

**UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS  
FACULTAD 2**



**Análisis y Diseño de un Data Mart para los procesos de  
Sistema Educativo del Sistema Penitenciario Cubano.**

**TRABAJO DE DIPLOMA PARA OPTAR POR EL TÍTULO DE  
INGENIERO EN CIENCIAS INFORMÁTICAS**

**Autor:** Carmen Leisa Rodríguez Vargas.

**Tutor:** Msc. Julio Omar Prieto Entenza.

La Habana, junio de 2011

“Año 53 de la Revolución



*"La tarea que los graduados de la UCI tienen por delante es grandiosa. Espero que la cumplan y la cumplirán".*

*Fidel Castro Ruz*

*Reflexión " El Robo de Cerebros " .17 de julio de 2007*



## ***Declaración de Autoría***

### **Declaración de Autoría**

Por este medio declaro ser la única autora de este trabajo y autorizo a la Universidad de las Ciencias Informáticas para que haga el uso que estime pertinente con este trabajo.

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

Carmen Leisa Rodríguez Vargas

Julio Omar Prieto Entenza

---

Firma del Autor

---

Firma del Tutor



**Datos de Contacto**

**Síntesis del Tutor:** Msc. Julio Omar Prieto Entenza.

**Correo Electrónico:** jprieto@uci.cu

**Categoría Científica:** Máster.

**Categoría Docente:** Instructor.

- Graduado de Ingeniero en Ciencias Informáticas en el año 2007.
- Ha participado en eventos como UCIENCIA e Informática 2009.
- Ha dirigido varias Tesis de Diploma.
- Vinculado a proyectos productivos y de investigación como desarrollador y arquitecto.



## **Dedicatoria**

*Este trabajo es resultado del esfuerzo y la motivación por superarme y poder regalarle este título a todos los que me han ayudado en mi formación.*

*Quiero dedicar especialmente cada una de las palabras que acá leerán a mis padres, mis luces e inspiración, a mis abuelas por haberme regalado dos padres maravillosos y ser tan especiales, a mi familia, a mis mejores amigos y al Comandante Fidel.*



## Agradecimientos

*En busca de un sueño dejé atrás ciudades y aulas hasta llegar a la UCA. En busca de ese sueño he estado siempre acompañada de personas que han marcado mi vida y han hecho posible ver este anhelo realizado.*

*Agradezco desde lo más profundo de mi corazón:*

*A mis padres María del Carmen y Julio Rafael, Tota y Titu, los dos amores más lindos que he tenido en mi vida, las personas más tiernas, inteligentes y especiales que me han guiado por el camino del bien, la verdad y la justicia. Por vivir conmigo los momentos más bellos, por su confianza y la seguridad de que un día llegaríamos hasta aquí. Por hacerme una persona feliz.*

*A mi mami en especial la gratitud por convertirme en una persona perseverante, por su entrega desmedida y su amor insaciable. Por su fervor revolucionario a Fidel.*

*A mi papi por todo su amor desde las primeras lecturas y juegos, por quererme con tanta pasión y vivir orgulloso de mí; yo también lo estoy de ti.*

*A mis tres abuelas María, Carmen y Adelfa, que han llenado de colores mi vida, por tanto amor, por haberlo compartido todo conmigo y estar pendientes de mí.*

*A mi abuelito Vargas, con quien tanto jugué dominó y siempre me hizo saber cuán especial era para él.*

*A mis tías, a Cefe, a mis primos y a Yami por las muestras de cariño.*



## Agradecimientos

*A mis amigos Marelis y Troya, y a todos los muchachos que hicieron de mis días en la Vocacional de Holguín la mejor etapa.*

*A Jenny, mi fiel amiga con quien he pasado momentos inolvidables.*

*A Zule y Pedro por su amistad sin límites y considerarme especial en sus vidas.*

*A Gali por su incondicionalidad cuando más lo necesité, por ser una amiga especial, por el cariño que me transmite y haberme hecho parte de su familia.*

*A Rivi, Dayi, Tito, Maruis y Yei por tantos años y momentos juntos.*

*A Yordany por su sencillez y su bondad.*

*A todos los profes con los cuales compartí.*

*A mi tutor Julio por su confianza y seguridad.*

*A los muchachos y profesores del proyecto, en especial a Yanet, Kraysia y Leandro por la ayuda en el desarrollo de este trabajo.*

*A los colegas de la FEM y la WJC, en especial al profe Silvano por apoyarme en tantas aventuras.*

*A Guillermo Cabrera, quien tecla en mi corazón las más grandes ocurrencias.*

*A todos los que me ayudaron en la realización de este trabajo.*

*A mis compañeros de grupo y a todos los que me han regalado parte de su vida.*

*Muchas gracias, los quiero*

*Carmen Leisa.*



## **RESUMEN**

Entre los procesos y actividades que desarrolla el Sistema Penitenciario Cubano, la toma adecuada de decisiones y el análisis de la información es primordial. En estos momentos, esta entidad del Estado no cuenta con una herramienta informática disponible para apoyar tales análisis.

El presente trabajo realiza una investigación acerca de los Almacenes de Datos, analiza los procesos principales que se realizan en el área de Sistema Educativo del Sistema Penitenciario Cubano y diseña un Data Mart como herramienta para la toma de decisiones.

La metodología Hefesto es la utilizada para diseñar el Data Mart. Su construcción se basa en el enfoque de Ralph Kimball de un modelo dimensional y en una arquitectura de tres capas. Se utiliza el Gestor de Base de Datos Oracle 11g y el Oracle Warehouse Builder como herramienta de desarrollo.

## **PALABRAS CLAVES**

Almacén de Datos, Data Mart, base de datos, decisiones, indicadores.



## ÍNDICE

INTRODUCCIÓN.....	1
CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA.....	4
1.1 Tendencias actuales en los Sistemas Penitenciarios.....	4
1.2 Almacenes de Datos.....	6
1.2.1 Características de un Almacén de Datos.....	7
1.2.2 Data Mart.....	8
1.2.3 Diferencias entre OLTP y Almacenes de Datos.....	9
1.2.4 Estructura de un Almacén de Datos.....	9
1.3 Base de Datos Multidimensional.....	10
1.3.1 Esquema en estrella.....	11
1.3.2 Esquema copo de nieve.....	11
1.3.3 Esquema constelación.....	11
1.4 Cubos multidimensionales.....	12
1.4.1 Cubos OLAP.....	12
1.5 Arquitectura de un Almacén de Datos.....	13
1.5.1 Arquitectura conceptual de los datos.....	13
1.5.2 Elementos de una arquitectura para el Almacén de Datos.....	15
1.6 Modelado de los datos.....	16
1.7 Herramientas de creación de un Almacén de Datos.....	17
1.7.1 Sistema de Gestión de Base de Datos Oracle.....	17
1.7.2 Oracle Warehouse Builder (OWB). Características.....	18
1.8 Metodologías para la construcción de un Almacén de Datos.....	18
1.8.1 Metodología DWEP. Características.....	18
1.8.2 Metodología Hefesto. Características.....	19
1.8.3 Pasos para la aplicación de la metodología propuesta.....	20
1.9 Conclusiones.....	20
CAPÍTULO 2: ANÁLISIS DEL DATA MART.....	22
2.1 Análisis de los procesos.....	22
2.2 Análisis de los requerimientos.....	25
2.2.1 Indicadores y perspectivas.....	25
2.2.2 Modelo Conceptual.....	33
2.2.3 Conformación de indicadores.....	36



