



Universidad de las Ciencias Informáticas

Título: Diseño e Implementación de un Mercado de Datos para el área de Tratamiento Educativo.

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

Autor(es): Yosley Rodríguez García
Yisenia Pérez Quintero

Tutor: Ing. Yanet del Risco Batista
Co-Tutor: Ing. Krysia Llull Céspedes

Ciudad de La Habana

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.

Yosley Rodríguez García

Firma del Autor

Yisenia Pérez Quintero

Firma del Autor

Ing. Yanet del Risco Batista

Firma del Tutor

Ing. Krysia Llull Céspedes

Firma del Co-Tutor

Agradecimientos:

Dedicatoria:

Resumen:

La presente investigación surgió como parte de la colaboración que existe entre la Universidad de las Ciencias Informáticas y la Dirección de Establecimientos Penitenciarios. El área de Tratamiento Educativo de la Dirección de Establecimientos Penitenciarios tiene la finalidad de presentar un conjunto de indicadores y cifras estadísticas que reflejan el comportamiento de los internos en relación con su reinserción a la sociedad. Muchos de los análisis deseados sobre esta área no pueden ser realizados, pues no cuentan con el soporte para llevarlos a cabo. En este trabajo se propuso una solución basada en el desarrollo de un mercado de datos que contribuyó al almacenamiento homogéneo de los datos, posibilitando un adecuado análisis de la información y ayudando en el proceso de toma de decisiones de los directivos del área de Tratamiento Educativo. Para ello se realizó el análisis, diseño, implementación y prueba de la aplicación que constituyó el resultado principal de este trabajo de diploma. Conjuntamente se hizo uso de varias herramientas y de una metodología de desarrollo para guiar el proceso de implementación. Se obtuvo un mercado de datos poblado y funcional, con una capa de inteligencia de negocio que brinda vistas de análisis actualizadas, permitiéndole a los especialistas del área de Tratamiento Educativo un mejor análisis de la información.

ÍNDICE

Introducción	1
Capítulo 1: Fundamentación teórica	5
Aplicaciones Informáticas.....	5
Almacén de Datos	7
Mercado de Datos	9
Base de Datos Multidimensional	9
Modo de almacenamiento de Datos.....	11
Arquitectura conceptual de los datos	12
Modelado de Datos	13
Metodologías estudiadas para el desarrollo.....	14
Herramientas de creación de un Almacén de Datos	17
Herramientas de modelado	18
Conclusiones Parciales	18
Capítulo 2: Análisis y diseño.....	20
Definición de los requisitos del negocio	20
Requisitos de información	20
Técnicas de identificación de requisitos	23
Temas analíticos	24
Validación de requisitos	28
Especificación de requerimientos:.....	25
Modelo dimensional	28
Granularidad.....	29
Matriz dimensional	30

Patrones de diseño para Almacenes de Datos	31
Modelado de datos	33
Diseño de la ETL	35
Capítulo 3: Implementación y pruebas	42
Implementación	42
Implementación de las ETL	42
Configuración de la herramienta de Administración	47
Herramientas de inteligencia de negocio	49
Tipos de procesamiento de la información	51
Pruebas	53
Lista de chequeo	53
Resultados de las pruebas	55
Conclusiones Generales	57
Recomendaciones	58
Bibliografía	59
Referencias bibliográficas	60
Anexos	61

Introducción

«Una de las características de la sociedad contemporánea es el papel central del conocimiento en los procesos productivos. Asistimos a la emergencia de un nuevo paradigma económico-productivo en el cual el factor más importante no es ya la disponibilidad de capital, mano de obra, materias primas o energía, sino el uso intensivo del conocimiento y la información.» Tünnermann y Chauí (2003). **(1)**

Actualmente las grandes organizaciones y negocios cuentan con sistemas operacionales, que como parte de los sistemas de información, interactúan entre sí con el fin de apoyar las actividades que se realizan en la empresa. A través del uso de estos sistemas y de las Tecnologías de la Información y las Comunicaciones (TIC), se han logrado importantes mejoras, como la automatización de los procesos operativos y el contar con una plataforma de información necesaria para apoyar la toma de decisiones.

El almacenamiento e interpretación de un gran cúmulo de información son dos funciones importantes en el ámbito empresarial por lo que se han desarrollado disímiles soluciones para operar con este valioso recurso que es la información, ejemplo de estos son los Almacenes de Datos, que permiten extraer conocimientos a partir de la información histórica almacenada de la organización.

En Cuba en el año 2002, tras la Orden 43/99 del Viceministro Primero del Interior de automatizar los procesos principales que se realizan en los centros penitenciarios, se desarrolló el Sistema Automatizado para el Control del Recluso (SACORE). Luego de implantado, se verificó que este sistema no logró abarcar todos los procesos que se ejecutaban en las prisiones, por lo que fue necesario desarrollar otras dos aplicaciones, surgiendo así el Sistema Automatizado de Incidencias de la Dirección de Establecimientos Penitenciarios (SAIDEP) y Sistema Automatizado de Capacidades de la Dirección de Establecimientos Penitenciarios (SACDEP).

A partir del Plan 20 x 50 y en colaboración con la Universidad de las Ciencias Informáticas, nace el proyecto Prisiones Cuba, el cual tiene como objetivo fundamental unir los tres sistemas que existen en la Dirección de Establecimientos Penitenciarios y adicionar un conjunto de funcionalidades necesarias para el funcionamiento de los procesos en los centros penitenciarios, aprovechando el desarrollo de las tecnologías de la informática y las comunicaciones que se está llevando a cabo en el país.

El área de Tratamiento Educativo del Sistema Penitenciario Cubano es la encargada de tomar las decisiones sobre la reinserción de los internos. La misma abarca distintos procesos como son: Trabajo, Actividades Educativas, Administración de la Biblioteca, Permisos, Emulación y Beneficios.

El departamento de Dirección es el encargado de realizar el análisis de la información de todas las áreas y generar los reportes sobre el comportamiento de los indicadores en un determinado período. Actualmente

el departamento está compuesto por un grupo reducido de especialistas que son los encargados de generar los reportes y brindar información a la jefatura para la toma de decisiones.

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

A los directivos de esta área le resulta muy complicado consultar cualquier información que necesite ser analizada por encontrarse la misma descentralizada.

No se cuenta con un mecanismo capaz de darle seguimiento a los objetivos estratégicos para apoyar la toma de decisiones a todos los niveles y que permita llevar el control efectivo de todas las actividades y procesos que se desarrollan en el área de Tratamiento Educativo.

Dada la situación anterior surge el siguiente **problema a resolver**: ¿Cómo contribuir a la toma de decisiones en el área de Tratamiento Educativo del Sistema Penitenciario Cubano?

Se define como **objeto de estudio** los procesos para la toma de decisiones en los sistemas penitenciarios Cubano, se plantea como **objetivo general**: Desarrollar un Mercado de Datos para apoyar el proceso de toma de decisiones sobre el área de Tratamiento Educativo del Sistema Penitenciario Cubano, enfocándose como **campo de acción** los procesos para la toma de decisiones en el área de Tratamiento Educativo del Sistema Penitenciario Cubano.

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

1. Analizar los requisitos del mercado de datos para el área de Tratamiento Educativo.
2. Realizar el modelado dimensional del mercado de datos para el área de Tratamiento Educativo.
3. Implementar los procesos de extracción, transformación y carga de los datos del mercado de datos para el área de Tratamiento Educativo.
4. Implementar el mercado de datos para el área de Tratamiento Educativo.

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

Objetivo 1

- ✓ Elaborar el marco teórico de la investigación.
- ✓ Identificación de temas analíticos del mercado de datos para el área de Tratamiento Educativo.
- ✓ Identificación de procesos de negocio del mercado de datos para el área de Tratamiento Educativo.
- ✓ Identificación de los requisitos del mercado de datos para el área de Tratamiento Educativo.

- ✓ Priorización de requisitos del mercado de datos para el área de Tratamiento Educativo.

Objetivo 2

- ✓ Identificación de los niveles de granularidad del mercado de datos para el área de Tratamiento Educativo.
- ✓ Identificación de las dimensiones, tablas de hechos y medidas del mercado de datos para el área de Tratamiento Educativo.
- ✓ Identificación de atributos de las dimensiones y tablas de hechos del mercado de datos para el área de Tratamiento Educativo.
- ✓ Construcción de la matriz Procesos-Dimensiones.
- ✓ Diseño del modelo dimensional detallado del mercado de datos para el área de Tratamiento Educativo.
- ✓ Revisión y validación del modelo dimensional del mercado de datos para el área de Tratamiento Educativo.

Objetivo 3

- ✓ Elaboración de las correspondencias con las fuentes de origen.
- ✓ Diseño de los procesos de extracción, transformación y carga del mercado de datos para el área de Tratamiento Educativo.
- ✓ Implementación de los procesos de extracción, transformación y carga para cargar las dimensiones y tablas de hechos del mercado de datos para el área de Tratamiento Educativo.

Objetivo 4

- ✓ Configuración de la herramienta de administración del mercado de datos para el área de Tratamiento Educativo.
- ✓ Implementación de los reportes del mercado de datos para el área de Tratamiento Educativo.
- ✓ Desarrollo del tablero de mando del mercado de datos para el área de Tratamiento Educativo.
- ✓ Realizar las pruebas al mercado de datos para el área de Tratamiento Educativo.

Posibles resultados:

- ✓ Diseño e implementación del mercado de datos de Tratamiento Educativo.
- ✓ Implementación de los Reportes de Tratamiento Educativo.
- ✓ Tablero de Mando de Tratamiento Educativo.

El presente documento consta de la siguiente distribución:

Capítulo 1: Marco teórico de la investigación

En el capítulo se realiza el marco teórico conceptual relacionado con el desarrollo de los almacenes de datos, se describen los conceptos fundamentales que serán tratados a lo largo de la investigación, así como la descripción de la metodología, técnicas y herramientas que serán utilizadas.

Capítulo 2: Análisis y diseño

En este capítulo se hará un estudio preliminar del negocio con el objetivo de definir necesidades de información, identificación de dimensiones, hechos y medidas, desarrollo de la matriz dimensional, modelo de datos. Se realizará el análisis y diseño de los datos del área de Tratamiento Educativo de la Dirección de Establecimientos Penitenciarios.

Capítulo 3: Implementación y prueba

En este capítulo se procederá a la implementación de la solución propuesta, se implementará el modelo de datos, se diseñarán los reportes del área de Tratamiento Educativo. Se aplicarán las pruebas apoyándose en las listas de chequeo con el objetivo de validar la solución propuesta.

Capítulo 1: Fundamentación teórica

Introducción

En este capítulo se hace un análisis relacionado con las aplicaciones informáticas que usan algunos sistemas penitenciarios para la ayuda de toma de decisiones. Además se abordan los conceptos, objetivos y características de los almacenes de datos, se realiza una breve descripción de los procesos analíticos del área de Tratamiento Educativo y de las herramientas a utilizar para desarrollar el mercado de datos.

Aplicaciones Informáticas:

A través de los grandes avances en las tecnologías diversos sistemas penitenciarios se han inclinado por la utilización de aplicaciones informáticas para el procesamiento de la información, generando grandes volúmenes de datos que son almacenados en bases de datos para su posterior análisis.

A continuación se mencionan algunas aplicaciones informáticas existentes en los sistemas penitenciarios, las cuales fueron valoradas para realizar este trabajo, teniendo presente las funcionalidades que brindan y de que manera realizan el análisis de la información estadística.

Sistema de Gestión Penitenciaria de Venezuela (SIGEP):

El Sistema de Gestión Penitenciaria (SIGEP), desplegado en Venezuela y desarrollado en la Universidad de Ciencias Informáticas, se creó con el objetivo de contar con un sistema informático capaz de soportar las decisiones estratégicas de la Dirección Nacional de Servicios Penitenciarios de Venezuela.

En este sistema como parte de la solución se implementó un mercado de datos sobre el Control Penal de la Sala Situacional del SIGEP para facilitar el proceso de toma de decisiones de los directivos del sistema de justicia venezolano. El mismo almacena la información referente a los individuos, sus expedientes, los centros penitenciarios, las decisiones judiciales, los procesos legales, las fases de los procesos legales, las ejecuciones de los procesos legales, los delitos relacionados con los procesos legales y las penas impuestas, 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. Este mercado de datos nunca llegó a ponerse en explotación.

Sistema de Gestión Penitenciaria de Ecuador (e-Sigpen):

Este sistema facilita el control y manejo de las actividades que realizan los centros de rehabilitación social en cada una de sus áreas, logrando de esta manera tener un control adecuado y oportuno de la información de las personas privadas de libertad. El Ministerio de Justicia, Derechos Humanos y Cultos conjuntamente con la Dirección Nacional de Rehabilitación Social están en continuo monitoreo de las actividades realizadas en los centros de rehabilitación social por medio del sistema e-Sigpen, el cual permite obtener datos estadísticos reales de la situación en que se encuentran los internos, logrando de esta manera tomar decisiones eficaces y oportunas para mejorar el ambiente carcelario del país.

Sistema Automatizado para el Control del Recluso (SACORE):

Con el objetivo de reorganizar y facilitar el trabajo dentro de las Instituciones Penales del país, garantizando la gestión y control de los datos y asegurando las leyes vigentes para la Dirección de Establecimientos Penitenciarios se crea el Sistema Automatizado para el Control del Recluso (SACORE).

Este sistema presenta reportes predefinidos lo que trae consigo una traba para la correcta toma de decisiones por parte de la administración, pues siempre que surja la necesidad de hacer un nuevo reporte los informáticos son los únicos capaces de darle solución a este problema. Además de que utilizan bases de datos operacionales, las cuales no están preparadas para realizar análisis de tendencias ya que los datos se actualizan constantemente y pueden ser eliminados.

A pesar de que este sistema resolvió una gran problemática en el momento de su creación, en la actualidad no resuelve todas las necesidades de información necesarias para el Departamento de Dirección de la DEP.

Luego de valorados los sistemas se tiene como conclusión que:

Las aplicaciones informáticas no tienen un mercado de datos que analiza el tratamiento educacional que se fomenta dentro de las instituciones penales a nivel mundial, solo están sujetas a la situación legislativa y judicial imperante en un país, lo cual las convierte en sistemas no portables para otros sistemas penitenciarios. Además estos sistemas tienen como peculiaridad que el almacenamiento de sus datos está sostenido en base de datos operacionales, las cuales no están preparadas para realizar análisis de tendencias. Los datos se actualizan constantemente y pueden ser eliminados. Por lo que utilizar un almacén de datos para guardar toda esta información histórica, propiciaría el apoyo a la buena toma de decisiones.

Almacén de Datos:

Un Almacén de Datos es una colección de datos orientada a un determinado ámbito, integrado, no volátil y variable en el tiempo, que ayuda al proceso de toma de decisiones en la entidad en la que se utiliza. **(2)**

Entre sus principales características tenemos (Bill Inmon): (3)

Orientado a temas: En el almacén de datos, la información se clasifica en base a los aspectos que son de interés para la organización, sin embargo en el ambiente operacional la estructura de los datos se diseña en torno a las aplicaciones. Las diferencias entre las funciones de las aplicaciones y la orientación a temas, radican en el contenido de los datos a nivel detallado. En el almacén de datos se excluye la información que no será usada por el proceso de sistemas de soporte de decisiones, mientras que la información de las orientadas a las aplicaciones, contiene datos para satisfacer de inmediato los requerimientos funcionales, que pueden ser usados o no por el analista de soporte de decisiones.

