

Universidad de las Ciencias Informáticas



Facultad 2

Título: Desarrollo de un Mercado de Datos para el Área de Seguridad Penitenciaria en el Sistema Penitenciario Cubano.

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

Autor(es):

Yudisleidy Pérez Lambert

Nadian Alemán Bauta

Tutor(es):

Ing. Yanet del Risco Batista

Ing. Krysia Llull Céspedes

La Habana, 2012

“Año 54 de la Revolución”



"Se puede adquirir conocimiento y conciencia a lo largo de toda la vida, pero jamás en ninguna otra época de su existencia una persona volverá a tener la pureza y el desinterés con que, siendo joven, se enfrenta a la vida."

Fidel Castro



Declaración de Autoría

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.

Yudisleidy Pérez Lambert

Firma del Autor

Nadian Alemán Bauta

Firma del Autor

Ing. Yanet del Risco Batista

Firma del Tutor

Ing. Krysia Llull Céspedes

Firma del Tutor



Datos de Contacto

DATOS DE CONTACTOS

Tutor: Ing. Yanet del Risco Batista.

Correo Electrónico: ydelrisco@uci.cu

Tutor: Ing. Krysia Llull Céspedes.

Correo Electrónico: klull@uci.cu

Autora: Yudisleidy Pérez Lambert.

Correo electrónico: ylambert@estudiantes.uci.cu

Autor: Nadian Alemán Bauta.

Correo electrónico: naleman@estudiantes.uci.cu



DEDICATORIA

Al Comandante en Jefe Fidel Castro Ruz, por hacer realidad nuestro sueño de forjarnos como profesionales en esta universidad.

De Yudisleidy Pérez Lambert:

A mis padres y mis familiares.

A mis amigos y amistades.

A todos los que han hecho posible que hoy esté aquí.

De Nadian Alemán Bauta:

A mis padres y mi hermanita.

A mis familiares y amigos.

A todos los que han hecho posible que hoy esté aquí.

Muchas gracias a todos.



Agradecimientos

AGRADECIMIENTOS

De Yudisleidy Pérez Lambert:

A mis abuelos María, Eduardo, Clara y Rolando por quererme y dedicarse a mis cuidados. Además por a verme regalado mis dos tesoros más preciados.

A mis padres Raidiluz y Elio por ser mi razón de existir, por ser las dos personas más especiales e importantes en mí vida. Por regalarme confianza cuando por uno que otro motivo, no la he tenido. Por ser mis amigos incondicionales y demostrarme cada día cuanto me quieren y cuánto significo para ellos. Por levantarse todos los días orgullosos de mí, como yo lo estoy de ellos.

A las tutoras Yanet y Krysia por a verme ayudado a ser realidad mi sueño y el de mis padres.

A todos los profesores del proyecto que de una forma u otra han hecho posible mi sueño.

A mis amigos Daniubis, Darién, Fabio y Maikel por estar a mi lado en estos 5 años.

A mi grupo, por ser parte de mi sueño.



Agradecimientos

De Nadian Alemán Bauta:

Agradecimientos especiales a aquellos dos que me trajeron el mundo y siempre me han brindado su apoyo: mi mamá y mi papá.

A todos mis familiares y a mi novia.

A mis amigos, que juntos hemos compartido cinco años inolvidables.

A mi tutores, que han aportado su granito de arena para la elaboración del presente trabajo.

A todos los que de una forma u otra han tenido influencia en mi desarrollo como persona y futuro profesional.



RESUMEN

El Sistema Penitenciario Cubano posee una base de datos operacional que acumula más de una década de datos para los centros penitenciarios y el acceso a estos se realiza de forma manual. El sistema penitenciario cuenta con reportes predefinidos para la toma de decisiones y al surgir una nueva solicitud de información, se hace necesaria la creación de un nuevo reporte. Los informáticos encargados de la construcción de los reportes deben responder a una gran cola de peticiones, a la que no se le puede dar solución en el tiempo requerido, dificultando el análisis de la información a los especialistas. Como consecuencia de esta situación no se puede realizar una adecuada toma de decisiones.

El propósito fundamental de este trabajo es desarrollar un mercado de datos para apoyar la toma de decisiones en el área de Seguridad Penitenciaria del Sistema Penitenciario Cubano.

Se obtuvo como resultado una herramienta de inteligencia de negocio que facilita el proceso de toma de decisiones en la Dirección de Establecimientos Penitenciarios (DEP). Esto posibilita que los procesos sean analizados desde distintas perspectivas, con el objetivo de obtener información histórica y operativa. Además, los usuarios pueden construir nuevos reportes, consultarlos, visualizarlos y modificar los ya existentes.

Palabras claves: almacén de datos, inteligencia de negocio, mercado de datos, seguridad penitenciaria.



ÍNDICE GENERAL

INTRODUCCIÓN	1
CAPÍTULO 1: MARCO TEÓRICO DE LA INVESTIGACIÓN	7
Introducción	7
1.1. Sistemas informáticos actuales en los Sistemas Penitenciarios.....	7
1.1.1. Sistema de Gestión Penitenciaria (e-Sigpen).....	7
1.1.2. Dirección General de Sistemas Penitenciarios de Panamá (DGSP)	8
1.1.3. Sistema de Gestión Penitenciaria (SIGEP)	8
1.1.4. Sistema Automatizado para el Control de Recluso (SACORE)	9
1.2. Almacén de datos	10
1.2.1. Características de un almacén de datos	11
1.2.2. Mercado de datos	14
1.2.3. Ventajas de los mercados de datos	14
1.3. Modelo multidimensional	15
1.3.1. Tablas de hechos	15
1.3.2. Tablas de dimensiones	15
1.4. Cubo multidimensional.....	18
1.4.1. Base de datos OLTP y OLAP	19
1.4.2. Tipo de modelo de almacenamiento	20
1.5. Arquitectura conceptual de los datos	20
1.6. Metodología.....	20
1.7. Herramientas de creación del mercado de datos	23
1.7.1. Sistema Gestor de Base de Datos Oracle 11g.....	23
1.7.2. Embarcadero ERStudio 8.0	23
1.7.3. Oracle Warehouse Builder 11g.....	23
1.7.4. Oracle Business Intelligence Interactive Dashboards 10.1.3.4.....	23
1.7.5. Administration Tool 10.1.3.4	24
1.7.6. Oracle Business Intelligence Answers 10.1.3.4	24
Conclusiones Parciales	24

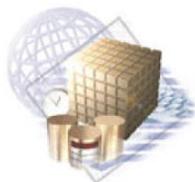


CAPÍTULO 2: ANÁLISIS Y DISEÑO DE LA SOLUCIÓN	25
Introducción	25
2.1. Análisis de requisitos	25
2.1.1. Técnicas de captura de requisitos	25
2.1.2. Identificación de temas analíticos y procesos de negocio.....	25
2.2. Especificación de requisitos.....	28
2.2.1. Requisitos de información.....	28
2.2.2. Especificación de requisitos de información.....	31
2.2.3. Matriz dimensional	35
2.2.4. Técnica de validación de requisitos	36
2.3. Modelado dimensional	36
2.3.1. Granularidad.....	36
2.3.2. Tablas de dimensiones	36
2.3.3. Tablas de hechos y medidas	37
2.3.4. Modelo dimensional detallado.....	37
2.3.5. Patrones de diseño utilizados	39
2.3.6. Modelos de datos	40
2.4. Diseño de la ETL	41
Conclusiones Parciales.....	45
CAPÍTULO 3: IMPLEMENTACIÓN Y PRUEBA	47
Introducción	47
3.1. Implementación de la ETL	47
3.1.1. Proceso de ETL para cargar las dimensiones.....	47
3.1.2. Proceso de ETL para las tablas de hechos.....	51
3.2. Configuración de la herramienta de administración	55
3.3. Especificación y desarrollo de aplicaciones de inteligencia de negocio	57
3.4. Tipos de procesamiento utilizados	61
3.5. Prueba.....	61
3.5.1. Descripción de los métodos de prueba a realizar.....	61



Índice General

Conclusiones Parciales	62
CONCLUSIONES GENERALES	64
REFERENCIA BIBLIOGRÁFICA	65
GLOSARIO DE TÉRMINOS	66



ÍNDICE DE FIGURAS

Figura 1: Arquitectura Bottom-Up.....	10
Figura 2: Almacén de datos, integrado.....	11
Figura 3: Almacén de datos, variante en el tiempo.....	12
Figura 4: Almacén de datos, orientado al tema.	13
Figura 5: Almacén de datos, no volátil.	14
Figura 6: Tipo de esquema, estrella.	16
Figura 7: Tipo de esquema, copo de nieve.	17
Figura 8: Tipo de esquema, constelación.....	18
Figura 9: Diagrama del Ciclo de Vida Dimensional del Negocio.....	21
Figura 10: Modelo dimensional detallado del proceso Incidencia.....	38
Figura 11: Modelo dimensional detallado del proceso Indisciplina.	38
Figura 12: Modelo de datos del proceso Incidencia.	40
Figura 13: Modelo de datos del proceso Indisciplina.....	41
Figura 14: Diseño de la ETL de la dimensión Incidencia.....	42
Figura 15: Diseño de la ETL de la dimensión Indisciplina.	42
Figura 16: Diseño de la ETL de la dimensión Nivel de Mando.	43
Figura 17: Diseño de la ETL de la dimensión Interno.....	45
Figura 18: Correspondencia de la tabla de dimensión Nivel de Mando.....	48
Figura 19: Correspondencia de la tabla de dimensión Tiempo.....	49
Figura 20: Correspondencia de la tabla de dimensión Incidencia.....	50
Figura 21: Correspondencia de la tabla de dimensión Indisciplina.	51
Figura 22: Correspondencia de la tabla <i>staging area</i> de Incidencia.....	52
Figura 23: Correspondencia de la tabla de hecho Incidencia.....	53
Figura 24: Correspondencia de la tabla <i>staging area</i> Indisciplina.....	54
Figura 25: Correspondencia de la tabla de hecho Indisciplina.....	55
Figura 26: Capa física.....	56
Figura 27: Capa de negocio.....	56
Figura 28: Capa de presentación.....	57
Figura 29: Ambiente de trabajo para modificar o crear reportes (Pestaña “Criterios”).....	58



Índice de Figuras

Figura 30: Ambiente de trabajo para modificar o crear reportes (Pestaña “Resultados”).....	58
Figura 31: Reportes operativos del proceso de Incidencia.	59
Figura 32: Reportes históricos del proceso Incidencia.	60
Figura 33: Reportes históricos del proceso Incidencia (filtro de fecha aplicado).....	60
Figura 34: Prueba de caja negra.....	62



ÍNDICE DE TABLAS

Tabla 1: Especificación de requisitos de información del proceso Incidencia.	33
Tabla 2: Especificación de requisitos de información del proceso Indisciplina.	35
Tabla 3: Matriz dimensional.	35
Tabla 4: Atributos extraídos para la dimensión Incidencia.....	41
Tabla 5: Atributos extraídos para la dimensión Indisciplina.	42
Tabla 6: Atributos extraídos para la dimensión Nivel de Mando.	43
Tabla 7: Atributos extraídos para la dimensión Interno.....	44



INTRODUCCIÓN

El transcurso de los años ha traído cambios al modo de vida actual, debido a numerosos avances que han transformado la forma en que operan las organizaciones. Muchos de estos avances han propiciado que se obtengan reformas significativas que contribuyen a suplir la necesidad del acceso a la información existente. La administración de la información es un proceso cada vez más dependiente de la tecnología. El volumen de información digital que se procesa o registra en las bases de datos de diferentes sistemas crece vertiginosamente, por lo que se hace necesario realizar una eficiente manipulación que facilite el acceso dinámico a dicha información.

A partir de la necesidad de lograr un manejo íntegro de los datos almacenados en las bases de datos, surge la Inteligencia de Negocio (*Business Intelligence* o, como se conoce internacionalmente por sus siglas en inglés, BI); la cual beneficia en gran medida la toma de decisiones en las empresas o instituciones que la pongan en práctica, debido a que *“integra por un lado el almacenamiento y por otro el procesamiento de grandes cantidades de datos, con el principal objetivo de transformarlos en conocimiento y en decisiones en tiempo real, a través de un sencillo análisis y exploración.”* (1)

Para obtener beneficio de la BI, se hace necesario integrarla a un almacén de datos (*Data Warehouse* o, como se conoce internacionalmente por sus siglas en inglés, DW). Este es un sistema que se caracteriza por integrar y refinar información de una o más fuentes de datos diferentes. Esto permite su estudio desde diferentes perspectivas y extraer conocimientos a partir de los datos históricos almacenados en la organización. Los almacenes de datos pueden estar formados por uno o varios mercados de datos (*Data Mart* o, como se conoce internacionalmente por sus siglas en inglés, DM), estos no son más que *“la implementación de un almacén de datos con alcance restringido a un área funcional, problema en particular, departamento, tema o grupo de necesidades”.* (1)

Cuba no se encuentra ajena a la evolución que están teniendo las Tecnologías de la Información y las Comunicaciones (TIC) a nivel mundial. Son numerosos los esfuerzos que se realizan para lograr la informatización del país, tratando de abarcar la mayor cantidad de sectores. En el año 1989 se dan los primeros pasos para la informatización del sistema penitenciario a nivel nacional, con el desarrollo de una aplicación que permitía el manejo de algunos datos de los internos. *“En el 2002, para darle cumplimiento a la orden 43/99 del Viceministro Primero del Interior, se realiza el Sistema Automatizado para el Control del Recluso (SACORE), el cual se comenzó a explotar en enero del 2003.”*(2) Con la puesta en



Introducción

funcionamiento de este sistema, se garantizaban respuestas a las solicitudes de información de los diferentes órganos e instituciones del Sistema Penitenciario Cubano; sin embargo, no se brindaba un soporte completo a las necesidades de información del MININT respecto a la población penitenciaria. Debido a esta situación, se hace necesario la creación de dos nuevas aplicaciones: el Sistema Automatizado de Incidencias de la Dirección de Establecimientos Penitenciarios (SAIDEP) y el Sistema Automatizado de Capacidades de la Dirección de Establecimientos Penitenciarios (SACDEP).

Como consecuencia de los años de explotación de las aplicaciones antes mencionadas, el Sistema Penitenciario acumula muchos años de información importante para la institución. El acceso a esta información se realiza de forma manual: se extrae de una página web, se exportan a documentos Excel y de ahí son generadas las diferentes gráficas para mantener las estadísticas de los temas referidos al sistema penitenciario. Otro inconveniente es que solo una parte de la información relacionada con la Seguridad Penitenciaria es analizada diariamente, lo que imposibilita un mejor control y seguridad en los centros penitenciarios.

Además, los cierres informativos son realizados cada tres meses por el Departamento de Dirección. Este departamento cuenta con una cantidad reducida de especialistas, que son los encargados de procesar toda la información. Esto impide que el Departamento de Dirección pueda realizar el análisis inmediato de la información para la toma de decisiones, provocando que los cierres informativos se extiendan aproximadamente 15 días. Debido a esto el proceso de toma de decisión se torna engorroso, por lo que existe la necesidad de la búsqueda de una vía que facilite el acceso a la misma con mayor rapidez.

Dada la situación anterior surge como **problema a resolver**: ¿Cómo facilitar la toma de decisiones para el área de Seguridad Penitenciaria del Sistema Penitenciario Cubano?

Se define como **objeto de estudio** de la investigación: el proceso de toma de decisiones en los sistemas penitenciarios, enfocándose en el **campo de acción**: el proceso de toma de decisiones en el área de Seguridad Penitenciaria del Sistema Penitenciario Cubano.

La **idea a defender** durante la investigación es: el desarrollo de un mercado de datos para el análisis de la información referente al área de Seguridad Penitenciaria del Sistema Penitenciario Cubano, facilitará el proceso de toma de decisiones en el área de Seguridad Penitenciaria.



Introducción

Se plantea como **objetivo general**: desarrollar un mercado de datos para la toma de decisiones en el área de Seguridad Penitenciaria del Sistema Penitenciario Cubano. El mismo se divide 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 Seguridad Penitenciaria.
3. Realizar el modelado dimensional del mercado de datos para el área de Seguridad Penitenciaria.
4. Implementar el proceso de Extracción, Transformación y Carga (ETL) del mercado de datos para el área de Seguridad Penitenciaria.
5. Implementar el mercado de datos para el área de Seguridad Penitenciaria.
6. Realizar las pruebas al mercado de datos para el área de Seguridad Penitenciaria.

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

Objetivo 1:

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

Objetivo 2:

- Identificación de temas analíticos del mercado de datos para el área de Seguridad Penitenciaria.
- Identificación de procesos de negocio del mercado de datos para el área de Seguridad Penitenciaria.
- Construcción de la matriz dimensional.
- Identificación de los requisitos del mercado de datos para el área de Seguridad Penitenciaria.
- Priorización de proceso del mercado de datos para el área de Seguridad Penitenciaria.

Objetivo 3:

- Identificación de los niveles de granularidad del mercado de datos para el área de Seguridad Penitenciaria.
- Identificación de las dimensiones, tablas de hechos y medidas del mercado de datos para el área de Seguridad Penitenciaria.



Introducción

- Identificación de atributos de dimensiones y tablas de hechos del mercado de datos para el área de Seguridad Penitenciaria.
- Diseño del modelo dimensional detallado del mercado de datos para el área de Seguridad Penitenciaria.
- Diseño del modelo datos del mercado de datos para el área de Seguridad Penitenciaria.
- Revisión y validación del modelado dimensional del mercado de datos para el área de Seguridad Penitenciaria.

Objetivo 4:

- Diseño del proceso de ETL del mercado de datos para el área de Seguridad Penitenciaria.
- Realización del proceso de ETL de las dimensiones y tablas de hechos del mercado de datos para el área de Seguridad Penitenciaria.

Objetivo 5:

- Configuración de la herramienta de administración del mercado de datos para el área de Seguridad Penitenciaria.
- Implementación de los reportes del mercado de datos para el área de Seguridad Penitenciaria.
- Desarrollo del tablero de mando del mercado de datos para el área de Seguridad Penitenciaria.

Objetivo 6:

- Selección de las técnicas de pruebas a realizar en el mercado de datos para el área de Seguridad Penitenciaria.
- Realización de las pruebas del mercado de datos para el área de Seguridad Penitenciaria.
- Recolección y clasificación de las no conformidades.

Resultados esperados:

- Desarrollo del Mercado de Datos del Área de Seguridad Penitenciaria.
- Implementación de los reportes del área de Seguridad Penitenciaria.
- Tablero de mando del área de Seguridad Penitenciaria.



Durante la investigación se utilizaron los siguientes **métodos científicos**:

Métodos teóricos

- Histórico - lógico: se utiliza para realizar el estudio del estado del arte, haciendo una valoración de las soluciones similares.
- Analítico - sintético: se utiliza para analizar las herramientas, tecnologías y la metodología a usar en el desarrollo del mercado de datos. Además, se identifican las características de estas y se realiza una valoración de cada una de ellas.
- Modelación: se emplea en la creación de los modelos realizados durante el desarrollo del mercado de datos. Estos modelos representan la propuesta de solución y permiten observar desde diferentes perspectivas el mercado de datos a desarrollar.

Métodos empíricos

- Entrevista: se pone en práctica en las entrevistas cerradas que se realizaron con el cliente, con el objetivo de conocer sobre el negocio, identificar los procesos y definir requisitos de información necesarios para la realización del mercado de datos.
- Análisis documental: con el objetivo de obtener un mayor conocimiento del negocio, se realiza un análisis de la bibliografía relacionada con los procesos que conforman el área de Seguridad Penitenciaria. También se estudió la bibliografía referente a las herramientas, tecnologías y la metodología a emplear en la solución.

El presente documento consta de la siguiente distribución: resumen, introducción, tres capítulos, conclusiones, recomendaciones, referencias bibliográficas, bibliografía, glosario de términos y anexos.