2.3 Conclusiones .....	40
CAPÍTULO 3: DISEÑO DEL DATA MART .....	41
3.1 Modelo Lógico del Data Mart .....	41
3.1.2 Tablas de dimensiones.....	41
3.1.3 Tablas de hechos .....	42
3.2 Diseño de los procesos de Extracción, Transformación y Carga (ETL).....	45
3.2.1 Procesos ETL.....	45
3.2.2 Procesos ETL para cargar las dimensiones.....	46
3.2.3 Procesos ETL para las tablas de hechos.....	51
3.2.4 Carga inicial.....	57
3.3 Conclusiones .....	60
CONCLUSIONES GENERALES.....	61
RECOMENDACIONES .....	62
REFERENCIAS BIBLIOGRÁFICAS .....	63
BIBLIOGRAFÍA.....	65



**ÍNDICE DE FIGURAS**

**Figura 1:** Arquitectura de datos de Una sola capa..... 14

**Figura 2:** Arquitectura de datos de Dos Capas..... 14

**Figura 3:** Arquitectura de datos de Tres Capas..... 15

**Figura 4:** Modelo Conceptual Proceso Trabajo. .... 34

**Figura 5:** Modelo Conceptual Proceso Características Socio – Antropológicas..... 35

**Figura 6:** Modelo Conceptual Proceso Actividades Educativas..... 36

**Figura 7:** Tablas de dimensiones del Data Mart..... 41

**Figura 8:** Tablas de hechos del Data Mart..... 42

**Figura 9:** Tabla de hechos FACT\_TRABAJO..... 43

**Figura 10:** Tabla de hechos FACT\_CARACT\_SOCIO-ANTROP..... 44

**Figura 11:** Tabla de hechos FACT\_ACTIVIDADES\_EDUCATIVAS..... 44

**Figura 12:** Proceso ETL..... 46

**Figura 13:** ETL DIM\_INTERNO..... 47

**Figura 14:** ETL DIM\_CENTRO..... 48

**Figura 15:** ETL DIM\_TIEMPO..... 49

**Figura 16:** ETL DIM\_UNIDAD\_PRODUCTIVA..... 50

**Figura 17:** ETL DIM\_ACTIVIDAD..... 51

**Figura 18:** ETL STG\_TRABAJO..... 52

**Figura 19:** ETL FACT\_TRABAJO..... 53

**Figura 20:** ETL STG\_CARAC\_SOCIO\_ANTRO..... 54

**Figura 21:** ETL FACT\_CARAC\_SOCIO\_ANTRO..... 55

**Figura 22:** ETL STG\_ACTIVIDADES\_EDUCATIVAS..... 56

**Figura 23:** ETL FACT\_ACTIVIDADES\_EDUCATIVAS..... 57

**Figura 24:** Flujo de Procesos para la Carga Inicial..... 58

**Figura 25:** Flujo de procesos para cargar las dimensiones comunes..... 58

**Figura 26:** Flujo de Procesos para cargar FACT\_TRABAJO..... 59

**Figura 27:** Flujo de Procesos para cargar FACT\_CARACT\_SOCIO\_ANTRO..... 59

**Figura 28:** Flujo de Procesos para cargar FACT\_ACTIVIDADES\_EDUCATIVAS..... 60



**ÍNDICE DE TABLAS**

<b>Tabla 1:</b> Perspectivas del Data Mart.....	26
<b>Tabla 2:</b> Análisis OLTP Perspectiva Tiempo.....	27
<b>Tabla 3:</b> Análisis OLTP Perspectiva Interno. ....	28
<b>Tabla 4:</b> Indicadores y Perspectivas Proceso Trabajo. ....	30
<b>Tabla 5:</b> Indicadores y Perspectivas Proceso Características Socio - Antropológicas. ....	32
<b>Tabla 6:</b> Indicadores y Perspectivas Proceso Actividades Educativas.....	33



### INTRODUCCIÓN

El paso de los años ha traído numerosos avances en el modo de vida actual, dada la rapidez con que se asumen y afrontan cada día los retos. Las Tecnologías de la Información y las Comunicaciones (TIC) han asumido en estos cambios un papel protagónico, debido a la importancia que se les están conferidas por su gran aporte al conocimiento y ser precursoras de la información.

Ante este vertiginoso desarrollo, las personas y empresas no han podido quedar inmóviles, sino que han encontrado en la información que manejan un valioso recurso, que de utilizarse correctamente puede guiarlos por la clave del éxito, o por el contrario, conducirlos al fracaso.

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 TIC, se han logrado importantes mejoras, como la automatización de los procesos operativos y el contar con una plataforma de información necesaria para la toma de decisiones.

Los Sistemas de Soporte a las Decisiones (DSS por sus siglas en inglés), son los encargados de que en las empresas se tengan en cuenta también actividades como las de previsión, planificación y administración, vitales para el desarrollo de los negocios en el presente y futuro. Para ello es necesario un adecuado análisis de los datos y una coherente toma de decisiones.(1)

Por la importancia que esto tiene en el ámbito empresarial, se han desarrollado herramientas que se basan en estos sistemas para operar, tal es el caso de 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)(2). 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 del Departamento de Establecimientos Penitenciarios (SAIDEP) y Sistema Automatizado de Capacidades del Departamento de Establecimientos Penitenciarios (SACDEP).

Actualmente surge la idea de implementar un único sistema informático, que abarque todos los procesos que convergen dentro del Sistema Penitenciario en Cuba, uniendo toda la información que generan los tres sistemas informáticos que están siendo explotados hasta el momento y agregando



nuevas funcionalidades que hoy todavía no están automatizadas, como son algunos de los procesos que se realizan en el área de Sistema Educativo.

El área de Sistema Educativo del Sistema Penitenciario Cubano es la encargada de tomar las decisiones sobre la reinserción de los privados de libertad. La misma abarca distintos procesos como son: Comportamiento Socio-Antropológico, Instrucción Escolar, Superación Técnica, Trabajo y Actividades Educativas.

A los directivos de esta área le resulta engorroso consultar cualquier información que necesite ser analizada por encontrarse la misma descentralizada y en su mayoría ser procesada de forma manual.

No es posible analizar el comportamiento existente en la población interna respecto a características como la edad, el color de la piel, el estado civil, el nivel cultural, etc. Tampoco se pueden tomar decisiones referentes al desempeño de los internos al incorporarse a la Instrucción Escolar, la Superación Técnica y el Trabajo. A su vez no se lleva un registro de la participación de los internos en las distintas actividades educativas que se realizan, lo cual afecta cualquier información que se pueda brindar o decisión a tomar.

Todo lo anterior se dificulta por no contar con un mecanismo eficiente, capaz de darle seguimiento a los objetivos estratégicos de manera que apoye la toma de decisiones a todos los niveles y permita tener un control efectivo de todas las actividades y procesos que se desarrollan.

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

Se define como **objeto de estudio** la toma de decisiones del Sistema Penitenciario Cubano, enfocándose como **campo de acción** la gestión de la información para la toma de decisiones relacionadas con el área de Sistema Educativo.

Se plantea como **objetivo general**: Realizar el Análisis y Diseño de un Data Mart para la toma de decisiones en el área de Sistema Educativo del Sistema Penitenciario Cubano. El mismo se divide en los siguientes **objetivos específicos**:

- Elaborar el marco teórico de la investigación.
- Seleccionar la metodología apropiada para la construcción de los Almacenes de Datos.
- Establecer los requisitos del área Sistema Educativo para el diseño del Data Mart.
- Definir la arquitectura del sistema.
- Diseñar el Data Mart.