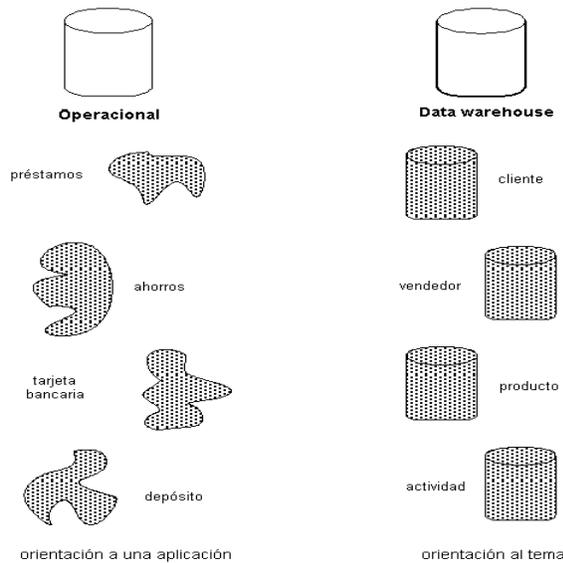


Figura 1: Orientado a temas.

Integrado: A través de los años los diseñadores y programadores de las diferentes aplicaciones han tomado sus propias decisiones sobre cómo se debería construir una aplicación. Por lo cual, cada uno por su parte ha dejado en cada aplicación, módulo y tabla, su propio estilo personalizado, confluyendo de esta manera en la creación de modelos muy inconsistentes e incompatibles entre sí. La integración implica que los datos de diversas fuentes que son producidos por distintos departamentos, secciones y aplicaciones, tanto

internas como externas se consoliden en una instancia antes de ser agregados al almacén de datos, y deben por lo tanto ser analizados para asegurar su calidad y limpieza.

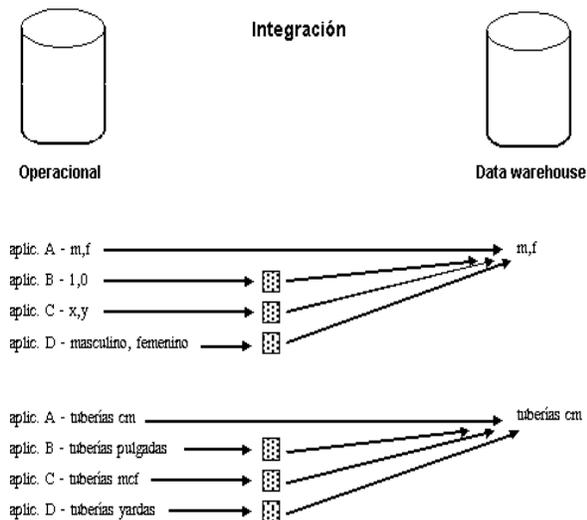


Figura 2: Integrado

Variante en el tiempo: En el almacén de datos toda la información es requerida en algún momento, esta información puede ser consultada con el objetivo de ser utilizada en comparaciones, tendencias y previsiones. Esta característica básica de los datos en un depósito, es muy diferente de la información encontrada en el ambiente operacional. En éstos, la información se requiere al momento de acceder. En otras palabras, en el ambiente operacional, cuando usted acceda a una unidad de información, usted espera que los valores requeridos se obtengan a partir del momento de acceso.

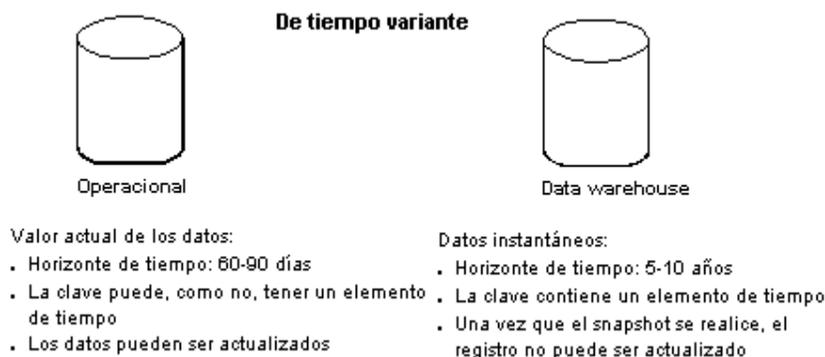


Figura 3: Variante en el tiempo

No volátil: La información de un almacén de datos existe para ser leída, pero no modificada. La información es por tanto permanente, significando la actualización del almacén de datos, la incorporación de los últimos valores que tomaron las distintas variables contenidas en él, sin ningún tipo de acción sobre lo que ya existía.

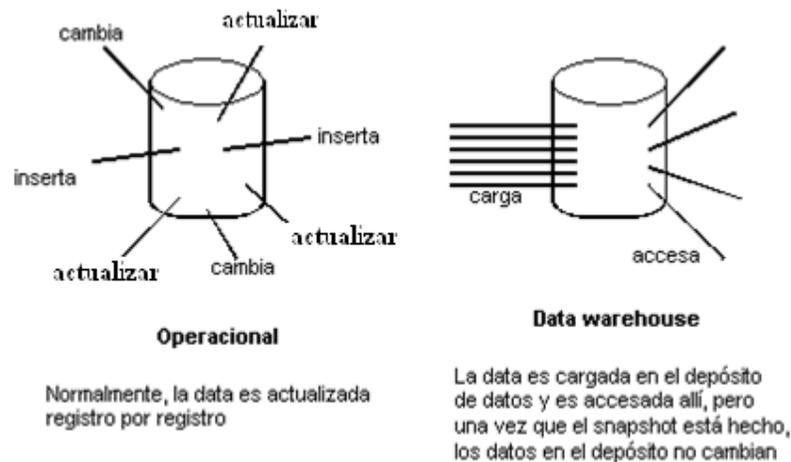


Figura 4: No volátil

Objetivos y Características de un Almacén de Datos:(4)

Los objetivos más importantes de un Almacén de Datos son:

- Proveer una única visión de los clientes a través de toda la compañía.
- Proveer la mayor cantidad de información a la mayor cantidad de personas posibles dentro de la organización.
- Mejorar el tiempo de emisión de informes.

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 la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. Un Mercado de Datos puede ser alimentado desde los datos de un almacén de datos, o integrar por sí mismo un compendio de distintas fuentes de información. **(5)**

Base de Datos Multidimensional

A diferencia de una base de datos que tiene una estructura bidimensional, donde cada columna proporciona un dato para un elemento de cada fila, en la construcción de los almacenes de datos se utiliza el modelado dimensional propuesto por Kimball en su libro "El Juego de Herramientas del Almacén de

Datos”(6). Este modelo garantiza una búsqueda rápida de los datos y la información que va a almacenarse a través de las tablas de hechos y las tablas de dimensiones.

Las tablas de hechos son las que contienen los hechos, indicadores o medidas a analizar por los analistas del negocio para apoyar la toma de decisiones.

Las tablas de dimensiones son las que permiten manipular la información contenida en las tablas de hechos definiendo como están organizados lógicamente los datos.

Existen tres variantes posibles de modelado de las bases de datos multidimensionales:

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

Esquema en estrella

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

Entre sus ventajas más significativas está que es el esquema más simple de interpretar, posee los mejores tiempos de respuesta, su diseño es fácil de modificar y simplifica el análisis. Sin embargo, es el menos robusto para la carga y es el más lento de construir. (7)

Esquema copo de nieve

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

Tiene como desventaja que de existir muchas tablas de dimensiones, cada una de ellas con varias jerarquías, pueden crearse bastantes tablas llegando a ser inmanejables. Además su desempeño puede verse reducido si existen muchas uniones y relaciones entre tablas. (7)

Esquema constelación

Este modelo está compuesto por una serie de esquemas en estrella y lo conforman una tabla de hechos principal y una o más tablas de hechos auxiliares. Estas tablas yacen en el centro del modelo y están relacionadas con sus respectivas tablas de dimensiones, vinculándose las tablas de hechos auxiliares con

algunas dimensiones asignadas a la tabla de hechos principal y también con nuevas tablas de dimensiones.

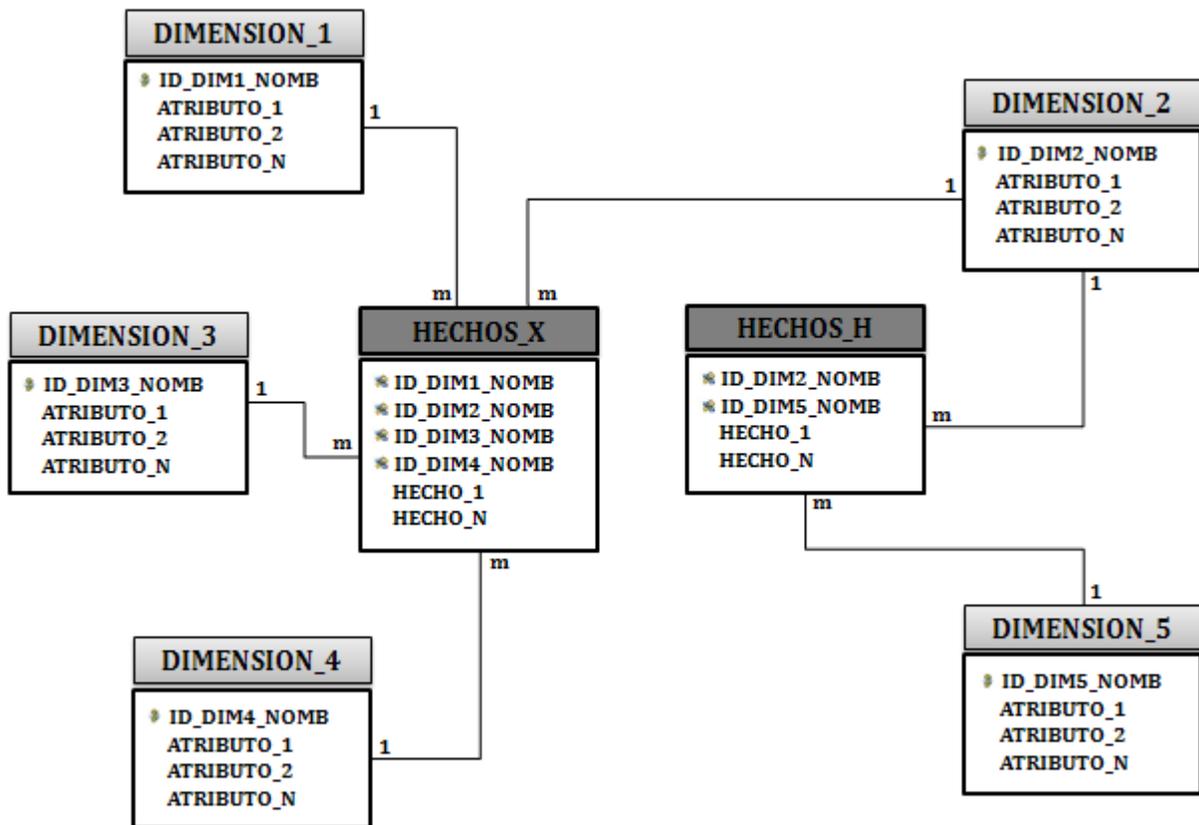


Figura 5: Ejemplo de un esquema constelación

Este esquema se destaca porque al tener más de una tabla de hechos facilita el análisis de más aspectos importantes del negocio y contribuye a la reutilización de las tablas de dimensiones, ya que pueden utilizarse para varias tablas de hechos. (7)

Se utilizará el modelado de datos dimensional donde los procesos del negocio estarán representados en sus respectivos esquemas en estrella, para luego unirlos formando el esquema constelación del mercado de datos del área de Tratamiento Educativo.

Modo de almacenamiento de Datos

ROLAP (Relational Online Analytical Processing en español Procesamiento Analítico en Línea Relacional)

Este modelo presenta los datos a los usuarios en forma de dimensiones de negocio. Con el fin de ocultar las estructuras de almacenamiento y presentar los datos dimensionalmente es creada la semántica de las etiquetas de los metadatos.

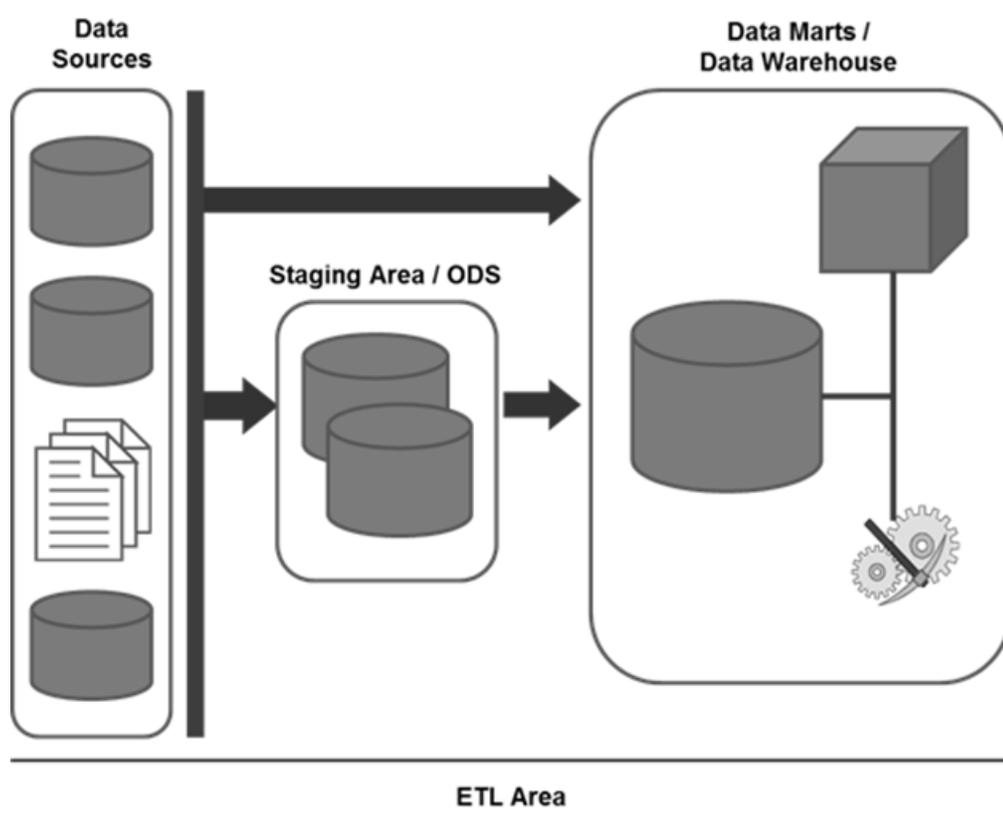
El modelo ROLAP es usado fundamentalmente sobre información que no se consulta frecuentemente debido a que no es muy óptimo en este sentido. Por ejemplo, información histórica de muchos años de antigüedad.

Arquitectura conceptual de los datos

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

En un almacén de datos suelen representarse varias capas a través de las cuales circulan los datos, de modo que estos se obtienen a partir de los de la capa previa, nombrándose de acuerdo al número de capas que abarcan. Para realizar el diseño del almacén de datos el proyecto definió usar la arquitectura de tres capas.

Arquitectura de datos de Tres Capas: La primera capa que presenta esta arquitectura es la relacionada con el origen de los datos necesarios para poblar las tablas dimensión y las tablas de hechos. En esta arquitectura para llevar a cabo la transformación de los datos de tiempo real a datos derivados, se requiere de una capa intermedia, donde se solucionan los problemas de inconsistencias, datos nulos y se realizan las transformaciones pertinentes a los datos que lo requieran. Por último se tiene la tercera capa que es la de presentación de los datos que fueron cargados en las tablas de hechos, y es donde se crean los reportes que se muestran al usuario final.



Modelado de Datos

Es un conjunto de conceptos que pueden servir para describir la estructura de una base de datos. Tanto en las bases de datos tradicionales como en los almacenes de datos existen tres niveles de modelado de los datos: conceptual, lógico y físico. Debido al tipo de consultas que realizan los almacenes de datos y el rendimiento esperado, hacen que las pautas de diseño y el modelado de los datos sean diferentes.

Modelo de datos conceptual: Es un esquema conceptual a partir de los requerimientos recopilados. El esquema conceptual es una descripción concisa de los requerimientos de información de los usuarios y contiene descripciones detalladas de los tipos de datos, vínculos y las restricciones. Este modelo refleja más el espacio del problema, que el espacio de la solución.

