'Femenino' o 'Masculino', y en caso de que esté vacío, toma valor 'Por Definir'.
- **FUNC_ESTADO_EXPEDIENTE:** recibe como parámetro de entrada los atributos activo y habilitado que viene del origen con tipo de dato numérico, y lo transforma en 'Abierto Habilitado', 'Abierto No Habilitado' o 'Cerrado'.
- **FUNC_SIN_SANCION:** dado el dato de entrada sanción, verifica que no esté vacío o tenga valor nulo, en dicho caso devuelve 0, en caso contrario el mismo valor de entrada.

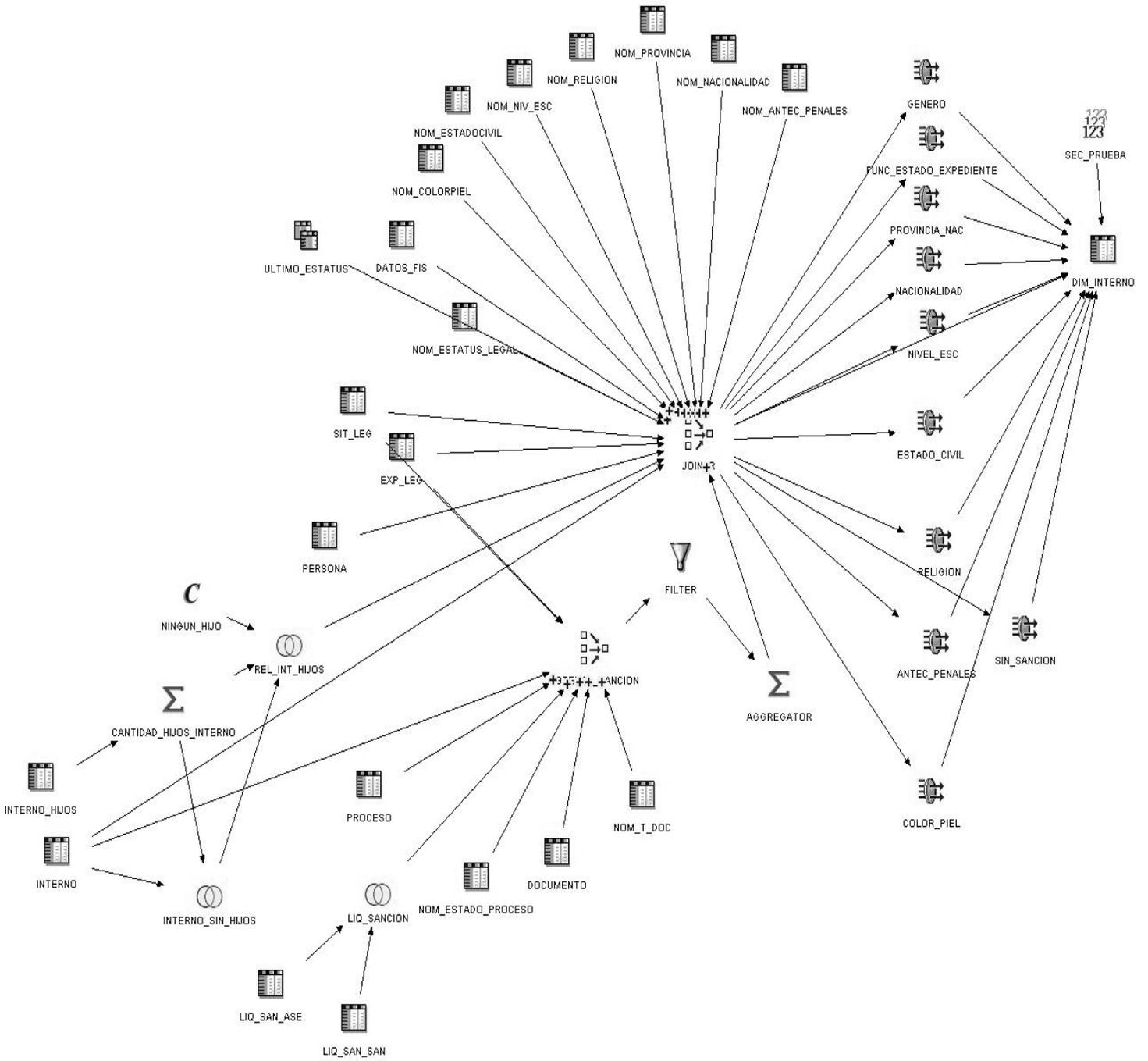


Figura 3.5: ETL DIM_INTERNO

- ETL para el STG_UBICACION

Para la carga del STG_UBICACION se toma como fuente de entrada los datos necesarios que se encuentran en las tablas: COLECTIVO, SIT_LEG, EXP_LEG, INTERNO, UBICACION, CAMA, LITERA, LOCAL_CON_LITERA, LOCAL, COLECTIVO_LOCAL, CENTRO_PENIT, ESTRUCTURA, NOM_PROVINCIA, NOM_T_CENTRO, NOM_T_EST, NOM_ESTATUS_LEGAL, AREA y la vista ULTIMO_STATUS.

Se utiliza un operador JOINER, nombrado UB_INT_CP, con el objetivo de relacionar las tablas antes mencionadas, para obtener la ubicación del interno dentro del centro penitenciario.