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

1. Estudio del arte sobre los Almacenes de Datos como herramientas para la toma decisiones.
2. Selección de las metodologías y herramientas.
3. Levantamiento de los requisitos del Subsistema Sistema Educativo para el diseño del Data Mart.
4. Análisis y definición de la arquitectura del sistema.
5. Diseño del modelo conceptual del Data Mart.
6. Diseño del modelo lógico del Data Mart.
7. Diseño del modelo físico del Data Mart.
8. Diseño de los procesos de Extracción, Transformación y Carga (ETL).

El presente documento cuenta de 3 capítulos:

### **Capítulo 1: “Fundamentación Teórica”.**

En este capítulo se realiza un estudio sobre los sistemas informáticos implantados en Centros Penitenciarios de varios países. Se profundiza en los conceptos, objetivos, características y principales enfoques de los Almacenes de Datos. Se fundamentan además las herramientas y metodologías a utilizar para el desarrollo del presente trabajo.

### **Capítulo 2: “Análisis del Almacén de Datos”.**

En este capítulo se analizan los procesos principales que se desarrollan en el área de Sistema Educativo, identificando los requisitos de los usuarios, indicadores y perspectivas y definiendo el modelo conceptual.

### **Capítulo 3: “Diseño del Almacén de Datos”.**

En este capítulo se define el modelo lógico del Data Mart, se diseñan las tablas de dimensiones y de hechos y se realizan los procesos de Extracción, Transformación y Carga (ETL).



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

En este capítulo se realiza un estudio sobre los sistemas informáticos implantados en Centros Penitenciarios de varios países. Se profundiza en los conceptos, objetivos y características de los Almacenes de Datos. Se fundamentan además las herramientas y metodologías a utilizar para el desarrollo del presente trabajo.

### **1.1 Tendencias actuales en los Sistemas Penitenciarios.**

Como parte del desarrollo tecnológico, son varios los países que cuentan con sistemas informáticos capaces de apoyar la toma de decisiones y la gestión de los procesos que se realizan en los Centros Penitenciarios.

- **Sistema Penitenciario de España.**

España, por ejemplo, cuenta con el EPVNET (Establecimiento Penitenciario Virtual en Red), un proyecto de software libre para la gestión informatizada de los Centros Penitenciarios. Esta aplicación iniciada en el 2002, se convirtió en el 2004 en un proyecto de software libre bajo la licencia GNU/GPL y desde el año 2007 funciona en el Establecimiento Penitenciario de Valencia y en otros centros de la región. (3)

- **Sistema Penitenciario de Panamá.**

En Panamá existe también un Sistema Penitenciario creado a principios del año 1997, como resultado de un proyecto financiado por las Naciones Unidas y el gobierno español, en coordinación con la Dirección General de Sistemas Penitenciarios de Panamá (DGSP). Esta aplicación permite mantener almacenado en una base de datos los registros de los internos que están detenidos en los centros penales a nivel nacional.(4)

- **Sistema de Gestión Penitenciaria de Ecuador.**

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

Estos tres sistemas aunque apoyan el análisis de la información y la toma de decisiones en cada nivel, presentan la desventaja de que sus procesos fundamentales están sujetos a leyes y particularidades



específicas de cada país. Además la información es procesada en bases de datos relacionales viéndose limitado el número de análisis que se pudieran obtener.

- **Sistema de Gestión Penitenciaria de Venezuela.**

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 del Ministerio del Interior y Justicia y de la Dirección General de Custodia y Rehabilitación del Recluso.(6)

La implantación del sistema y el correspondiente crecimiento de los datos generados trajeron como consecuencia que se viera afectado el rendimiento de la aplicación a mayor escala, al incluir reportes que precisan la búsqueda en varias tablas o varios módulos al mismo tiempo. Con el objetivo de darle una solución a este problema se crearon vistas materializadas para tratar de hacer más eficiente el rendimiento de la aplicación de manera que las consultas a la base de datos se realizaran de forma más rápida.

Esta variante no fue fructífera ya que las vistas se actualizaban sólo una vez al día y al sistema penitenciario le interesaba que la información fuera operativa, o sea, que a la hora de ser consultada se tuvieran los datos en tiempo real, y con las vistas no se lograba este objetivo al estar las mismas en muchas ocasiones desfasadas con respecto a los datos reales. Por otra parte en el momento de actualizar las vistas disminuía el rendimiento de la aplicación al demorarse la actualización de los datos.

De igual manera al establecer los requisitos funcionales se pensaron una cantidad de reportes predefinidos lo que trajo consigo una traba para la correcta toma de decisiones por parte de la administración ya que siempre que surja la necesidad de hacer un nuevo reporte los desarrolladores son los únicos capaces de darle solución a este problema.

- **Sistema Penitenciario Cubano.**

En el año 1989 comienza la automatización del Sistema Penitenciario Cubano a nivel nacional, con un sistema de control penal que permitía la gestión de los datos principales del recluso y algunos aspectos del trabajo de la especialidad.

En cumplimiento de la orden 43/99 del Vice Ministro Primero del Ministerio del Interior, como medida para mejorar la eficacia de los procedimientos dentro del Sistema Penitenciario, surge el primer sistema informático que gestiona la labor realizada en estas áreas, el Sistema Automatizado para el



Control del Recluso (SACORE). El mismo fue implementado con las tecnologías existentes en el momento (Computadoras 486, sin conectividad interna, ni externa, programación en Developer y Gestor de Base Datos Oracle).

La explotación del SACORE inició en enero de 2003 y almacenaba la información en tres módulos principales conocidos como: Control Penal, Tratamiento Educativo y Orden Interior.

Entre las principales funcionalidades que brinda están contempladas las siguientes:

- Respuestas inmediatas a las solicitudes de información de los diferentes órganos e instituciones del estado como son: Jefatura del MININT, Ministerio de Justicia, Tribunales, Fiscalías, entre otros.
- Recoge prácticamente la totalidad de la información de los reclusos en todas las especialidades.
- Los partes que se emiten son obtenidos de forma automatizada.
- Permite el traslado automático de todos los datos del recluso al nivel nacional.
- Permite salidas estadísticas de:
  - El comportamiento de las incidencias que se cometen en el sistema penitenciario.
  - Las medidas disciplinarias que se le imponen a los internos luego de haber infringido lo establecido en el reglamento penitenciario.
  - De las ocupaciones de objetos no permitidos por lo establecido en el reglamento penitenciario.

Aunque el SACORE almacena datos de 8 años atrás, no es suficiente con este sistema para un mayor aprovechamiento de la información y tomar decisiones que respondan a los intereses del Sistema Penitenciario Cubano. Es necesario hacer uso de otras herramientas, tal como han hecho los desarrolladores del Sistema de Gestión Penitenciaria de Venezuela, que por iniciativa propia para solucionar el problema presentado con las vistas materializadas, han comenzado la construcción de un Almacén de Datos para apoyar la toma de decisiones usando la Herramienta de Inteligencia de Negocio (Pentaho).

### **1.2 Almacenes de Datos.**

Los Almacenes de Datos, Data Warehouse en inglés, o depósito de datos como también se les conoce, son actualmente el centro de atención de las grandes instituciones, porque proveen un ambiente para que las organizaciones hagan un mejor uso de la información que está siendo administrada por diversas aplicaciones operacionales.



*“Un almacén de datos se crea al extraer datos desde una o más bases de datos de aplicaciones operacionales. La data extraída es transformada para eliminar inconsistencias y resumir si es necesario y luego ser cargadas en el almacén. El proceso de transformar, crear el detalle de tiempo variante, resumir y combinar los extractos de datos, ayudan a crear el ambiente para el acceso a la información Institucional. Este nuevo enfoque ayuda a las personas individuales, en todos los niveles de la empresa, a efectuar su toma de decisiones con más responsabilidad.”(7)*

Los **objetivos principales** de los Almacenes 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.
- Monitorear el comportamiento de los clientes.
- Predecir el nivel de consumo que tendrán los productos.
- Mejorar la capacidad de respuesta a las cuestiones del negocio.
- Mejorar la productividad.