Modelo de datos lógico: El modelo lógico es un puente intermedio entre el nivel conceptual y el físico. Este modelo describe los datos en detalle, generalmente incluye todas las entidades y las relaciones entre ellas, atributos, tipos de datos y llaves, sin tener en cuenta cómo serán implementados físicamente en la base de datos.

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

Metodologías estudiadas para el desarrollo

Existen diversas metodologías que pretenden dar un acercamiento a una propuesta ideal para el desarrollo de almacenes de datos. Cada autor la orienta a la optimización del rendimiento y a su visión de los principales procesos que se deben tener en cuenta para construir un almacén de datos flexible y dinámico. En dependencia de la problemática presentada es la política de selección por una u otra metodología.

Metodología Kimball

Para el desarrollo de este trabajo se decidió por parte del grupo de arquitectura de SIDEP (Sistema de Información para la Dirección de Establecimiento Penitenciario) utilizar la metodología propuesta por Ralph Kimball en 1992(8)

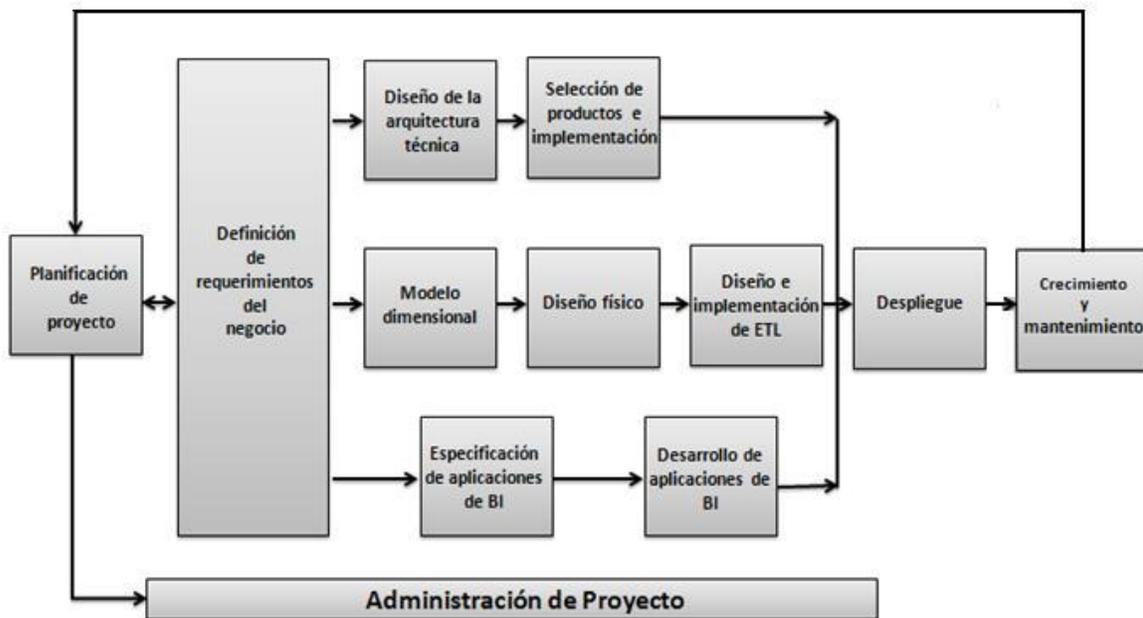


Figura 6: Ciclo de vida dimensional

Planificación

En esta etapa se determina el propósito del proyecto del almacén de datos, sus objetivos específicos y el alcance del mismo, los principales riesgos y una aproximación inicial a las necesidades de información.

Definición de los requisitos del negocio

En esta fase mediante entrevistas, recogida de información y documentos se establecerán las funcionalidades del futuro almacén de datos, siempre desde el punto de vista del negocio.

Modelado Dimensional

En esta fase a partir de los requerimientos se construye el modelo dimensional que dará soporte al almacén de datos. Comienza con un modelo dimensional de alto nivel obtenido a partir de los procesos priorizados. Luego se completan las tablas con una lista de atributos bien formada. Posteriormente se identifican todos los atributos útiles y sus ubicaciones, definiciones y reglas de negocios asociadas que especifican cómo se cargan estos datos. Finalmente se revisa el modelo con diferentes audiencias, cada una con diferentes conocimientos técnicos y del negocio.

Diseño Físico

El modelo dimensional diseñado en la fase anterior debe ser traducido al diseño físico para el motor de base de datos específico que hayamos seleccionado. En esta etapa se establecen las estrategias de agregación, así como la creación inicial de índices en el área de almacenamiento.

Diseño y desarrollo de los procesos de transformación y carga

En este estado se diseñan y desarrollan las áreas temporales de almacenamiento previo a la transformación. En esta etapa se diseñan y desarrollan también los procedimientos para combinar los datos, cualificarlos, identificar cambios, gestionar las claves, crear valores de análisis y manejar errores.

Diseño de la arquitectura técnica

En esta etapa se diseña la integración entre las distintas tecnologías, existentes con el nuevo proyecto, se analiza el impacto en los sistemas así como las necesidades que surgirán para soportar el proyecto. También se define la línea de tiempo para las distintas partes del proyecto. El diseño de la arquitectura técnica depende de los requisitos del negocio. En esta fase se deben tener en cuenta las políticas de tecnologías de la información, las arquitecturas tecnológicas existentes, la evolución futura tecnológica de la empresa, así como sinergias con otras empresas del grupo.

Selección de productos e Instalación

Esta etapa tiene como objetivo evaluar, seleccionar e instalar los productos para el almacén de datos. El primer paso antes de seleccionar nuevos productos, es comprender el proceso de aprobación de compras interno de la compañía. Luego se realiza una lista de productos candidatos y se evalúa cada uno. Finalmente se seleccionan e instalan los productos para el almacén de datos.

Especificación de Aplicaciones de Inteligencia de Negocio

Siguiendo los requerimientos del negocio establecidos en la primera fase, en esta fase se establece el diseño y funcionalidades de las distintas aplicaciones que explotarán los datos del almacén. Como no todos los usuarios del almacén necesitan tener acceso ad hoc¹, se recomienda definir un conjunto de aplicaciones estándar con los informes que se generan con mayor frecuencia para ser usados por estos.

Desarrollo de Aplicaciones de Inteligencia de Negocio

Se realiza el desarrollo de las aplicaciones diseñadas en la fase anterior. Esto implica la configuración de la herramienta de metadatos y la elaboración de reportes específicos.

Desarrollo

En esta fase convergen las fases en las que se crearon los procedimientos y áreas de extracción y transformación de datos y las de desarrollo de las aplicaciones de inteligencia de negocios. En esta fase se concluye la integración entre las distintas áreas y se implementa el sistema en producción, con las debidas pruebas preliminares. En esta fase también será necesario formar a los usuarios en las herramientas de explotación de los datos, así como la generación de manuales de uso.

Mantenimiento y Crecimiento

La creación de un DWH es un proceso (de etapas bien definidas, con comienzo y fin, pero de naturaleza espiral) que acompaña a la evolución de la organización durante toda su historia. Se necesita continuar con las actualizaciones de forma constante para poder seguir la evolución de las metas por conseguir. Al contrario de los sistemas tradicionales, los cambios en el desarrollo deben ser vistos como signos de éxito. Es importante establecer las prioridades para poder manejar los nuevos requerimientos de los usuarios y de esa forma poder evolucionar y crecer.

¹ Consultas personalizadas

Gestión del Proyecto

La gestión del proyecto asegura que las actividades del ciclo de vida se lleven a cabo de manera sincronizada. La gestión del proyecto acompaña todo el ciclo de vida. Entre sus actividades principales se encuentra la monitorización del estado del proyecto y el acoplamiento entre los requerimientos del negocio y las restricciones de los sistemas de información para poder manejar correctamente las expectativas en ambos sentidos.

Herramientas de creación de un Almacén de Datos²

En la actualidad se han desarrollado diversas herramientas con el fin de dar un acercamiento a la automatización del diseño, construcción, implementación y mantenimiento de los almacenes de datos, los que se implementan sobre un sistema gestor de bases de datos y se utilizan herramientas de ETL para integrar los datos obtenidos desde distintas base de datos.

Sistema Gestor de Base de Datos Oracle

Es el sistema gestor de base de datos (SGBD) líder en la industria, también es considerado uno de los sistemas de bases de datos más completos, destacando su soporte de transacciones, estabilidad, disponibilidad, escalabilidad, fiabilidad y seguridad. Oracle es el motor de base de datos relacional más usado a nivel mundial, puede ejecutarse en todas las plataformas y cuenta con un lenguaje de diseño de base de datos muy completo (PL\SQL).

Oracle Warehouse Builder (OWB)

Es una herramienta que permite realizar el diseño completo de un almacén de datos y luego cargarlo en un motor de base de datos Oracle mediante el diseño de procesos de ETL. La aplicación incluye una potente interfaz gráfica que permite crear dimensiones, cubos, correspondencias y procesos. Esta herramienta permite tomar los datos desde diferentes fuentes y sistemas e integrarlos para proporcionar información de alta calidad que luego será analizada. Está concebida para trabajar con la tecnología de Base de Datos Oracle. **(9)**

Oracle Business Intelligence Server

El servidor de BI (Business Intelligence, en español: Inteligencia de Negocio) permite representar el negocio de manera unificada. En este podemos disponer de todas las fuentes de datos a los que se desea que accedan las herramientas de visualización y transformarlas de forma que la manera en que el usuario

² Son las herramientas usadas para crear el mercado de datos de Tratamiento Educativo

las ve sea natural según su dominio del negocio. Además, funciona como motor de cálculo e integración de datos. **(9)**

Administration Tool

Es la herramienta utilizada para administrar el servidor de Inteligencia de Negocio de Oracle. Permite al administrador del servidor crear y editar repositorios desde modo en línea o desconectado.

Oracle BI Answers

BI Answer es una poderosa herramienta de visualización con una interfaz Web. Esta es capaz de conectarse con el servidor de BI permitiendo al usuario realizar consultas de manera sencilla, simplemente arrastrando elementos. Además, en ella es posible la creación de diferentes tipos de gráficos y tablas según las necesidades del cliente. **(9)**

Oracle Business Intelligence Interactive Dashboards

BI Dashboards permite crear pizarrones virtuales donde los usuarios del sistema pueden ingresar y ver consultas a las que necesitan acceder comúnmente y con un formato establecido que puede ser un gráfico u otra representación. Todo esto puede hacerse ingresando a una aplicación web desde cualquier navegador. **(9)**

Herramientas de modelado

Aplicaciones Informáticas que se utilizan durante todo el ciclo de vida de desarrollo del Software en tareas como realizar diseños y modelar diferentes tipos de diagramas.

Embarcadero ERStudio 8.0

Esta herramienta brinda la posibilidad de visualizar los datos desde su nacimiento hasta su destino, mediante el modelado de datos, permitiendo efectuar un análisis consistente con el objetivo de estudiar y documentar visualmente el flujo de datos. Embarcadero ERStudio 8.0 ayuda a los modeladores y arquitectos de datos a solucionar problemas de negocio en torno a la integración y consolidación de los datos.

Conclusiones Parciales.

Las aplicaciones informáticas utilizadas por distintos sistemas penitenciarios tanto extranjeros como nacionales que se estudiaron en el presente capítulo, no tienen un almacén de datos en explotación para generar las estadísticas que sustentarán la toma de decisiones y la mayoría basan sus análisis en cuestiones legales y judiciales del interno y no al tratamiento educacional del mismo. A partir de esto se definió la necesidad de desarrollar un mercado de datos que apoye la toma de decisiones en la Dirección

de Establecimientos Penitenciarios. El análisis de los conceptos de almacén de datos y mercado de datos identificó las principales características deseables de la propuesta de solución. Para su desarrollo se utilizó las herramientas descritas y la metodología Kimball con una arquitectura de datos de tres capas. El diseño del esquema del mercado de datos utilizado fue constelación basado en el enfoque dimensional con una estructura de almacenamiento ROLAP.

Capítulo 2: Análisis y diseño

Introducción

En este capítulo se presenta el análisis realizado en la especialidad de Tratamiento Educativo de la Dirección de Establecimientos Penitenciarios, se realiza un estudio de los procesos que son de mayor interés para el departamento y las perspectivas que son usadas en el análisis. Con los datos se identifican los indicadores que se analizan para realizar el diseño del mercado de datos correspondiente.

Definición de los requisitos del negocio

Para dar cumplimiento a los pasos planteados por la metodología, se analizaron los requisitos que son de gran importancia para el negocio que se evalúa. Se realizaron entrevistas con los usuarios, en este caso el departamento de dirección y a los jefes del área de Tratamiento Educativo de la DEP (Dirección de Establecimientos Penitenciarios) en busca de necesidades de información. Se analizó cada uno de los procesos que intervienen en el negocio del área y los indicadores que se tienen en cuenta para la toma de decisiones.

Para la captura de requisitos del almacén de datos se introduce un nuevo tipo de requisitos llamados requisitos de información.

Requisitos de información

Estos requisitos representan toda la información que debe estar disponible para el usuario final y constituyen la entrada principal para el proceso de inteligencia de negocio y para futuros reportes bases. A continuación se especifican los requisitos de información que fueron identificados agrupados por procesos para una mejor estructuración.

Administración de la Biblioteca:

- R11. Mostrar el total de préstamos por nivel de mando.
- R12. Mostrar el total de asistencias por interno a la biblioteca.
- R13. Mostrar el total de libros existentes en la biblioteca.
- R14. Mostrar el total de libros donados por instituciones.
- R15. Mostrar el total de actividades realizadas por las bibliotecas.
- R16. Mostrar el total de solicitudes de un libro determinado.
- R17. Mostrar el total de libros existentes en la biblioteca por género.

Emulación:

- RI8. Mostrar el total de internos con 100 puntos en la emulación por nivel de mando.
- RI9. Mostrar el total de estímulos recibidos por interno.
- RI10. Mostrar el total de internos destacados por nivel de mando.
- RI11. Mostrar el total de estímulos individuales por nivel de mando.
- RI12. Mostrar el total de estímulos colectivos por nivel de mando.
- RI13. Mostrar el total de internos que han recibido más de tres estímulos.
- RI14. Mostrar el total de internos que nunca han recibido un estímulo.
- RI15. Mostrar el total de internos fuera de categoría.
- RI16. Mostrar el total de colectivos destacados por nivel de mando.
- RI17. Mostrar el total de colectivos con más de dos estímulos.

Permisos:

- RI18. Mostrar el total de permisos por tipo de permiso en un rango de fecha.
- RI19. Mostrar el total de permisos suspendidos por tipo de permiso.
- RI20. Mostrar el total de violaciones por tipo de permiso.
- RI21. Mostrar el total de permisos de salida con más de 72 horas por tipo de permiso.
- RI22. Mostrar el por ciento que representa las violaciones por tipo de permiso.
- RI23. Mostrar el por ciento que representa la cantidad de salidas con más de 72 horas por tipo de permiso.

Trabajo:

- RI24. Mostrar el total de internos incorporados al trabajo por condiciones.
- RI25. Mostrar el total de internos incorporados al trabajo por esfera.
- RI26. Mostrar el total de internos incorporados al trabajo por actividad.
- RI27. Mostrar el total de internos no incorporados al trabajo, por causales.
- RI28. Mostrar el total de unidades productivas por nivel de mando.
- RI29. Mostrar el total de unidades productivas por tipo.
- RI30. Mostrar el total de contratos de trabajo por nivel de mando.
- RI31. Mostrar el total de contratos con violaciones.
- RI32. Mostrar el salario promedio de los internos por entidad empleadora.
- RI33. Mostrar salario promedio de los internos por unidad productiva.
- RI34. Mostrar el total de internos con salario pendiente de pago por unidad.

- RI35. Mostrar el total de internos con salario pendiente de pago por entidad empleadora.
- RI36. Mostrar el total de internos con salario retenido de pago por unidad.
- RI37. Mostrar la cantidad de internos con salario retenido de pago por entidad empleadora.
- RI38. Mostrar el total internos con responsabilidad civil, que trabajan y no se les descuenta.
- RI39. Mostrar el total de internos que reciben salario y continúan recibiendo ayuda económica.
- RI40. Mostrar el total de internos con responsabilidad civil que no trabajan.