CAPÍTULO 1: “MARCO TEÓRICO DE LA INVESTIGACIÓN”.

En este capítulo se realiza un estudio sobre los sistemas informáticos que dan soporte al proceso de toma de decisiones en centros penitenciarios de diversos países y en Cuba. Además, se profundiza en los conceptos, objetivos y características de los almacenes de datos, como elementos a tener en cuenta para el desarrollo del mercado de datos para el área de Seguridad Penitenciaria de la DEP. Finalmente se describe la metodología y las herramientas que serán utilizadas para su implementación.



CAPÍTULO 2: “ANÁLISIS Y DISEÑO DE LA SOLUCIÓN”.

En el presente capítulo se realiza el análisis de requisitos del área de Seguridad Penitenciaria. En él se identifican los temas analíticos con sus respectivos procesos, así como la construcción de la matriz dimensional. Se realiza la especificación de requisitos y el modelado dimensional, incluyendo el diseño de los modelos dimensionales detallados y los modelos de datos de cada proceso, haciendo uso de algunos patrones de diseño de almacenes de datos. Además, se diseña el proceso de ETL de los datos.

CAPÍTULO 3: “IMPLEMENTACIÓN Y PRUEBA DEL MERCADO DE DATOS”.

En el presente capítulo se especifica la configuración de la Herramienta de Administración (*Administration Tool*). También se describe la implementación del proceso de ETL para cargar las dimensiones y las tablas de hechos. Se especifica la aplicación de inteligencia de negocio para el usuario final. Además, se realizan pruebas a la aplicación antes mencionada, utilizando las listas de chequeo y se muestran los resultados obtenidos.



Capítulo 1: Marco Teórico de la Investigación

CAPÍTULO 1: MARCO TEÓRICO DE LA INVESTIGACIÓN

Introducción

En este capítulo se realiza un estudio sobre los sistemas informáticos que dan soporte al proceso de toma de decisiones en centros penitenciarios de diversos países y en Cuba. Además, se profundiza en los conceptos, objetivos y características de los almacenes de datos, como elementos a tener en cuenta para el desarrollo del mercado de datos para el área de Seguridad Penitenciaria de la DEP. Finalmente se describe la metodología y las herramientas que serán utilizadas para su implementación.

1.1. Sistemas informáticos actuales en los Sistemas Penitenciarios

Debido a los avances existentes en las tecnologías, ha sido posible que diversas organizaciones hayan optado por la utilización de aplicaciones informáticas, con el fin de mejorar el procesamiento y la manipulación de la información. Esto se debe, en gran medida, a que en la actualidad se generan grandes cantidades de datos y estos a su vez pueden ser almacenados en bases de datos para su posterior análisis. A continuación se realiza un estudio de algunos sistemas informáticos que son utilizados en los centros penitenciarios de distintos países.

1.1.1. Sistema de Gestión Penitenciaria (e-Sigpen)

El Sistema de Gestión Penitenciaria de Ecuador, *“facilita el control y manejo de las actividades que realizan los Centros de Rehabilitación Social (CRS) en cada una de sus áreas, logrando de esta manera tener un control adecuado y oportuno de la información de las Personas Privadas de Libertad (PPL). El Ministerio de Justicia, Derechos Humanos y Cultos conjuntamente con la Dirección Nacional de Rehabilitación Social están en continuo monitoreo de las actividades realizadas en los CRS por medio del sistema e-Sigpen, el cual permite obtener datos estadísticos reales de la situación en que se encuentran las PPL, logrando de esta manera tomar decisiones eficaces y oportunas para mejorar el ambiente carcelario del país.”* (3)

A pesar de que esta aplicación facilita la gestión de la información y apoya en el análisis de la información, tiene como desventaja que los datos son procesados en bases de datos operacionales. Por esta razón, se ven limitados en cuanto al número de análisis a realizar para llevar a cabo un adecuado proceso de toma de decisiones.



Capítulo 1: Marco Teórico de la Investigación

1.1.2. Dirección General de Sistemas Penitenciarios de Panamá (DGSP)

“La Dirección General de Sistemas Penitenciarios de Panamá (DGSP) fue creado a principio del año 1997, como resultado de un proyecto financiado por las Naciones Unidas y el gobierno español. Esta aplicación permite mantener almacenado en una base de datos Oracle los registros de los internos que están detenidos en los centros penales a nivel nacional. Esta información solo es almacenada durante el tiempo de detención del interno en el centro penitenciario; posterior a la libertad, el expediente del mismo es eliminado.” (4)

Debido a que el expediente del interno es eliminado cuando este es puesto en libertad, el sistema no guarda información histórica. Esto imposibilita que se puedan realizar análisis de tendencias u obtener informes históricos que puedan ser útiles en la toma de decisiones.

1.1.3. Sistema de Gestión Penitenciaria (SIGEP)

El Sistema de Gestión Penitenciaria fue desplegado en Venezuela y desarrollado en la Universidad de las Ciencias Informáticas. El SIGEP tiene implementada una Sala Situacional que obtiene de la aplicación los datos de todas las áreas del sistema y muestra los resultados mediante reportes predefinidos.

Como iniciativa de los desarrolladores del SIGEP, se desarrollaron dos mercados de datos para facilitar el proceso de toma de decisiones de los directivos del sistema de justicia venezolano, pero no fueron desplegados y no forman parte de la solución. Estos mercados de datos se hicieron sobre el Control Penal y sobre Clasificación y Atención Integral de la Sala Situacional del SIGEP. A continuación se muestran algunas de las características de estos mercados.

“El Mercado de Datos sobre el Control Penal de la Sala Situacional del SIGEP, además de traer algunos reportes predefinidos, brinda tres variantes diferentes para la exploración de los datos. Los usuarios finales pueden explorar centrándose en las estancias, este punto de vista muestra los individuos, sus expedientes y las decisiones judiciales relacionados con sus estancias en los centros penitenciarios modelados en el tiempo. También se puede explorar teniendo en cuenta las penas impuestas a los individuos. Esta exploración incluye los individuos, sus expedientes, las decisiones y las ejecuciones de proceso legal relacionados con las penas a lo largo del tiempo. Por último permite la exploración por los procesos legales, en este punto de vista se pueden ver los individuos, sus expedientes y las decisiones relacionados con los procesos legales en el transcurrir del tiempo.” (5)



Capítulo 1: Marco Teórico de la Investigación

“El Mercado de Datos sobre la Clasificación y la Atención Integral de la Sala Situacional del SIGEP, además de los reportes predefinidos, ofrece 5 variantes diferentes para la exploración de los datos. Los usuarios finales podrán explorar basándose en la clasificación, en este se muestran los individuos, sus expedientes y la clasificación de los individuos en cada uno de los centros penitenciarios modelados en el tiempo. De igual manera se puede explorar teniendo en cuenta la actividad cultural; esta exploración incluye individuos, sus expedientes, el nivel y el tipo de manifestación cultural practicada en el centro por el individuo a lo largo del tiempo. También se puede enfocar la exploración centrándose en la actividad productiva, incluyendo la información de los individuos, sus expedientes, el tipo de actividad productiva que realiza en el centro a lo largo del tiempo de estancia en el penal. Otra manera de exploración es por la actividad deportiva, incluyendo individuos, sus expedientes, el nombre y tipo de disciplina deportiva en las que participa el individuo modelado en el tiempo. Por último los usuarios pueden explorar centrándose en la actividad educativa, la cual muestra a los individuos, sus expedientes y la clasificación de las actividades deportivas desarrolladas en el centro a través del tiempo.” (6)

A pesar de que se desarrollaron los mercados de datos antes mencionados; estos no presentan reportes relacionados con la seguridad penitenciaria ni dan la posibilidad de crearlos.

1.1.4. Sistema Automatizado para el Control de Recluso (SACORE)

El Sistema Penitenciario Cubano utiliza el SACORE desde el año 2003. “En sus 12 años de explotación se ha ido perfeccionando a partir de los requerimientos y solicitudes de los usuarios, conteniendo como módulos principales Control Penal, Tratamiento Educativo y Orden Interior (los dos últimos con varias opciones y salidas pendientes) y además se desarrollaron otros dos sistemas automatizados que complementan la información del SACORE: el SAIDEP y el SACDEP.” (2) Aún con el desarrollo de estos sistemas para llevar a cabo los análisis estadísticos, este proceso se retrasaba debido al procedimiento empleado para transformar los datos almacenados. Dichas transformaciones se hacen hasta obtener un reporte final; esto último, es lo utilizado por los directivos en la toma de decisiones.

Los tres sistemas descritos anteriormente, en su conjunto, almacenan varios años de información; pero sus funcionalidades no procesan o gestionan la misma en su totalidad, limitando el proceso de toma de decisiones de los funcionarios del Sistema Penitenciario Cubano. Otro inconveniente es que no muestran toda la información relacionada con la seguridad penitenciaria de los centros penitenciarios.



Capítulo 1: Marco Teórico de la Investigación

Luego de analizados los sistemas informáticos antes mencionados, se impone entonces desarrollar una solución que facilite el análisis de la información para apoyar la toma de decisiones, por ejemplo: un almacén de datos. Esta solución unida a herramientas de BI, dará respuesta a las necesidades de información del Sistema Penitenciario Cubano.

1.2. Almacén de datos

“Un almacén de datos es una colección de datos orientada al tema, integrada, variante en el tiempo y no volátil para el soporte del proceso de toma de decisiones de la empresa.” (7) La creación de los almacenes de datos es, desde el punto de vista técnico, el camino que siguen las grandes empresas para efectuar una solución de inteligencia de negocio. Para esto, toman de las bases de datos toda aquella información relevante o importante para la toma de decisiones.

Los almacenes de datos pueden adoptar diferentes tipos de arquitecturas en dependencia del enfoque con que se quiera implementar. Entre las arquitecturas más conocidas se encuentran *Top-Down* y *Bottom-Up*. Debido a las necesidades que se presentan, el arquitecto del SIDEPE-DWH¹ determinó utilizar la arquitectura *Bottom-Up*, que propone implementar previamente los mercados de datos para posteriormente integrarlos en un almacén de datos.

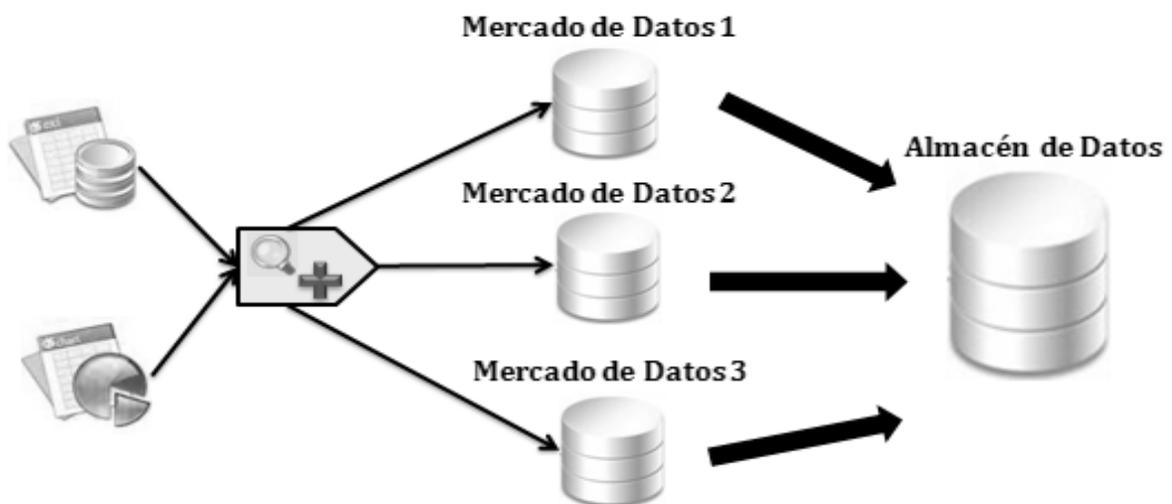


Figura 1: Arquitectura Bottom-Up.

¹ SIDEPE-DWH es un subproyecto del proyecto Prisiones Cuba, dicho proyecto es el encargado de desarrollar el Sistema Informativo de la Dirección de Establecimientos Penitenciarios (SIDEPE).



Capítulo 1: Marco Teórico de la Investigación

1.2.1. Características de un almacén de datos

Un almacén de datos debe cumplir con las siguientes características:

- Integrado
- Variante en el tiempo
- Orientado al tema
- No volátil

A continuación se dará una breve explicación de cada una de las características anteriormente mencionadas.

▪ Integrado

“La integración implica que todos los datos de diversas fuentes que son producidos por distintos departamentos, secciones y aplicaciones, tanto internos como externos, deben ser consolidados en una instancia antes de ser agregados al almacén de datos” (1); para que después puedan ser limpiados y transformados. “La integración de datos resuelve diferentes tipos de problemas relacionados con las convenciones de nombre, unidades de medidas, codificaciones, fuentes múltiples” (1) e incompatibilidad entre los tipos de datos.

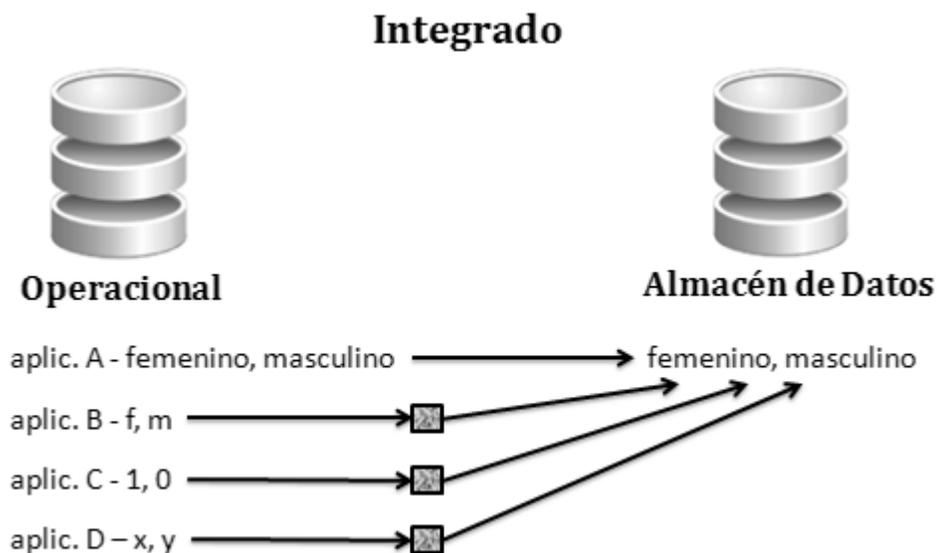


Figura 2: Almacén de datos, integrado.



Capítulo 1: Marco Teórico de la Investigación

▪ Variante en el tiempo

Toda estructura clave de un almacén contiene implícita o explícitamente un elemento del tiempo, constituyendo una de las principales ventajas del almacén de datos, pues los datos son almacenados junto a un sello de tiempo.

“Esta cualidad que no se encuentra en fuentes de datos operacionales garantiza poder desarrollar análisis de la información, siendo procesada como una serie de instantáneas, cada una representando un período de tiempo. Es decir, que gracias al sello de tiempo se podrá tener acceso a diferentes versiones de la misma información”. (1)

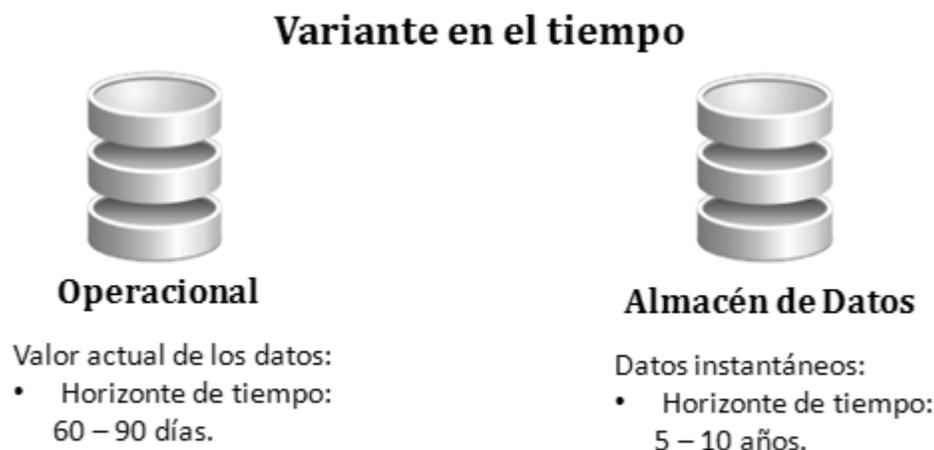


Figura 3: Almacén de datos, variante en el tiempo.

▪ Orientado al tema

En el almacén de datos *“la información se clasifica en base a los aspectos que son de interés para la organización.”* (1) Por esta razón, *“la estructura del mismo difiere considerablemente a la de los clásicos procesos operacionales orientados a las aplicaciones.”* (1) Las mayores diferencias entre la orientación de procesos y funciones de las aplicaciones y la orientación a temas, radican en el contenido de la información a nivel detallado.

“El almacén de datos excluye la información que no será utilizada exclusivamente en el proceso de toma de decisiones; mientras que en los procesos orientados a las aplicaciones, se incluyen todos aquellos datos que son necesarios para satisfacer de manera inmediata los requerimientos funcionales de la actividad que soporten.” (1)



Capítulo 1: Marco Teórico de la Investigación



Figura 4: Almacén de datos, orientado al tema.

- **No volátil**

“La información es útil para el análisis y la toma de decisiones solo cuando es estable. Los datos operacionales varían momento a momento, en cambio, los datos una vez que entran en el almacén de datos no cambian.” (1)

Las operaciones insertar, borrar y modificar se realizan habitualmente en las bases de datos operacionales. En los almacenes la manipulación básica de los datos es mucho más simple, debido a que hay dos únicos tipos de operaciones: la carga de datos y el acceso a estos.



Capítulo 1: Marco Teórico de la Investigación

No volátil

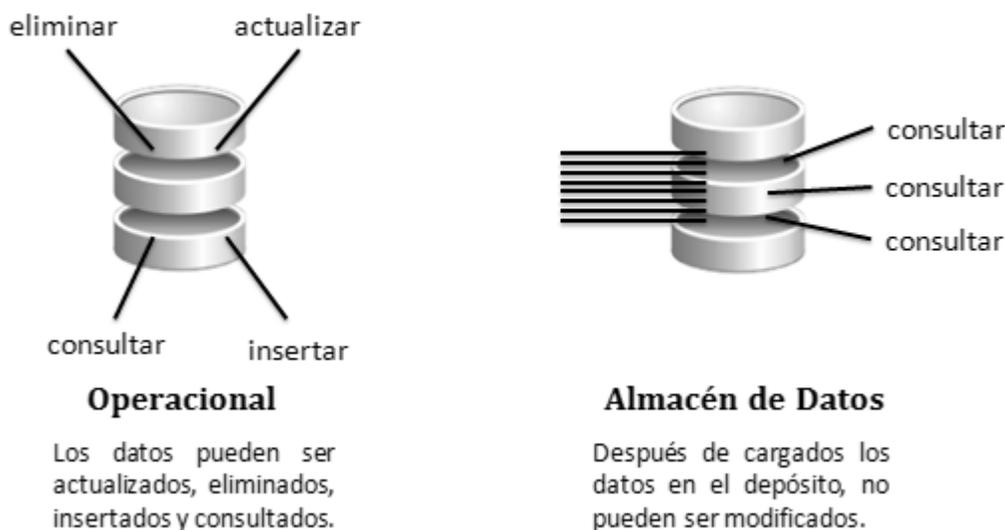


Figura 5: Almacén de datos, no volátil.

1.2.2. Mercado de datos

Un mercado de datos es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer de una estructura de datos, que posibilita analizar la información al detalle desde las perspectivas que afecten a los procesos de dicho departamento.