En los almacenes el análisis de los datos constituye un pilar fundamental, varios de sus objetivos se centran en la efectividad que ha de tener la información para la empresa y los clientes. Las características de este depósito se basan también en el uso correcto que se ha de darle a la información.

### 1.2.1 Características de un Almacén de Datos.

William H. Inmon y Ralph Kimball (8) son figuras de referente consulta para una mayor comprensión de los Almacenes de Datos, pues desarrollaron sus propios modelos, enfoques y arquitecturas. Kimball es el principal promotor del enfoque dimensional para el diseño de los almacenes, mientras Inmon, considerado como el padre de los Data Warehouse en su libro “Building the Data Warehouse” (9) propone su construcción basándose en el modelo Entidad – Relación, por ser más adaptable y rico que el dimensional.

Precisamente es Inmon quien definió las características principales de los Almacenes de Datos, estas son (10):

**Orientado a Temas:** Una primera característica de los Almacenes de Datos es que la información es clasificada según los aspectos que son de interés para la empresa. Se excluye la información que no



es necesaria o no va a ser utilizada. Los almacenes son organizados por temas, porque esto facilita su acceso y entendimiento por parte de los usuarios finales.

**Integrado:** El aspecto más importante del ambiente del Almacén de Datos es que la información encontrada al interior esté siempre integrada. Esta integración puede verse de muchas maneras: en convenciones de nombres consistentes, en la medida uniforme de variables, en la codificación de las estructuras, en atributos físicos de los datos, en las fuentes múltiples y otros.

**Variante en el Tiempo:** La información contenida en el almacén es siempre solicitada en algún momento, de ahí que los datos almacenados sean llamados de “tiempo variante”. En un ambiente operacional la información se representa en una línea de tiempo de sesenta a noventa días, mientras en el almacén es de cinco a diez años. A su vez cada estructura clave en el Almacén de Datos contiene, implícito o explícitamente, un elemento de tiempo como día, semana, mes y año. Una vez que se registra correctamente la información en el depósito, esta no puede ser actualizada.

**No volátil:** En los Almacenes de Datos para el análisis y la toma de decisiones se requiere de una base de datos estable. En los ambientes operacionales se realizan operaciones de insertar, borrar y modificar los datos. Los Almacenes de Datos solo permiten la carga inicial de los datos y el acceso a los mismos.

Tener bien claras las características principales de los Almacenes de Datos permite una construcción coherente del depósito, estas particularidades son válidas también para cuando se desea diseñar un Data Mart.

### 1.2.2 Data Mart

*“Un Data Mart 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 Data Mart puede ser alimentado desde los datos de un Data Warehouse, o integrar por sí mismo un compendio de distintas fuentes de información”. (12)*

En el momento de crear el Data Mart de un área funcional de una empresa, es necesario definir la estructura óptima para el análisis de su información. Esta estructura puede estar tanto en una base de datos OLAP (On Line Analytical Processing), como en una OLTP (On Line Transaction Processing); la selección ha de estar en consecuencia con los requisitos y las características propias de cada organización.



La creación de los Data Mart OLAP se basa en los cubos OLAP, los cuales se construyen agregando las dimensiones e indicadores necesarios de cada cubo relacional.

A su vez los Data Mart OLTP pueden basarse en una simple extracción del Almacén de Datos, aunque lo común es introducir mejoras en su rendimiento, siendo las agregaciones y los filtrados las operaciones más usuales. Para esto se aprovechan las características particulares de cada área de la empresa.

### 1.2.3 Diferencias entre OLTP y Almacenes de Datos.

*“Los OLTP son diseñados para soportar el procesamiento de información diaria de las empresas, y el énfasis recae en maximizar la capacidad transaccional de sus datos. Su estructura es altamente normalizada, para brindar mayor eficiencia a sistemas con muchas transacciones que acceden a un pequeño número de registros y están fuertemente condicionadas por los procesos operacionales que deben soportar, para la óptima actualización de sus datos. Esta estructura es ideal para llevar a cabo el proceso transaccional diario, brindar consultas sobre los datos cargados y tomar decisiones diarias, en cambio los esquemas de DW están diseñados para poder llevar a cabo procesos de consulta y análisis para luego tomar decisiones estratégicas y tácticas de alto nivel.” (10)*

Las diferencias entre los OLTP y los Almacenes de Datos se basan en que los primeros soportan la información diaria de los negocios, maximizan la capacidad transaccional de sus datos y brindan consultas sobre los datos cargados y las decisiones diarias, mientras los almacenes son diseñados para poder llevar a cabo procesos de consulta y análisis que sirven luego para tomar decisiones importantes dentro de la organización.

Las diferencias entre los Almacenes de Datos y las fuentes OLTP quedan evidenciadas en los criterios analizados anteriormente, y a su vez, en la estructura que ambos poseen, pues en el depósito de datos es diferente a la de los sistemas operacionales.

### 1.2.4 Estructura de un Almacén de Datos.

Los Almacenes de Datos tienen una estructura distinta a las base de datos operacionales, hay niveles diferentes de esquematización y detalle que delimitan el depósito. La estructura de los mismos es la siguiente:

- **Detalle de datos actuales.**



El interés más importante radica en estos datos debido a que reflejan las ocurrencias más actuales. Son de gran volumen y se almacenan en disco, al más bajo nivel de granularidad para facilitar el acceso. Su administración suele ser costosa y compleja.

- **Detalle de datos antiguos.**

Los datos antiguos son aquellos que se almacenan sobre alguna forma de almacenamiento masivo. No son frecuentemente accedidos y se almacenan a un nivel de detalle, consistente con los datos detallados actuales. Es poco usual utilizar el disco como medio de almacenamiento.

- **Datos ligeramente resumidos.**

Los datos ligeramente resumidos son aquellos que provienen desde un bajo nivel de detalle encontrado al nivel de detalle actual. Este nivel del Almacén de Datos casi siempre se almacena en disco.

- **Datos completamente resumidos.**

Estos datos son compactos y fácilmente accesibles. A veces se encuentran en el ambiente del Almacén de Datos y en otros, fuera del límite de la tecnología que ampara al almacén.

- **Metadata.**

El componente final del Almacén de Datos es el de la metadata. De diversas maneras la metadata se sitúa en una dimensión diferente al de otros datos del almacén, esto es debido a que su contenido no es tomado directamente desde el ambiente operacional.

Otra de las peculiaridades de los Almacenes de Datos en cuanto a su estructura, es que esta es representada en un modelo dimensional, mientras en las bases de datos comunes es en una estructura bidimensional.

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

Las tablas de hechos 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 dimensionales:

- Esquema en estrella (Star Scheme).
- Esquema copo de nieve (Snowflake Scheme).
- Esquema constelación (Starflake Scheme).

### 1.3.1 Esquema en estrella.

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

Entre sus ventajas más significativas está que es el esquema más simple de interpretar, posee los mejores tiempos de respuesta, su diseño es fácil de modificar y simplifica el análisis. Sin embargo, es el menos robusto para la carga y es el más lento de construir.(10)

### 1.3.2 Esquema copo de nieve.

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

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.(10)

### 1.3.3 Esquema constelación.

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



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.(10)

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

### 1.4 Cubos multidimensionales.

Unas de las estructuras más utilizadas para representar los datos del Almacén de Datos es el cubo multidimensional, el cual convierte los datos que se encuentran en las filas y columnas, en una matriz de N dimensiones.