Actividades Educativas:

- RI41. Mostrar el total de internos incorporados a las actividades educativas.
- RI42. Mostrar el total de internos no incorporados a las actividades educativas.
- RI43. Mostrar el total de internos que participaron en una actividad como participantes pasivos.
- RI44. Mostrar el total de internos que participaron en una actividad como participantes activos.
- RI45. Mostrar el total de familiares que participaron en una actividad.
- RI46. Mostrar el total de actividades realizadas por nivel de mando.
- RI47. Mostrar el total de convenios educativos en un periodo de tiempo por nivel de mando.
- RI48. Mostrar el total de colaboradores que participan en una actividad.
- RI49. Mostrar el total de instructores que participan en una actividad.
- RI50. Mostrar el total de actividades estimadas por planes.
- RI51. Mostrar el total de actividades realizadas por planes.
- RI52. Mostrar el total de actividades penales realizadas.
- RI53. Mostrar el total de actividades extrapenales realizadas.

Beneficio:

- RI54. Mostrar el total de internos con propuestas de rebaja de sanción.
- RI55. Mostrar el total de internos con propuestas de rebaja de sanción por motivo.
- RI56. Mostrar el total de internos con propuestas de libertad condicional.
- RI57. Mostrar el total de internos con propuestas de suspensión de medida de seguridad.
- RI58. Mostrar el total de internos con suspensión de TCCI.
- RI59. Mostrar el total de internos con sustitución de sanción por subsidiaria.
- RI60. Mostrar el total de internos para progresión de régimen.
- RI61. Mostrar el total de internos para progresión de fase.
- RI62. Mostrar el total de internos para regresión de régimen.
- RI63. Mostrar el total de internos para regresión de fase.

- RI64. Mostrar el total de internos con tránsito anticipado a Media.
- RI65. Mostrar el total de internos con tránsito anticipado a Mínima.
- RI66. Mostrar el total de rebajas de sanción aprobadas.
- RI67. Mostrar el total de rebajas de sanción denegadas.
- RI68. Mostrar el por ciento que representan los permisos denegados.
- RI69. Mostrar el total de permisos aprobados.
- RI70. Mostrar el total de internos con más de dos beneficios.
- RI71. Mostrar el total de internos con licencia extrapenal.

Generales:

- RI72. Descargar documentos en diferentes formatos.
- RI73. Imprimir documentos.
- RI74. Modificar visualización de gráficos.

Técnicas de identificación de requisitos

La captura de requisitos tiene como ventaja que al participar el usuario garantiza que el proyecto se adecue a sus necesidades, siendo muy importante tener bien claro las diferentes actividades a realizar para obtener los beneficios planteados. De las técnicas para la captura de requisitos propuestas por la metodología se utilizaron el análisis de la documentación y las entrevistas donde se recomendaba primero hacer un estudio de la documentación para tener una preparación y tener una visión propia del negocio antes de iniciar el proceso de recopilación de requisitos. Luego seleccionar el grupo de entrevistados de manera horizontal que pueden ser individuos o grupos muy pequeños y elaborar un conjunto de cuestionarios antes de comenzar la entrevista la cual fomentará un alto grado de participación de los entrevistados lo que generará datos a un mejor nivel de detalle.

Análisis de la información: Se estudiaron los documentos y resoluciones del área de Tratamiento Educativo. Entre los documentos analizados se encuentran las leyes y normativas a seguir en los centros penitenciarios, para los distintos procesos que se tienen en cuenta en el área de Tratamiento Educativo.

Entrevistas: Se realizó una serie de preguntas acerca del sistema. Se utilizó con el propósito de conocer cuáles son las metas y objetivos del negocio para el Área de Tratamiento Educativo, sus prioridades, los procesos que abarca y cuál es el personal que trabaja en dicha área, las responsabilidades que le competen a cada uno, así como se identificaron sus necesidades de información.

Temas analíticos:

La metodología de Kimball propone para una mejor organización de las áreas del negocio agrupar por temas analíticos los procesos del negocio, donde cada tema agrupa más de un proceso.

A partir de las entrevistas realizadas y del análisis de la documentación, se identificaron temas analíticos y procesos de negocio en el área de Tratamiento Educativo.

El área de Tratamiento Educativo se dividió en dos temas analíticos, tratamiento penitenciario y beneficio, donde todas sus funciones están relacionadas y sus salidas permitirán conocer cómo marchan las estadísticas con respecto a los diferentes indicadores que se manejan en esta área:

Tratamiento penitenciario: Este tema se encarga de todo el proceso de superación del interno con el objetivo de reintegrarlo a la sociedad. Agrupa los procesos Emulación, Actividades Educativas, Administración de la Biblioteca y Trabajo.

Administración de la Biblioteca: Es el proceso encargado de manipular fácilmente el proceso de préstamos y devoluciones de obra literarias que se les facilitan a los internos que se encuentran en el centro penitenciario.

Actividades Educativas: Este proceso es el encargado de la gestión de la información referente a aquellas actividades que tributan a la superación profesional y cultural de los internos, las mismas pueden realizarse dentro o fuera de los centros penitenciarios a los que pertenecen los internos. Estas actividades educativas también pueden ser organizadas por la biblioteca del centro penitenciario.

Emulación: Este proceso se ocupa de gestionar toda la información referente a la puntuación que reciben los internos en su tránsito por el sistema penitenciario como parte del tratamiento educativo que se les dispensa. Los internos deben cumplir con una serie de parámetros que se le evalúan mensualmente y le van sumando puntos de estímulo hasta llegar a la suma de 100 puntos cada mes, también en caso de incurrir en indisciplinas se les restan puntos a este acumulado y pierden la condición de destacado en ese período que cometió la indisciplina. Al final del mes si el acumulado está entre 90 y 100 puntos, el interno obtiene la condición de destacado en ese período.

Trabajo: Trabajo permite registrar las actividades productivas que realizan los sancionados y asegurados, registrándose la unidad productiva o empresa que solicitó el contrato: datos de la unidad productiva, datos del jefe de brigada destinado a la misma.

Beneficio: Este tema es el encargado de operar con los beneficios que se les pueden asignar a los internos. Agrupa los procesos de Beneficios y Permisos.

Beneficios: Es el proceso encargado de gestionar todos los beneficios que se les puede otorgar a un interno o a los cuales tiene derecho el mismo, se entiende por promoción de los internos aquellos derechos que implican un cambio esencial de su situación legal: rebaja de sanción, progresión en régimen, propuestas de libertad condicional, licencia extrapenal, sustitución de la sanción privativa de libertad por subsidiarias de esta, suspensión o cambio de la medida de seguridad, suspensión de la sanción de trabajo correccional con internamiento y suspensión de la sanción de privación de libertad.

Permisos: Este proceso es el encargado de generar un listado con todos aquellos internos que podrán ser analizados en el consejo de promoción para este beneficio.

Especificación de requerimientos:

Nº	Nombre	Descripción
RI1	Mostrar el total de préstamos por nivel de mando.	Representa la cantidad de préstamos que ha hecho la biblioteca visualizándolo en una gráfica de barra donde cada una de ellas representa una provincia permitiendo ver dicho reporte en forma de tabla.
	Campos	Tipos de Datos
	Nivel de Mando	Cadena de caracteres
	Tiempo	Fecha
	Total de préstamos	Número entero
	Nombre	Descripción
RI2	Mostrar el total de asistencias por interno a la biblioteca.	Representa el comportamiento de la asistencia de un interno a la biblioteca visualizándolo en una gráfica de barra

		donde cada una de ellas representa una provincia permitiendo ver dicho reporte en forma de tabla.
	Campos	Tipos de Datos
	Interno	Cadena de caracteres
	Tiempo	Fecha
	Nivel de Mando	Cadena de caracteres
	Total de asistencias	Número entero
	Nombre	Descripción
RI3	Mostrar el total de libros existentes en la biblioteca.	Representa la cantidad de libros existentes en la biblioteca visualizándolo en una gráfica de barra donde cada una de ellas representa una provincia permitiendo ver dicho reporte en forma de tabla.
	Campos	Tipos de Datos
	Tiempo	Fecha
	Nivel de Mando	Cadena de caracteres
	Total de libros	Cadena de caracteres
	Nombre	Descripción
RI4	Mostrar el total de libros donados por instituciones.	Representa la cantidad de libros por procedencia visualizándolo en una gráfica de barra donde cada una de ellas representa una provincia permitiendo ver dicho reporte en forma de tabla.

	Campos	Tipos de Datos
	Tiempo	Fecha
	Procedencia	Cadena de caracteres
	Nivel de Mando	Cadena de caracteres
	Total de libros	Número entero
	Nombre	Descripción
RI5	Mostrar el total de actividades realizadas por las bibliotecas.	Representa la cantidad de actividades realizadas en la biblioteca por tipo visualizados en una gráfica de barra donde cada una de ellas representa una provincia permitiendo ver dicho reporte en forma de tabla.
	Campos	Tipos de Datos
	Tiempo	Fecha
	Nivel de Mando	Cadena de caracteres
	Tipo	Cadena de caracteres
	Total de actividades	Número entero
	Nombre	Descripción
RI6	Mostrar el total de solicitudes de un libro determinado.	Representa la cantidad de solicitudes que ha tenido un libro específico visualizándolo en una gráfica de barra donde cada una de ellas representa una provincia permitiendo ver dicho reporte en forma de tabla.

	Campos	Tipos de Datos
	Tiempo	Fecha
	Nivel de Mando	Cadena de caracteres
	Tipo	Cadena de caracteres
	Total de solicitudes	Número entero
	Nombre	Descripción
R17	Mostrar el total de libros existentes en la biblioteca por género.	Representa la cantidad de libros existentes en la biblioteca por género en una gráfica de barra donde cada una de ellas representa una provincia, permitiendo ver dicho reporte en forma de tabla.
	Campos	Tipos de Datos
	Tiempo	Fecha
	Nivel de Mando	Cadena de caracteres
	Género	Cadena de caracteres
	Total de libros	Número entero

Tabla 1: Especificación de requisitos

Validación de requisitos

El objetivo de la validación de requisitos es verificar todos los requisitos existentes en los documentos para asegurarse que presentan una descripción aceptable del sistema a entregar.

La metodología propone como técnica de validación de requisitos un prototipo con el objetivo principal de confirmar que el diseño y las tecnologías reunirán los requisitos del negocio antes de invertir tiempo y dinero en el desarrollo del sistema de transformación de datos completo. Se propone seleccionar cuidadosamente un subconjunto pequeño pero representativo de datos a utilizar para la creación de

prototipos. Para el área de Tratamiento Educativo se utilizaron los prototipos de los mercados de datos de tesis anteriores.

Modelo dimensional

Dentro del Modelo de Datos se encuentra el “modelo dimensional”, que tiene como elementos principales las dimensiones y los hechos, los cuales recogen de los procesos que abarca el subsistema, los atributos necesarios para dar respuesta a los requisitos de información. En las dimensiones se describe los detalles del hecho, y el hecho es la relación existen entre las diferentes dimensiones. Es importante conocer la composición de ambos y la relación que existe entre ellos para la eficiente función del Mercado de Datos.

Granularidad

La granularidad representa el nivel de detalle al que se desea almacenar la información sobre el negocio que se esté analizando. Mientras mayor sea el nivel de detalle de los datos, se tendrán mayores posibilidades analíticas, porque los mismos podrán ser resumidos o sumariados. Para todos los procesos el nivel de granularidad del tiempo será diario y el nivel de mando será hasta centro penitenciario.

Tablas de dimensiones

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

- **dim_actividad_biblioteca:** Contiene las características de las actividades que se realizan en la biblioteca.
- **dim_libro:** Contiene los datos del libro.
- **dim_asistencia_biblioteca:** Contiene los datos referentes a las asistencias de una persona a la biblioteca.
- **dim_interno:** Contiene los atributos comunes del interno que se van a utilizar en todos los procesos.
- **dim_tiempo:** Contiene los datos relacionados con el tiempo.
- **dim_nivel_mando:** Contiene los atributos comunes de los niveles de mando que se van a utilizar en todos los procesos.

Para profundizar en la composición de cada una de las tablas dimensión se puede apoyar en el documento "Especificaciones del Modelo de Datos Dimensional" anexo al expediente del proyecto Prisiones Cuba.

Tablas de hechos

Una tabla de hechos es una representación de un proceso de negocio. Esta tabla se encuentra rodeada de tablas de dimensiones que determinan como están distribuidos y relacionados los datos lógicamente en el ambiente del negocio. En el proceso administración de la biblioteca podemos encontrar las siguientes tablas de hechos:

- **fact_prestamo:** En esta tabla quedarán registrados los datos referentes a un préstamo que se realice en la biblioteca.
- **fact_actividad_biblioteca:** En esta tabla quedarán registrados los datos referentes a las actividades que organiza la biblioteca.
- **fact_asistencia_biblioteca:** En esta tabla quedará registrados los datos referentes a la asistencia de personas a la biblioteca.

Medidas

Las medidas que se recogen en el proceso administración de la biblioteca y sus respectivos subprocesos son:

- **Total de préstamos:** Representa la cantidad de préstamos que tiene una biblioteca, un libro, un interno.
- **Total de libros:** Representa la cantidad de libros existentes en la biblioteca, por tipo de ingreso, género.
- **Total de actividades:** Representa la cantidad de actividades realizadas por la biblioteca por nivel de mando, por tipo.
- **Total de asistencias:** Representa la cantidad de asistencias por interno, por actividad.

Matriz dimensional

La matriz dimensional representa las relaciones que se establecen entre las áreas temáticas y las dimensiones identificadas. En la matriz se observa la existencia de dimensiones comunes y las que son específicas por cada tema analítico.

Dimensiones	Administración Biblioteca	Emulación	Beneficio	Permiso	Trabajo	Actividad Educativa
Interno	x	x	x	x	x	x
Tiempo	x	x	x	x	x	x
Nivel de Mando	x	x	x	x	x	x
Asistencia biblioteca	x					
Libro	x					
Préstamo	x					
Actividad	x					
Emulación		x				
Estímulo		x				
Beneficio			x			
Permisos				x		
Unidad Productiva					x	
Actividad Educativa						x

Tabla 2: Matriz dimensional

Patrones de diseño para Almacenes de Datos

En la actualidad los almacenes de datos suelen ser muy grandes y en ocasiones el uso de los patrones de diseño facilita el trabajo dando una solución al problema de manera simple. Para el diseño del mercado de datos del área Tratamiento Educativo se utilizaron los siguientes patrones:

Llaves subrogadas: En el mercado de datos del área de Tratamiento Educativo se hace necesario utilizar este tipo de llaves ya que las dimensiones no presentan un identificador claro del negocio. Este patrón soluciona el problema anterior, pues plantea que se debe generar una llave primaria única para cada dimensión en vez de usar un atributo identificador en el contexto dado. Para la creación de estas, se utilizan números enteros, ya que está demostrado que estos no se repiten o se repiten con una probabilidad extremadamente baja. Esta llave subrogada permite que las tablas dimensión del mercado de datos sean más fáciles de consultar por el identificador, dado que se conoce el mismo tipo de datos (numeric o number) de todos en cada tabla dimensión.

Dimensiones lentamente cambiantes: Este tipo de dimensión gestiona los cambios que se producen en los atributos de las dimensiones, este tipo de dimensión establece varios métodos para su procesamiento:

SCD Tipo 1: Sobrescribir: cuando hay un cambio en los valores de un atributo, sobrescribimos el valor antiguo con el nuevo sin registrar una historia. Esto significa perder toda la historia del dato, y cuando hagamos un análisis veremos la información histórica desde el punto de vista actual.

SCD Tipo 2: Añadir fila: cuando hay un cambio, creamos un nuevo registro en la tabla. El nuevo registro tiene una nueva clave subrogada, de forma que una entidad de sistema operacional (por ejemplo, un interno), puede tener varios registros en la tabla de la dimensión según se van produciendo los cambios. Estamos gestionando un versionado, que además puede incluir unas fechas para indicar los periodos de validez, numerador de registros o un indicador de registro activo o no.