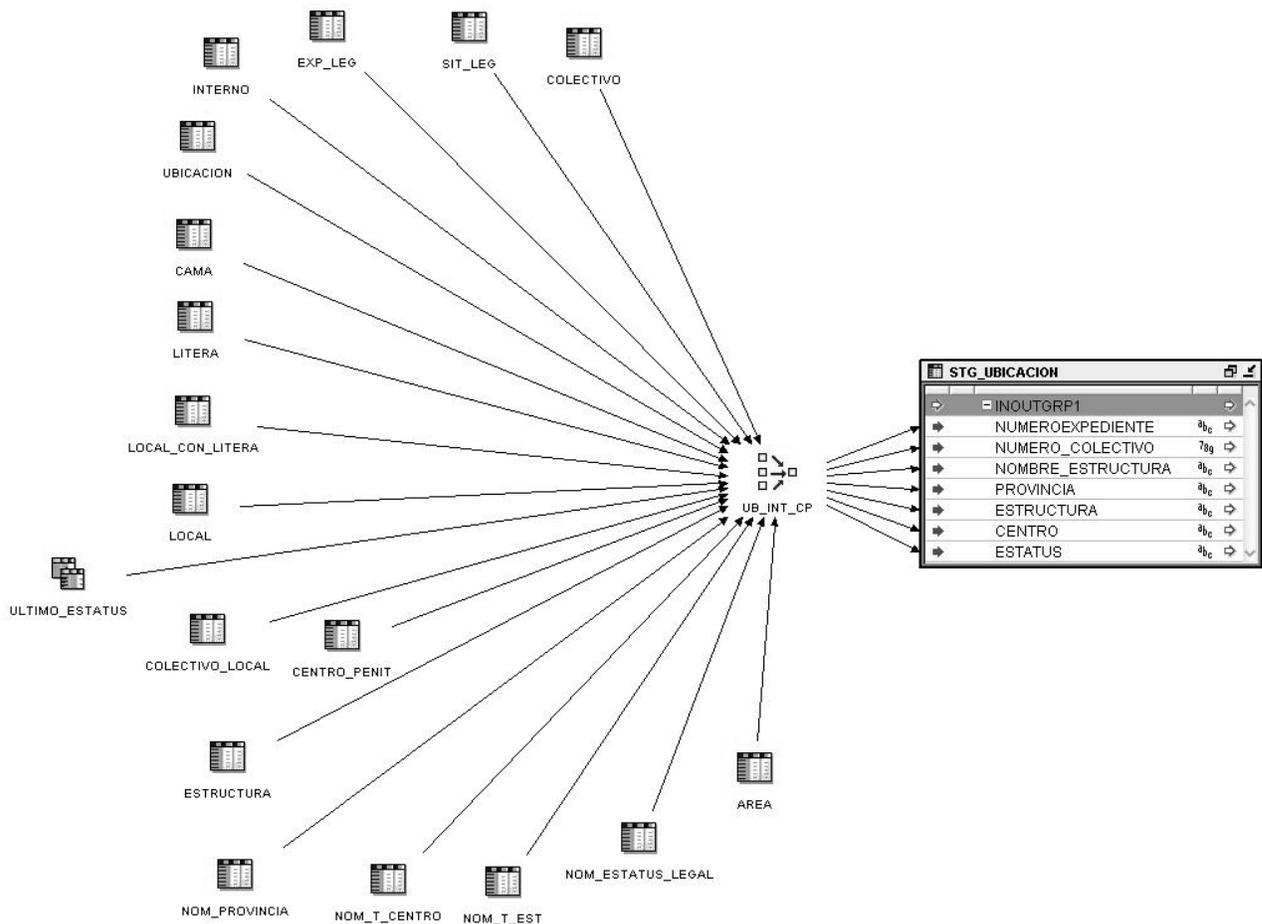


Figura 3.6: ETL STG_UBICACION

Capítulo 3: Implementación y Prueba del Mercado de Datos

- ETL para el STG_INGRESO_DELITO

Para la carga del STG_INGRESO_DELITO se toma como fuente de entrada los datos necesarios que se encuentran en las tablas: INTERNO, EXP_LEG, SIT_LEG, PROCESO, INGRESO_PROCESO, INGRESO, INGRESO_REAL y NOM_MOT_ING y DOCUMENTO, para obtener los datos relacionados al ingreso, y las tablas REMISION_ACUSADO, DOC_RAD_CAU, AUTO_CAM_LUGAR, CONS_DEN, NOM_DELITO, DELITO, COP_SEN_DELITO, COP_SEN, AUTO_DEC_INS y DOC_FISC, para obtener los documentos donde se registran los delitos asociados al proceso de ingreso.

Se utiliza un operador JOINER, por cada documento que registra delito, el cual es nombrado según el tipo de documento que sea, para tomar todos los posibles delitos que puede tener asociado un interno, los cuales posteriormente son unificados por el operador SET, nombrado DEL_DOC. Para relacionar los documentos que recogen el delito y que a su vez están asociados al proceso de ingreso, se utiliza un operador JOINER, nombrado DATOS_ING_DEL, pero como los internos que ingresan como “Asegurados” no se les registra delito, el operador de unión va a devolver todos los ingresos aunque no todos tengan asociados delitos. Al final los datos de salidas del operador de unión pasan por un DESDUPLICADOR para eliminar los datos que se repitan.

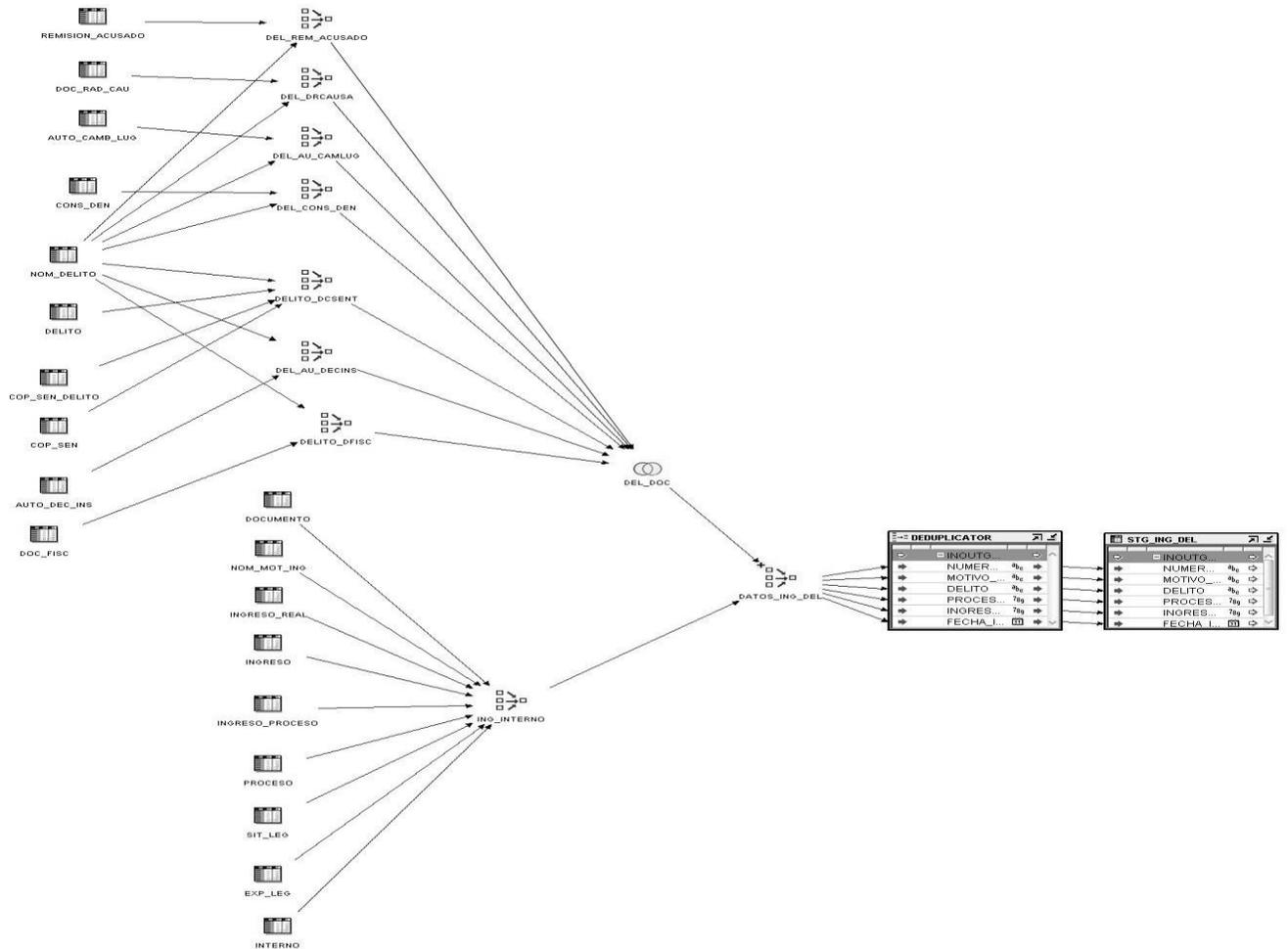


Figura 3.7: ETL STG_INGRESO_DELITO

- ETL para la FACT_INGRESO_DELITO

Para la carga de la FACT_INGRESO_DELITO se utiliza los datos de entrada del STG_UBICACION y el STG_INGRESO_DELITO, donde a partir del STG_UBICACION se obtienen los internos ubicados en los centros penitenciarios, y en el STG_INGRESO_DELITO se obtienen los procesos de ingreso que tienen los internos con el delito registrado para cada ingreso, en caso de los asegurados el valor del delito toma valor nulo.