*“Los cubos son elementos claves en OLAP (On Line Analytical Processing), una tecnología que provee rápido acceso a datos en un Almacén de Datos (Data Warehouse). Los cubos proveen un mecanismo para buscar datos con rapidez y tiempo de respuesta uniforme independientemente de la cantidad de datos en el cubo o la complejidad del procedimiento de búsqueda.” (14)*

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

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

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

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

#### 1.4.1 Cubos OLAP

*“Los sistemas OLAP son bases de datos orientadas al procesamiento analítico. Este análisis suele implicar, generalmente, la lectura de grandes cantidades de datos para llegar a extraer algún tipo de información útil: tendencias de ventas, patrones de comportamiento de los consumidores, elaboración de informes complejo, etc. Este sistema es típico de los Data Mart.” (15)*

Los sistemas OLAP se clasifican en las siguientes categorías(16):

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

Los datos se mantienen en la base de datos relacional y las agregaciones se almacenan en tablas dentro de la base de datos donde se encuentra el Data Mart. Las ventajas de su uso están dadas



porque no se duplican los datos, brinda mucha flexibilidad y no hay limitación en el uso de dimensiones.

- **MOLAP (Multidimensional On Line Analytic Processing).**

Los datos y las agregaciones se almacenan en estructuras multidimensionales. Ofrece mejor respuesta a las consultas debido a que posee los datos calculados y no hay celdas vacías. Sin embargo tiene la desventaja que para modificar un cubo, es necesario calcularlo todo nuevamente, lo que incide directamente en la flexibilidad.

- **HOLAP (Hybrid on Line Analytic Processing).**

Constituye un sistema híbrido entre MOLAP y ROLAP, combinando estas dos implementaciones para almacenar algunos datos en un motor relacional y otros en una base de datos multidimensional. Tiene la ventaja de que sus consultas son más ágiles al utilizar motor OLAP y consumen menos espacio.

Una vez definido el esquema a implementar y la estructura para representar los datos, es prudente definir la arquitectura del Almacén de Datos, teniendo en cuenta la manera en que operan los datos en el depósito y los elementos que la integran.

### 1.5 Arquitectura de un Almacén de Datos

*“Una Arquitectura Data Warehouse (Data Warehouse Architecture - DWA) es una forma de representar la estructura total de datos, comunicación, procesamiento y presentación, que existe para los usuarios finales que disponen de una computadora dentro de la empresa.” (7)*

#### 1.5.1 Arquitectura conceptual de los datos

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

*“...la arquitectura incluye todo lo que se necesita para preparar y guardar los datos. Por otro lado, también contiene todos los recursos para distribuir la información desde el Data Warehouse. Está compuesta más allá de reglas, procedimientos y funciones que permiten al Almacén de Datos trabajar y cumplir los requisitos de la empresa... Define las normas, medidas, diseño general, y técnicas de apoyo”. (17)*

En la **Arquitectura de Datos de Una Sola Capa** la información se guarda una sola vez en el Almacén de Datos, almacenándose solo los datos de tiempo real, sobre los que actúan los sistemas



informacionales y operacionales. Esto puede traer inconvenientes, ya que ambos sistemas operan sobre el mismo conjunto de datos y en instantes cuando se precisen pueden no estar disponibles para fines operacionales por estar siendo consultados y no poder actualizarse. (Ver Figura 1).

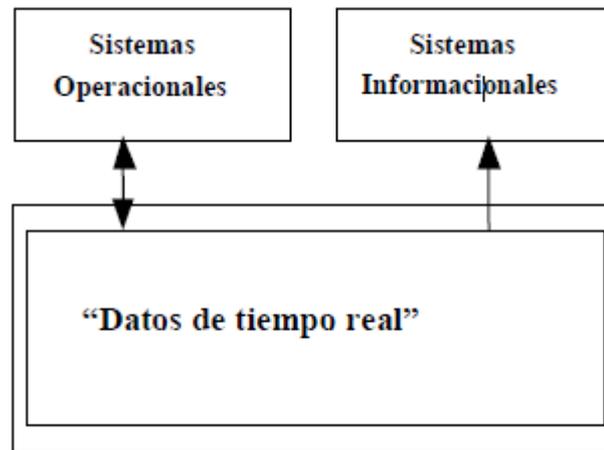


Figura 1: Arquitectura de datos de Una sola capa.

La **Arquitectura de Datos de Dos Capas**, garantiza el acceso a los datos por los sistemas operacionales en cualquier momento. La misma contiene una capa inferior, donde se almacenan los datos de tiempo real utilizados por las aplicaciones operacionales en modo lectura/escritura y una capa superior para los datos derivados utilizados por las aplicaciones informacionales. (Ver Figura 2).

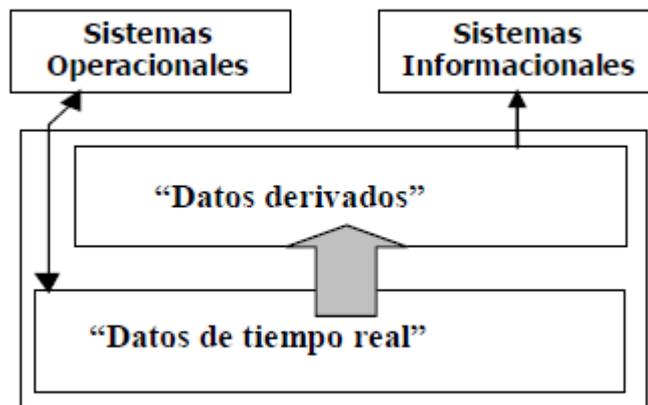


Figura 2: Arquitectura de datos de Dos Capas.



En la **Arquitectura de Datos de Tres Capas** para llevar a cabo la transformación de los datos de tiempo real a datos derivados, se requiere de una capa intermedia, conocida como capa de datos reconciliados, donde se solucionan los problemas de inconsistencias y se realiza el procesamiento de los distintos conjuntos de datos de tiempo real adecuadamente. (Ver Figura 3).

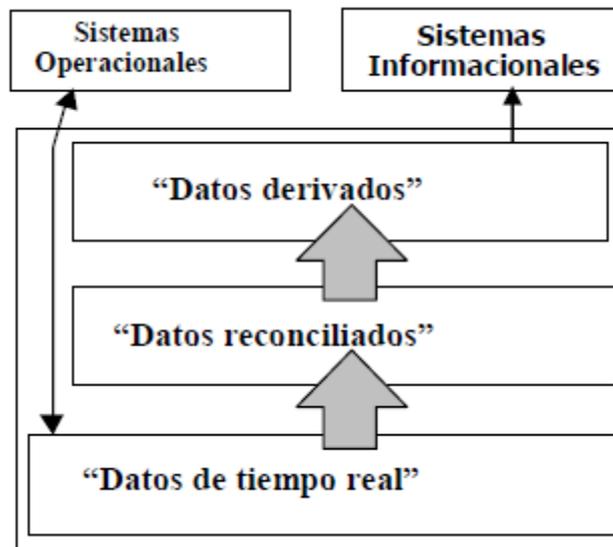


Figura 3: Arquitectura de datos de Tres Capas.

### 1.5.2 Elementos de una arquitectura para el Almacén de Datos.

La arquitectura se constituye de un número de partes interconectadas: (18)

- **Base de datos operacional / Nivel de base de datos externo.**

Una desventaja de los sistemas operacionales y en especial de las bases de datos operacionales históricas es su enfoque limitado y las dificultades que presenta cuando se desea acceder a los datos con otros fines. El Almacén de Datos se encarga de liberar la información que es almacenada en las bases de datos operacionales y agruparla con la de las fuentes externas.