En ocasiones los almacenes de datos son iniciados por la implementación de un mercado de datos, con el propósito de minimizar riesgos y que la entrega a los clientes se realice en el tiempo reglamentado. En el momento que el mercado de datos haya tenido éxito en su desarrollo, su alcance puede ir ampliándose hasta llegar a conformar un almacén de datos.

1.2.3. Ventajas de los mercados de datos

Algunas de las ventajas de aplicar un mercado de datos a un negocio, son las siguientes:

- Transforma los datos orientados a las aplicaciones en información dirigida a la toma de decisiones.
- Conllevan poco tiempo de construcción y puesta en marcha.
- Provee la capacidad de analizar y explotar las diferentes áreas de trabajo y de realizar un análisis inmediato de las mismas.



Capítulo 1: Marco Teórico de la Investigación

- Reflejan rápidamente sus beneficios y cualidades.

1.3. Modelo multidimensional

Para la creación de los mercados de datos, uno de los modelos más utilizados es el modelo multidimensional propuesto por Ralph Kimball². Debido a este tipo de modelo es que surgen las bases de datos multidimensionales. Estas brindan un acceso rápido a los datos, para analizar sus relaciones y consiguientes resultados y están orientadas a temas. Tienen este nombre porque la información se almacena en forma multidimensional; a través de tablas de hechos y tablas de dimensiones.

1.3.1. Tablas de hechos

Las tablas de hechos contienen los datos que serán utilizados para la creación de los indicadores y apoyar el proceso de toma de decisiones. La llave primaria de estas tablas es la unión de las llaves primarias de cada una de las tablas de dimensiones conectadas a la tabla de hecho. *“Debido a que una tabla de hechos se encuentra interrelacionada con sus respectivas tablas de dimensiones, permite que los hechos puedan ser accedidos, filtrados y explorados por los valores de los campos de estas tablas de dimensiones, obteniendo de este modo una gran capacidad analítica.”* (1)

1.3.2. Tablas de dimensiones

“Las tablas de dimensiones definen como están los datos organizados lógicamente y proveen el medio para analizar el contexto del negocio.” (1) Las mismas contienen, generalmente, una llave simple y atributos que la describen. Existe una dimensión fundamental en todo mercado de datos: la dimensión tiempo. Esto ocurre porque todo registro que se incluya constituye la ocurrencia de un fenómeno en un instante de tiempo definido. Dicha dimensión es la que establece uno de los objetivos fundamentales de la construcción de un mercado de datos, la conservación de un “histórico”. La calidad del modelado dimensional, depende en gran medida de cuan descriptivos y manejables sean los atributos dimensionales escogidos.

Existen 3 tipos de esquemas dimensionales de un almacén de datos.

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

² Ralph Kimball es conocido como el "Padre de la Inteligencia de Negocio".



Capítulo 1: Marco Teórico de la Investigación

A continuación se dará una breve explicación de los tipos de esquemas dimensionales mencionados anteriormente.

- **Esquema en estrella**

El esquema en estrella es el esquema dimensional más fácil de modificar y simple de interpretar. El mismo “consta de una tabla de hechos central y de varias tablas de dimensiones relacionadas a esta, a través de sus respectivas claves.” (1) Además, este esquema “es soportado por casi todas las herramientas de consulta y análisis.” (1)



Figura 6: Tipo de esquema, estrella.

- **Esquema copo de nieve**

El esquema copo de nieve es una ampliación del esquema en estrella con la particularidad de que las dimensiones se encuentran normalizadas. En este tipo de esquema “*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.*” (1) La ventaja fundamental que proporciona este esquema, es que se ocupa menor espacio de almacenamiento; sin embargo, aumenta el número de tablas con las que el usuario debe interactuar e incrementa la complejidad de las consultas a realizar.



Capítulo 1: Marco Teórico de la Investigación

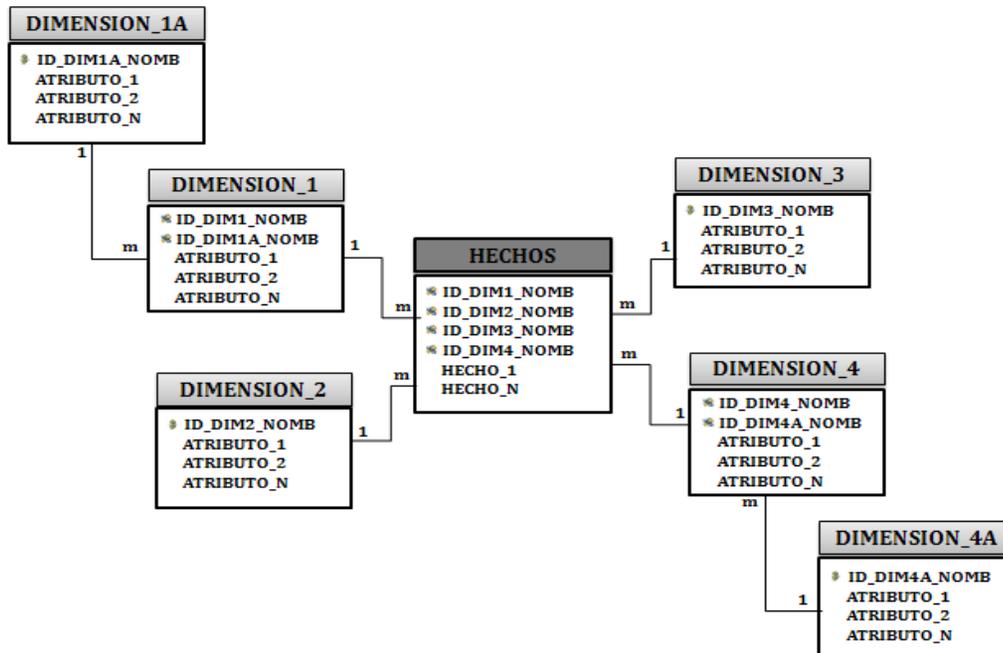


Figura 7: Tipo de esquema, copo de nieve.

▪ Esquema constelación

El esquema constelación o copo de estrellas está formado por dos o más esquemas de estrellas. Este esquema contiene dos o más tablas de hechos, lo que permite realizar un análisis más detallado del negocio. Estas tablas de hechos se encuentran relacionadas con sus correspondientes tablas de dimensiones. El esquema constelación “contribuye a la reutilización de las tablas de dimensiones, ya que una misma tabla de dimensión puede utilizarse para varias tablas de hechos.” (1)



Capítulo 1: Marco Teórico de la Investigación

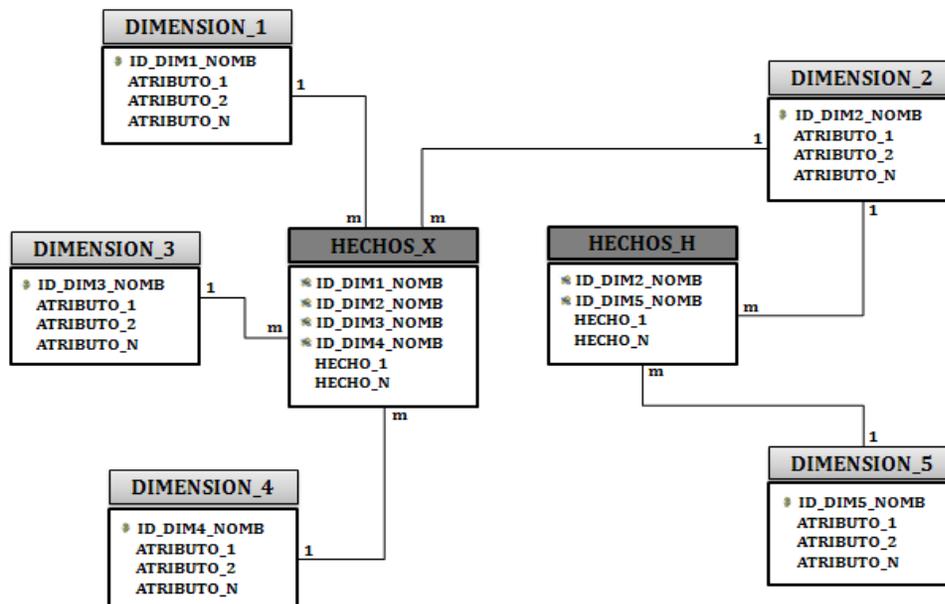


Figura 8: Tipo de esquema, constelación.

Entre los tres esquemas dimensionales especificados anteriormente, el SIDEPE-DWH estableció hacer uso del esquema de estrella para el modelado de los datos de cada uno de los procesos por separados y el esquema de constelación para el modelado de los datos del mercado de datos en general.

1.4. Cubo multidimensional

Los cubos multidimensionales son los objetos principales del Proceso Analítico en Línea (*On Line Analytic Processing* o, como se conoce internacionalmente por sus siglas en inglés, OLAP); esta es una tecnología que proporciona un rápido acceso a la información de un almacén de datos. Los cubos brindan una vía fácil para consultar los datos. La tabla central del esquema es la tabla de hechos y las demás son las tablas de dimensiones. Esta estructura de datos es una de las más utilizadas.



Capítulo 1: Marco Teórico de la Investigación

Los cubos multidimensionales contienen 3 elementos fundamentales:

Medidas: “son sumalizaciones efectuadas sobre algún hecho o expresiones basadas en sumalizaciones, que serán incluidos en algún cubo multidimensional, con el fin de analizar los datos almacenados en el mercado de datos.” (1)

Atributos: “constituyen los criterios de análisis que se utilizarán para analizar los indicadores dentro de un cubo multidimensional. Los mismos se basan, en su gran mayoría, en los campos de las tablas de dimensiones y/o expresiones.” (1)

Jerarquías: “representan una relación lógica entre dos o más atributos pertenecientes a un cubo multidimensional; siempre y cuando posean su correspondiente relación padre-hijo.” (1)

1.4.1. Base de datos OLTP y OLAP

Procesamiento de Transacciones en Línea

Los sistemas de Procesamiento de Transacciones en Línea (*On Line Transaction Processing* o, como se conoce internacionalmente por sus siglas en inglés, OLTP) son bases de datos orientadas al procesamiento de transacciones. Estas pueden implicar diferentes operaciones, como por ejemplo: inserción, eliminación y modificación de datos. Este proceso es típico de las bases de datos operacionales. En estos sistemas el acceso a los datos está optimizado para tareas frecuentes de escrituras, los formatos de los datos no son necesariamente uniformes y el historial de datos suele limitarse a los datos actuales o recientes.

Proceso Analítico en Línea (OLAP)

Los sistemas de Proceso Analítico en Línea (*On Line Analytic Processing* o, como se conoce internacionalmente por sus siglas en inglés, OLAP) son bases de datos orientadas al procesamiento analítico, debido a que proporcionan información útil para la toma de decisiones. En este tipo de sistema el acceso a la información acostumbra a ser de solo lectura. Se realizan pocas actualizaciones y eliminaciones, las inserciones son a través de la carga de datos y estos son agrupados de acuerdo a las áreas de negocios. Las bases de datos OLAP toman la información de varias fuentes de datos mediante el proceso de ETL.



Capítulo 1: Marco Teórico de la Investigación

1.4.2. Tipo de modelo de almacenamiento

Existen tres tipos de modelos de almacenamiento de los datos, estos son: *Relational On Line Analytic Processing* (ROLAP), *Multidimensional On Line Analytic Processing* (MOLAP) y *Hybrid On Line Analytic Processing* (HOLAP). El equipo de arquitectura del SIDEP-DWH definió utilizar ROLAP como el tipo de modelo de almacenamiento.

▪ ROLAP:

ROLAP es un tipo de modelo que se construye sobre tecnología relacional; pero a la vez, brinda varias posibilidades que ayudan a mejorar el rendimiento. Cuenta con todas las ventajas de un Sistema Gestor de Base de Datos (SGBD) Relacional. Además, es flexible y soporta grandes volúmenes de datos.

1.5. Arquitectura conceptual de los datos

La arquitectura conceptual de los datos es la forma de representar la organización total de datos, comunicación, procesamiento y presentación. Contiene los recursos para distribuir la información en el almacén de datos. En los almacenes de datos comúnmente se representan varias capas y es a través de estas que circula la información.

Se utilizará la arquitectura de tres capas; pues para realizar las transformaciones de los datos, se necesita una capa intermedia en la cual se corrijan los problemas de inconsistencias. Aquí se realizaría el procesamiento de los distintos grupos de datos de tiempo real correctamente.

Para llevar a cabo el diseño del mercado de datos, el arquitecto del SIDEP-DWH definió la arquitectura mencionada con anterioridad; debido a que cumple con las características y funcionalidades necesarias.

1.6. Metodología

La metodología definida por el SIDEP-DWH para el ciclo de desarrollo del mercado de datos es la de Kimball. Las características de esta responden al enfoque Bottom-Up especificado anteriormente. Además, se basa en lo que Ralph Kimball denomina Ciclo de Vida Dimensional del Negocio (Business Dimensional Lifecycle).

Para la construcción del mercado de datos, la metodología de Kimball propone una serie de etapas a las que hay que dar cumplimiento durante su ciclo de vida. La misma se encuentra organizada en tres áreas o caminos distintos.



Capítulo 1: Marco Teórico de la Investigación

- ✓ Camino superior o Tecnología: este camino comprende las tareas que se encuentren relacionadas con el software a ser utilizado para el desarrollo del mercado de datos.
- ✓ Camino del medio o Datos: en este camino es donde se diseñan y se implementan todas las tareas relacionadas con el modelado dimensional y con el desarrollo de la ETL de los datos de origen al almacén de datos.
- ✓ Camino inferior o Aplicaciones de Inteligencia de Negocio: este camino es el encargado de manejar todas las tareas de diseño y de desarrollo de las aplicaciones de negocios para los usuarios finales.

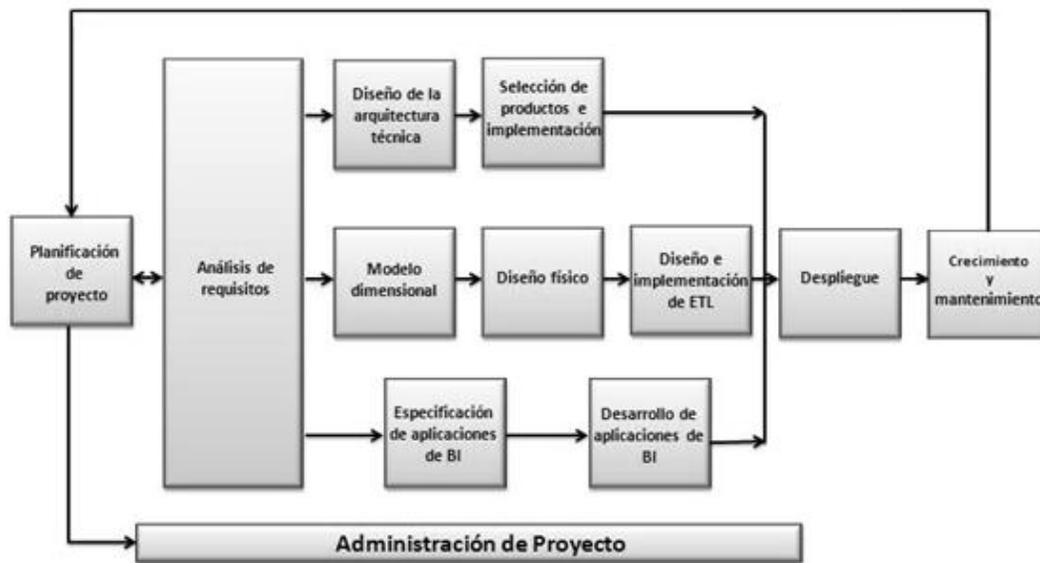


Figura 9: Diagrama del Ciclo de Vida Dimensional del Negocio.

El presente trabajo se enmarca en las etapas siguientes: Planificación, Análisis de requisitos, Modelado dimensional, Diseño físico, Diseño e implementación de ETL y Especificación y desarrollo de aplicaciones de inteligencia de negocio. A continuación se hace una breve descripción de estas etapas:

▪ Planificación

La planificación del proyecto del almacén de datos involucra todo el ciclo de vida de Kimball, desde su inicio hasta su despliegue. En esta etapa se determinan los objetivos específicos y la meta a alcanzar con el almacén de datos, así como determinar la necesidad de información de la empresa.



Capítulo 1: Marco Teórico de la Investigación

- **Análisis de requisitos**

En esta etapa se desarrolla un proceso de encuestas, estudio de documentación relacionada con el negocio y entrevistas realizadas al personal técnico, con el objetivo de recoger los requisitos que establezcan las funcionalidades a tener en cuenta para la realización del almacén de datos. Los requisitos deben de establecer la importancia del mismo.

- **Modelado dimensional**

El modelado dimensional comienza con un modelo de alto nivel obtenido a partir de los procesos priorizados. Este depende en gran medida del análisis de requisitos descrito anteriormente. Posteriormente se define el nivel de granularidad que especifica el nivel de detalle que posee cada registro de una tabla de hecho; esto es posible a partir de los datos actuales de la base de datos. Luego se pasa a elegir las tablas de dimensiones, brindando una forma de análisis sobre una medida en una tabla de hecho. Después se identifican las medidas y las tablas de hechos. Cada tabla de hecho tiene como atributo una o más medidas de acuerdo a los requisitos. Por último; se diseña el modelo dimensional detallado.

- **Diseño físico**

El diseño físico es elaborado luego de realizarse el modelado dimensional. En esta etapa se decide la cantidad de índices a tener en cuenta, el tipo de índice y donde se guardan los mismos. También se especifica el tipo de particionamiento (horizontal o vertical) y se definen los nombres que tendrán los índices, las particiones y las llaves.

- **Diseño e implementación de ETL**

El proceso de ETL es el que se encarga de extraer los datos de la base de datos origen hacia el almacén de datos. Además, realiza las transformaciones aplicando las distintas pautas, con el objetivo de incrementar la calidad y la firmeza de los datos y por último, se ejecuta la carga de los datos en el almacén de datos.

- **Especificación y desarrollo de aplicaciones de inteligencia de negocio**

Una de las partes fundamentales de los proyectos que utilizan almacenes de datos e Inteligencia de Negocio, es la de proporcionar a los usuarios una forma más estructurada de acceder a la información y facilitar el proceso de toma de decisiones. Esto se realiza a través de la aplicación de Inteligencia de Negocio.



Capítulo 1: Marco Teórico de la Investigación

1.7. Herramientas de creación del mercado de datos

Para el desarrollo del mercado de datos, se hace necesario tener en cuenta un conjunto de herramientas que le proporcione al desarrollador facilidades en el trabajo a ser realizado. Todas las herramientas a utilizar fueron definidas por el equipo de arquitectura del SIDE-P-DWH y por el cliente.

1.7.1. Sistema Gestor de Base de Datos Oracle 11g

“El Sistema Gestor de Base de Datos Oracle 11g es una plataforma completa de base de datos para almacén de datos e inteligencia empresarial, que combina rendimiento y escalabilidad líder en la industria, análisis con alta integralidad, calidad de datos e integración incorporada, en una plataforma única. Al incorporar OLAP, Minería de Datos y funciones estadísticas directamente en la base de datos, Oracle ofrece toda la funcionalidad de un motor analítico independiente. Puesto que la integración de datos es un requisito central de cualquier almacén de datos, Oracle Database 11g incluye Oracle Warehouse Builder, una herramienta de ETL líder que utiliza las capacidades de acceso de datos heterogéneos y transformación de datos escalable de Oracle.” (8)

1.7.2. Embarcadero ERStudio 8.0

“ERStudio es una herramienta de modelado de datos para el diseño y la comprensión de las bases de datos, que ayuda a las empresas a descubrir, documentar y reutilizar los activos de datos.” (9) Además, se utiliza para la creación de particiones, índices, realizar ingeniería inversa y generar la base de datos.