Debido a que los internos que ingresan como “Asegurados” no tienen delitos asociados al ingreso, por lo que la salida del STG_INGRESO_DELITO, va a tener valores nulos para el delito, se utilizan 2 filtros,

Capítulo 3: Implementación y Prueba del Mercado de Datos

el nombrado DEL_CON_VALOR, devuelve los ingresos que tienen delitos registrados, y el nombrado DEL_VACIO, devuelve los ingresos que no tienen asociados delitos. Posteriormente se utiliza un operador SET, nombrado INGRESOS_ASEG, que devuelve los procesos de ingresos de los “Asegurados”. La salida de este SET es la entrada para el operador de unión EXP_LEG_ASEG_DEL, que se encarga de retornar todos los expedientes legales de los internos asegurados con sus respectivos procesos de ingresos. Mediante el SET VALORES_DEL_ING, se obtienen todos los internos con sus ingresos y el delito asociado a ellos, donde para el caso de los “Asegurados” se registra “No aplica delito por ser asegurado”.

Para recoger los valores que van a poblar el hecho, se utiliza los datos de salida del operador VAL_POBLAR_HECHO, que son necesarios para hacer la comparación con los operadores de consulta de llaves, con el objetivo de obtener los datos de las dimensiones que coincidan con los valores cargados de la base de datos fuente, para finalmente registrarlo en el hecho.

La constante CANT, es la medida que se registra en el hecho, que en este caso siempre va a tomar valor 1, ya que este proceso representa un hecho sin hecho.

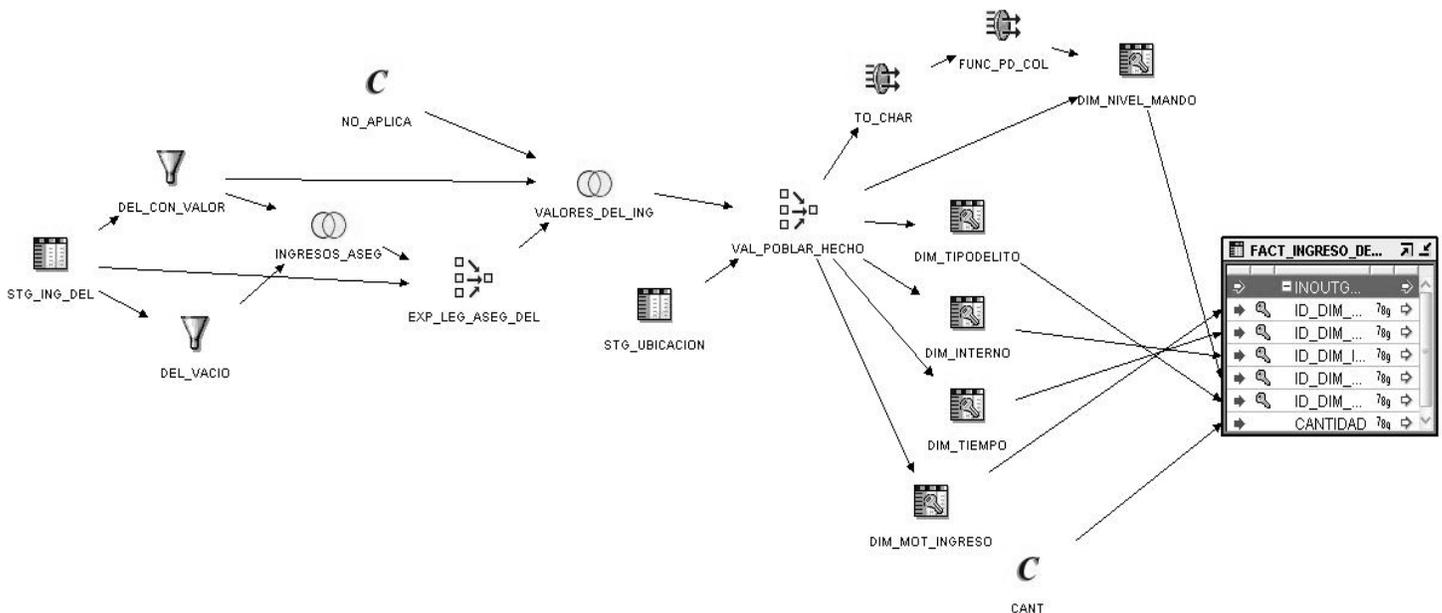


Figura 3.8: ETL FACT_ INGRESO_DELITO

➤ ETL para el cubo Características Socio-Antropológicas

El cubo de Características Socio-Antropológicas solo tiene relación con las dimensiones comunes, debido que en este proceso no hay ocurrencia de hecho, pero se modela para darle respuesta a los requisitos de información definidos por la Dirección de Establecimientos Penitenciarios, para conocer el comportamiento de las características de los internos en los centros. El proceso ETL de estas dimensiones ya fueron especificadas anteriormente.

- ETL para el STG_CARACT_SA

Para la carga del STG_CARACT_SA se toma como fuente de entrada los datos necesarios que se encuentran en las tablas: EXP_LEG e INTERNO.

Se utiliza un operador JOINER, con el objetivo de relacionar las tablas antes mencionadas, para obtener los expedientes de los internos.

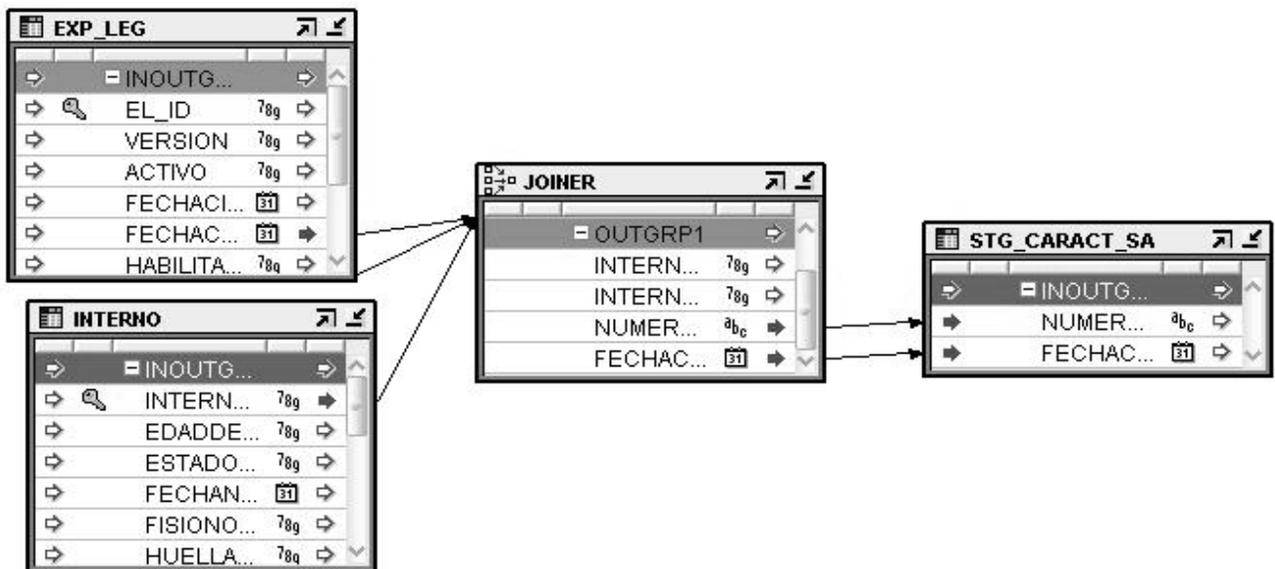


Figura 3.9: ETL STG_CARACT_SA

- ETL para el FACT_CARACT_SA

Para la carga de la FACT_CARACT_SA se utiliza los datos de entrada del STG_UBICACION y el STG_CARACT_SA.