- **Nivel de acceso a la información.**

El nivel de acceso a la información es el nivel del que los usuarios finales se encargan directamente. Representa las herramientas que se utilizan y todo el software y hardware necesario para mostrar las informaciones que serán analizadas.

- **Nivel de acceso a los datos.**



Este nivel es el encargado de que los usuarios puedan acceder a todos los datos desde los sistemas operacionales, sin tener en cuenta las herramientas de acceso o ubicación disponible de los datos. Es el responsable entonces de las interfaces entre las fuentes de información y los sistemas operacionales. El lenguaje SQL es el más usado con este fin.

- **Nivel de Directorio de Datos (Metadata).**

La metada es toda la información relacionada con los datos de un negocio: tipos de datos, consultas, validaciones, etc. Para que los usuarios tengan acceso a todos los datos desde cualquier sistema operacional es necesario mantener alguna información sobre las bases de datos operacionales o un repositorio de la información metadata.

- **Nivel de Gestión de Procesos.**

El nivel de gestión de procesos está asociado a la programación de diversas tareas y procedimientos que se realizan en aras de construir y mantener actualizado el Almacén de Datos.

- **Nivel de Mensaje de la Aplicación.**

El nivel de mensaje de la aplicación consiste en el transporte de la información que se realiza alrededor de la red de la empresa. Se puede usar para aislar aplicaciones operacionales estratégicas, recolectar transacciones o mensajes y luego entregarlos de forma segura.

- **Nivel del Almacén de Datos (Físico).**

En un Almacén de Datos físico se almacenan de forma flexible copias de datos operacionales y externos, los cuales son de fácil acceso.

- **Nivel de Organización de Datos.**

Es el nivel final dentro de la arquitectura y abarca todos los procesos necesarios como editar, seleccionar, resumir y cargar los datos en el almacén y acceder a la información desde bases de datos operacionales y externas. Permite además el análisis de la calidad de los datos y filtros que identifican modelos y estructuras de datos, dentro de los sistemas operacionales existentes.

### 1.6 Modelado de los datos.

Al igual que en las bases de datos tradicionales, en los Almacenes de Datos existen tres niveles de modelado de datos: conceptual, lógico y físico. Debido al tipo de consultas que se realizan en los almacenes y el rendimiento que se espera de ellos, las pautas para el diseño y modelado de los datos, hacen que sean diferentes.



- **Modelo de datos conceptual.**

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

- **Modelo de datos lógico.**

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

- **Modelo de datos físico.**

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

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

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

El Almacén de Datos se implementa sobre un Sistema Gestor de Base de Datos y se utilizan herramientas de Extracción, Transformación y Carga (ETL) para integrar los datos obtenidos desde distintas base de datos.

#### **1.7.1 Sistema de Gestión de Base de Datos Oracle.**

Un Sistema Gestor de Bases de Datos (SGBD) es un conjunto de programas que brindan la posibilidad a los usuarios de crear y mantener una base de datos con el fin de utilizarla para diversas aplicaciones.

Las principales funciones que debe cumplir un SGBD se relacionan con la creación y mantenimiento de la base de datos, el control de accesos, la manipulación de datos de acuerdo con las necesidades del usuario, el cumplimiento de las normas de tratamiento de datos, evitar las redundancias e inconsistencias y mantener la integridad.

El proyecto “Humanización Penitenciaria Cuba” de la Universidad de las Ciencias Informáticas, es el encargado de la automatización de los procesos del Sistema Penitenciario Cubano. A solicitud de sus



clientes este proyecto trabaja con el Gestor de Base de Datos Oracle, consideración por la cual el Data Mart para el departamento de Sistema Educativo también utilizará este SGBD.

Oracle es el motor de base de datos relacional más usado a nivel mundial. Se considera uno de los sistemas de bases de datos más completos, destacando su soporte de transacciones, estabilidad y escalabilidad. (19)

Oracle puede ejecutarse en todas las plataformas y cuenta con un lenguaje de diseño de base de datos muy completo (PL/SQL) que permite procedimientos almacenados, con una integridad referencial declarativa bastante potente. Otra considerable ventaja es que el software del servidor puede correr en multitud de sistemas operativos, además es la base de datos con más orientación hacia Internet.

### **1.7.2 Oracle Warehouse Builder (OWB). Características.**

Oracle provee una herramienta para el diseño del Almacén de Datos llamada Oracle Warehouse Builder (OWB) (20), la cual permite el diseño del proceso de Extracción, Transformación y Carga de los datos (ETL), dando soporte a todo el flujo de datos necesarios para poblar el almacén.

Esta herramienta permite tomar los datos desde diferentes fuentes y sistemas e integrarlos, para proporcionar información de alta calidad que luego será analizada. Proporciona una amplia biblioteca de transformaciones para datos de tipo texto, numérico, fechas y otros.

Posee un conjunto de interfaces gráficas de usuario, siendo el Centro de Diseño una de las principales y la que permite importar los objetos de origen y realizar el diseño de las ETL.

### **1.8 Metodologías para la construcción de un Almacén de Datos**

DWEP (Data Warehouse Engineering Process) y Hefesto son metodologías de desarrollo para la construcción de los Almacenes de Datos.

#### **1.8.1 Metodología DWEP. Características.(21)**

Esta metodología es independiente de la implementación que se realice, tanto relacional, como dimensional u orientada a objetos. Está basada en el Proceso Unificado de Desarrollo de Software (RUP) y utiliza el Lenguaje Unificado de Modelado (UML) para visualizar los modelos realizados.

La misma propone la estructuración del Almacén de Datos en cinco etapas y tres niveles.

**Etapas:**



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

### Niveles:

- Conceptual.
- Lógico.
- Físico.

Como el DWEP es una instanciación de RUP, para la construcción de los Almacenes de Datos, define también el ciclo de vida del proyecto en cuatro fases: Inicio, Elaboración, Construcción y Transición, y en cinco flujos de trabajo fundamentales: Requerimientos, Análisis y Diseño, Implementación y Prueba. Además adiciona dos nuevas actividades: Mantenimiento y Revisión Pos-desarrollo. En cada uno de los flujos de trabajo se utilizan diferentes diagramas UML para modelar y documentar el proceso de desarrollo.

### 1.8.2 Metodología Hefesto. Características. (10)

*“Hefesto es una metodología propia, cuya propuesta está fundamentada en una muy amplia investigación, comparación de metodologías existentes y experiencias propias en procesos de confección de Almacenes de Datos”.*

Entre las características más notables de Hefesto está que se basa en los requerimientos de los usuarios, quienes se involucran en cada etapa de la realización del Almacén de Datos. Se adapta con facilidad a los cambios que ocurren en el negocio, generando como resultados modelos conceptuales y lógicos que son fáciles de comprender y cada fase constituye el punto de partida para iniciar otra.

A su vez esta metodología es independiente de las herramientas que se utilicen para la implementación del almacén y de las estructuras físicas que presente.

Teniendo en cuenta estas características que son aplicables tanto para la construcción del Almacén de Datos como para el Data Mart y basándose en que el diseño del almacén conlleva en todas sus fases una comunicación directa con el usuario final, se propone utilizar esta metodología.



### 1.8.3 Pasos para la aplicación de la metodología propuesta.

La metodología propuesta es Hefesto y se basa en cuatro pasos fundamentales, ellos son: (10)

- **Paso1: Análisis de Requerimientos.**

En este paso se analiza el negocio en profundidad, determinándose los requerimientos de los usuarios y los indicadores y perspectivas a tener en cuenta para la construcción del almacén. Al finalizar se confecciona un modelo conceptual donde se puede visualizar el resultado obtenido en este paso.