1.7.3. Oracle Warehouse Builder 11g

Oracle Warehouse Builder (como se conoce internacionalmente por sus siglas en inglés, OWB) es una herramienta que posibilita la integración de los datos, el diseño e implementación del proceso de ETL y además integra todos los componentes necesarios para realizar este proceso. Esta herramienta *“proporciona todas sus capacidades dentro de un repositorio común y una interfaz de usuario”*. (8) Además, *“efectúa todas las transformaciones dentro de una base de datos Oracle, aprovechando la escalabilidad y el rendimiento de la plataforma de base de datos”*. (8)

1.7.4. Oracle Business Intelligence Interactive Dashboards 10.1.3.4

El Oracle Business Intelligence Interactive Dashboards le ofrece al usuario una interfaz web a través de la cual puede acceder a la información de forma dinámica. Además, posibilita la visualización de los datos mediante gráficas y tablas; teniendo la completa facultad de navegar, transformar e interactuar con los



Capítulo 1: Marco Teórico de la Investigación

resultados obtenidos de estos. Esta herramienta permite además añadir ficheros compartidos y repositorios de documentos.

1.7.5. Administration Tool 10.1.3.4

El Administration Tool es una herramienta de administración para importar los metadatos de bases de datos y otras fuentes de datos, simplificando y reorganizando los datos importados en un modelo de negocio. Además estructura el modelo de negocio para su presentación a los usuarios que solicitan información de inteligencia empresarial, a través de interfaces de usuario de Oracle BI; tales como: Oracle BI Answers y Oracle BI Interactive Dashboards.

1.7.6. Oracle Business Intelligence Answers 10.1.3.4

El Oracle Business Intelligence Answers propone un entorno web que les permite a los usuarios un mejor intercambio con los datos que se encuentran almacenados. Esta herramienta ofrece la posibilidad de crear consultas y mostrar la información en diferentes formatos y representaciones como: gráficos, tablas y tablas dinámicas. Además; permite crear selectores para lograr que sea aún más interactivo en sus solicitudes de Business Intelligence.

Conclusiones Parciales

En este capítulo se estudiaron varios sistemas informáticos utilizados en centros penitenciarios de diversos países y de Cuba. Se analizaron sus características y el modo en que muestran la información referente a la toma de decisiones. Además, se realizó un análisis de las desventajas de cada uno de ellos, lo que permitió llegar a la conclusión de que el desarrollo de un mercado de datos para apoyar el proceso de toma de decisiones en los centros penitenciarios cubanos, es la solución más factible para realizar un mejor análisis de la información.



Capítulo 2: Análisis y Diseño de la solución

CAPÍTULO 2: ANÁLISIS Y DISEÑO DE LA SOLUCIÓN

Introducción

En el presente capítulo se realiza el análisis de requisitos del área de Seguridad Penitenciaria. En él se identifican los temas analíticos³ con sus respectivos procesos, así como la construcción de la matriz dimensional. Se realiza la especificación de requisitos y el modelado dimensional, incluyendo el diseño de los modelos dimensionales detallados y los modelos de datos de cada proceso, haciendo uso de algunos patrones de diseño de almacenes de datos. Además, se diseña el proceso de ETL de los datos.

2.1. Análisis de requisitos

En esta etapa se hace uso de algunas técnicas de captura de requisitos, se identifican los temas analíticos y los procesos. Además, a partir del análisis realizado se construye la matriz dimensional.

2.1.1. Técnicas de captura de requisitos

Para capturar los requisitos del cliente son empleadas algunas técnicas que permiten identificar y definir las necesidades que presentan los mismos. A continuación, se especifican y se describen las técnicas que fueron utilizadas para la captura de requisitos del mercado de datos.

- Análisis de documentación: con el objetivo de obtener un mayor conocimiento del negocio se estudiaron las leyes, resoluciones y los reglamentos del Sistema Penitenciario Cubano. Para el diseño e implementación del proceso de ETL fue necesario analizar la base de datos del SIDEP. Este estudio se orientó principalmente a los aspectos relacionados con el área de Seguridad Penitenciaria.
- Entrevista: se realizaron varias entrevistas a los especialistas del Departamento de Dirección de la DEP, en el que se recogieron opiniones, criterios y descripciones, que aportaron elementos esenciales para el desarrollo del mercado de datos.

2.1.2. Identificación de temas analíticos y procesos de negocio

A partir de la captura de requisitos, se identificaron los temas analíticos y los procesos de negocio del área de Seguridad Penitenciaria. Esta es la encargada del análisis y administración de la información que le es

³ Los temas analíticos agrupan los procesos comunes entre sí, de acuerdo a la información de estos.



Capítulo 2: *Análisis y Diseño de la solución*

tomada al interno, en relación al control de las actividades de custodia y de orden interior de los establecimientos penitenciarios.

A continuación se describen cada uno de los temas analíticos y procesos de negocio; haciendo énfasis en sus particularidades.

➤ **Descripción de los procesos por tema analítico**

Dentro del área de Seguridad Penitenciaria se identificaron varios procesos por temas analíticos. Los datos almacenados en estos, son los que permiten el análisis, modelado e implementación del mercado de datos.

A continuación se describen estos procesos por los temas analíticos identificados en el área de Seguridad Penitenciaria.

1. Estado Disciplinario

El tema analítico Estado Disciplinario recoge la información referente a los hechos que ocurren en los centros penitenciarios y las medidas disciplinarias que se aplican a las indisciplinas. Incluye los procesos: Incidencia, Indisciplina y Medida Disciplinaria.

- **Incidencia:** este proceso permite almacenar las incidencias que alteran el orden interior y afectan la seguridad de los establecimientos penitenciarios. El diseño de este proceso posibilita analizar las incidencias ocurridas en los centros penitenciarios, los tipos de incidencias, los internos que cometieron estas incidencias, los instantes de tiempo en que ocurrieron y los lugares donde más ocurren. Esto posibilita que el proceso de toma de decisiones se beneficie en gran medida; pues el análisis de esta información ayuda a prevenir este tipo de hechos.
- **Indisciplina:** este proceso permite almacenar y analizar todas las indisciplinas cometidas por los privados de libertad y la información de relevancia relacionada con estas. Las indisciplinas están categorizadas teniendo en cuenta su nivel de gravedad (Graves, Menos Graves y Leves). Este proceso se diseña con el objetivo de generar los reportes necesarios para evaluar las diferentes indisciplinas que cometen los internos. Con la realización de este proceso se mejorará el estudio y la detección de patrones de comportamiento de los internos y se potenciará la toma de decisiones.
- **Medida Disciplinaria:** las medidas disciplinarias son impuestas a los internos según las indisciplinas que son registradas y que no hayan sido sancionadas con anterioridad, y de acuerdo



Capítulo 2: *Análisis y Diseño de la solución*

al nivel de gravedad de estas. Con la realización de este proceso, se podrá almacenar y estudiar la información referente a las medidas disciplinarias que se le imponen a los internos y las indisciplinas asociadas a estas.

2. Visitas a los internos

El tema analítico Visitas a los internos, incluye el proceso Visita y almacena los principales datos relacionados con las diferentes visitas que reciben los internos.

- **Visita:** existen dos tipos de visitas: las familiares y las institucionales. Dentro de las familiares entran aquellas que realizan ciudadanos que ejerzan una influencia positiva para los internos, algún familiar y/o cónyuge. Forman parte de las institucionales, las llevadas a cabo por: jueces, fiscales, abogados, funcionarios y representantes de organismos de la administración central del estado. Con el diseño de este proceso se permite que tanto para visitas familiares como institucionales, sea posible realizar búsquedas sobre las visitas realizadas a través del registro histórico de visitantes.

3. Acciones de Control

El tema analítico Acciones de Control, incluye el proceso Requisa. En este tema se almacenan los datos más importantes de las acciones que se realizan en los centros penitenciarios en pos de lograr una mayor seguridad, tanto para los internos como para los oficiales.

- **Requisa:** las requisas son las inspecciones que se realizan en los centros penitenciarios. Estas pueden ser efectuadas a privados de libertad, locales o visitantes. De cada requisa se registran los objetos incautados y el lugar donde se realizó. Los objetos incautados a un privado de libertad quedan reflejados en su expediente legal. Este proceso permite mantener actualizado el registro de requisas realizadas en los establecimientos penitenciarios y generar reportes que facilitan el control y análisis de este tipo de información. Lo antes mencionado conduce a un fortalecimiento en la toma de decisiones, pues de esta manera se puede estudiar la información sobre varios de los aspectos que influyen en la seguridad de estos centros.



Capítulo 2: Análisis y Diseño de la solución

2.2. Especificación de requisitos

Las necesidades de información para los funcionarios de la DEP, son aquellas especificaciones que dan cumplimiento a sus tareas internas. Con la intención de controlar estas necesidades, se han especificado un conjunto de medidas e indicadores que muestran su comportamiento. Se identificaron las necesidades de la organización y los requisitos a través del estudio de los objetivos y de los indicadores de cada proceso, en los que fueron identificados 69 requisitos de información.

2.2.1. Requisitos de información

“Los requisitos de información describen qué información debe almacenar el sistema para satisfacer las necesidades de los clientes y usuarios a la hora de hacer las consultas.” (10) Componen la principal entrada para el proceso de inteligencia de negocio y para la creación de futuros reportes.

A continuación se especifican los requisitos de información que fueron identificados. Estos se muestran agrupados por procesos para lograr una mayor organización de los mismos.

Incidentes

- R11. Mostrar cantidad de incidencias por nivel de mando.
- R12. Mostrar cantidad de incidencias por tipo de incidencias.
- R13. Mostrar cantidad de incidencias por tipo de centro.
- R14. Mostrar cantidad de incidencias por meses.
- R15. Mostrar cantidad de incidencias por área.
- R16. Mostrar cantidad de incidencias por tipo de incidencias por meses.
- R17. Mostrar cantidad de incidencias por seguimiento por nivel de mando.
- R18. Mostrar cantidad de incidencias por seguimiento MININT por nivel de mando.
- R19. Mostrar cantidad de incidencias por ampliación por nivel de mando.
- R110. Mostrar cantidad de internos con incidencias por nivel de mando.
- R111. Mostrar cantidad de internos con incidencias contra cantidad de incidencias por nivel de mando.
- R112. Mostrar internos con records de incidencias por nivel de mando.
- R113. Mostrar internos con records de incidencias por tipo de incidencias.
- R114. Mostrar centros con records de incidencias por tipo de incidencias.



Capítulo 2: Análisis y Diseño de la solución

RI15. Mostrar tasa de participantes en un período por centro.

Indisciplina

RI16. Mostrar cantidad de indisciplinas por denominación por nivel de mando.

RI17. Mostrar cantidad de indisciplinas por nivel de mando.

RI18. Mostrar cantidad de indisciplinas por tipo de indisciplinas.

RI19. Mostrar cantidad de indisciplinas por área.

RI20. Mostrar cantidad de indisciplinas por tipo de centro.

RI21. Mostrar cantidad de indisciplinas por meses.

RI22. Mostrar cantidad de indisciplinas por tipo de indisciplinas por meses.

RI23. Mostrar internos con records de indisciplinas por nivel de mando.

RI24. Mostrar cantidad de internos con indisciplinas contra cantidad de indisciplinas por nivel de mando.

RI25. Mostrar cantidad de internos con indisciplinas por tipo de indisciplinas.

RI26. Mostrar internos con records de indisciplinas por tipo de indisciplinas.

RI27. Mostrar centros con records de indisciplinas por tipo de indisciplinas.

RI28. Mostrar centros con records de indisciplinas por nivel de mando.

Medidas disciplinaria

RI29. Mostrar cantidad de medidas disciplinarias por nivel de mando.

RI30. Mostrar cantidad de medidas disciplinarias por tipo de medida.

RI31. Mostrar cantidad de medidas disciplinarias por tipo de indisciplinas.

RI32. Mostrar cantidad de medidas disciplinarias por denominación de la indisciplina por nivel de mando.

RI33. Mostrar cantidad de medidas disciplinarias por reclamaciones por provincia.

RI34. Mostrar cantidad de días que impusieron y cantidad de días que fueron efectivos por nivel de mando, en caso que la medida disciplinaria sea celda disciplinaria.

RI35. Mostrar cantidad de internos con medidas disciplinarias por nivel de mando.



Capítulo 2: Análisis y Diseño de la solución

RI36. Mostrar cantidad de internos con medidas disciplinarias contra medidas disciplinarias por nivel de mando.

Requisa

RI37. Mostrar cantidad de requisas generales realizadas por nivel de mando.

RI38. Mostrar cantidad de requisas parciales planificadas por nivel de mando.

RI39. Mostrar cantidad de requisas parciales no planificadas por nivel de mando.

RI40. Mostrar cantidad de requisas generales por área.

RI41. Mostrar cantidad de requisas visita realizadas por nivel de mando.

RI42. Mostrar cantidad de requisas internos realizados por nivel de mando.

RI43. Mostrar cantidad de requisas paquetes realizadas por nivel de mando.

RI44. Mostrar cantidad de objetos encontrados en requisas de paquete por nivel de mando.

RI45. Mostrar cantidad de objetos encontrados por nivel de mando.

RI46. Mostrar cantidad de objetos encontrados por área.

RI47. Mostrar cantidad de objetos encontrados por tipo de requisa.

RI48. Mostrar cantidad de objetos encontrados en requisas generales por nivel de mando.

RI49. Mostrar cantidad de objetos encontrados en requisas parciales por nivel de mando.

RI50. Mostrar cantidad de objetos encontrados en requisas visita por nivel de mando.

RI51. Mostrar cantidad de objetos encontrados en requisas parciales planificadas por nivel de mando.

RI52. Mostrar cantidad de objetos encontrados en requisas parciales no planificadas por nivel de mando.

RI53. Mostrar cantidad de objetos encontrados en requisas interno por nivel de mando.

Visita

RI54. Mostrar cantidad de internos visitados por nivel de mando.

RI55. Mostrar cantidad de internos visitados contra visitas por nivel de mando.

RI56. Mostrar cantidad de visitas por nivel de mando.

RI57. Mostrar cantidad de visitas familiares realizadas por nivel de mando.



Capítulo 2: Análisis y Diseño de la solución

- RI58.** Mostrar cantidad de visitas conyugales realizadas por nivel de mando.
- RI59.** Mostrar cantidad de visitas por tipo visita por nivel de mando.
- RI60.** Mostrar cantidad de visitas por estímulo por nivel de mando.
- RI61.** Mostrar cantidad de visitas autorizadas fuera del plan de visitas por nivel de mando.
- RI62.** Mostrar cantidad de visitas al ingreso por nivel de mando.
- RI63.** Mostrar cantidad de visitas familiares por estímulos por nivel de mando.
- RI64.** Mostrar cantidad de visitas conyugales por estímulos por nivel de mando.
- RI65.** Mostrar cantidad de visitas institucionales por nivel de mando.
- RI66.** Mostrar cantidad de visitas de la comunidad por nivel de mando.
- RI67.** Mostrar cantidad de visitas consulares por nivel de mando.
- RI68.** Mostrar cantidad de visitas proyecto por nivel de mando.
- RI69.** Mostrar cantidad de internos que nunca han recibido visita por nivel de mando.

2.2.2. Especificación de requisitos de información

A continuación se describen algunos requisitos de información de los procesos Incidencia e Indisciplina. Estos se muestran agrupados por procesos para lograr una mayor organización de los mismos. Las demás descripciones se encuentran en el expediente del Proyecto “Prisiones Cuba”.

Nº	Nombre	Descripción	Complejidad	Prioridad para el cliente
R11	Mostrar cantidad de incidencias por provincia.	<p>Muestra el total de incidencias ocurridas por provincia, dando la posibilidad de cambiar la vista al reporte mediante una lista desplegable (gráfica de pastel, gráfica de barras o tabla dinámica).</p> <p>Por defecto, se muestra la información en una gráfica de pastel, donde cada cuña representa el total de incidencias de los últimos 5 años por provincia.</p> <p>Al seleccionar una de ellas, se muestran las incidencias por centro y al seleccionar un centro, se muestran las</p>	Alta	Alta



Capítulo 2: Análisis y Diseño de la solución

		incidencias por colectivo.		
	Campos	Tipos de datos	Reglas o restricciones	
	Cantidad	Numérico	<ul style="list-style-type: none"> Para los internos que tengan el expediente abierto habilitado. 	
	Provincia	Cadena de caracteres		
	Observaciones			
R12	Mostrar cantidad de incidencias por tipo de centro	<p>Muestra el total de incidencias ocurridas en cada uno de los tipos de centro, dando la posibilidad de cambiar la vista al reporte mediante una lista desplegable (gráfica de barras o tabla dinámica).</p> <p>Por defecto, se muestra la información en una gráfica de barras, donde cada barra constituye una provincia con las incidencias ocurridas por el tipo de centro.</p> <p>Al seleccionar una provincia, se muestran las incidencias por centro por tipo de centro y al seleccionar un centro, se muestran las incidencias por colectivo por tipo de centro.</p>	Alta	Alta
	Campos	Tipos de datos	Reglas o restricciones	
	Cantidad	Numérico	<ul style="list-style-type: none"> Para los internos que tengan el expediente abierto habilitado. 	
	Provincia	Cadena de caracteres		
	Tipo de centro	Cadena de caracteres		
		Observaciones		
R13	Mostrar cantidad de incidencias por meses.	<p>Muestra el total de incidencia por meses, dando la posibilidad de cambiarle la vista al reporte mediante una lista desplegable (gráfica de barras o tabla dinámica).</p> <p>Por defecto, se muestra la información en una gráfica de barras, donde cada barra representa el total de incidencias de los últimos 5 años por meses por</p>	Alta	Alta



Capítulo 2: Análisis y Diseño de la solución

	provincia. Al seleccionar una provincia, se muestran las incidencias por centro por meses y al seleccionar un centro, se muestran las incidencias por colectivo por meses. Al seleccionar un mes, se muestra el total de incidencias y el día en que ocurrieron.		
Campos	Tipos de datos	Reglas o restricciones	
Cantidad	Numérico	<ul style="list-style-type: none"> Para los internos que tengan el expediente abierto habilitado. 	
Provincia	Cadena de caracteres		
Meses	Cadena de caracteres		
Observaciones			

Tabla 1: Especificación de requisitos de información del proceso Incidencia.

Nº	Nombre	Descripción	Complejidad	Prioridad para el cliente
R11	Mostrar cantidad de indisciplinas por provincia.	<p>Muestra el total de indisciplinas ocurridas por provincia, dando la posibilidad de cambiar la vista al reporte mediante una lista desplegable (gráfica de pastel, gráfica de barras o tabla dinámica).</p> <p>Por defecto, se muestra la información en una gráfica de pastel, donde cada cuña representa el total de indisciplinas de los últimos 5 años por provincia.</p> <p>Al seleccionar una de ellas, se muestran las indisciplinas por centro y al seleccionar un centro, se muestran las indisciplinas por colectivo.</p>	Alta	Alta
	Campos	Tipos de datos	Reglas o restricciones	
	Cantidad	Numérico	<ul style="list-style-type: none"> Para los internos que tengan el expediente abierto habilitado. 	
	Provincia	Cadena de caracteres		
	Observaciones			



Capítulo 2: Análisis y Diseño de la solución

RI2	Mostrar cantidad de indisciplinas por meses.	<p>Muestra el total de indisciplinas por meses, dando la posibilidad de cambiarle la vista al reporte mediante una lista desplegable (gráfica de barras o tabla dinámica).</p> <p>Por defecto, se muestra la información en una gráfica de barras, donde cada barra representa el total de indisciplinas de los últimos 5 años por meses por provincia.</p> <p>Al seleccionar una provincia, se muestran las indisciplinas por centro por meses y al seleccionar un centro, se muestran las indisciplinas por colectivo por meses. Al seleccionar un mes, se muestra el total de indisciplinas y el día en que ocurrieron.</p>	Alta	Alta
	Campos	Tipos de datos	Reglas o restricciones	
	Cantidad	Numérico	<ul style="list-style-type: none"> Para los internos que tengan el expediente abierto habilitado. 	
	Provincia	Cadena de caracteres		
	Meses	Cadena de caracteres		
	Observaciones			
RI3	Mostrar cantidad de indisciplinas por denominación.	<p>Muestra el total de indisciplinas ocurridas por denominación (gravedad), dando la posibilidad de cambiarle la vista al reporte mediante una lista desplegable (gráfica de barras o tabla dinámica).</p> <p>Por defecto, se muestra la información en una gráfica de barras, donde cada barra representa el total de indisciplinas de los últimos 5 años por la denominación de la indisciplinación (grave, menos grave y leve) por provincia.</p> <p>Al seleccionar una provincia, se muestran las indisciplinas por centro</p>	Alta	Alta



Capítulo 2: Análisis y Diseño de la solución

		por denominación y al seleccionar un centro, se muestran las indisciplinas por colectivo por denominación. Al seleccionar una denominación, se muestra el total de indisciplinas por tipo de indisciplina.	
Campos	Tipos de datos	Reglas o restricciones	
Cantidad	Numérico	<ul style="list-style-type: none"> Para los internos que tengan el expediente abierto habilitado. 	
Provincia	Cadena de caracteres		
Denominación	Cadena de caracteres		
Observaciones			

Tabla 2: Especificación de requisitos de información del proceso Indisciplina.