Posteriormente se utiliza un operador de unión, nombrado DATOS_CARACT_SA, para relacionar las tablas antes mencionadas.

Los datos de salida de este operador, se utilizan para hacer la comparación con los operadores de consulta de llaves, pero primeramente se le aplican algunas transformaciones a los datos inconsistentes que vienen del origen. El objetivo de utilizar los operadores de consulta de llaves, es obtener los datos de las dimensiones que coincidan con los valores cargados de la base de datos fuente, para finalmente registrarlo en el hecho.

La constante CANTIDAD, es la medida que se registra en el hecho, que en este caso siempre va a tomar valor 1, porque no hay ocurrencia de hecho.

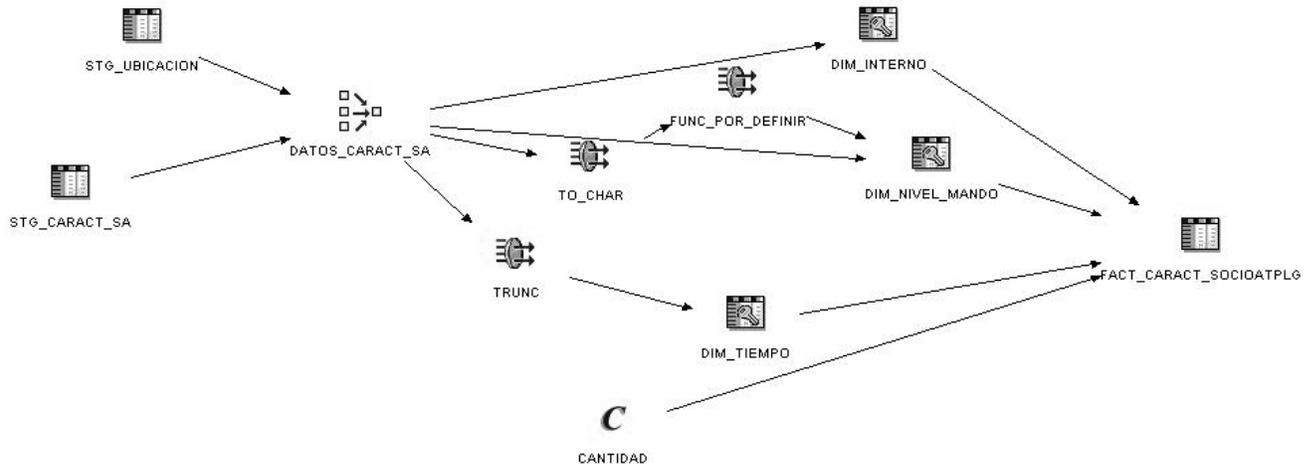


Figura 3.10: ETL FACT_CARACT_SA

3.1.2 Configuración de la herramienta de Administración

Luego de realizar el proceso de Extracción, Transformación y Carga de los datos, se configura el repositorio, que es el encargado de relacionar las herramientas de consulta con los datos almacenados en el mercado. Para realizar la configuración se utiliza la herramienta de administración (Administration Tool).

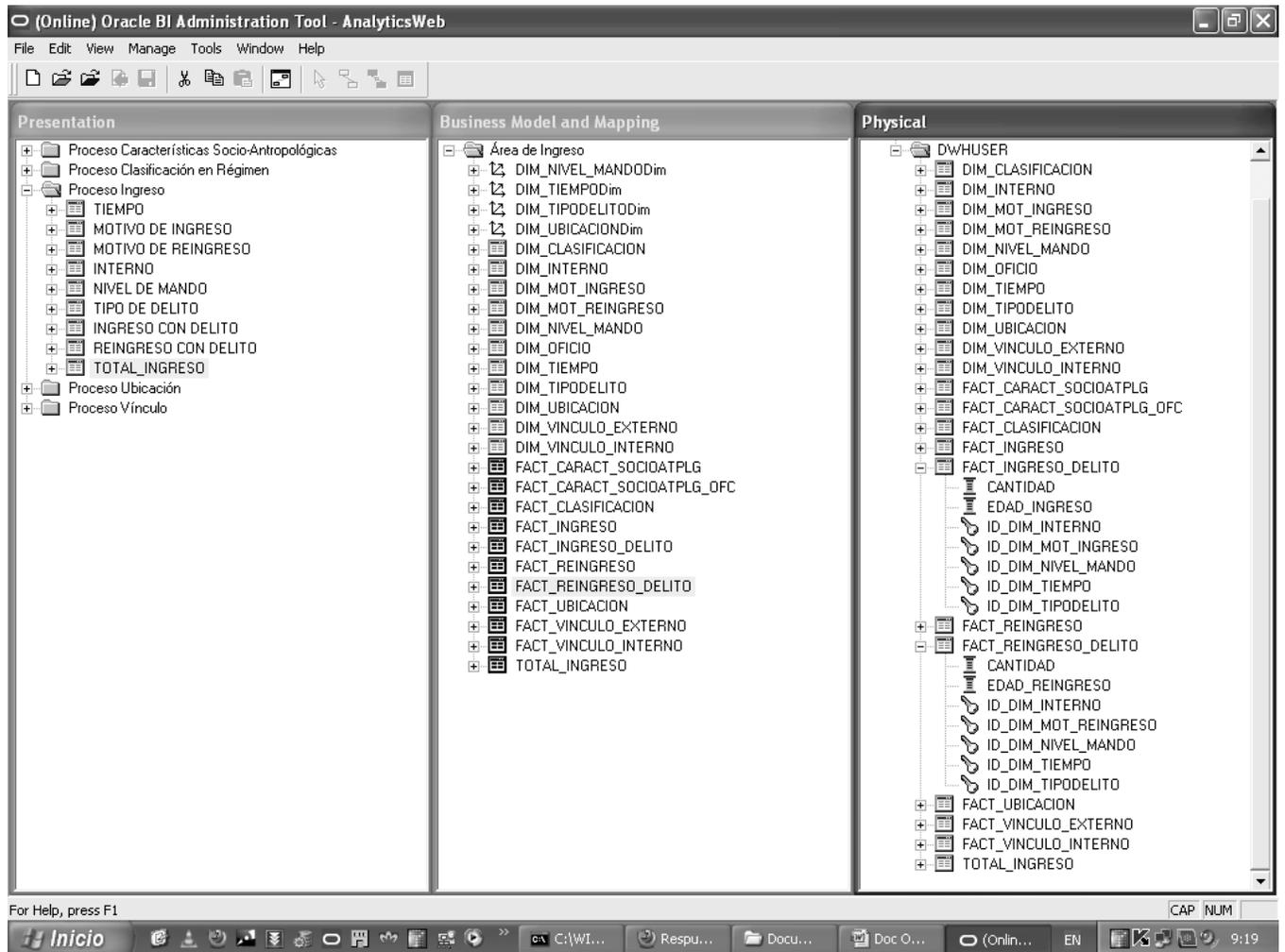


Figura 3.11 Administration Tool

El repositorio está compuesto por tres capas, las cuales son: la capa física, la del modelo de negocio y la capa de presentación donde residen los datos que serán consultados para la creación de los reportes. Como primera dirigida a crear el repositorio, se importan las tablas de los cubos pertenecientes al mercado de datos, en la capa física, importando a su vez las restricciones de llaves primarias y llaves foráneas que permiten las relaciones entre las tablas. Los valores que sean necesarios para la creación de los reportes que no se encuentren almacenados en el mercado, debido a que son valores derivados, se agregan en esta capa, para posteriormente editar la fórmula en la capa de negocio. En este caso se agrega el campo edad en las tablas FACT_INGRESO_DELITO y

Capítulo 3: Implementación y Prueba del Mercado de Datos

FACT_REINGRESO_DELITO, que corresponde con la edad de ingreso, o reingreso según sea el caso, además de la tabla TOTAL_INGRESO con el objetivo de crear un atributo que calcule el total de ingresos y reingresos.