- **Paso 2: Análisis de los OLTP.**

Una vez creado el modelo conceptual se analizan las fuentes de datos disponibles para determinar cómo serán calculados los indicadores y establecer las correspondencias entre este modelo y los OLTP, definiendo qué campos se incluirán en cada perspectiva.

- **Paso 3: Modelo Lógico del Almacén de Datos.**

Se confecciona el modelo lógico de la estructura del Almacén de Datos, teniendo en cuenta el modelo conceptual creado. Primeramente se define el tipo de modelo a utilizar y luego se llevan a cabo las acciones propias para el diseño.

- **Paso 4: Procesos ETL.**

Una vez construido el modelo lógico, se procede a probarlo con datos, a través de procesos ETL. Para ello se debe tener en cuenta cuál es la información que se desea almacenar.

### 1.9 Conclusiones.

En este capítulo se abordó sobre las principales características de los Almacenes de Datos con el fin de tener un amplio dominio de la aplicación a desarrollar. Se propone utilizar el enfoque dado por Kimball de la estructura dimensional al ofrecer mayor facilidad al usuario para la exploración y la búsqueda de la información y analizar con profundidad las áreas críticas.

Se propone la arquitectura ROLAP por la flexibilidad que brinda, así como la arquitectura de datos de Tres Capas por ser la más completa, pues aunque existe actualmente una única fuente de datos origen, se prevé que en el futuro el Sistema Penitenciario Cubano pueda integrarse a otras entidades como la ONE y la Fiscalía Nacional. A su vez este modelo evita que se presenten problemas con los datos, al desaparecer las inconsistencias de los mismos y garantizar que los usuarios no reciban una información incorrecta.



Se decide utilizar la metodología Hefesto porque a través de la misma se realiza un análisis completo que permite abarcar los procesos principales del área y es la propuesta de la Universidad de Ciencias Informáticas para el diseño de los almacenes de datos. Además examina e interpreta de forma óptima las necesidades de información del negocio y apoya una mejor construcción y adaptación del Data Mart.

Como herramienta de desarrollo para el Data Mart se propone Oracle Warehouse Builder ya que permite todo el diseño del proceso de Extracción, Transformación y Carga de los datos (ETL), dando soporte a todo el flujo de datos necesarios para poblar el almacén.

Se utilizará como Sistema Gestor de Base de Datos Oracle 11g por ser capaz de manejar eficientemente todo el volumen de información del Data Mart y proporcionar el soporte necesario para la construcción y mantenimiento del mismo.



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

En este capítulo se presenta el análisis realizado a los diferentes departamentos del área de Sistema Educativo del Sistema Penitenciario Cubano, guiado por la metodología Hefesto para llevar a cabo todos los pasos correspondientes al diseño del Data Mart. Se identifican los principales procesos de cada área y se definen los indicadores y perspectivas necesarios para el diseño del almacén.

### **2.1 Análisis de los procesos.**

El área de Sistema Educativo es la encargada de tomar decisiones sobre la reinserción de los privados de libertad. Entre sus misiones principales está la de ejecutar el proceso de estudio, observación, evaluación y diagnóstico del interno. Se encarga además de dirigir el tratamiento educativo que se les dispensa a los sancionados y lograr una educación en la formación de valores, que contribuyan a reforzar las cualidades positivas, logrando normas de convivencia social y una cultura general integral en el proceso de resocialización. (22)

Las decisiones o medidas que toman los funcionarios autorizados de esta área están dirigidas a influir directamente sobre un interno o colectivo de ellos, enfatizando en las conductas positivas y eliminando las actitudes negativas contrarias a la disciplina y al régimen penitenciario.

Entre los principales procesos que se desarrollan se encuentran:

- **Proceso Características Socio – Antropológicas:**

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

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

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

- **Proceso Instrucción Escolar:**



La Instrucción Escolar se imparte en tres niveles: Enseñanza Obrera Campesina (EOC), Secundaria Obrero Campesina (SOC) y Facultad Obrero Campesina (FOC). Los internos con necesidades educativas especiales reciben la Enseñanza Especial. Se imparten además programas de Enseñanza Superior que incluyen especialidades del INDER y el MINSAP.

Esta instrucción es obligatoria para todos los internos que no tienen el nivel que se imparte en el centro donde están internados y los que tienen hasta 20 años de edad y los jóvenes adultos. Al concluir el curso escolar se le entrega al interno el documento acreditativo de haberlo aprobado.

Para los especialistas de esta área es importante poder analizar cuántos internos están incorporados a la instrucción y cuántos no, los resultados que se obtienen por evaluaciones, cuáles son los cursos donde se presentan mayores problemas y las asignaturas con más bajos índices de promoción.

Toda esta información deberá ser emitida en reportes e informes estadísticos, que permitan comparar un curso escolar con otro, períodos del año determinado, para poder tomar decisiones relacionadas con el desempeño de los internos, la incorporación a la instrucción y la efectividad de los cursos impartidos.

- **Proceso Superación Técnica:**

Este proceso consiste en la capacitación de oficios a la que los internos se integran. La misma se ejecuta a pie de obra, en forma de adiestramiento o en cursos teóricos prácticos que se imparten. Para ello se tienen en cuenta tanto la formación vocacional del interno, como las necesidades de cada territorio donde se encuentran los centros penitenciarios.

Los cursos tienen una duración de no más de 6 meses, exceptuando los de técnico medios. Al concluir el curso de capacitación se realiza la evaluación correspondiente a los internos.

Todos aquellos internos que posean severas medidas y manifiestan una alta peligrosidad no ingresan a la capacitación, como tampoco los que no puedan concluir el curso programado, presenten prescripción facultativa o sean enfermos crónicos.

Tener un control estadístico de los internos incorporados a la capacitación, los que están en condiciones de sumarse y no lo han hecho, los oficios que más se ejercen y las evaluaciones obtenidas son indicadores que necesitan conocer los trabajadores del área. Contar con esta información posibilita comparar tales resultados con los mostrados en cursos anteriores, valorar cuán efectivo es el proceso de capacitación, el nivel de aceptación por parte de los internos así como el



impacto en los territorios de la incorporación de los sancionados al trabajo y en base a ello tomar decisiones más certeras que eleven la calidad del proceso.

- **Proceso Trabajo:**

El Trabajo Socialmente Útil (TSU) tiene carácter formativo, creador y conservador de hábitos laborales, con el fin de preparar a los internos para su futura reinserción a la sociedad.

La incorporación a este proceso es un deber del interno y no tiene un carácter aflictivo, ni se aplica como medida disciplinaria. Los sancionados han de cumplir con eficiencias las normas laborales, de producción y disciplinarias, manteniendo una conducta correcta en los puestos de trabajo.

Obtener información relacionada con aquellos internos que están incorporados al trabajo y los que no se han integrado, las unidades donde se prestan servicios y las tareas productivas que se realizan, sería muy útil en los análisis a efectuar, pues se tomarían decisiones encaminadas a una masiva incorporación de los internos a esta actividad y se podrían comparar los resultados con igual período de tiempo en años anteriores.

- **Actividades Educativas:**

Las actividades y eventos educativos que se realizan persiguen el objetivo de lograr la formación ética de los internos y un desarrollo físico, psicológico y cultural. A su vez tienen un contenido educativo y están orientadas esencialmente a estimular la participación masiva de los internos en las mismas. Además de desarrollar hábitos de convivencia en grupo, de disciplina, solidaridad y pertenencia.

Existen talleres de creación y apreciación de las diversas manifestaciones artísticas y culturales como grupos de teatros, musicales, solistas y encuentros literarios. Además se efectúan competencias de juegos deportivos, se proyectan películas y se disfruta de la programación televisiva. Las actividades son planificadas y organizadas mensualmente en todos los colectivos.