SCD Tipo 3: Añadir columna: cuando hay un cambio, nos guardamos el valor anterior en una columna distinta, actualizando el campo con el nuevo valor (para cada campo, tendremos una tupla valor anterior, valor actual). Solo nos vamos a guardar, por tanto, los dos últimos valores.

Para el mercado de datos se decidió utilizar específicamente las dimensiones lentamente cambiantes de tipo 2, debido a que es de interés conservar todos los datos históricos en el mercado de datos. Un ejemplo de dimensión lentamente cambiante es la dimensión Interno, ya que contiene atributos como son el estado civil, el número de hijos entre otros los cuales tienden a variar en el tiempo y se hace

necesario conservar en el mercado de datos todos los estados por los que han ido transitando los datos del interno.

Modelado de datos

Para darle respuesta a los requisitos de información se realizó el modelado de datos, compuesto por sus tablas de hechos y sus respectivas dimensiones. A continuación se muestran imágenes del modelo de datos para los subprocesos Asistencia a la biblioteca, Actividad de la biblioteca y Préstamos correspondientes al proceso Administración de la biblioteca.

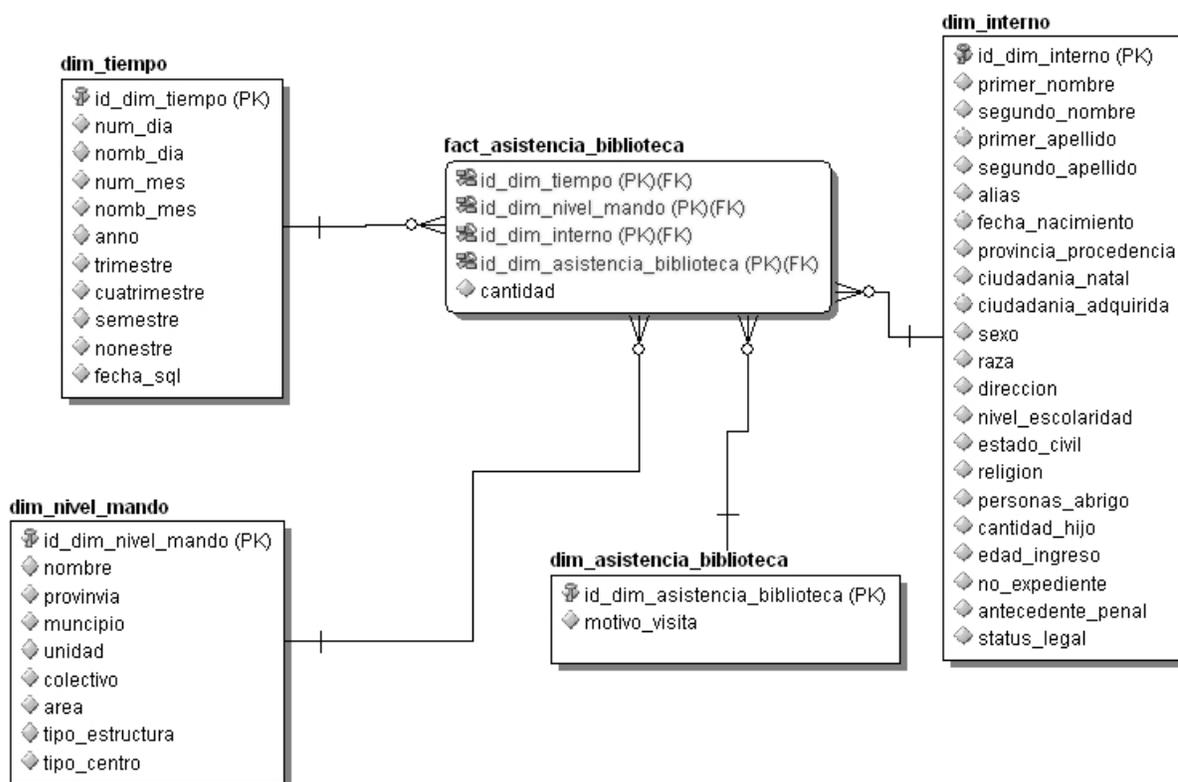


Figura 7: Modelo de datos del paquete "Asistencia biblioteca"

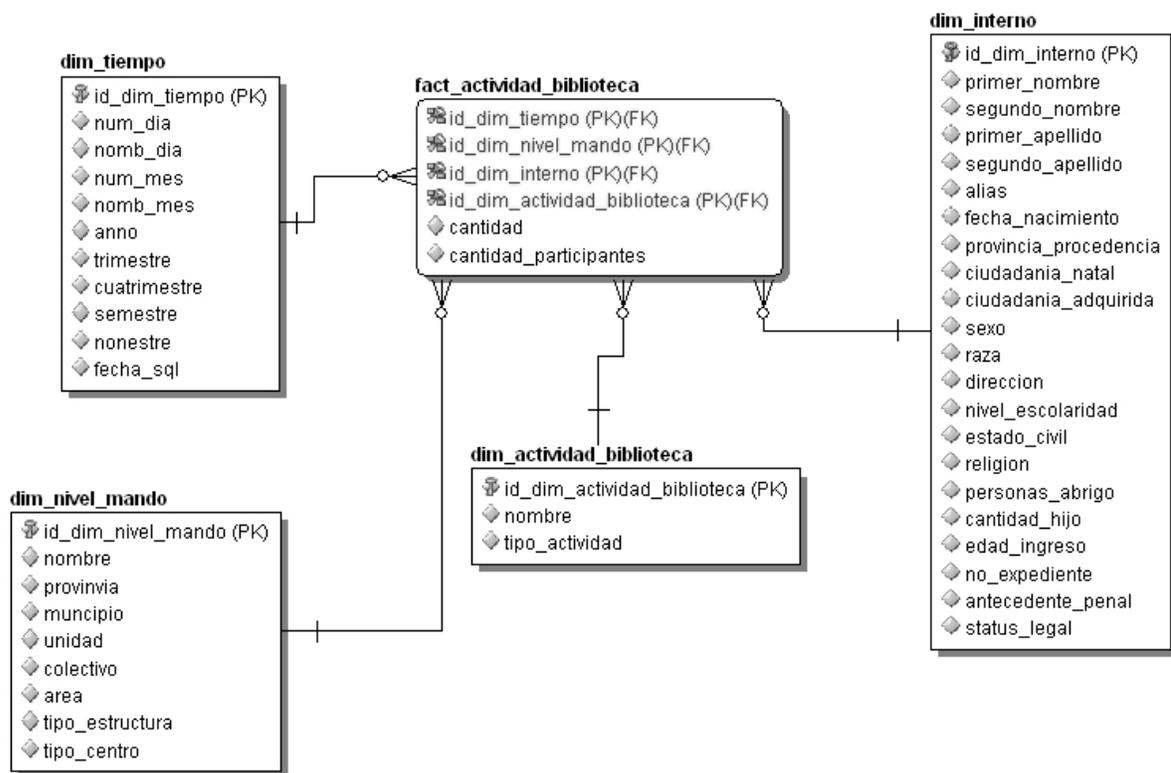


Figura 8: Modelo de datos del paquete "Actividad biblioteca"

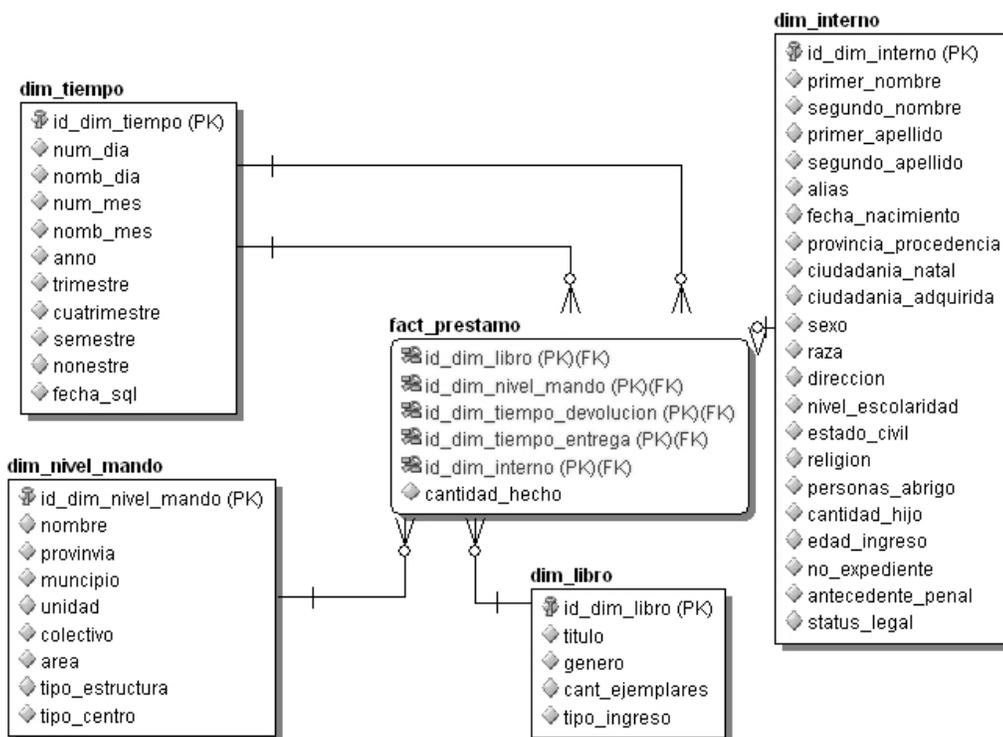


Figura 9: Modelo de datos del paquete "Préstamo"

Diseño de la ETL

La realización de los procesos de ETL desde la fuente de los sistemas operacionales hasta el propio Almacén de datos conlleva a una correcta correspondencia entre los elementos, en aras de que el almacén no cuente con valores vacíos o incorrectos. Por ello es adecuado identificar la mejor fuente de datos para cada registro, estandarizar códigos y formatos de representación y usar conversiones y combinaciones para generar nuevos datos y corregirlos.

Para la realización del diseño de estos procesos y en correspondencia con la arquitectura de tres capas se identificó primeramente el origen de los datos necesarios para la carga inicial de las tablas dimensión, posteriormente se crearon las Staging Area (Áreas de almacenamiento temporal), las cuales constituyen una línea divisoria entre el sistema operacional y el depósito de datos, las mismas van a almacenar todos los datos necesarios de las tablas de la base de datos de origen, y por último se procede a cargar los datos en las tablas de hechos.

Dimensión dim_interno

Una vez realizado el diseño de la dimensión interno, se analizaron las tablas de la base de datos origen para garantizar que para cada atributo de la dimensión interno en el almacén de datos exista la correspondencia con la base de datos fuente. En este paso se identificaron las tablas INTERNO, NOM_COLORPIEL, NOM_PROVINCIA, NOM_NIV_ESC, NOM_ESTADOCIVIL, EXP_LEGAL entre otras como se puede apreciar en la imagen siguiente, la cual constituye solo una representación del origen de los datos para una mejor comprensión y organización de las distintas fuentes de datos:

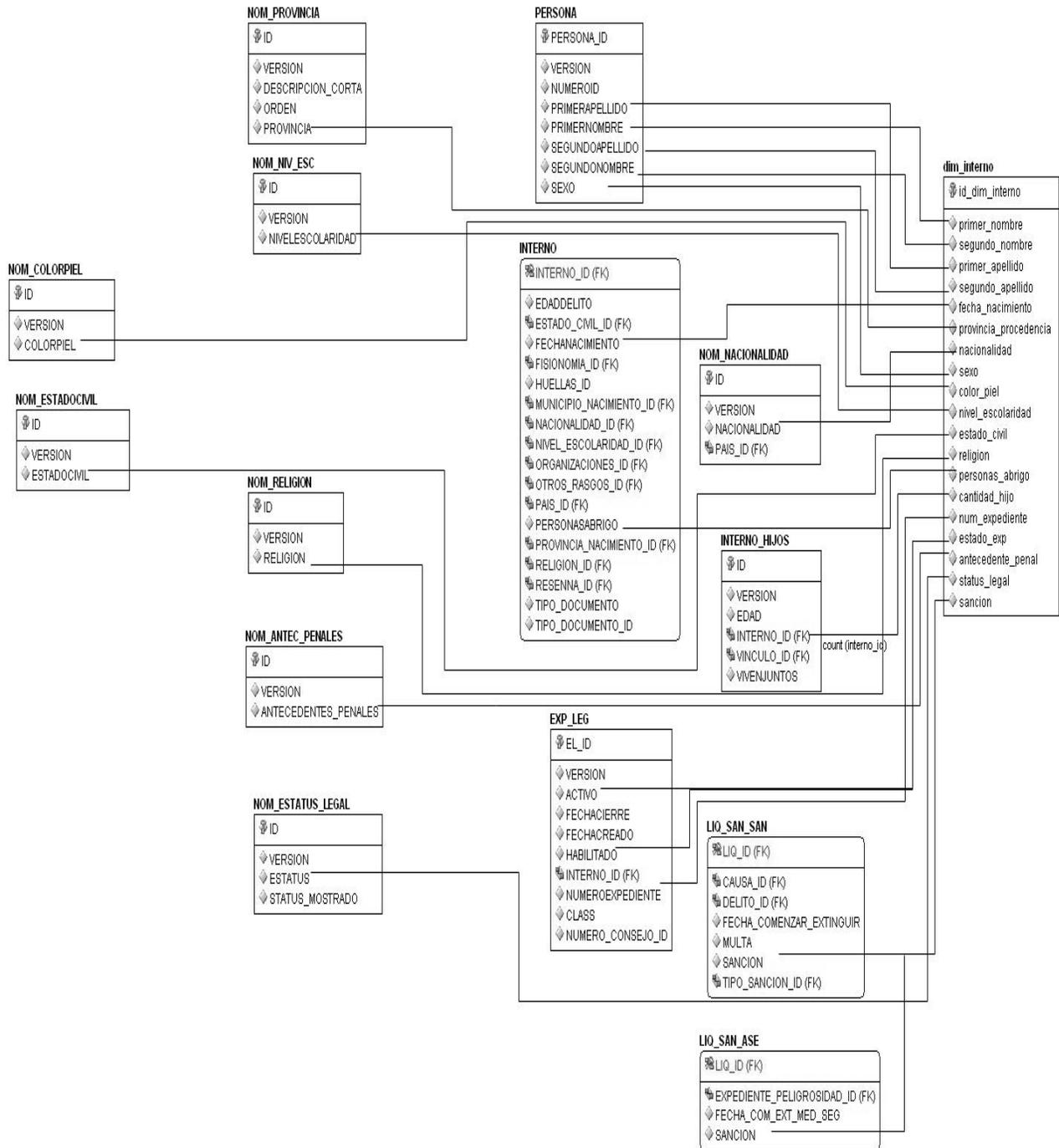


Figura 10: Correspondencia de la dimensión Interno

Dimensión dim_nivel_mando

Una vez realizado el diseño de la dimensión nivel de mando, se analizaron las tablas de la base de datos origen para garantizar que para cada atributo de la dimensión nivel de mando en el almacén de datos exista la correspondencia con la base de datos fuente. En este paso se identificaron las tablas ESTRUCTURA, NOM_PROVINCIA, COLECTIVO entre otras como se puede apreciar en la imagen siguiente, la cual constituye solo una representación del origen de los datos para una mejor comprensión y organización de las distintas fuentes de datos:

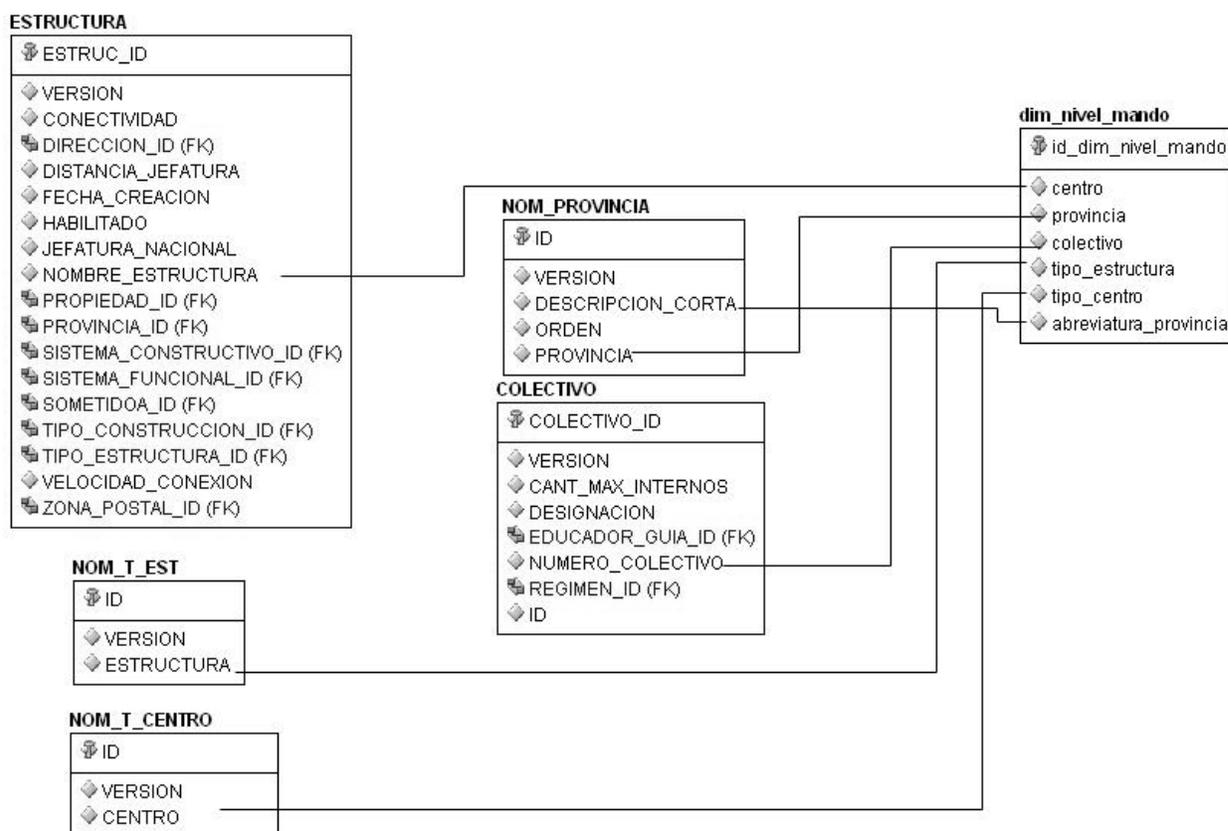


Figura 11: Correspondencia de la dimensión Nivel de mando

Dimensión dim_actividad_biblioteca

Una vez realizado el diseño de la dimensión actividad biblioteca, se analizaron las tablas de la base de datos origen para garantizar que para cada atributo de la dimensión actividad biblioteca en el almacén de datos exista la correspondencia con la base de datos fuente. En este paso se identificaron las tablas ACTIVIDAD_BIBLIOTECA y NOM_TIPO_ACTIVIDAD como se puede apreciar en la imagen siguiente, la cual constituye solo una representación del origen de los datos para una mejor comprensión y organización de las distintas fuentes de datos:

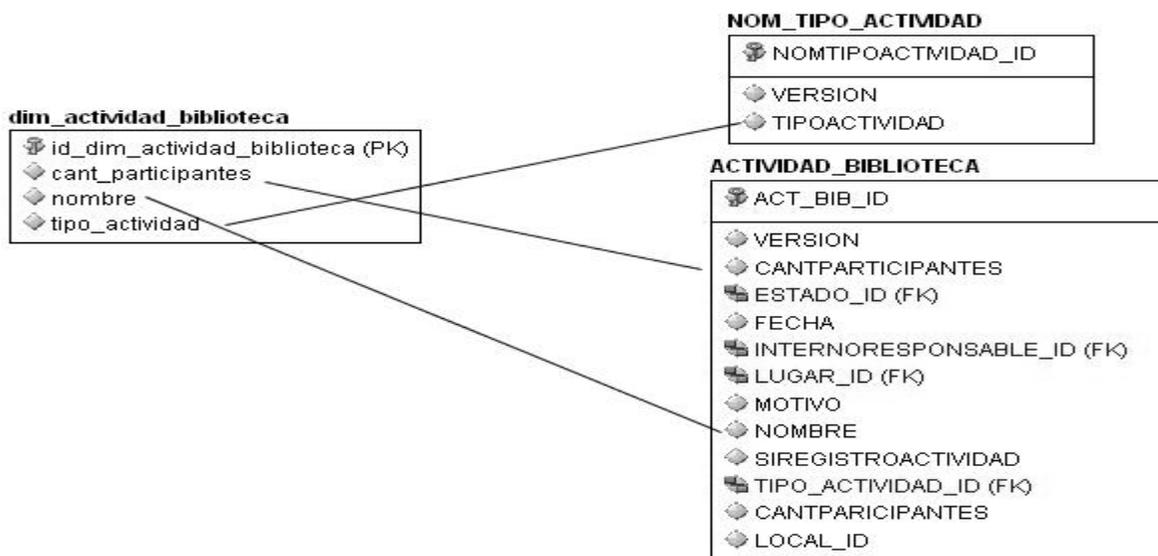


Figura 12: Correspondencia de la dimensión Actividad biblioteca

Dimensión dim_libro

Una vez realizado el diseño de la dimensión libro, se analizaron las tablas de la base de datos origen para garantizar que para cada atributo de la dimensión libro en el almacén de datos exista la correspondencia con la base de datos fuente. En este paso se identificaron las tablas OBRA_LITERARIA, NOM_GENERO y NOM_TIPO_INGRESO como se puede apreciar en la imagen siguiente, la cual constituye solo una representación del origen de los datos para una mejor comprensión y organización de las distintas fuentes de datos:

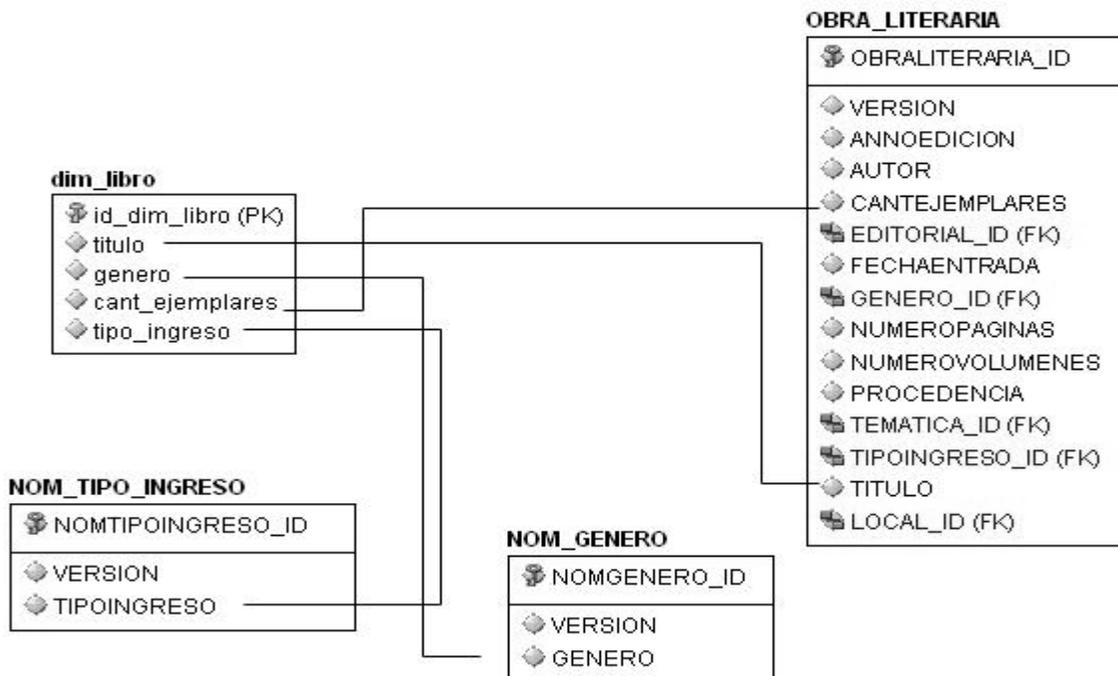


Figura 13: Correspondencia de la dimensión Libro

Dimensión asistencia_biblioteca

Una vez realizado el diseño de la dimensión asistencia biblioteca, se analizaron las tablas de la base de datos origen para garantizar que para cada atributo de la dimensión asistencia biblioteca en el almacén de datos exista la correspondencia con la base de datos fuente. En este paso se identificó como única tabla necesaria de la base de datos origen el nomenclador NOM_MOTIVO_VISITA como se puede apreciar en la imagen siguiente, la cual constituye solo una representación del origen de los datos para una mejor comprensión y organización de las distintas fuentes de datos:

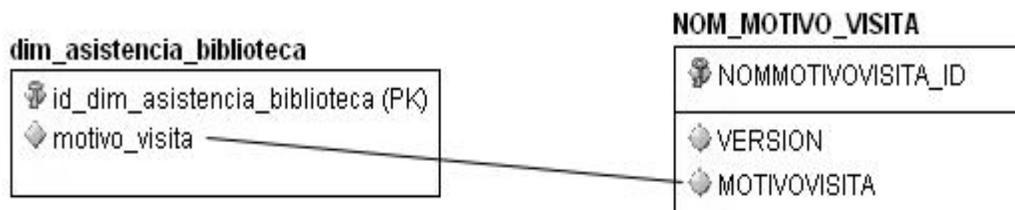


Figura 14: Correspondencia de la dimensión Asistencia biblioteca

Conclusiones Parciales

En este capítulo se definieron los principales procesos que se llevaron a cabo en el desarrollo del análisis y diseño del mercado de datos para el área de Tratamiento Educativo de la Dirección de Establecimientos Penitenciarios, determinando de esta manera los temas analíticos a los cuales están asociados los mismos. Se especificaron las dimensiones y las tablas de hechos que componen el modelo lógico para el proceso descrito y se diseñó la matriz dimensional. Se elaboraron las correspondencias con la base de datos origen realizando el diseño del proceso de extracción transformación y carga.

Capítulo 3: Implementación y pruebas

Introducción

En este capítulo se procederá a la implementación de la solución propuesta, se llevará a cabo los procesos de ETL, se configurará la herramienta de administración así como se implementarán los reportes y se diseñará el tablero de mando. También se realizará la validación del mercado de datos de Tratamiento Educativo aplicando el método de prueba de caja negra.

Implementación

Para la implementación se hace necesario realizar la extracción, transformación y carga de los datos desde la fuente, después la configuración de la herramienta de administración, la creación de los reportes y por último la confección del tablero de mando.

Implementación de las ETL.

La implementación de las ETL fue realizada utilizando la herramienta Oracle Warehouse Builder. El proceso de las ETL se le realizan a cada cubo, es decir a las tablas de hechos y sus respectivas tablas dimensiones.

Carga inicial

Es el proceso encargado de que se carguen las dimensiones comunes y de poblar los cubos correspondientes.

A continuación se especifica el proceso ETL del cubo Asistencia a la biblioteca:

- **ETL Dimensión asistencia biblioteca**

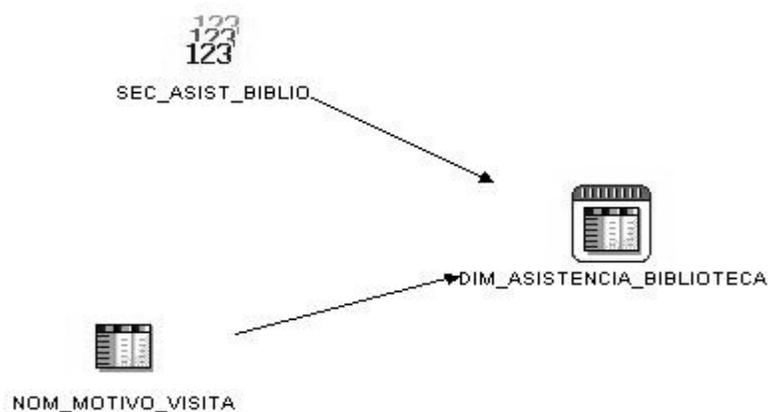


Figura 15: Procesos ETL de la dimensión Asistencia biblioteca

A continuación se especifican los pasos para cargar la DIM_ASISTENCIA_BIBLIOTECA:

Se tomará como fuente de entrada los datos contenidos en la tabla NOM_MOTIVO_VISITA y se utiliza el operador de secuencia para generar el id secuencial de la dimensión.

Dimensiones Comunes

- ETL Dimensión tiempo

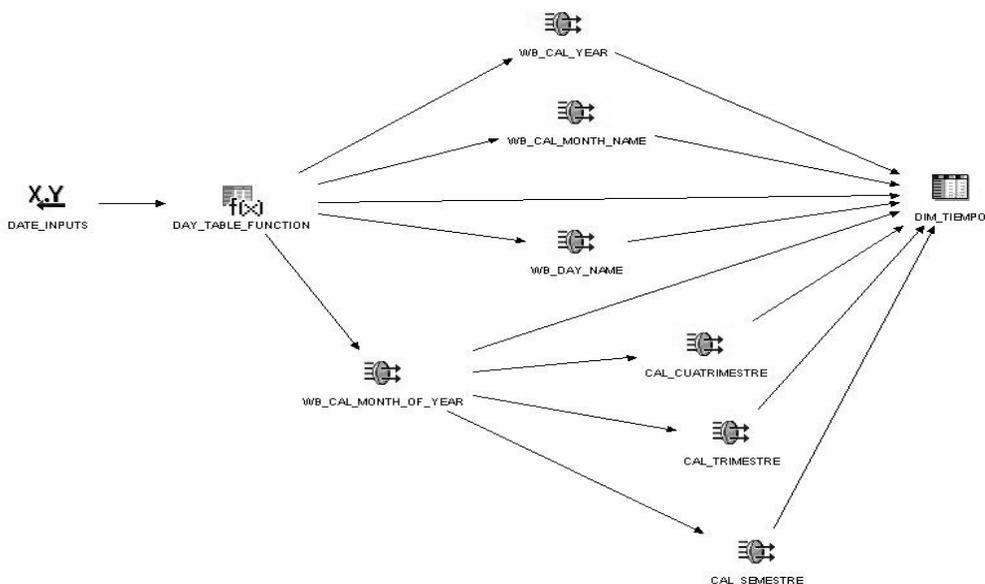


Figura 16: Procesos ETL de la dimensión Tiempo

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

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

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

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

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

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

- ETL Nivel de mando

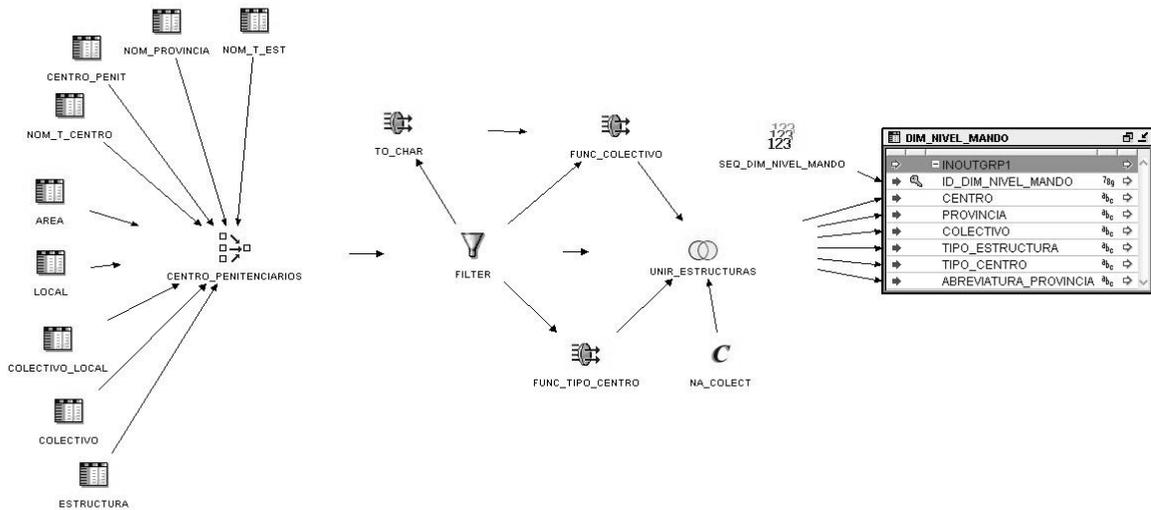


Figura 17: Procesos ETL de la dimensión Nivel de mando

Para la carga de la dimensión dim_nivel_mando se toma como fuente de entrada los datos necesarios que se encuentran en las tablas: estructura, colectivo, colectivo_local, local, area, nom_t_centro, centro_penit, nom_provincia y nom_t_est.