La capa intermedia es la del modelado del negocio donde se especifican las correspondencias entre el modelo de negocio y los esquemas de la capa física. En esta capa se le asigna a los datos el enmascaramiento de negocio de los usuarios, con una nueva definición de nombre distinta al valor real que se almacena en la BD. Además se definen las jerarquías de las dimensiones NIVEL DE MANDO, TIEMPO, TIPO DELITO y UBICACION para la navegabilidad en los reportes, y a la medida cantidad de las tablas fact se le asigna sus respectivas reglas de agregaciones. También se edita la función para calcular la edad de ingreso o reingreso, la cual se obtiene a partir de los atributos fecha de nacimiento de la dimensión interno, y de la fecha en que se le realizó el ingreso o reingreso obtenida de la dimensión tiempo. Otro atributo que se calcula es el total de ingresos, a partir del atributo cantidad de las tablas fact correspondientes al ingreso y reingreso.

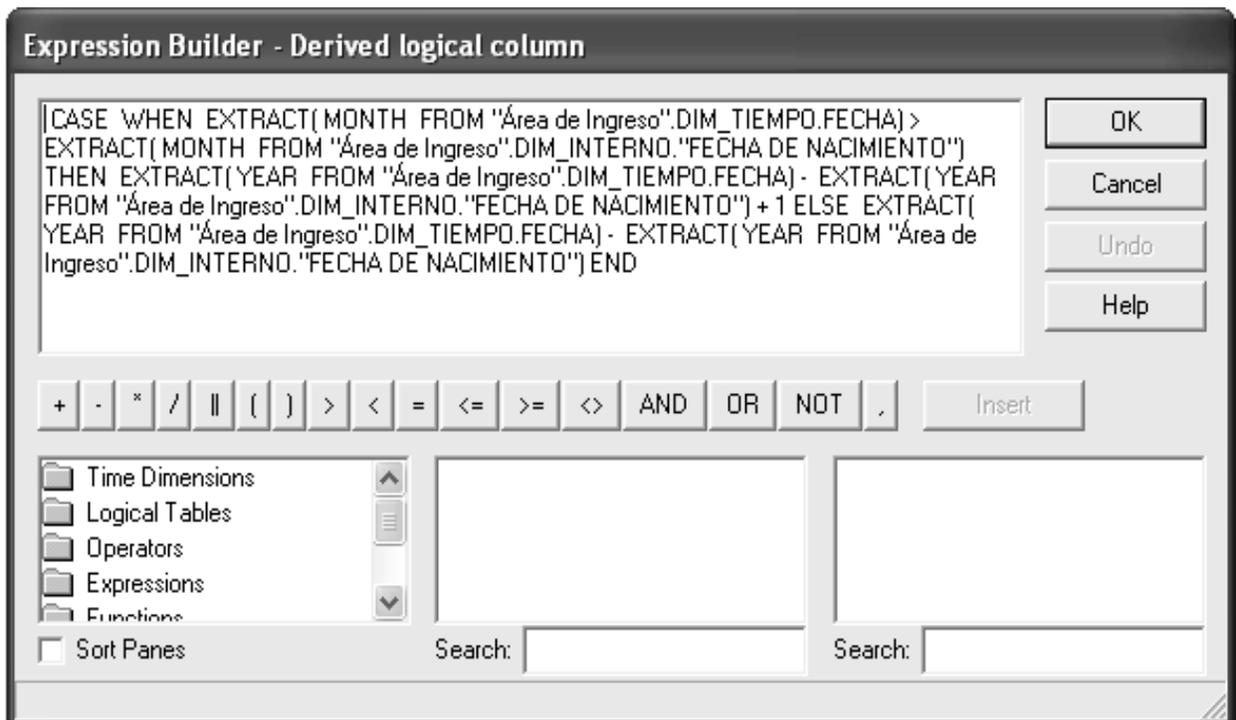


Figura 3.12 Función para calcular la edad

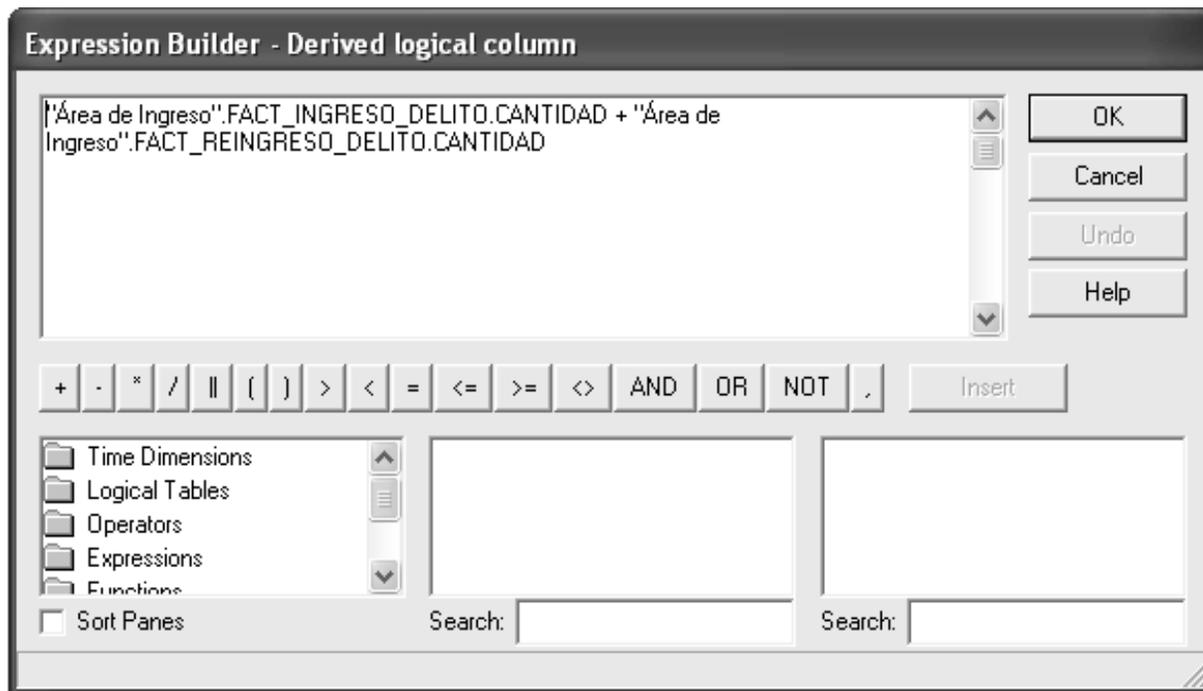


Figura 3.12 Función para calcular el total de ingresos

La última capa es la de presentación, en la cual se muestran cómo serán vistos los datos por los usuarios finales a través de las herramientas de consulta. Permite organizar las tablas por áreas temáticas, renombran las tablas y columnas con valores representativos del negocio.