2.2.3. Matriz dimensional

A partir del análisis de los requisitos realizado anteriormente, se procede a construir la matriz dimensional. Este artefacto es realizado con el objetivo de darle prioridad a los procesos de negocios más críticos, de acuerdo al criterio del cliente, y de tener el control de los cambios que puedan ocurrir en una dimensión o proceso determinado.

Dimensiones	Procesos				
	Incidencia	Indisciplina	Medida Disciplinaria	Requisa	Visita
Area	x	x		x	
Interno	x	x	x		x
Tiempo	x	x	x	x	x
Nivel de mando	x	x	x	x	x
Incidencia	x				
Indisciplina		x	x		
Medida Disciplinaria			x		
Requisa				x	
Visita					x

Tabla 3: Matriz dimensional.

La matriz dimensional presenta en sus columnas los procesos de negocio identificados y en sus filas las dimensiones del mercado de datos. Una X en la intersección de las filas y las columnas, representa las



Capítulo 2: Análisis y Diseño de la solución

relaciones que existen entre los procesos que constituyen el área de Seguridad Penitenciaria y las dimensiones que los componen.

2.2.4. Técnica de validación de requisitos

La metodología utilizada para la implementación del mercado de datos, propone como técnica de validación de requisitos el desarrollo de un prototipo funcional. Este se hace con el objetivo de mostrarle al cliente una versión reducida del producto final, permitiéndole interactuar y conformar nuevas ideas para la implementación del mercado de datos. Para realizar el prototipo del mercado de datos para el área de Seguridad Penitenciaria, se tuvieron en cuenta los ejemplos que ofrece la Suite de Oracle Business Intelligence.

2.3. Modelado dimensional

El modelado dimensional es un proceso iterativo. Este se obtiene a partir de los procesos priorizados de la matriz dimensional descrita anteriormente. Luego del análisis de los requisitos y de especificados los temas analíticos, se identificaron 5 procesos y se tomó la decisión, en conjunto con la DEP, de descartar el proceso de Control Físico y Control de Equipamiento debido a que no aportaban información relevante para el proceso de toma de decisiones.

2.3.1. Granularidad

El primer paso para conformar el modelado dimensional es especificar la granularidad. La misma representa el nivel de detalle que posee cada registro de una tabla de hechos. Para el desarrollo del mercado de datos se tuvo en cuenta el mayor nivel de detalle posible, con el propósito de realizar agrupamientos al nivel deseado a partir de los datos actuales.

Los niveles de granularidad a tener en cuenta para el modelo de datos, son: en el tiempo se llega hasta el nivel de detalle diario y en el nivel de mando se identifica hasta el colectivo.

2.3.2. Tablas de dimensiones

Las tablas de dimensiones tienen un conjunto de atributos que brindan una perspectiva o forma de análisis sobre una medida en una tabla hechos. Estos atributos dimensionales permiten filtrar los indicadores que se quieren obtener.

Luego de detallado el nivel de granularidad de los datos, se identificaron 9 tablas de dimensiones. Entre las tablas de dimensiones existentes, hay 4 que son comunes debido a que son utilizadas por varios



Capítulo 2: *Análisis y Diseño de la solución*

procesos. A continuación, se hace referencia a las tablas de dimensiones de los procesos Incidencia e Indisciplina:

Tablas de dimensiones comunes:

- **DIM_INTERNO:** guarda los datos relacionados con los internos.
- **DIM_TIEMPO:** almacena la información relacionada con el tiempo.
- **DIM_NIVEL_MANDO:** guarda la información referente a los distintos niveles de mando existentes en los centros penitenciarios.
- **DIM_AREA:** almacena los datos relacionados con las distintas áreas existentes en los centros penitenciarios.

Tablas de dimensiones específicas:

- **DIM_INCIDENCIA:** guarda la información de las incidencias ocurridas en los establecimientos penitenciarios.
- **DIM_INDISCIPLINA:** almacena los datos pertenecientes a las indisciplinas ocurridas en los centros penitenciarios.

2.3.3. Tablas de hechos y medidas

La tabla de hechos representa la ocurrencia de un determinado suceso dentro de la organización y generalmente almacenan medidas numéricas. Los hechos son filtrados, agrupados y explorados a través de atributos definidos en las tablas de dimensiones.

Se identificaron 7 tablas de hechos. A continuación se hará referencia a las tablas de hechos a utilizar en los procesos Incidencia e Indisciplina:

- **FACT_INCIDENCIA:** guarda la información referente a la ocurrencia de una incidencia.
- **FACT_INDISCIPLINA:** almacena los datos relacionados con la ocurrencia de una indisciplina.

2.3.4. Modelo dimensional detallado

Se concluye el modelado dimensional del mercado de datos con la construcción del modelo dimensional detallado. Este es una descripción conceptual de la estructura de la base de datos y representa la información a través de relaciones y atributos. Por medio de este modelo quedan especificados y relacionados los procesos con sus posibles tablas de dimensiones. Posteriormente, en la realización de



Capítulo 2: Análisis y Diseño de la solución

los modelos de datos, los nombres de los atributos, los nombres de las tablas y la cantidad de atributos puede cambiar; todo en dependencia de las necesidades del cliente.

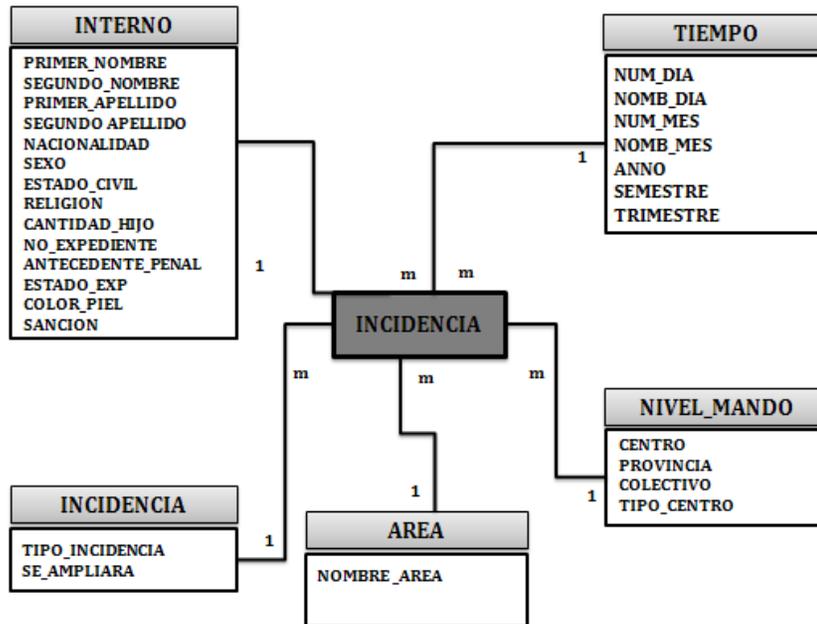


Figura 10: Modelo dimensional detallado del proceso Incidencia.

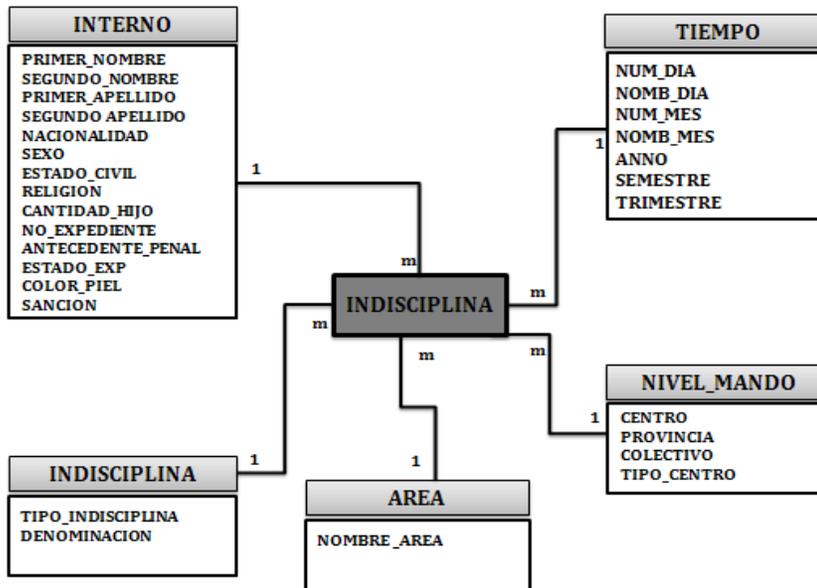


Figura 11: Modelo dimensional detallado del proceso Indisciplina.



Capítulo 2: *Análisis y Diseño de la solución*

2.3.5. Patrones de diseño utilizados

A continuación se mencionan los patrones de diseño utilizados para la realización de los modelos lógicos. Estos patrones se emplearon para resolver problemas comunes en el desarrollo del mercado de datos, debido a que generalmente indican la mejor forma de darle solución a estos.

- **Llaves subrogadas:**

El patrón llaves subrogadas se aplica para darle solución al problema de que los identificadores de las tablas de dimensiones no se pueden repetir. Esto se resuelve mediante la creación de secuencias; es decir, se genera una llave primaria única para cada entidad. Esta solución garantiza que estos identificadores no se repitan. Además, ocupan menos espacio en la base de datos y tienen más rendimiento que las tradicionales llaves naturales.

- **Tablas de hechos sin hechos:**

Este patrón se utiliza cuando hay tablas de hechos que no tienen medidas y representan la ocurrencia de un evento determinado. Las tablas de hechos sin hechos solo tienen como atributos los identificadores de las dimensiones con las que está relacionada.

- **Dimensiones lentamente cambiante:**

Las dimensiones lentamente cambiantes (SCD), son dimensiones que tienden a modificar sus atributos a través del tiempo. Cuando ocurren estos cambios en las dimensiones se pueden establecer varias estrategias a seguir para su procesamiento:

- ✓ **SCD Tipo 1 (Sobrescribir):** cuando existen cambios en algún valor de un atributo de la tabla de dimensión, se sobrescribe el valor antiguo con el nuevo. Esto significa que no guarda los cambios históricos de los datos, por lo que este tipo es usado solo en los casos en donde la información histórica no sea importante.
- ✓ **SCD Tipo 2 (Añadir fila):** cuando existen cambios en algún valor de un atributo de la tabla de dimensión, se añade a la tabla de dimensión una nueva fila con el valor del atributo a cambiar. Permite guardar el registro histórico de los datos.
- ✓ **SCD Tipo 3 (Añadir columna):** cuando existen cambios en algún valor de un atributo de la tabla de dimensión, se agrega a la tabla de dimensión una columna adicional con el nuevo valor.



Capítulo 2: Análisis y Diseño de la solución

Para el desarrollo del mercado de datos se decidió utilizar la dimensión lentamente cambiante de tipo 2, debido a que posibilita almacenar los valores históricos de los datos. El mercado de datos tiene como dimensión lentamente cambiante, la dimensión interno. Esta posee atributos que pueden cambiar con el tiempo, como por ejemplo: número de hijos y estado civil. Por esta razón se hace necesario para el proceso de toma de decisiones, conservar los diferentes valores que pueden tomar los datos de los internos.

2.3.6. Modelos de datos

Los modelos de datos de los procesos, forman parte de la solución a los requisitos de información descritos anteriormente. Seguidamente, se pueden ver los modelos de datos de los procesos Incidencia e Indisciplina.

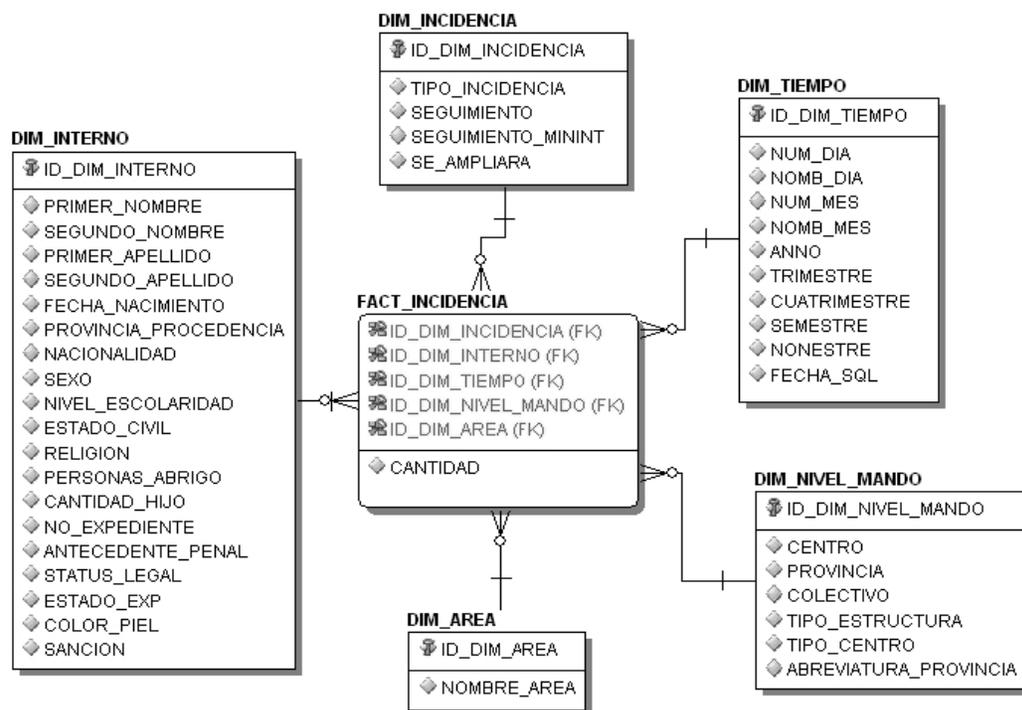


Figura 12: Modelo de datos del proceso Incidencia.



Capítulo 2: Análisis y Diseño de la solución

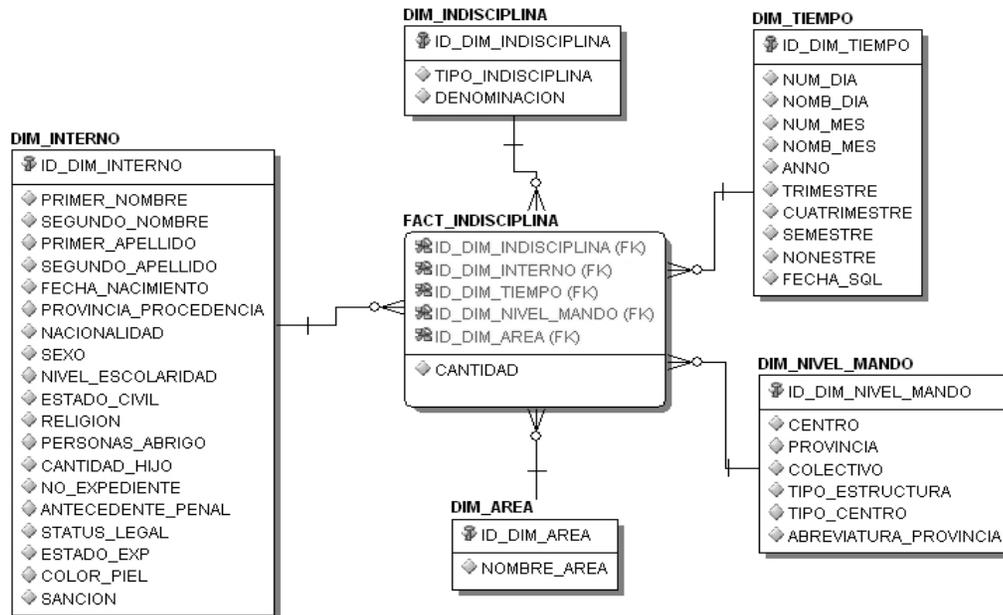


Figura 13: Modelo de datos del proceso Indisciplina.

Estos modelos se prueban contra los requisitos de información. Esto se hace comprobando si el modelo de datos da respuesta a diferentes preguntas, por ejemplo: ¿Se puede obtener la cantidad de incidencias por meses para cada área? ¿Se puede obtener la cantidad de indisciplinas por tipo de indisciplinas por meses por área?

2.4. Diseño de la ETL

Una vez diseñado el modelo de datos del área de Seguridad Penitenciaria, se procede a realizar el diseño de la ETL de los datos. En este se identifican los datos necesarios de la fuente origen para poblar las tablas de dimensiones. La fuente de origen utilizada para la obtención de los datos es la base de datos del SIDE. A continuación se muestran los atributos extraídos para llenar cada dimensión y su respectivo diseño de la ETL.

✓ Dimensión Incidencia

Tabla	Atributos extraídos
NOM_INCIDENCIA	<ul style="list-style-type: none"> INCIDENCIA
INCIDENCIA	<ul style="list-style-type: none"> SE_AMPLIARA SEGUIMIENTO SEGUIMIENTOMININT

Tabla 4: Atributos extraídos para la dimensión Incidencia.



Capítulo 2: Análisis y Diseño de la solución

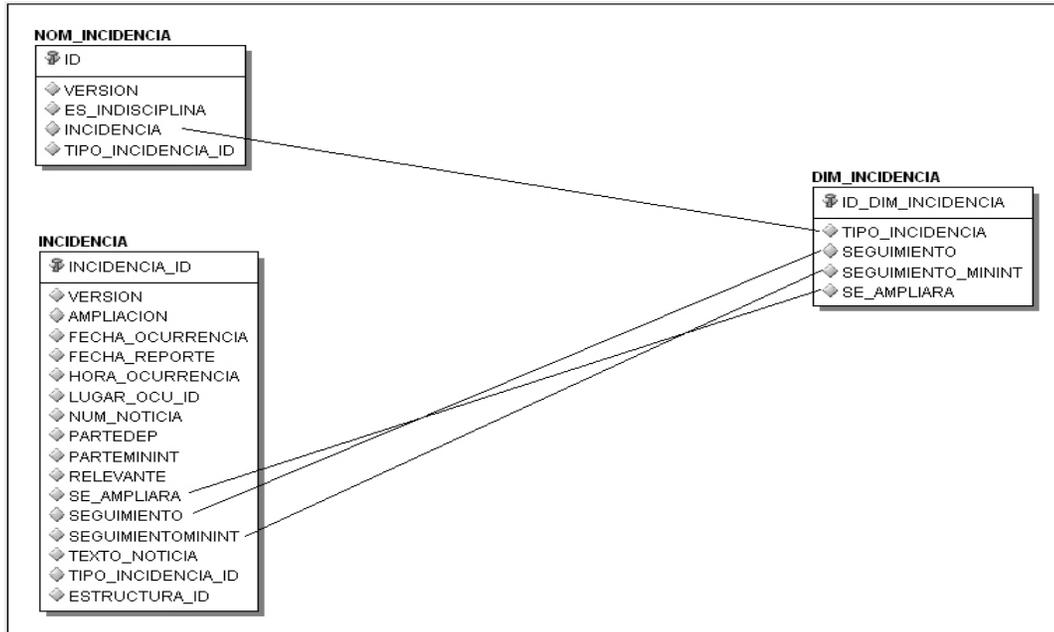


Figura 14: Diseño de la ETL de la dimensión Incidencia.