Se procede a utilizar el operador JOINER titulado “centro_penitenciaros”, encargado de unir las tablas antes mencionadas.

Luego se utiliza el operador FILTER, el objetivo del mismo es aplicar una condición de filtro definida, su salida es el resultado del filtro aplicado.

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

- **Func_tipo_centro:** dado el tipo de estructura y el tipo de centro devuelve el tipo de centro correspondiente.
- **Func_colectivo:** dado el tipo de estructura y el número de colectivo devuelve el número de colectivo correspondiente.

Posteriormente se emplea el operador SET_OPERATION titulado “unir_estructuras”, su función en este caso es la de combinar los resultados de las consultas y devolver el resultado de esto. Se utiliza el operador CONSTANT titulado “na_colect” para cargar datos que no se encuentran en la base de datos y que es necesario en la dimensión dim_nivel_mando. El identificador de la dimensión es el valor generado por la secuencia SEQ_DIM_NIVEL_MANDO.

Áreas de almacenamiento temporal

- ETL STG Ubicación

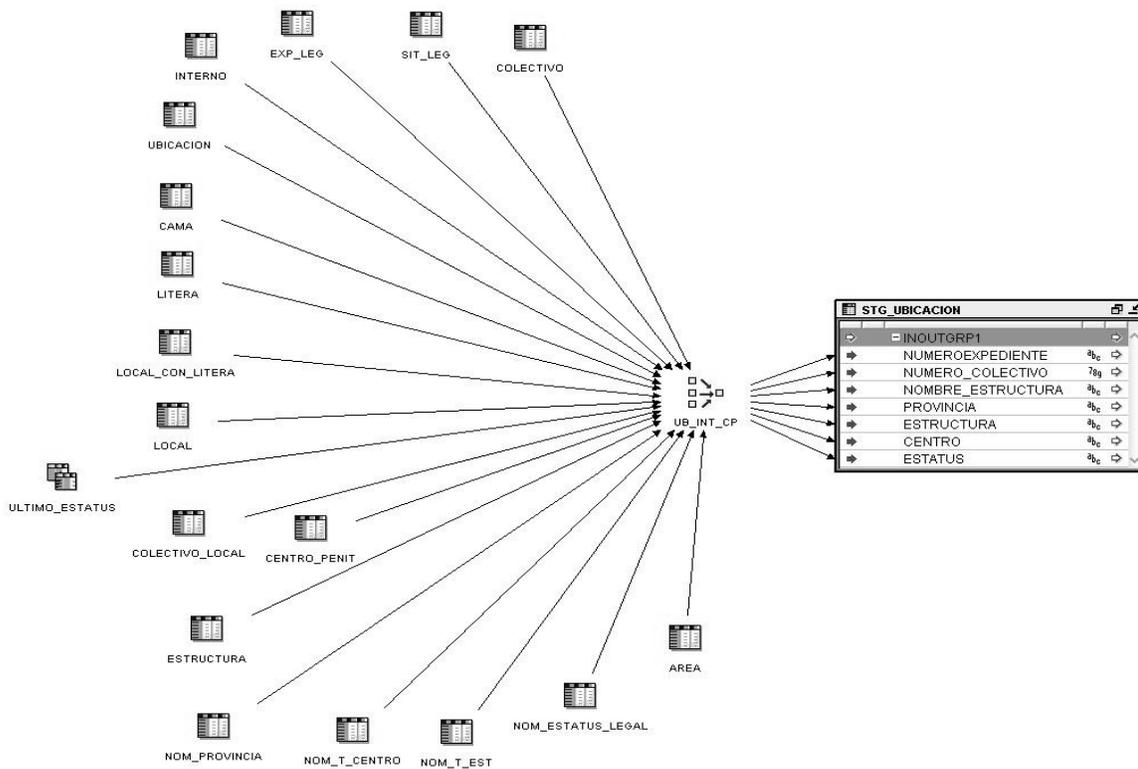


Figura 18: Procesos ETL del STG Ubicación

Para la carga del stg_ubicacion se toma como fuente de entrada los datos necesarios que se encuentran en las tablas: colectivo,sit_leg, exp_leg, interno, ubicación, cama, litera, local_con_litera, local, colectivo_local, centro_penit, estructura, nom_provincia, nom_t_centro, nom_t_est, nom_estatus_legal y area.

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

- STG Asistencia biblioteca

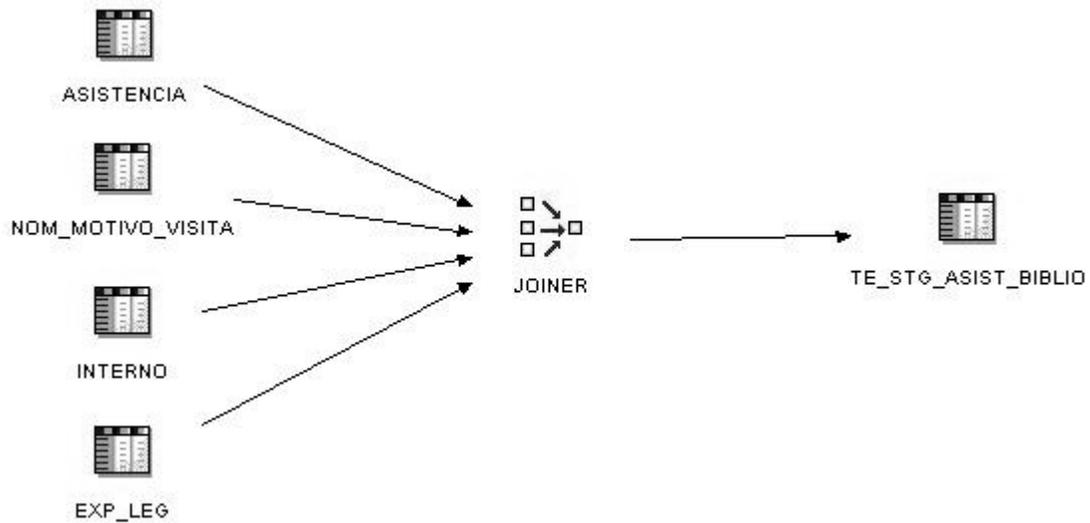


Figura 19: Procesos ETL del STG Asistencia biblioteca

A continuación se especifican los pasos para cargar el STG_ASIST_BIBLIO:

- Se toma como fuente de entrada los datos necesarios que se encuentren en las tablas: ASISTENCIA, NOM_MOTIVO_VISITA, INTERNO y EXP_LEG.
- Se procede a utilizar el operador Joiner, es el encargado de unir varios conjuntos de filas de diferentes fuentes con diferentes cardinalidades y produce una salida. En este caso es el encargado de unir las tablas antes mencionadas, para producir la salida encargada de poblar el STG_ASIST_BIBLIO.

• **ETL Hecho Asistencia biblioteca**

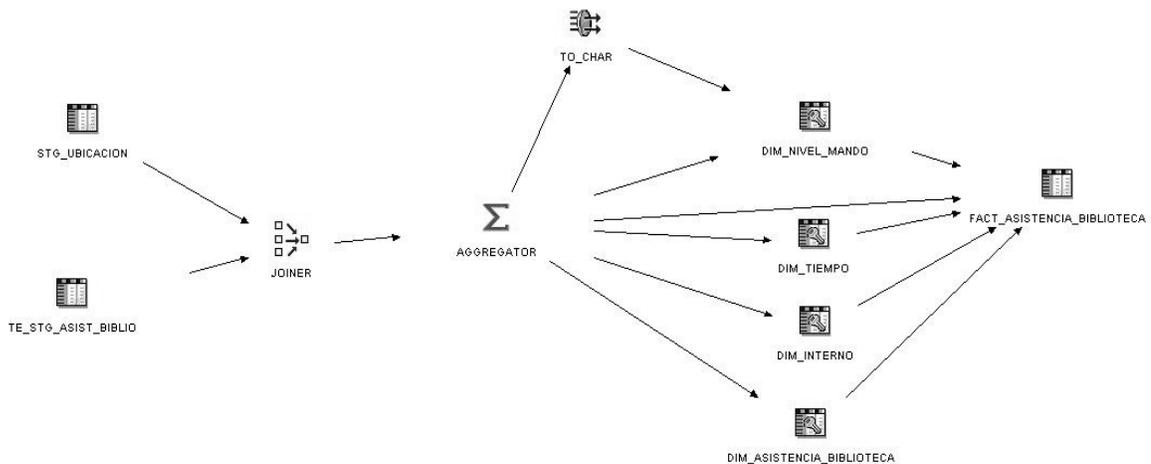


Figura 20: Procesos ETL del hecho Asistencia biblioteca

A continuación se especifican los pasos para cargar la FACT_ASISTENCIA_BIBLIOTECA:

- Se utiliza como datos de entrada el STG_ASIST_BIBLIO y el STG_UBICACION.
- Se procede a utilizar el operador Joiner para unir las tablas por el atributo NUMEROEXPEDIENTE que ambas poseen.
- Los datos de salida del Joiner se toman como datos de entrada en el operador Agregator, para luego utilizar la función COUNT y contar las veces que un mismo interno visita la biblioteca en un mismo día por el mismo motivo y guardar este valor en el campo Cantidad.
- Se utiliza el operador de transformación para convertir el campo NUMERO_COLECTIVO del tipo de dato NUMERIC a VARCHAR.
- El operador Key_Lookup(Consulta) se utiliza para leer los datos de las dimensiones correspondientes a través de su clave o llave. En este caso se obtienen como salida los datos de las dimensiones interno, tiempo, nivel de mando y asistencia biblioteca, cargándose de esta manera los datos en la FACT_ASISTENCIA_BIBLIOTECA.

Configuración de la herramienta de Administración

Una vez cargados todos los cubos pertenecientes al mercado de datos se configura el repositorio que relaciona las herramientas de consulta con los datos del mercado. En este paso se utiliza la herramienta de administración (Administration Tool) que brinda la suite ORACLE 11g.

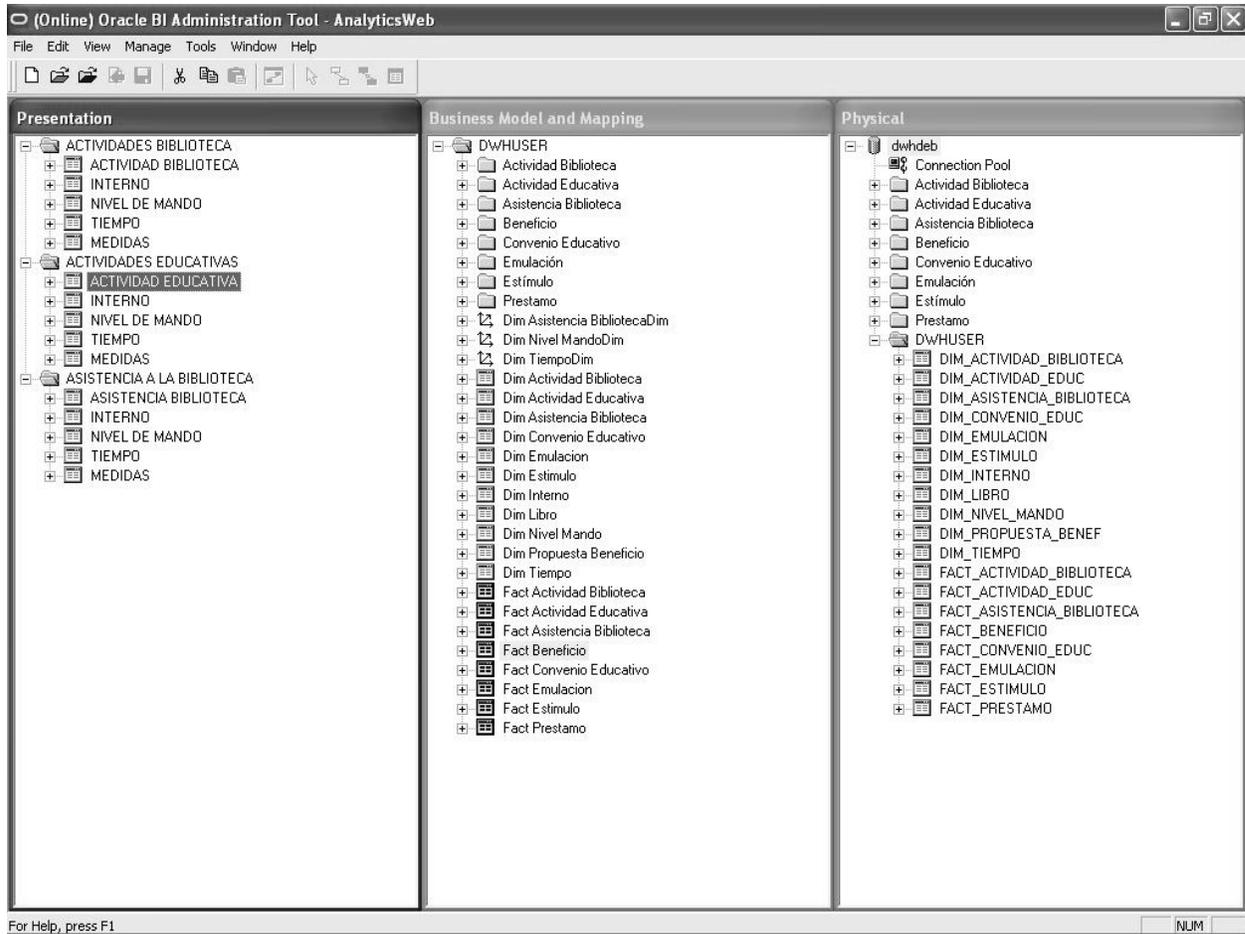


Figura 21: Configuración de la herramienta Administration Tool del OBI

El repositorio creado está conformado por tres capas: la capa física, modelo de negocio y mapeo, y la capa de presentación. Para la creación del repositorio primeramente se importan las tablas de los cubos pertenecientes al mercado de datos en la capa física, definiéndose las restricciones de llaves primarias y llaves foráneas que permiten las relaciones entre las tablas de hechos y sus respectivas tablas dimensión. En el modelo de negocio o capa intermedia se especifican las correspondencias entre el modelo de negocio y los esquemas de la capa física. Se define el modelado de los datos del negocio donde se le asigna a las fuentes físicas el vocabulario de negocio de los usuarios. Además se definen las jerarquías de las dimensiones y las medidas con sus respectivas reglas de agregaciones. Por último, se construye la capa de presentación, en la cual se propone como serán vistos los datos por los usuarios finales a través de las herramientas de consulta, permitiendo que se organicen las tablas por áreas temáticas, se renombren las tablas y columnas por valores representativos del negocio

Herramientas de inteligencia de negocio

Se definieron dos usuarios finales que consultarán los datos del mercado de datos del área de Tratamiento Educativo y un usuario administrador del mercado de datos. Cada usuario tendrá distintos privilegios administrativos sobre el mercado de datos por lo que se dividió la herramienta de respuesta en dos niveles administrativos, para el especialista se definió el tablero de mando donde se muestran los reportes predefinidos y para el oficial la herramienta de consultas ad-hoc.