3.1.3 Construcción de reportes en la herramienta de Respuesta

Una vez configurado el repositorio de datos necesario para realizar las consultas y el análisis de la información para los usuarios, se procede a crear los reportes correspondientes a los procesos de Ingreso y Características Socio-Antropológicas. Se utiliza como herramienta de trabajo el Oracle BI Answers en la cual se crearán los reportes, la cual permite crear gráficos interactivos, tablas dinámicas e informes. A continuación se muestran imágenes de algunos de los reportes:

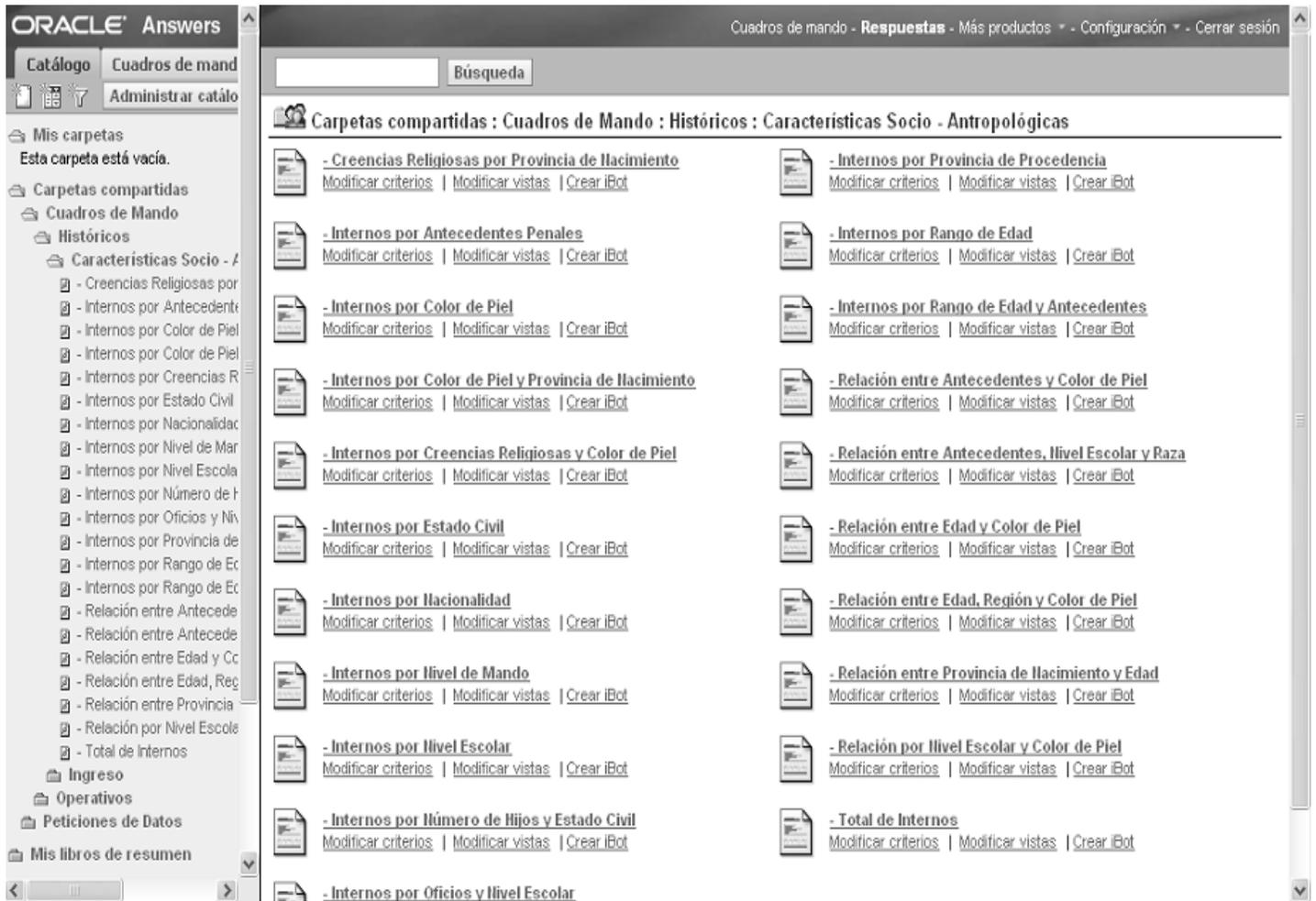


Figura 3.13: Pantalla del Oracle Answers, asociada a los reportes del proceso Características Socio-Antropológicas.

Una de las facilidades de esta herramienta es la posibilidad que le brinda al usuario para poder crear y configurar nuevos reportes cuando lo necesite, debido a que su creación puede ser a través de consultas ad-hoc, por lo que los usuarios del negocio no necesitan ser expertos. Permite además crear nuevos filtros, peticiónes de datos y editar fórmulas complejas. A continuación se muestra un ejemplo de fórmula editada para agrupar el rango de edad, y el resultado de un reporte a partir de una vista gráfica:

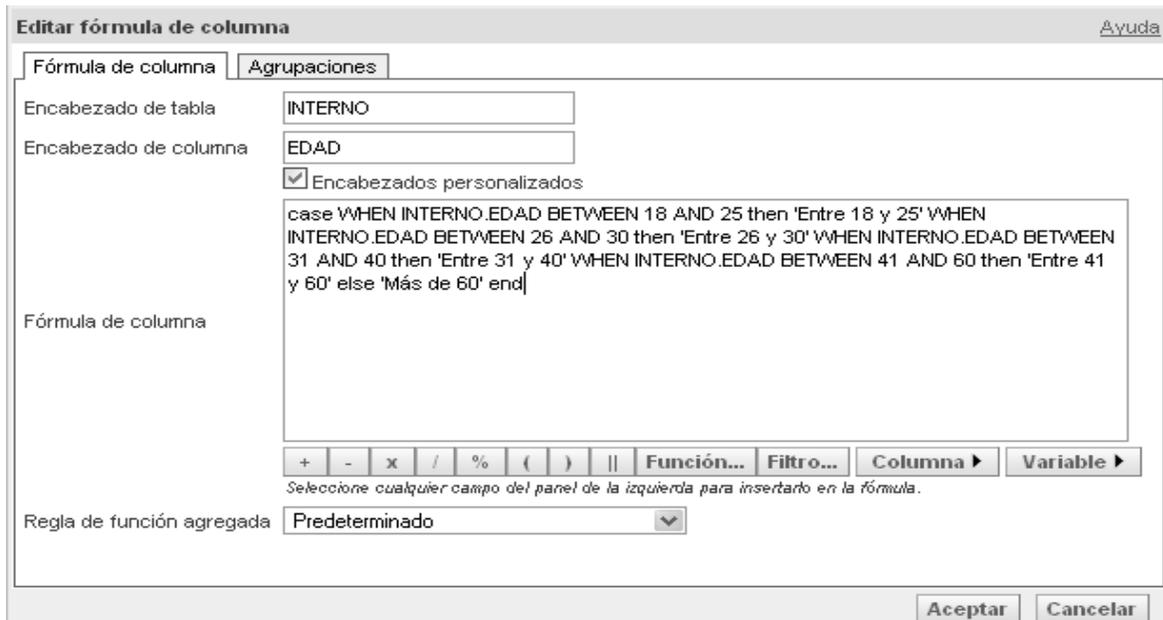


Figura 3.14: Fórmula para agrupar las edades por rango

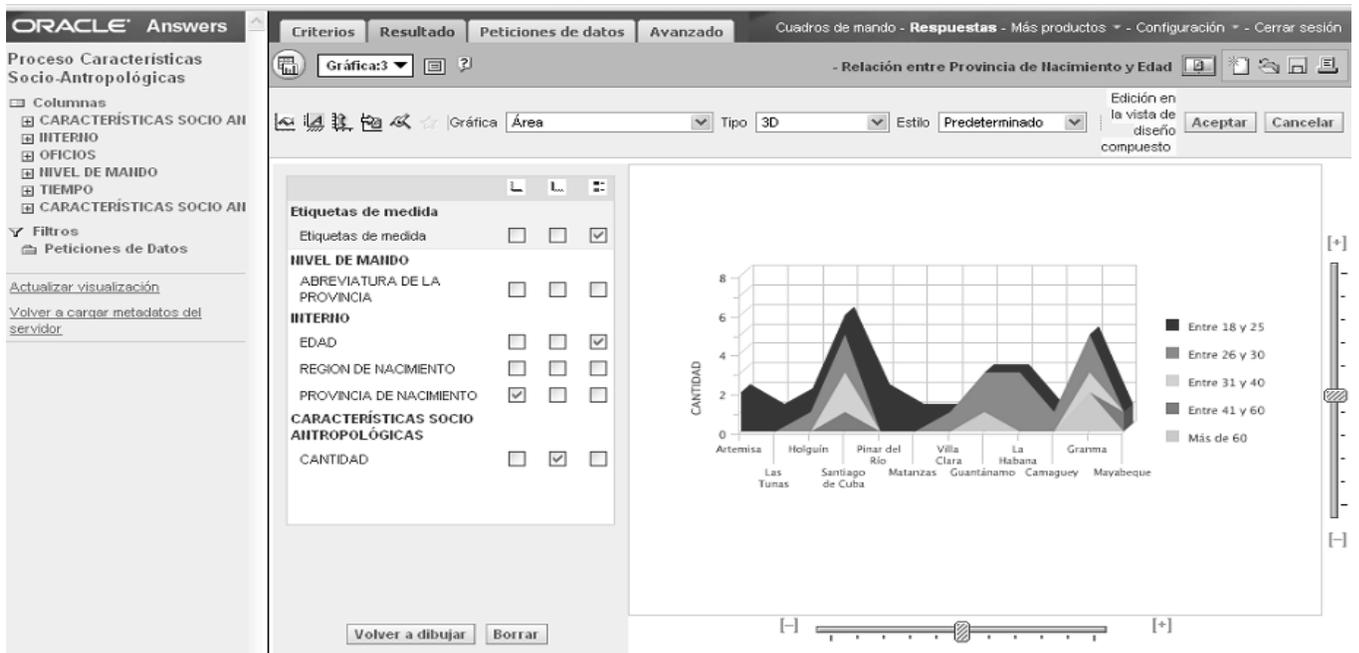


Figura 3.14: Vista Resultado de un reporte en el Oracle Answers, asociada a un gráfico de área, donde muestra el rango de edad de los internos por provincia de procedencia

3.1.4 Configuración del Tablero de Mando

Luego de crear todos los reportes correspondientes a los requisitos de información, se comienza con la configuración del tablero de mando, que será visualizado por los especialistas de dirección del Área de Ingreso. Para ellos se crearon dos tableros de mando, uno para mostrar los reportes históricos donde se analiza toda la información almacenada en el rango de tiempo de 2008-2012, el cual permite filtrar los reportes mediante un rango de fechas, y el otro para los reportes operativos, donde solo se mostrará el comportamiento de los indicadores analizados para una semana. A continuación se muestra el tablero de mando histórico:



Figura 3.15: Tablero de Mando Histórico para el Área de Ingreso

Capítulo 3: Implementación y Prueba del Mercado de Datos

Ambos tableros están compuestos por 7 páginas, correspondientes a los procesos que se analizan en el área. Se muestran los reportes predefinidos más relevantes de cada proceso, pero el usuario tiene la posibilidad de configurarlo según sus necesidades. En algunos gráficos se le da la posibilidad al usuario de cambiar la vista del mismo, además de que todos los reportes mostrados pueden ser filtrados por la petición de datos del encabezado.

3.1.5 Tipos de Procesamiento

Los tipos de procesamientos proporcionan realizar un análisis de los datos de manera interactiva a través de las dimensiones y sus jerarquías. A continuación se describen las utilizadas en el mercado de datos:

Drill-down: Permite apreciar los datos en un mayor detalle, moviéndose desde un nivel más general a un nivel de detalle en una jerarquía. Esto brinda la posibilidad de introducir un nuevo nivel o criterio de agregación en el análisis. Un ejemplo de jerarquía es: año, trimestre, mes y día en la dimensión tiempo.

Pivot: Permite girar los ejes de un cubo para proporcionar una representación alternativa de los datos y poder analizar la información desde diferentes perspectivas.

Slice y Dice: Permite aplicar restricciones a una dimensión de un cubo mediante la elección de un valor o un rango de valores. Dice es la combinación de varios slice, ya que se le aplica restricciones a dos o más dimensiones.

3.2 Pruebas

Las pruebas son un conjunto de actividades en las cuales un sistema es ejecutado bajo condiciones o requisitos específicos, donde los resultados son observados y registrados para dar una evaluación de algún aspecto del sistema determinar así la calidad del mismo.

Las pruebas fueron realizadas con el objetivo de comprobar que el sistema cumple con los requisitos de información detectados en etapas anteriores. Se realizan para identificar posibles fallos en la implementación del sistema.

- **Descripción de los métodos de prueba a realizar**

Los métodos de pruebas definen que estrategia seguir en cuanto a verificación y validación del sistema, ya que están diseñados con el propósito de descubrir fallos y no para demostrar que el software funciona.

Existen dos métodos de prueba: Pruebas de caja blanca, estas pruebas comprueban los caminos lógicos dentro del sistema y son derivadas a partir de las especificaciones internas del diseño o del código. Otro de los métodos son las Pruebas de caja negra las cuales pretenden demostrar que las entradas al software se aceptan de forma adecuada y se produce el resultado correcto. Dada la necesidad de que las pruebas abarquen los requerimientos, las funciones y las respuestas de la aplicación, se decide utilizar las pruebas de caja negra. Este método presenta varias técnicas de pruebas a emplear. Para realizar las pruebas al Mercado de Datos se utilizó la técnica lista de chequeo, que es la definida por la metodología utilizada, a continuación se describe con más detalle el método y la técnica:

- **Pruebas de Caja Negra o Funcionales**

Las pruebas de caja negra son las que se ejecutan sobre la interfaz del software, las mismas pretenden comprobar las funcionalidades del software, y si se obtiene un resultado correcto.