✓ Dimensión Indisciplina

Tabla	Atributos extraídos
NOM_INDISCIPLINA	▪ TIPO_INDISCIPLINA
NOM_GRAVE_INDIS	▪ GRAVEDAD

Tabla 5: Atributos extraídos para la dimensión Indisciplina.

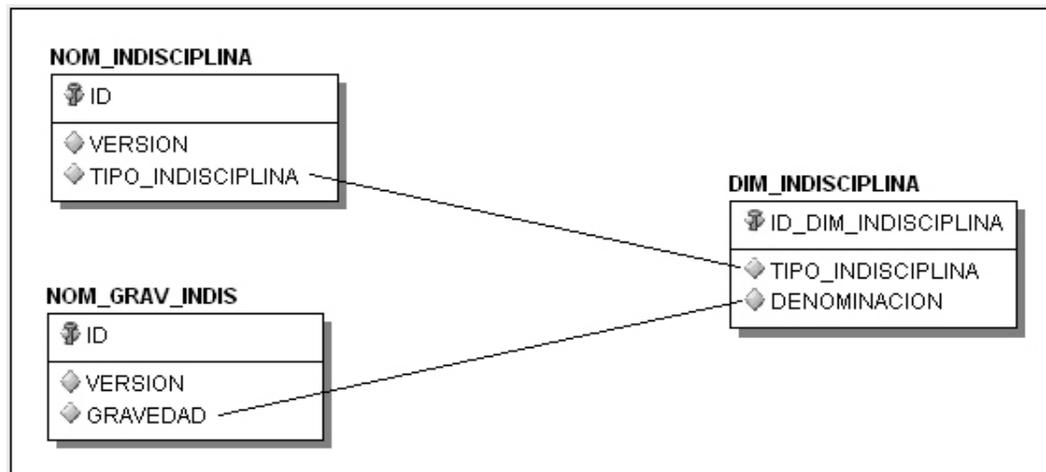


Figura 15: Diseño de la ETL de la dimensión Indisciplina.



Capítulo 2: Análisis y Diseño de la solución

✓ Dimensión Nivel de Mando

Tabla	Atributos extraídos
ESTRUCTURA	<ul style="list-style-type: none"> NOMBRE_ESTRUCTURA
NOM_PROVINCIA	<ul style="list-style-type: none"> DESCRIPCIÓN_CORTA PROVINCIA
NOM_T_CENTRO	<ul style="list-style-type: none"> CENTRO
NOM_T_EST	<ul style="list-style-type: none"> ESTRUCTURA
COLECTIVO	<ul style="list-style-type: none"> NUMERO_COLECTIVO

Tabla 6: Atributos extraídos para la dimensión Nivel de Mando.

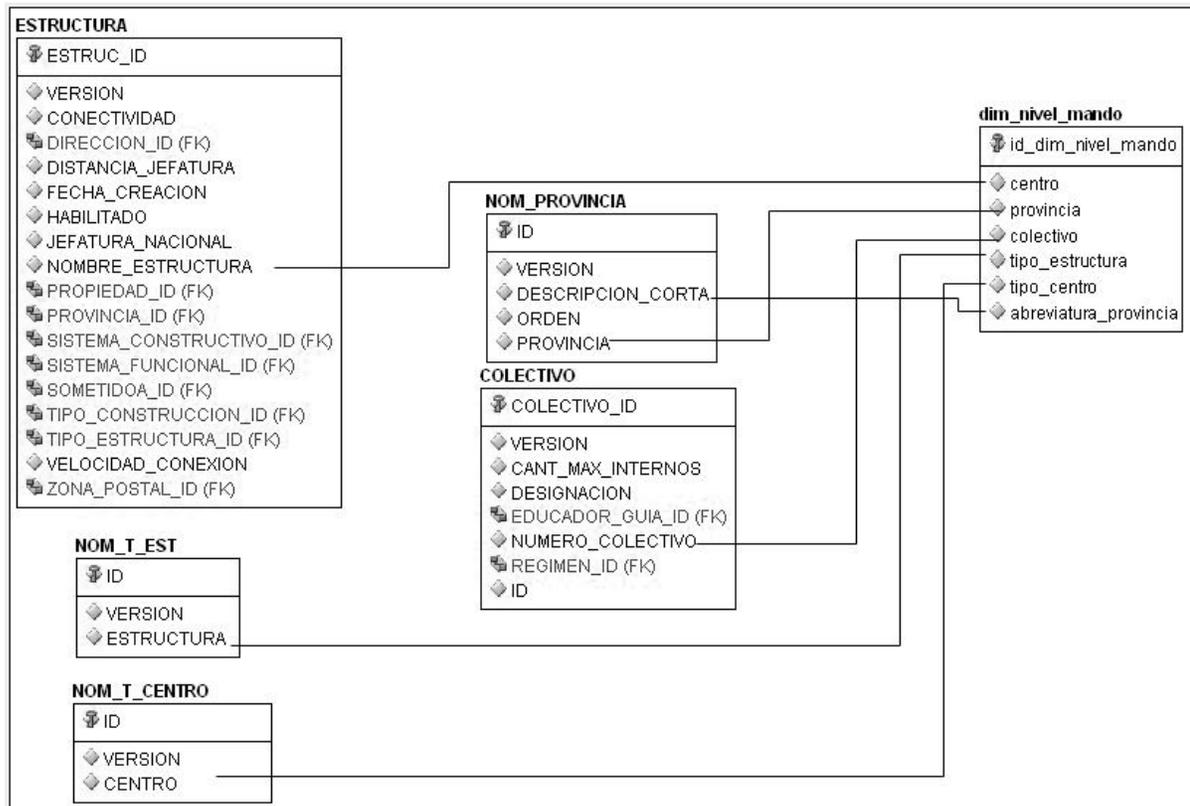


Figura 16: Diseño de la ETL de la dimensión Nivel de Mando.

✓ Dimensión Interno

Tabla	Atributos extraídos
PERSONA	<ul style="list-style-type: none"> PRIMERNOMBRE PRIMERPELLIDO



Capítulo 2: Análisis y Diseño de la solución

	<ul style="list-style-type: none"> ▪ SEGUNDONOMBRE ▪ SEGUNDOAPELLIDO ▪ SEXO
NOM_PROVINCIA	<ul style="list-style-type: none"> ▪ PROVINCIA
NOM_NIV_ESC	<ul style="list-style-type: none"> ▪ NIVELESCOLARIDAD
NOM_COLORPIEL	<ul style="list-style-type: none"> ▪ COLORPIEL
NOM_ESTADOCIVIL	<ul style="list-style-type: none"> ▪ ESTADOCIVIL
NOM_RELIGION	<ul style="list-style-type: none"> ▪ RELIGION
NOM_ANTEC_PENALES	<ul style="list-style-type: none"> ▪ ANTECEDENTES_PENALES
NOM_ESTATUS_LEGAL	<ul style="list-style-type: none"> ▪ ESTATUS
INTERNO	<ul style="list-style-type: none"> ▪ FECHANACIMIENTO ▪ PERSONASABRIGO
NOM_NACIONALIDAD	<ul style="list-style-type: none"> ▪ NACIONALIDAD
INTERNO_HIJOS	<ul style="list-style-type: none"> ▪ INTERNO_ID
EXP_LEG	<ul style="list-style-type: none"> ▪ ACTIVO ▪ HABILITADO ▪ NUMEROEXPEDIENTE
LIQ_SAN_SAN	<ul style="list-style-type: none"> ▪ SANCION
LIQ_SAN_ASE	<ul style="list-style-type: none"> ▪ SANCION

Tabla 7: Atributos extraídos para la dimensión Interno.



Capítulo 2: Análisis y Diseño de la solución

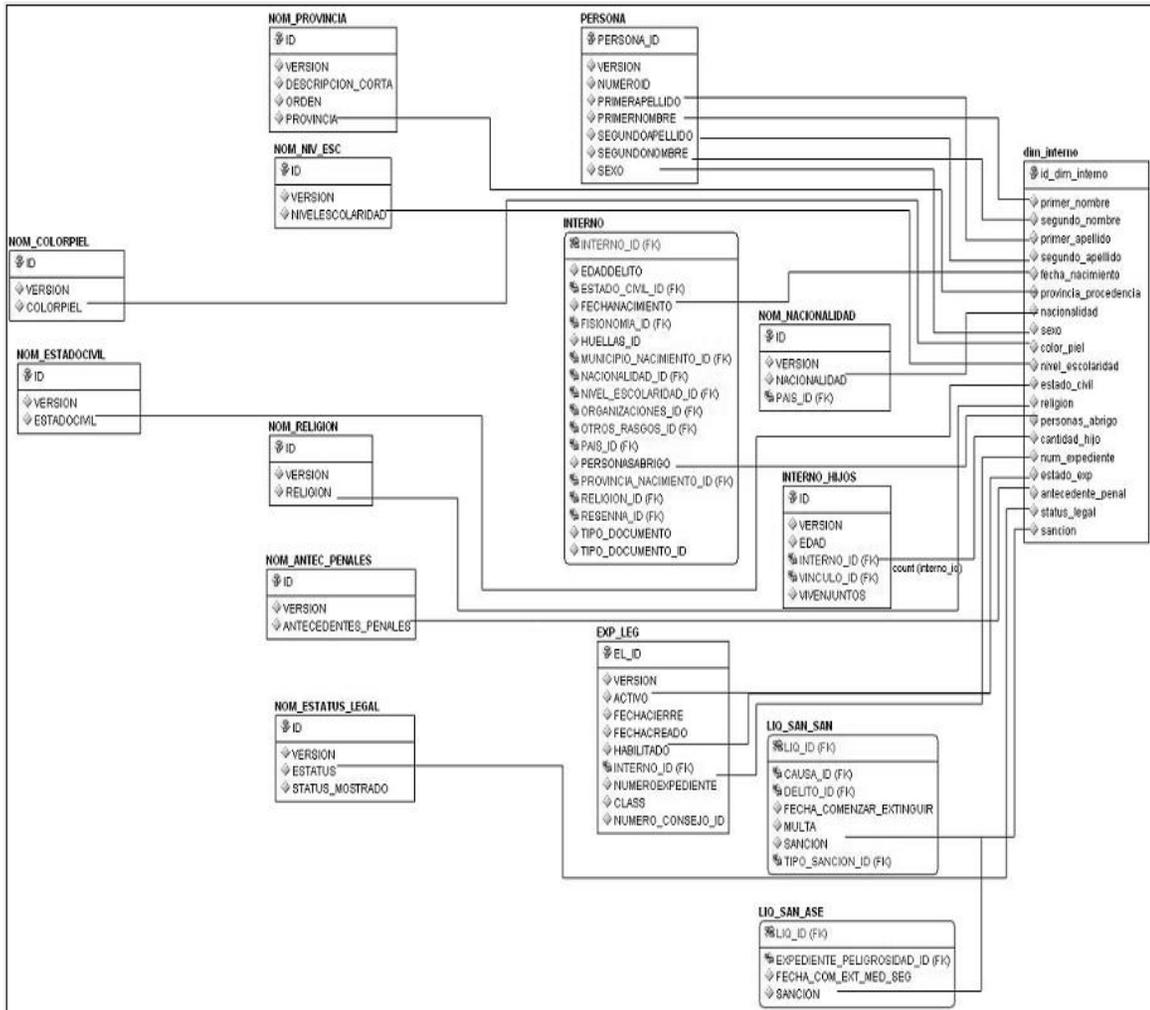


Figura 17: Diseño de la ETL de la dimensión Interno.

Conclusiones Parciales

En el presente capítulo se realizó el análisis y diseño de la propuesta de solución, para lo cual se identificaron los temas analíticos con sus correspondientes procesos de negocio. Esto dio lugar a la identificación y priorización de los requisitos de información del mercado de datos para el área de Seguridad Penitenciaria y la construcción de la matriz dimensional.

Para darle solución a los requisitos de información identificados, se realizó el modelado dimensional mediante la identificación de los niveles de granularidad, así como las tablas de dimensiones y tablas de hechos. Se realizaron los modelos dimensionales detallados y los modelos de datos de cada uno de los



Capítulo 2: Análisis y Diseño de la solución

procesos de negocio, haciendo uso de los esquemas dimensionales definidos con anterioridad. Además, se realizó el diseño de ETL de los procesos de negocio del área de Seguridad Penitenciaria.



Capítulo 3: Implementación y Prueba

CAPÍTULO 3: IMPLEMENTACIÓN Y PRUEBA

Introducción

En el presente capítulo se especifica la configuración de la Herramienta de Administración (*Administration Tool*). También se describe la implementación del proceso de ETL para cargar las dimensiones y las tablas de hechos. Se especifica la aplicación de inteligencia de negocio para el usuario final. Además, se realizan pruebas a la aplicación antes mencionada, utilizando las listas de chequeo y se muestran los resultados obtenidos.

3.1. Implementación de la ETL

Luego de diseñado el proceso de ETL se realiza la implementación del mismo, para cargar la información en el mercado de datos. Para llevar a cabo esta tarea y en correspondencia con la arquitectura de tres capas, se crearon las tablas *staging area* (área de almacenamiento temporal). Estas constituyen un puente entre el sistema operacional y el mercado de datos. Estas tablas se utilizan para limpiar y transformar los datos, antes de cargarlos en las tablas de hechos. La implementación del proceso de ETL se realiza en el Oracle Warehouse Builder (OWB).

3.1.1. Proceso de ETL para cargar las dimensiones

A continuación se especifica cómo se realiza el proceso de ETL para las dimensiones relacionadas con los procesos Incidencia e Indisciplina.

▪ ETL dimensión Nivel de Mando

Para realizar el proceso de carga de la tabla DIM_NIVEL_MANDO, se seleccionan todas las tablas correspondientes desde la fuente de datos y se toman como datos de entradas algunos de los atributos que se encuentran en dichas tablas (ESTRUCTURA, COLECTIVO, COLECTIVO_LOCAL, LOCAL, AREA, NOM_T_CENTRO, CENTRO_PENIT, NOM_PROVINCIA y NOM_T_EST). Luego de seleccionados los datos, se utiliza el operador JOINER (similar al JOIN en SQL), para unir las tablas escogidas.

Después de esto se emplea el operador FILTER, para filtrar los datos de entrada de acuerdo a una condición que se le define (la condición es con el objetivo de cargar solo los tipos de estructuras siguientes: Jefatura y Centros Penitenciarios). Este filtro devuelve como resultado la información que cumpla con la condición. Luego de obtenidos los niveles de mando, se usan varias funciones para transformar algunos de sus datos. Entre estas funciones se encuentran: FUNC_TIPO_CENTRO,



Capítulo 3: Implementación y Prueba

FUNC_TO_CHAR y FUNC_COLECTIVO. A continuación se procede a utilizar el operador SET_OPERATION, nombrado UNIR_ESTRUCTURAS, con el propósito de combinar lo resultante de las funciones y el filtro. Para finalizar, se hace uso del operador CONSTANT para llenar algunos datos que son importantes y que no se encuentran en la base de datos origen.

Los valores que tendrá el identificador de la dimensión son los generados por la secuencia SEQ_DIM_NIVEL_MANDO. Una vez obtenidos todos los datos, se procede a mapearlos con los de la tabla DIM_NIVEL_MANDO como se muestra en la Figura 18.

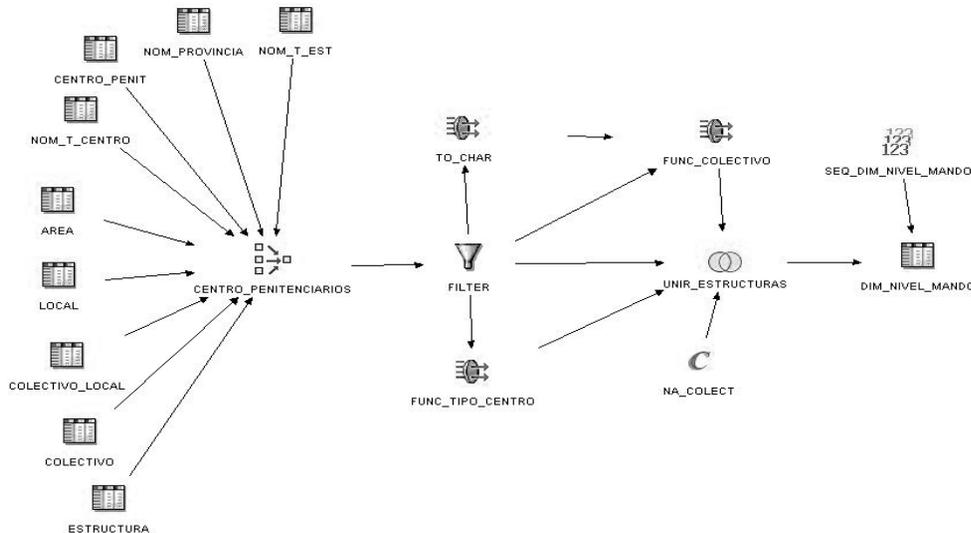


Figura 18: Correspondencia de la tabla de dimensión Nivel de Mando.

Al finalizar el proceso de ETL de la dimensión Nivel de Mando, en la base de datos quedan cargados los atributos de la tabla DIM_NIVEL_MANDO (provincia, centro, colectivo, tipo de estructura y tipo centro). Estos datos luego se usan para filtrar los indicadores de los procesos con los que se relaciona esta dimensión.

▪ ETL dimensión Tiempo

Para realizar la carga de la tabla DIM_TIEMPO, se define la fecha de inicio y la cantidad de años que van a ser generados. Para llevar a cabo este proceso se utiliza el componente INPUT_PARAMETER, nombrado DATE_INPUTS, que se conecta al operador de función DAY_TABLE_FUNCTION. Este operador es utilizado para la creación de las fechas que se encuentran dentro del intervalo escogido.



Capítulo 3: Implementación y Prueba

Posteriormente se utilizan varios operadores de transformación, algunos son propios del OWB y otros se implementaron para poder realizar las transformaciones necesarias.

Entre los operadores de transformación del OWB, están:

- **WB_CAL_YEAR:** dada una fecha devuelve el número del año correspondiente.
- **WB_CAL_MONTH_NAME:** dada una fecha devuelve el nombre del mes correspondiente.
- **WB_CAL_MONTH OF YEAR:** dada una fecha devuelve el número del mes correspondiente.
- **WB_DAY_NAME:** Dada una fecha devuelve el nombre del día correspondiente.

Entre los operadores implementados se encuentran:

- **CAL_CUATRIMESTRE:** recibe como parámetro el número de un mes y devuelve el número del cuatrimestre al que corresponde.
- **CAL_TRIMESTRE:** dado el número de un mes, devuelve el número del trimestre correspondiente.
- **CAL_SEMESTRE:** dado el número de un mes, devuelve el número del semestre correspondiente.