Herramienta para el usuario final (Tablero de mando)

Esta herramienta es la que permite mostrarle al Especialista del área de Tratamiento Educativo con menos privilegios administrativos sobre el mercado de datos, los reportes operativos e históricos previamente construidos por el administrador del sistema y el oficial del área de Tratamiento Educativo.



Figura 22: Herramienta Tablero de mando (Interactive Dashboards)

Herramienta de consultas ad-hoc

Esta herramienta además de mostrarle al Oficial del área de Tratamiento Educativo los reportes operativos e históricos previamente construidos por el administrador del sistema o por él mismo, permite que este pueda crear nuevos reportes así como consultas a la base de datos en la pestaña "Respuestas", nótese que el usuario especialista no tiene permisos para esta funcionalidad.



Figura 23: Herramienta de consultas ad-hoc (Answers)

El usuario oficial puede crear o modificar reportes, añadiendo los criterios y peticiones de datos que satisfagan los nuevos requisitos de información que necesite sobre un cubo determinado como se muestra en la siguiente imagen:



Figura 24: Creación de reportes en la herramienta Answers

Tipos de procesamiento de la información:

El almacén de datos del área de Tratamiento Educativo permite varios tipos de procesamiento de la información que contiene. Una de las características del acceso a la información que brinda es la posibilidad de representarla a diferentes niveles de agregación. Esto se logra mediante las características conocidas como Drill Down y Roll Up.³ Drill Down explora los hechos hacia los niveles más detallados de la jerarquía de dimensiones, Roll Up explora los hechos iterativamente hacia el nivel más alto de agregación. Desde el atributo Provincia se puede hacer Drill Down hasta llegar al nivel más bajo que es Colectivo, pasando previamente por el nivel intermedio que es Centro.

³ Estas son técnicas para navegar a través de distintos niveles de detalle de una jerarquía de datos, desde los de mayor nivel de agregación (también llamados datos sumariados) hasta los más detallados.



Figura 25: Procesamiento de la información utilizando Drill Down

Otra forma de procesar la información es mediante la técnica de Slice & Dice (Rebanar los cubos): Estos términos son utilizados para describir como los datos multidimensionales pueden ser mostrados aplicando diferentes filtros a los cubos de datos, los filtros creados permiten filtrar los datos por Año, Trimestre, Mes del año, Provincia, Centro, Tipo de centro y Tipo de estructura.

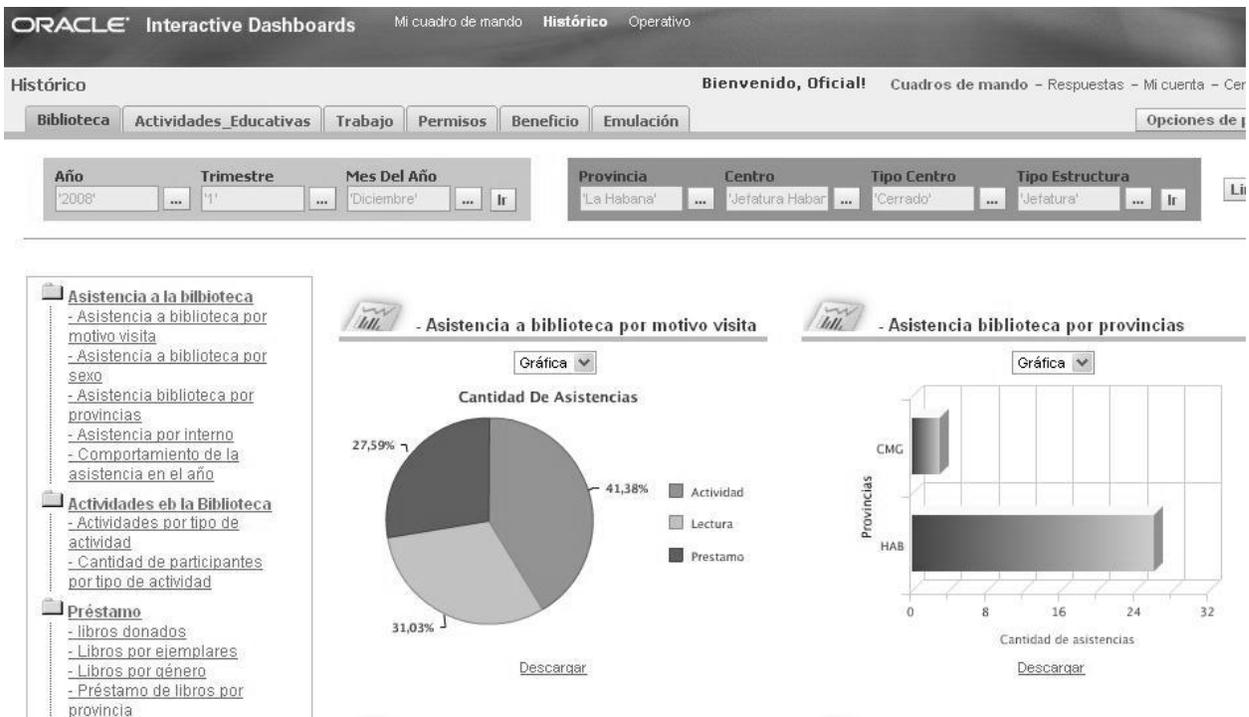


Figura 26: Procesamiento de la información utilizando Slice & Dice

Pruebas

Las pruebas son los procesos que se le realizan al producto para verificar su correcto funcionamiento y calidad. Las pruebas demuestran hasta qué punto las funciones del software parecen funcionar de acuerdo con las especificaciones y parecen alcanzarse los requisitos del negocio.

Como técnica de prueba utilizamos la establecida por la metodología que plantea comprobar la lógica de la aplicación, primeramente creando una consulta sencilla y verificar el resultado manualmente, luego comprobar que los resultados son los correctos a través de la aplicación. Cambiar los parámetros y volver a realizar el reporte. Ejecutar la consulta desde otro equipo e introducir datos erróneos para esos parámetros, una vez que la herramienta está funcionando correctamente y es probado por otra persona se verifican los datos apoyándose en la lista de chequeo propuesta por la metodología.

Lista de chequeo: Las listas de chequeo se realizan con el objetivo de evaluar la calidad de los artefactos que se generan en el análisis de la solución. Mediante estas listas se pueden cubrir todos los aspectos aplicables en los temas de análisis. La lista de chequeo propuesta por la metodología tiene como objetivo general evaluar el proceso de análisis de los datos contenidos en el mercado de datos a través de técnicas OLAP, permitiendo recoger los puntos eficientes e ineficientes que posea dicho proceso.

Peso	Criterio de evaluación	Eval	NP	Cantidad de elementos afectados	Comentarios
Crítico	¿Los subtítulos de las filas y las columnas corresponden a la descripción de sus atributos?				
Crítico	¿Los reportes son configurables a través de la interfaz del sistema?				
	¿La interfaz está orientada a facilitar el uso de las funciones del sistema por parte de los usuarios?				
Crítico	¿Los usuarios son capaces de manipular los resultados de manera que se ajusten a sus necesidades, conformando nuevos reportes?				
Crítico	¿El sistema muestra las funcionalidades diferentes de acuerdo al rol de los usuarios?				
Crítico	¿Se puede navegar a través de los niveles de granularidad de los reportes?				
Crítico	¿Los procesos pueden ser analizados desde todas sus dimensiones?				
Crítico	¿El título de la gráfica se corresponde con los datos mostrados?				
	¿Los elementos de la tabla siguen el mismo formato de letra?				

Crítico	¿Los parámetros de los requisitos se muestran en los reportes?				
	¿La interfaz tiene errores ortográficos?				
	¿El tipo de gráfica es el más representativo para mostrar la información?				

Tabla 3: Lista de chequeo

Resultados de las pruebas

Se realizaron 2 iteraciones detectándose 46 no conformidades en la primera iteración de ellas 26 altas, 13 medias y 7 bajas. En la segunda iteración se detectaron 17 no conformidades de ellas 7 altas, 6 medias y 4 bajas.

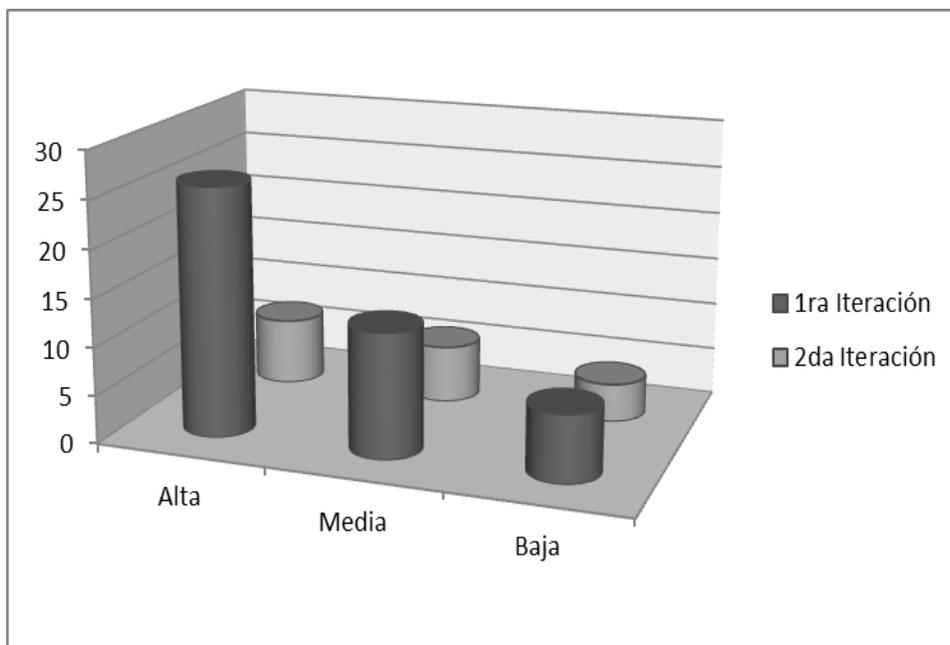


Figura 27: Resultado de las pruebas

Conclusiones Parciales:

En este capítulo se implementó el proceso de extracción, transformación y carga para todos los procesos del mercado de datos del área de Tratamiento Educativo, se configuró el repositorio de datos permitiendo a las herramientas de consulta el acceso a los datos. Las herramientas de consultas se conformaron por la herramienta de construcción de reportes y el tablero de mando como aplicación de usuario final. Además se realizaron las pruebas como parte de la validación de la solución apoyándose en la lista de chequeo para detectar los de errores.

Conclusiones Generales:

- Se desarrolló la herramienta de consultas ad-hoc del mercado de datos Tratamiento Educativo que permite a los usuarios la construcción de nuevas solicitudes o reportes.
- Se desarrolló el mercado de datos del área de Tratamiento Educativo como una herramienta de apoyo a la toma de decisiones.

Recomendaciones:

- Utilización de la información generada y la documentación para proyectos futuros similares, siempre cumpliendo con las normas y políticas de confidencialidad requeridos.
- Tomar como referencia la presente investigación con el objetivo de realizar un trabajo en los distintos organismos del país para capacitar a los clientes sobre las facilidades y ventajas de los Almacenes de Datos.
- Realizar la integración de otras fuentes de datos que complementen la información del Mercado de datos.
- Incentivar en la Universidad de las Ciencias Informáticas las investigaciones referentes al proceso de extracción, transformación y carga de los datos e Inteligencia de Negocios.
- Realizar un estudio acerca de técnicas de optimización, que puedan ser aplicadas al proceso de extracción, transformación y carga e Inteligencia de Negocios desarrollado.

Bibliografía

1. **Del Risco Batista, Yanet.** *Proyecto Técnico.* 2009.
2. **Llull Céspedes, Krysia y Méndez Gutiérrez, Yaimara.** *Sistema de Apoyo a la Toma de Decisiones del Sistema Penitenciario Cubano.* La Habana, Cuba: s.n., 2011.
3. **Rodríguez Vargas, Carmen Leisa.** *Análisis y Diseño de un Data Mart para los procesos de Sistema Educativo del Sistema Penitenciario Cubano.* La Habana, Cuba: s.n., 2011.
4. **Pedraza Fernández, Libisdey y Ramos Hernández, Alejandro.** *Diseño de un Almacén de Datos para apoyar la toma de decisiones en el Centro de Informática Médica.* La Habana, Cuba: s.n., 2010.
5. **Ing. Bernabeu, Ricardo Dario.** 2007. *Data Warehousing.* Hefesto. Córdoba, Argentina : s.n., 2010.
6. **De los Ángeles, María.** *Procesamiento Analítico en Línea.* Universidad Nacional del Nordeste. Trabajo de Adscripción. LU:30-505., 2006.
7. **Kimball, Ralph.** El juego de Herramientas del Almacén de Datos. New York: John Wiley y Sons.
Gloria Wolff, Carmen. Modelamiento Multidimensional
8. **Díaz Villanueva, Vladimiro.** Almacenes de Datos. Universidad de Valencia
9. **Casales Cabrera, María Evelia.** Almacenes de Datos. 2009
10. **Blasco Domínguez, Antonio.** Data Warehousing. Programación en Castellano. [En línea] [Citado el: 30 de noviembre de 2011.] http://www.programacion.com/articulo/data_warehousing_201/4.
11. **Kimball, Ralph y Ross, Margy.** Relentlessly Practical Tools for Data Warehousing and Business Intelligence. Published by Wiley Publishing, Inc., Indianapolis, Indiana, 2010.
12. Durán Toro, A. Ruiz Cortés, R. Corchuelo Gil y M. Toro Bonilla. *Identificación de Patrones de Reutilización de Requisitos de Sistemas de Información.* Facultad de Informática y Estadísticas, Universidad de Sevilla.

Referencias bibliográficas:

1. Carlos Tünnermann Bernhein: 10 de mayo de 1933, Managua, Nicaragua. Abogado y educador. Marilena De Souza Chaui: 4 de septiembre de 1941, São Paulo, Brasil. Filósofa y educadora.
2. Inmon, W. (1996). Building the Data Warehouse (2nd edition). New York: John Wiley & Sons, Inc.
3. 20 de julio, 1945, San Diego, California, USA: Es conocido como el padre del Data Warehouse (Almacén de Datos). Defiende una metodología descendente (top-down) a la hora de diseñar un almacén de datos, ya que de esta forma se considerarán mejor todos los datos corporativos.
4. Bressán, Griselda E. "Objetivos del Data Warehousing". <http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MineriaDatosBressan.htm#Objetivos%20del%20Data%20Warehouse>.
5. Business Intelligence + Informática estratégica. [En línea] [Citado el: 30 de noviembre de 2011.] http://www.sinnexus.com/business_intelligence/datamart.aspx.
6. Libro El Juego de Herramientas del Almacén de Datos, Ralph Kimball, editorial John Wiley & Sons. 1996, Traducción: M.Sc. Rosendo de Jesús Moreno Rodríguez
7. Bernabeu, Ricardo Dario. DATA WAREHOUSING: Investigación y Sistematización de Conceptos - HEFESTO: Metodología propia para la Construcción de un Data Warehouse. V2.0. 2010.
8. En el año 1998 dicha metodología se recoge como proceso a seguir en el desarrollo de un DWH con el libro: "*The Data Warehouse Lifecycle Toolkit*", editorial: Wiley, fecha de publicación: 13 de agosto, 1998.
9. Oracle Business Intelligence. [Online] http://www.neuronet.cl/egroupware/sitemgr/sitemgr-site/?page_name=OBI_ONE.
10. Robert E. Shannon (1988)

Anexos**Preguntas de las entrevistas****TEMA: RESPONSABILIDADES**

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

TEMA: ASUNTOS Y OBJETIVOS DE NEGOCIO

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

TEMA: ANÁLISIS DE LOS REQUERIMIENTOS

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