Las pruebas de caja negra procuran encontrar errores en las siguientes categorías:

- Funciones incorrectas o faltantes
- Errores de Interfaz
- Errores en estructuras de datos o en acceso a bases de datos externas
- Errores de comportamiento o desempeño
- Errores de inicialización y terminación

Estas pruebas tienden a aplicarse durante las últimas etapas de la prueba. Debido a que estas desatienden a propósito la estructura de control, la atención se concentra en el dominio de la información. (18)

Capítulo 3: Implementación y Prueba del Mercado de Datos

- **Lista de Chequeo**

La lista de chequeo realiza una serie de preguntas en forma de cuestionario para probar la solución. Su objetivo principal es dar una valoración de la calidad del producto.

La lista de chequeo que se utilizó fue:

Nivel de importancia	Criterio de evaluación	Eval	NP	Cantidad de elementos afectados	Comentarios
Alta	¿Los subtítulos de las filas y las columnas corresponden a la descripción de sus atributos?	9		2 elementos	
Alta	¿Los reportes son configurables a través de la interfaz del sistema?	10			
Baja	¿La interfaz está orientada a facilitar el uso de las funciones del sistema por parte de los usuarios?	9		2 elementos	
Media	¿Los usuarios son capaces de manipular los resultados de manera que se ajusten a sus necesidades, conformando nuevos reportes?	9		2 elementos	Solo tienen permiso de conformar nuevos reportes el usuario "Oficial".
Alta	¿El sistema muestra las funcionalidades diferentes de acuerdo al rol de los usuarios?	10			

Capítulo 3: Implementación y Prueba del Mercado de Datos

Alta	¿Se puede navegar a través de los niveles de granularidad de los reportes?	9		2 elementos	La granularidad significa especificar el nivel de detalle. Ej: Provincia, Centro, Colectivo.
Media	¿Los procesos pueden ser analizados desde todas sus dimensiones?	9		2 elementos	
Media	¿El título de la gráfica se corresponde con los datos mostrados?	9		1 elemento	
Baja	¿Los elementos de la tabla siguen el mismo formato de letra?	8		3 elementos	
Media	¿Los parámetros de los requisitos se muestran en los reportes?	9		1 elemento	
Baja	¿La interfaz tiene errores ortográficos?	7		4 elementos	
Baja	¿El tipo de gráfica es el más representativo para mostrar la información?	8		3 elementos	

3.2.1 Resultados de las Pruebas

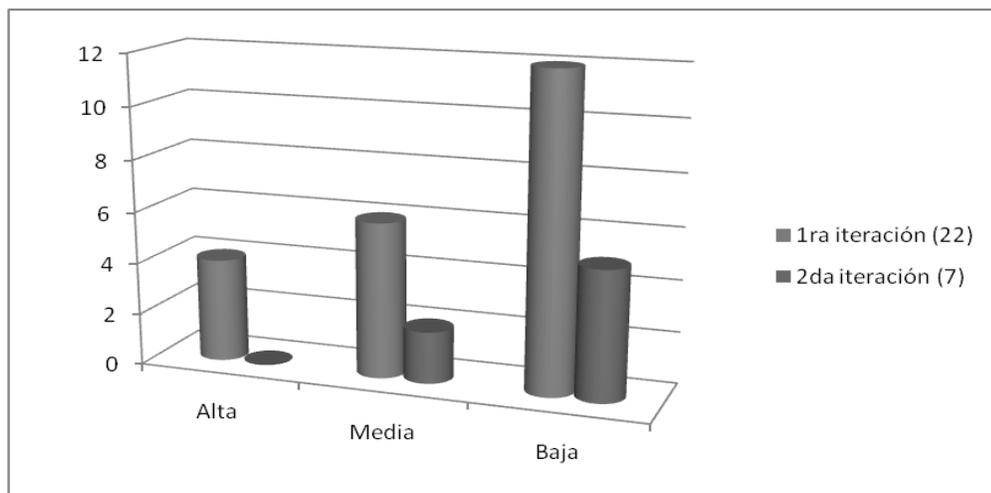
Para evaluar la solución desarrollada, la metodología propone realizar 2 iteraciones de prueba. Primeramente se realizó una iteración por parte del equipo de desarrollo, mostrando los resultados en

Capítulo 3: Implementación y Prueba del Mercado de Datos

la lista de chequeo propuesta, posteriormente se realizaron las pruebas de regresión con el objetivo de comprobar que las no conformidades detectadas en la primera iteración han sido corregidas, además de buscar nuevas no conformidades. Después de terminadas las pruebas de regresión, se procede a realizar otra iteración con personas ajenas al desarrollo del mercado de datos, y que no posean conocimiento de la solución, la cual se realizará con el cliente. Para medir la calidad del mercado de datos, se definió un criterio de evaluación en el intervalo de 1 a 10; donde de 1 a 5 representa el criterio de mala calidad, entre 6 y 7 baja calidad, entre 8 y 9 es aceptable y 10 que tiene alta calidad. Se abarcaron los métodos y técnicas de pruebas antes expuestos en cada una de las iteraciones, arrojando resultados visibles, los cuáles validan la calidad del producto construido. A continuación se muestran la cantidad de No Conformidades identificadas en la 1ra iteración y en las pruebas de regresión, la cual se nombra 2da iteración.

Clasificaciones	1ra iteración	2da iteración
Alta	4	0
Media	6	2
Baja	12	5

Se muestra de forma detallada un gráfico de las no conformidades encontrados durante las iteraciones realizadas:



Capítulo 3: Implementación y Prueba del Mercado de Datos

Luego de concluida la primera iteración y las pruebas de regresión, se analizaron, por parte del equipo de desarrollo, las No Conformidades encontradas para determinar cuáles realmente constituyeron errores del sistema. Las pruebas fueron realizadas, comprobando que hubiesen sido corregidos los no conformidades detectados en la iteración anterior, lo que contribuyó a mejorar la calidad y funcionalidad del software.

Conclusiones Parciales

En este capítulo se describió la implementación de la ETL para los procesos del caso de estudio del Área de Ingreso. Se configuró el repositorio de datos, el cual permitió relacionar los datos del mercado con las herramientas de consulta para la realización de los reportes. Se construyeron los reportes predefinidos correspondientes a los 84 requisitos de información, que responden a las necesidades del cliente. Se configuró el tablero de mando Histórico y Operativo para los directivos del Área de Ingreso. Además se realizaron las pruebas al mercado de datos mostrando los resultados obtenidos en las mismas, y se corrigieron las no conformidades detectadas.

CONCLUSIONES

- Se identificaron 84 requisitos de información que corresponden a las necesidades más urgentes de los directivos del Área de Ingreso.
- Se realizó el proceso de Extracción, Transformación y Carga de los datos, el cual permitió poblar el mercado de datos del Área de Ingreso para sustentar el análisis de los reportes.
- Se crearon los reportes predefinidos correspondientes a los requisitos de información, los cuales muestran el comportamiento de los indicadores más relevantes de los procesos del negocio, lo que le permite a los directivos tener un amplio conocimiento del estado en que se encuentra el Área de Ingreso.
- El usuario final puede crear nuevos reportes cuando lo necesite a través de consultas personalizadas.
- Se implementó el Mercado de Datos para el Área de Ingreso con el propósito de apoyar el proceso de toma de decisiones en la DEP.
- Se desarrolló una nueva herramienta de trabajo para los especialistas de dirección del Área de Ingreso del Sistema Penitenciario Cubano que les facilitará la realización de los cierres trimestrales.