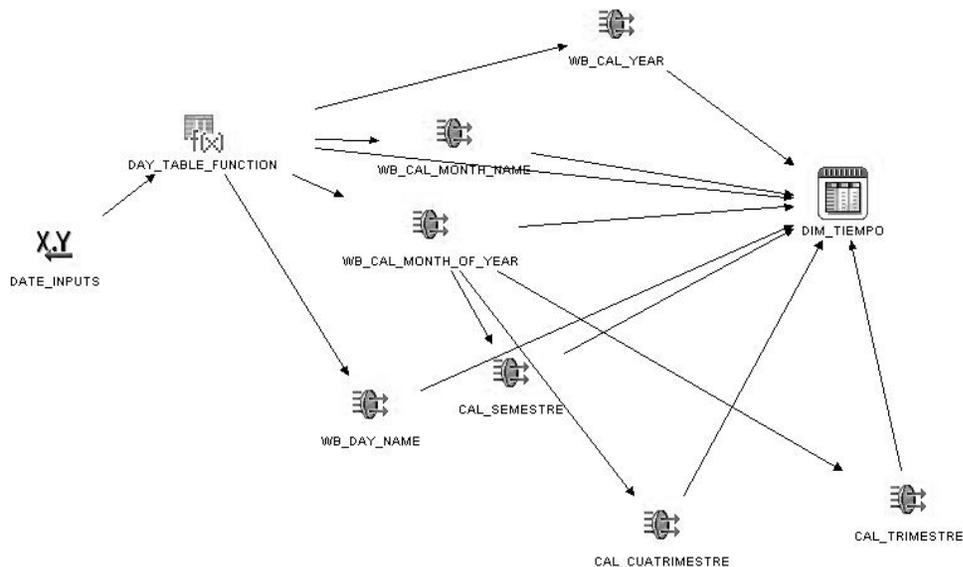


Figura 19: Correspondencia de la tabla de dimensión Tiempo.

Al finalizar el proceso de ETL de la dimensión Tiempo, en la base de datos queda la tabla DIM_TIEMPO con los atributos (día, nombre del día, mes, nombre del mes, año, semestre, trimestre y cuatrimestre) que permitirán filtrar los indicadores de todos los procesos.



Capítulo 3: Implementación y Prueba

▪ ETL dimensión Incidencia

Para realizar la carga de la tabla DIM_INCIDENCIA, se crean varias constantes mediante el operador CONSTANT. Esto es necesario hacerlo porque estos valores creados con las constantes, no existen en la base de datos origen y son de importancia para el mercado de datos. Luego se usan los operadores SET_OPERATION para unir los valores de las constantes y también se selecciona la tabla NOM_INCIDENCIA desde la fuente de datos. A continuación se usa el JOINER para unir las tablas necesarias a través de la condición de unión.

El identificador de la dimensión es un valor generado por la secuencia SEQ_DIM_INCIDENCIA. Una vez obtenidos todos los datos se procede a mapearlos con los de la tabla DIM_INCIDENCIA, como se muestra en la figura 20.

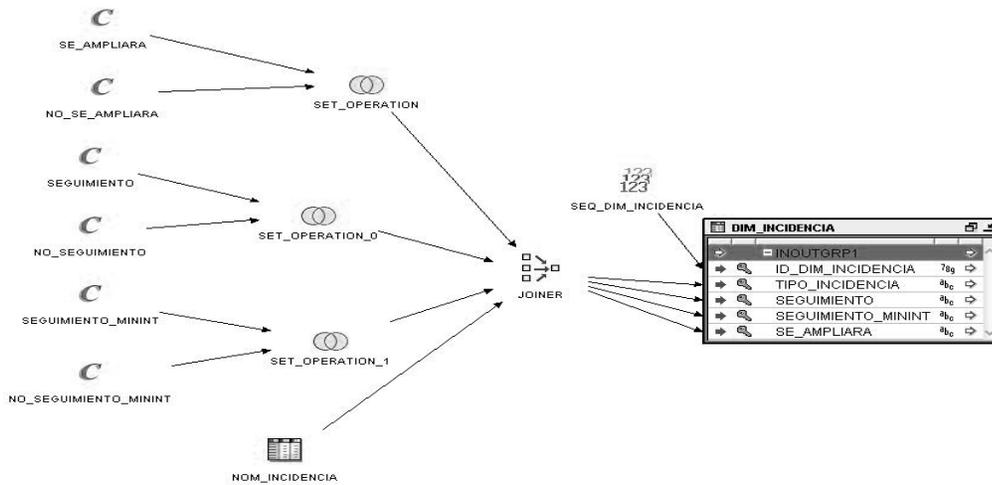


Figura 20: Correspondencia de la tabla de dimensión Incidencia.

Concluido el proceso de ETL de la dimensión Incidencia, en la base de datos queda la tabla DIM_INCIDENCIA con los atributos (incidencia, tipo de incidencia, seguimiento, seguimiento del MININT y se ampliará) que permitirán filtrar los indicadores de todos los procesos con los que se relaciona dicha dimensión.

▪ ETL dimensión Indisciplina

Con el propósito de cargar los datos de la tabla DIM_INDISCIPLINA, se seleccionan las tablas correspondientes desde la fuente de datos: NOM_GRAV_INDIS y NOM_INDISCIPLINA. Posteriormente se utiliza el componente JOINER para obtener los datos necesarios a través de la condición de unión.



Capítulo 3: Implementación y Prueba

El identificador de la dimensión es un valor generado por la secuencia SEQ_DIM_INDISCIPLINA. Una vez obtenidos todos los datos se procede a mapearlos con los de la tabla DIM_INDISCIPLINA, como se muestra en la figura 21.

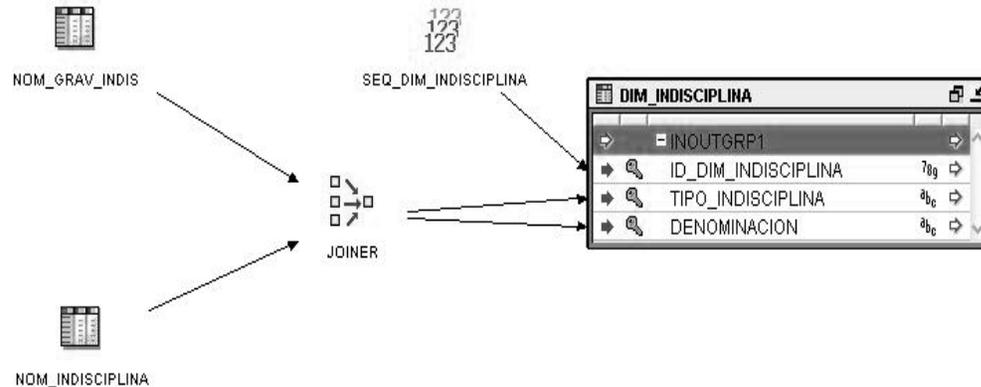


Figura 21: Correspondencia de la tabla de dimensión Indisciplina.

Concluido el proceso de ETL de la dimensión Indisciplina, en la base de datos queda la tabla DIM_INDISCIPLINA con los atributos (indisciplina, tipo de indisciplina y denominación) que permitirán filtrar los indicadores de todos los procesos con los que se relaciona esta dimensión.

3.1.2. Proceso de ETL para las tablas de hechos

Seguidamente se especifica cómo se realiza el proceso de ETL para las tablas de hechos y las tablas *staging area* relacionadas con los procesos Incidencia e Indisciplina.

▪ ETL tabla de hecho INCIDENCIA

Antes de efectuar el proceso de ETL de la tabla FACT_INCIDENCIA se realiza el de la *staging area* diseñada para este proceso, la STG_SP_INCIDENCIA. En esta se almacenan de manera intermedia los datos que luego contendrá la tabla de hecho del proceso Incidencia.

✓ ETL *staging area* INCIDENCIA

Las transformaciones para cargar la STG_SP_INCIDENCIA se inician seleccionando las tablas correspondientes. Luego se utiliza el operador JOINER para unir las tablas necesarias a través de la condición de unión (similar a una consulta SQL). Una vez obtenidos todos los datos se mapean con los atributos de la tabla STG_SP_INCIDENCIA, como se muestra en la figura 22.



Capítulo 3: Implementación y Prueba

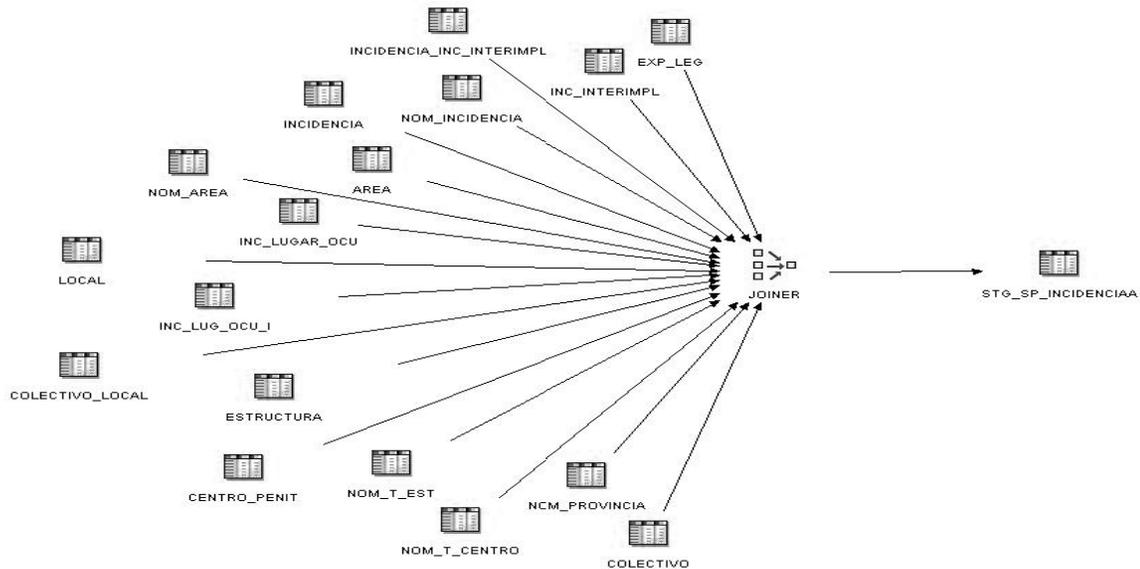


Figura 22: Correspondencia de la tabla *staging area* de Incidencia.

Cuando se llena la tabla STG_SP_INCIDENCIA, se procede a realizar el proceso de ETL de la tabla FACT_INCIDENCIA como muestra en la figura 23.

Primeramente se selecciona la tabla necesaria, en este caso es la STG_SP_INCIDENCIA. Luego se hace necesario realizar transformaciones a algunos de los datos extraídos, para poder integrarlos correctamente al mercado de datos. Las transformaciones son las siguientes:

- **FUNC_INCIDENCIA_AMPLIA:** recibe como parámetro un valor numérico y devuelve como resultado una cadena de texto. Este dato se extrae de la base de datos origen con valores de 0 o 1, luego de la modificación se carga como: “Se ampliara” si es 1 y “No se ampliara” si es 0.
- **FUNC_INCIDENCIA_SEG:** recibe como parámetro un valor numérico y devuelve como resultado una cadena de texto. Este dato se extrae de la base de datos origen con valores de 0 o 1, luego de la modificación se carga como: “Seguimiento” si es 1 y “No seguimiento” si es 0.
- **FUNC_INCIDENCIA_SEGMINI:** recibe como parámetro un valor numérico y devuelve como resultado una cadena de texto. Este dato se extrae de la base de datos origen con valores de 0 o 1, luego de la modificación se carga como: “Seguimiento MININT” si es 1 y “No seguimiento MININT” si es 0.



Capítulo 3: Implementación y Prueba

- **TRUNC:** recibe como parámetro de entrada una fecha y la devuelve truncada, de acuerdo al formato que se le especifique. Este dato se extrae de la base de datos origen con el formato 'mes/día/año hora', luego de la modificación se carga con el formato 'día/mes/año'. Esta función es propia del OWB.
- **TO_CHAR:** recibe como parámetro un valor numérico y lo devuelve convertido a cadena de texto. Esta función es propia del OWB.

Una vez que se transforman estos datos, se lleva a cabo el mapeo correspondiente entre estos y la tabla de hechos. Para realizar el proceso antes mencionado se utilizan los operadores de consulta de claves KEY_LOOKUP. Estos tienen como objetivo filtrar cada atributo de los *staging area* con sus correspondientes tablas de dimensiones y posteriormente, se pasan los resultados a la FACT_INCIDENCIA.

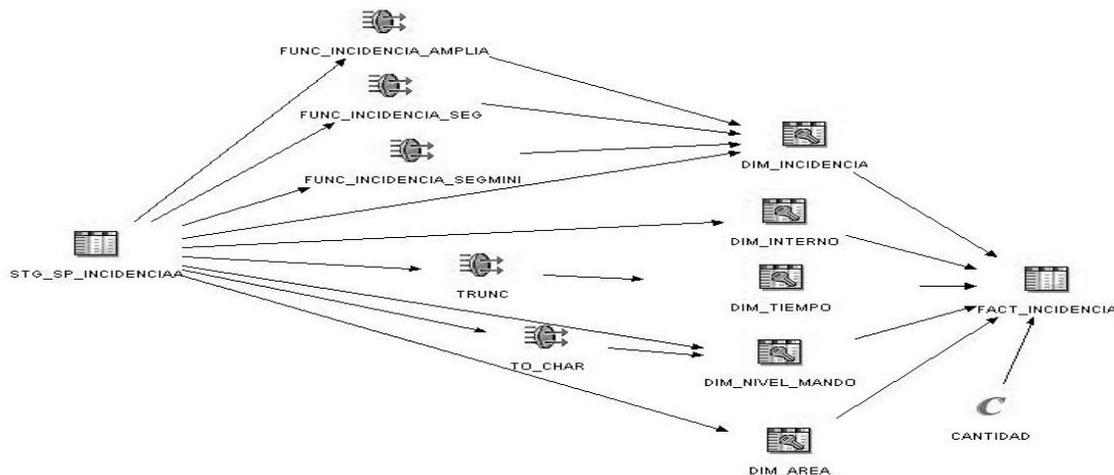


Figura 23: Correspondencia de la tabla de hecho Incidencia.

Al concluir el proceso de ETL de la tabla FACT_INCIDENCIA, queda cargada en la base de datos esta tabla de hecho con todos los identificadores de las dimensiones relacionadas a ella; lo que permite el análisis de los indicadores propuestos para el proceso Incidencia.

▪ ETL tabla de hecho INDISCIPLINA

Antes de efectuar el proceso de ETL de la tabla FACT_INDISCIPLINA se realiza el de la *staging area* diseñada para este proceso, la STG_SP_INDISCIPLINA. En esta tabla se almacenan de manera intermedia los datos que luego contendrá la tabla de hechos.



Capítulo 3: Implementación y Prueba

✓ ETL *staging area* INDISCIPLINA

Las transformaciones para cargar la STG_SP_INDISCIPLINA, se inician seleccionando las tablas correspondientes. Posteriormente se utiliza el JOINER para unir las tablas necesarias a través de la condición de unión. Una vez que se obtienen todos los datos, se mapean con los de la tabla STG_SP_INDISCIPLINA como indica la figura 24.

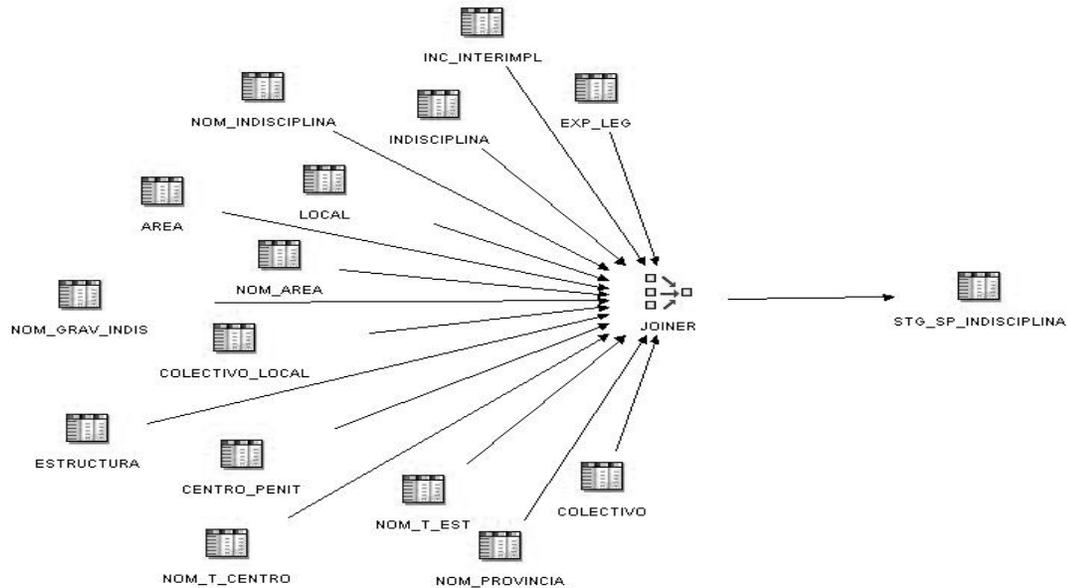


Figura 24: Correspondencia de la tabla *staging area* Indisciplina.

Al terminar la carga de datos de la tabla STG_SP_INDISCIPLINA, se procede a realizar el proceso de ETL de la tabla FACT_INDISCIPLINA; tal y como se muestra en la figura 25.

Para realizar la carga de datos de la tabla de hechos del proceso Indisciplina, primero se selecciona la tabla necesaria, la STG_SP_INDISCIPLINA. Seguidamente se hace necesario realizar transformaciones a algunos de los datos extraídos, para poder integrarlos correctamente al mercado de datos.

Una vez que se transforman los datos, se lleva a cabo el mapeo correspondiente entre estos y los restantes atributos de la *staging area* y la tabla de hechos. Para realizar el proceso antes mencionado se utilizan los operadores de consulta de claves KEY_LOOKUP. Estos tienen como objetivo filtrar cada atributo de los *staging area* con sus correspondientes tablas de dimensiones y posteriormente, se pasan los resultados a la FACT_INDISCIPLINA.



Capítulo 3: Implementación y Prueba

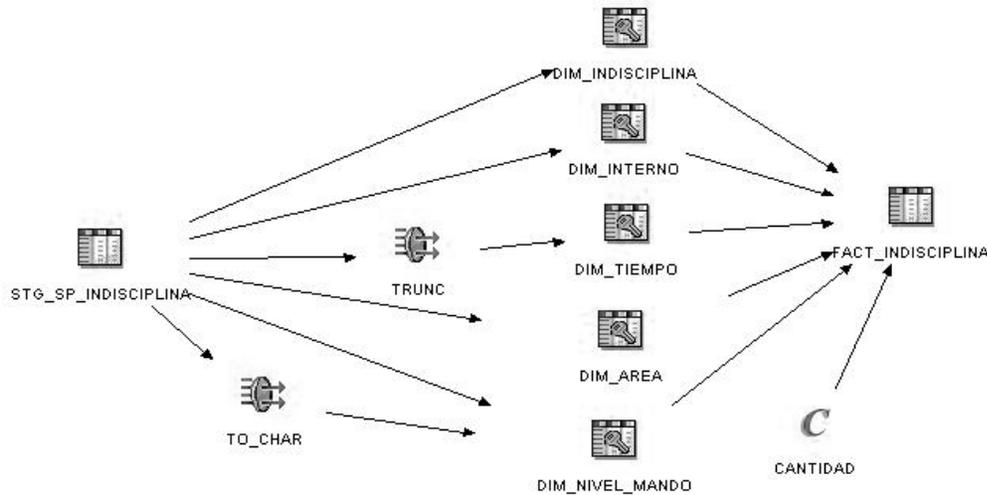


Figura 25: Correspondencia de la tabla de hecho Indisciplina.

Al concluir el proceso de ETL de la tabla `FACT_INDISCIPLINA`, queda cargada en la base de datos esta tabla de hecho con todos los identificadores de las dimensiones relacionadas a ella; lo que permite el análisis de los indicadores propuestos para el proceso Indisciplina.

3.2. Configuración de la herramienta de administración

Luego de realizado el proceso de ETL para cargar cada una de las dimensiones y las tablas de hechos, se procede a configurar el repositorio del mercado de datos para que pueda ser usado por las herramientas de consulta. En este repositorio es donde se definen las tablas de dimensiones y de hechos a utilizar y el mapeo entre estas. Aquí se crea también la navegabilidad de las dimensiones. Para todo lo antes mencionado es que se emplea el Administration Tool. Esta herramienta está compuesta por 3 capas; a continuación se describen cada una de ellas.

▪ Capa física

La capa física define las fuentes de datos a utilizar para cargar el mercado de datos. En esta capa puede haber una o más fuentes de datos, incluyendo bases de datos, hojas de cálculo y documentos XML. El método recomendado para llenar la capa física, es mediante la importación de la información; tanto de las bases de datos, como de otras fuentes. También brinda la posibilidad de importar esquemas o partes de estos de las fuentes de datos existentes. Además, se pueden crear objetos de forma manual, se definen las relaciones físicas de cada una de las tablas importadas o creadas y se especifican las llaves primarias y foráneas.



Capítulo 3: Implementación y Prueba

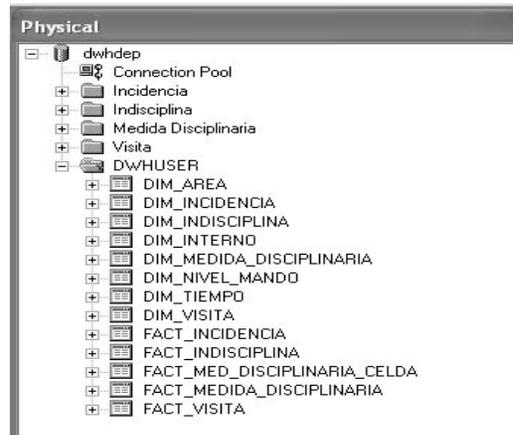


Figura 26: Capa física.

Después de construir la capa física, se procede a realizar el modelo de negocio en la capa de modelo de negocio y mapeo.

▪ Capa de modelo de negocio y mapeo

En la capa de modelo de negocio y mapeo es donde se define el modelo de negocio de los datos y se especifica el mapeo entre este y los esquemas de la capa física. Esta capa puede estar formada por uno o más modelos de negocio. Permite crear diferentes niveles de jerarquía, que luego serán utilizados por las herramientas de consulta, y relacionar las tablas que se tomaron de la capa física. El objetivo fundamental del modelo de negocio es plasmar lo que piensan los usuarios acerca de su negocio, utilizando su propio vocabulario.

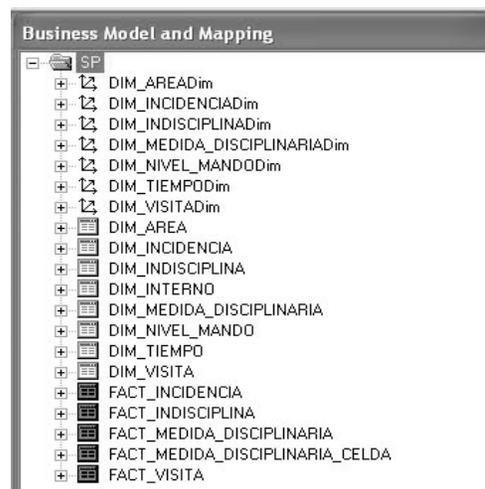


Figura 27: Capa de negocio.



Capítulo 3: Implementación y Prueba

Luego de realizada la capa física y la capa de modelo de negocio y mapeo, se realiza la capa de presentación.

▪ Capa de presentación

Esta capa es la vista de los datos que se le muestran al usuario final en las herramientas cliente y las aplicaciones, tal como Oracle BI Answers. Esta capa permite simplificar y modificar el modelo de negocio y la capa de mapeo para los usuarios finales.

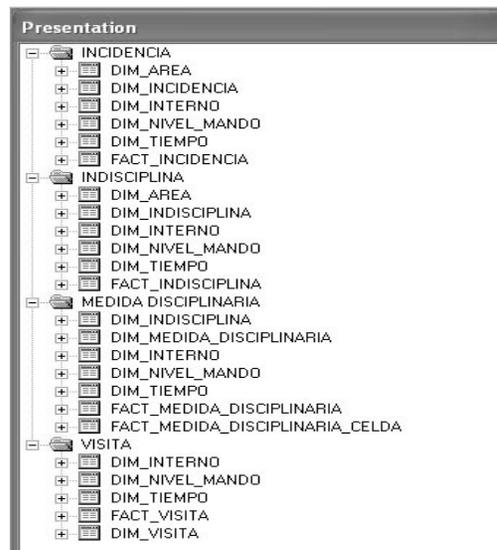


Figura 28: Capa de presentación.

Después de confeccionadas las tres capas mencionadas anteriormente, se procede a la creación de la aplicación de inteligencia de negocio.

3.3. Especificación y desarrollo de aplicaciones de inteligencia de negocio

Una parte importante de todo proyecto de inteligencia de negocio, es proporcionar al usuario final una forma fácil y rápida de acceder a la información del mercado de datos. Este acceso se realiza a través de lo que llamamos aplicaciones de inteligencia de negocio.

Las aplicaciones de inteligencia de negocio presentan diferentes tipos de informes y herramientas de análisis, de utilidad para los usuarios. Estas herramientas están clasificadas en diferentes categorías. Para el desarrollo del mercado de datos se usan dos de ellas, estas son: herramienta para consultas ad-hoc y herramienta para el usuario final.



Capítulo 3: Implementación y Prueba

▪ Herramienta para consultas ad-hoc (Oracle BI Answers)

La herramienta para consultas ad-hoc le brinda al usuario del área de Seguridad Penitenciaria, de acuerdo a sus permisos, la posibilidad de crear sus propios reportes, así como modificar los ya existentes según sus necesidades. Los reportes pueden ser filtrados por diferentes atributos y se pueden crear diferentes gráficas para mostrar los resultados, lo que facilita el análisis de los mismos. A continuación se muestra la interfaz de la herramienta.



Figura 29: Ambiente de trabajo para modificar o crear reportes (Pestaña “Criterios”).

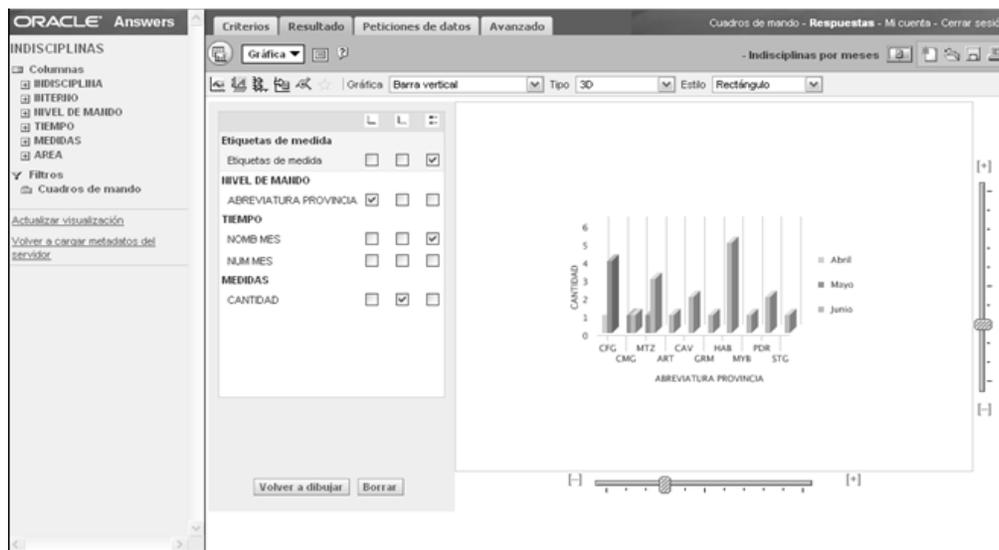


Figura 30: Ambiente de trabajo para modificar o crear reportes (Pestaña “Resultados”).



Capítulo 3: Implementación y Prueba

▪ Herramienta para el usuario final (Oracle BI Interactive Dashboards)

La herramienta para el usuario final, o tablero de mando, les permite a los usuarios del área de Seguridad Penitenciaria visualizar información acerca de lo que sucedió o está sucediendo en su área. Esta puede ser utilizada por los usuarios siguientes: Oficial de Dirección, Primer Especialista y Especialista.

Se crearon dos tableros de mando, uno para los reportes históricos y otro para los operativos. Los reportes operativos son aquellos que muestran los resultados actuales y por tanto, son informes que los usuarios usan día a día. Estos son los que permiten trabajar de manera proactiva. Los reportes históricos pueden mostrar toda la información almacenada o solo aquella definida por un rango de fechas. Permiten establecer patrones de comportamiento y tendencias. A continuación se muestran ejemplos de estos tipos de reportes.

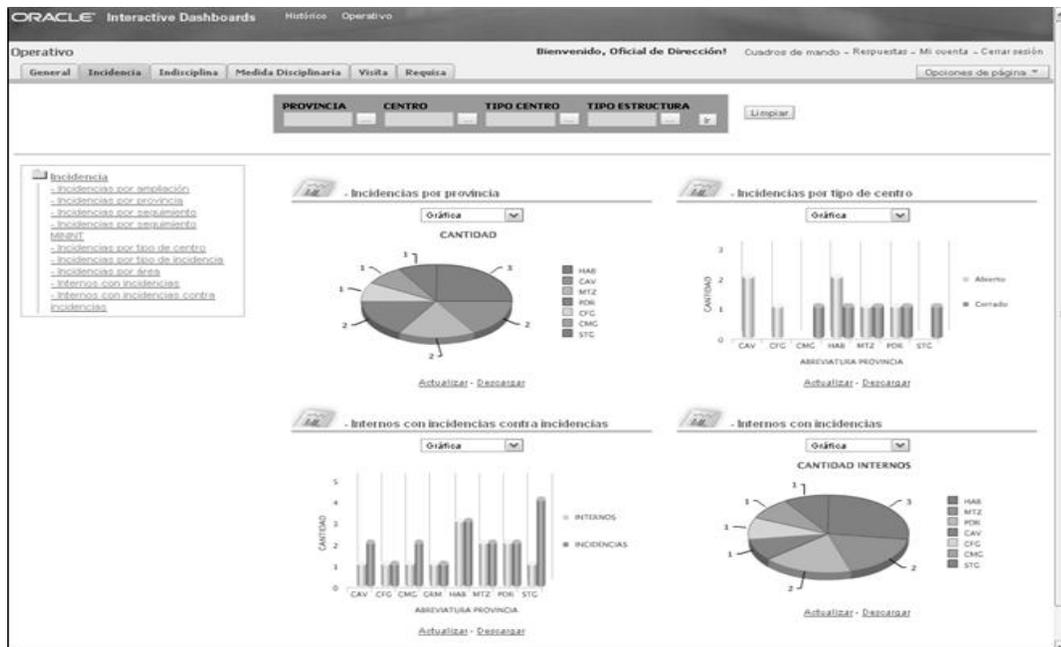


Figura 31: Reportes operativos del proceso de Incidencia.



Capítulo 3: Implementación y Prueba



Figura 32: Reportes históricos del proceso Incidencia.



Figura 33: Reportes históricos del proceso Incidencia (filtro de fecha aplicado).



Capítulo 3: Implementación y Prueba

3.4. Tipos de procesamiento utilizados

Una de las particularidades a tener en cuenta para el acceso a la información en los mercados de datos, es la posibilidad de representarla en diferentes niveles de agregación. Por esta razón se usan los tipos de procesamiento.

Para el desarrollo del mercado de datos se utilizaron estos: *Drill Down* y *Slice and Dice*. El *Drill Down* posibilita la navegación a través de distintos niveles de detalles de una jerarquía de datos, desde los de mayor nivel de agregación hasta los más detallados. Un ejemplo de esto se puede ver en la jerarquía de la dimensión Nivel de Mando. En esta, la provincia es el mayor nivel jerárquico, luego el centro y por último el colectivo, que es el más detallado.

El otro tipo de procesamiento es empleado para describir como los datos multidimensionales pueden ser mostrados aplicando diferentes filtros a los cubos. *Slice* conforma un conjunto de los datos de un cubo aplicándole una única restricción a una sola dimensión, mediante la elección de un atributo en particular. *Dice* es la combinación de varios *Slice*; o sea, se aplica más de una restricción y puede ser para una misma dimensión o para varias dimensiones.

3.5. Prueba

Una vez concluido con el diseño e implementación del Mercado de Datos para el Área de Seguridad Penitenciaria, se procede a la realización de las pruebas. Las pruebas se realizan con el objetivo de verificar que el sistema cumple con todas las funcionalidades y para encontrar la mayor cantidad de errores posibles. Además, se emplean para determinar la calidad e identificar fallas, faltas y errores en la implementación del sistema.

3.5.1. Descripción de los métodos de prueba a realizar

Existen diferentes métodos de pruebas a ser empleados para comprobar la calidad de los sistemas. Se tienen dos vertientes fundamentales: las pruebas de caja blanca y las pruebas de caja negra.

De los métodos mencionados anteriormente, se hizo uso del método de “*pruebas de caja negra, también denominadas, pruebas de comportamiento.*” (11) La prueba de caja negra es “*un enfoque complementario, que tiene probabilidades de descubrir una clase diferente de errores de los que se descubrirían con los métodos de caja blanca.*” (11) Esta se realiza sobre la interfaz del sistema y no es necesario conocer la lógica del programa, solamente las funcionalidades que debe realizar.



Capítulo 3: Implementación y Prueba



Figura 34: Prueba de caja negra.

Para darle cumplimiento al método de prueba especificado anteriormente, se hizo uso de la lista de chequeo, con el objetivo de comprobar el correcto funcionamiento del mercado de datos.

- **Lista de chequeo**

Las pruebas basadas en lista de chequeo (*checklist*, como se conoce internacionalmente en inglés) realizan una serie de preguntas en forma de cuestionario, con el objetivo de comprobar el grado de cumplimiento de los requisitos. Además, permiten identificar problemas existentes en la aplicación. La lista de chequeo empleada la proporcionó el Centro de Tecnologías de Gestión de Datos (DATEC).

La metodología de Kimball propone que primeramente las pruebas sean realizadas por el equipo de desarrollo y posteriormente, se debe realizar una segunda iteración con personas que no posean conocimientos del negocio.

Conclusiones Parciales

En este capítulo se realizó la ETL de los procesos del área de Seguridad Penitenciaria, lo que permitió cargar en el mercado de datos la información que luego será usada por las herramientas de consulta y análisis. También se creó y configuró el repositorio con la herramienta de administración de acuerdo a los requisitos del cliente, permitiendo el enlace entre el mercado de datos y las aplicaciones de inteligencia de negocio.

Luego de configurado el repositorio de datos, se trabajó con las aplicaciones de inteligencia de negocio. En estas se crearon los reportes y se configuraron los tableros de mando para el mercado de datos del área de Seguridad Penitenciaria y será mediante estos, que los funcionarios de la DEP harán el proceso



Capítulo 3: Implementación y Prueba

de toma de decisiones. Finalmente, se realizaron dos iteraciones de pruebas al mercado de datos, mediante el método de caja negra.



CONCLUSIONES GENERALES

Como resultado del presente trabajo se desarrolló el mercado de datos para la toma de decisiones en el área de Seguridad Penitenciaria del Sistema Penitenciario Cubano. Para ello se realizó un estudio del marco teórico en el que está sustentada la investigación, haciendo énfasis en las principales características y funcionalidades que debe poseer un mercado de datos.

Se realizó el análisis de requisitos del mercado de datos para el área de Seguridad Penitenciaria, lo que hizo posible la identificación de los procesos, los temas analíticos y los requisitos de información. Para darle solución a los requisitos de información se realizó el modelado dimensional, en el cual se definieron las dimensiones, medidas y tablas de hechos que conformaron el mercado de datos.

La realización del modelado dimensional fue la base para el diseño e implementación del proceso de ETL de los datos, el cual permite cargar en el mercado de datos la información para la toma de decisiones. Se implementaron los reportes que dan respuesta a las necesidades de información de la DEP y se configuró el tablero de mando que facilita el análisis y visualización de los resultados. Para comprobar la calidad del trabajo se aplicaron pruebas de caja negra, lo que permitió la identificación de las no conformidades, las cuales fueron resueltas en su totalidad.



REFERENCIA BIBLIOGRÁFICA

1. **Bernabeu, Ricardo Diario.** *Hefesto: Metodología para la Construcción de un Data Warehouse.* Córdoba, Argentina : s.n., 2010.
2. **del Risco Batista, Yanet.** *Proyecto Técnico.* La Habana : s.n., 2009.
3. Portal Iberoamericano de la Justicia Electrónica. [En línea] 2010.
<http://www.piaje.org/PT/Services/Paginas/statesservices.aspx?Country=Ecuador>.
4. Dirección General del Sistema Penitenciario. *República de Panamá.* [En línea] 10 de noviembre de 2007.
<http://www.sistemapenitenciario.gob.pa/>.
5. **Martínez Bravet , René y Lazo Pedraja, David .** “*Repositorio de Datos de la Sala Situacional del SIGEP, desde el punto de vista del Control Penal.*”. La Habana : s.n., 2010.
6. **García Izquierdo, Malena .** *Diseño del Repositorio de Datos de la Sala Situacional del SIGEP desde el punto de vista de Clasificación y Atención Integral.* La Habana : s.n., 2011.
7. **Inmon, Willian H.** *Building the Data Warehouse.* Canada : Wiley Computer Publishing, 2002.
8. **Lumpkin, George .** *Oracle Database 11g para Data Warehousing y Business Intelligence.* Estados Unidos : s.n., 2008.
9. Embarcadero. [En línea] [Citado el: 13 de febrero de 2012.] www.embarcadero.com .
10. **Durán Toro, A., y otros, y otros.** *Identificación de Patrones de Reutilización de Requisitos de Sistemas de Información.* Sevilla, España : s.n.
11. **Pressman, Roger S.** *La Ingeniería del Software. Un enfoque práctico.* s.l. : Mc Graw Hill, 2005.



GLOSARIO DE TÉRMINOS

- ✓ **Cubo:** Representación o visualización de una tabla de hecho y sus dimensiones correspondientes.
- ✓ **DEP:** La Dirección de Establecimientos Penitenciarios, es la encargada de la ejecución y el logro de los objetivos implícitos en las sanciones y medidas privativas de libertad.
- ✓ **Dimensiones:** Son elementos para ubicar datos que participan en el análisis, son datos que deben de tenerse en cuenta para poder tomar decisiones basadas en ellos.
- ✓ **ETL:** La extracción, transformación y carga, es el proceso que permite a las organizaciones mover datos de múltiples fuentes, limpiarlos y cargarlos en otras bases de datos para su posterior análisis.
- ✓ **Inteligencia de negocio:** Es una arquitectura y colección de herramientas que buscan mejorar a las organizaciones, proporcionando vistas de aspectos de negocio a todos los empleados (estratégico, táctico, operacional) para que tomen mejores y más relevantes decisiones en menos tiempo y con la mayor información posible.
- ✓ **Mercado de datos:** Un mercado de datos es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica.
- ✓ **MININT:** Ministerio del Interior.
- ✓ **Nivel de mando:** El nivel de mando es la de jerarquía que presentan los centros penitenciarios, como son: provincia, centro y colectivo.
- ✓ **Seguridad Penitenciaria:** La Seguridad Penitenciaria es el conjunto de medidas y acciones sistematizadas y relacionadas entre sí, con el propósito de asegurar el control físico, la vigilancia permanente y el orden disciplinario de los internos, así como la seguridad y protección de los lugares de internamiento y de todo el personal que en él permanece y lo visitan.
- ✓ **Sistema Penitenciario Cubano:** Garantiza el funcionamiento de los centros destinados al internamiento y el tratamiento a los internos.
- ✓ **Tabla de hechos:** Es una representación de un proceso de negocio. Esta tabla se encuentra rodeada de tablas de dimensiones que determinan cómo están distribuidos y relacionados los datos lógicamente en el ambiente del negocio.