Para un eficiente análisis de la información relacionada con este proceso, es necesario conocer cuántos internos se incorporan a las actividades, qué por ciento del total representa esto y cuáles son las actividades de mayor atracción por parte de los sancionados.

Una vez analizados los procesos fundamentales del área, se tiene una mayor claridad de la información que los usuarios necesitan consultar para poder tomar decisiones oportunas que favorezcan al desarrollo de estas actividades.



## 2.2 Análisis de los requerimientos.

Para realizar un análisis de los requerimientos se deben de identificar las perspectivas y los indicadores por los cuáles se obtendrá finalmente la información a consultar.

### 2.2.1 Indicadores y perspectivas.

Según Hefesto (10) los indicadores para que sean efectivos son, en general, valores numéricos y representan lo que se desea analizar concretamente, ejemplo de esto: saldos, promedios, cantidades, etc.

En cambio las perspectivas se refieren a los objetos mediante los cuales se quiere examinar los indicadores; llevadas al modelo relacional sería una tabla de la base de datos, que se identifica por atributos o características en dependencia del proceso al que pertenece para dar cumplimiento a los objetivos propuestos.

El Tiempo es usado comúnmente como una perspectiva, ya que en los análisis que se realizan deben haber indicadores asociados a años, meses, días, trimestre, etc. A las perspectivas que como el Tiempo, están presentes en todos los procesos, se les llama “perspectivas comunes”.

Las otras perspectivas que están presentes en cada proceso se seleccionan acorde a los requerimientos identificados.

Este documento recoge el análisis y diseño realizado a los procesos: Trabajo, Comportamiento Socio – Antropológico y Actividades Educativas del Sistema Penitenciario Cubano.

La Tabla 1 muestra todas las perspectivas que tendrá el Data Mart con una descripción de lo que representan.

<b>Perspectivas</b>	<b>Es una tabla con sus atributos correspondientes los cuales permitirán filtrar los indicadores propuestos.</b>
<b>Perspectivas Comunes</b>	
<b>Perspectivas</b>	<b>Descripción</b>
Tiempo	Tabla que indica especificaciones de la fecha en la que ocurre un evento relacionado con un proceso en específico.
<b>Perspectivas por proceso</b>	
<b>Nombre Proceso: Trabajo</b>	



Perspectivas	Descripción
Centro	Tabla que agrupa las características del centro penitenciario donde se encuentra el interno que se incorpora al trabajo.
Unidad Productiva	Tabla que agrupa las características de la unidad productiva donde se desarrolla el trabajo.
Interno	Tabla que agrupa las características del interno incorporado al trabajo.
<b>Nombre Proceso: Características Socio – Antropológicas</b>	
Perspectivas	Descripción
Centro	Tabla que agrupa las características del centro penitenciario donde se encuentra el interno.
Interno	Tabla que agrupa las características del interno.
<b>Nombre Proceso: Actividades Educativas</b>	
Perspectivas	Descripción
Nivel	Tabla que agrupa las características del nivel que realiza la actividad.
Actividad	Tabla que agrupa las características de la actividad que se realiza.
Interno	Tabla que agrupa las características del interno que participa en la actividad.

**Tabla 1:** Perspectivas del Data Mart.

Luego de definidas las perspectivas para cada uno de los procesos, se identifican los indicadores que se analizarán y los atributos que contendrán cada perspectiva.

De manera general cada proceso estará constituido por una serie de indicadores y de perspectivas.

Para definir las perspectivas es necesario el análisis de los OLTP y establecer las correspondencias entre los datos que deben contener y la información disponible en la fuente de datos.

Para la perspectiva Tiempo se determinan los atributos que tendrá y la descripción de los mismos como muestra la Tabla 2. En este caso solo es necesario establecer la correspondencia con el campo FECHA\_SQL que existe en la base de datos, pues los restantes son calculados.



Análisis OLTP: Perspectiva Tiempo	
Atributo destino	Descripción
codigo_tiempo	Código de una fecha en particular.
num_dia	Número del día en que se desarrolla el proceso.
num_mes	Número del mes en que se desarrolla el proceso.
num_anno	Número del año en que se desarrolla el proceso.
nomb_dia	Nombre del día en que se desarrolla el proceso.
nomb_mes	Nombre del mes en que se desarrolla el proceso.
trimestre	Período en que se desarrolla el proceso, en este caso de tres meses.
cuatrimestre	Período en que se desarrolla el proceso, en este caso de cuatro meses.
semestre	Período en que se desarrolla el proceso, en este caso de seis meses.
fecha_sql	Fecha completa del día en que se desarrolla el proceso y es mediante este atributo que se establecerá la correspondencia con la fuente de datos.
num_dia_semana	Número de la semana en que se desarrolla el proceso.

Tabla 2: Análisis OLTP Perspectiva Tiempo.

Para la perspectiva Interno se determinan y describen los atributos como muestra la Tabla 3, especificando para cada uno la fuente de origen, acorde a la correspondencia realizada.

Análisis OLTP: Perspectiva Interno		
Atributo destino	Tablas origen	Descripción
id_interno	Interno	Es la clave primaria de la tabla "Interno" y representa a un Interno en particular.



nombre	Interno, Persona	Nombre del interno.
primer_apellido	Interno, Persona	Primer apellido del interno.
segundo_apellido	Interno, Persona	Segundo apellido del interno.
fecha_nacimiento	Interno, Persona	Fecha de nacimiento del interno.
sexo	Interno, Persona	Sexo del interno.
color_piel	Datosfisionomicos, Nom_color_piel	Color de la piel del interno.
oficio	Interno_oficio, Nom_oficios	Oficio del interno.
nivel_escolaridad	Interno, Nom_nivel_escolaridad	Nivel de escolaridad del interno.
estado_civil	Interno, Nom_estado_civil	Estado civil del interno.
religión	Interno, Nom_religion	Religión que profesa el interno.
personas_abrigo	Interno	Cantidad de personas que tiene a su abrigo el interno.
número_hijos	Interno_hijos	Número de hijos del interno.

**Tabla 3:** Análisis OLTP Perspectiva Interno.

La Tabla 4 recoge la información relacionada con el Proceso Trabajo, los indicadores y perspectivas definidas para el mismo, así como las correspondencias establecidas a cada una de sus perspectivas.

Proceso	Trabajo	
Indicadores	Descripción	
Total de internos incorporados.	Representa la cantidad de internos que están incorporados a las actividades productivas.	
Por ciento de internos incorporados al trabajo.	Representa el Por ciento de internos que están incorporados a las actividades productivas	
Total de internos no incorporados al trabajo.	Representa la cantidad de internos que no están incorporados a las actividades productivas.	
Por ciento de internos no incorporados al trabajo.	Representa el Por ciento de internos que no están incorporados a las actividades productivas.	
<b>Análisis OLTP: Perspectiva Centro</b>		



Atributo Destino	Tablas Origen	Descripción
id_centro	Centro_penitenciario	Es la clave primaria de la tabla “Centro_Penitenciario” y representa un Centro Penitenciario en particular.
nombre_centro	Estructura	Nombre del centro.
municipio	Estructura, Direcciones, Nom_municipio	Nombre del municipio donde se encuentra el centro.
provincia	Estructura, Direcciones, Nom_municipio, Nom_provincia	Nombre de la provincia donde se encuentra el centro.
tipo_centro	Centro, Tipo_centro	Indica el tipo de centro.
tipo_estructura	Estructura, Nom_tipo_estructura	Tipo de estructura a la cual pertenece el centro.

**Análisis OLTP: Perspectiva Unidad Productiva**

Atributo destino	Tablas origen	Descripción
id_unidad_prod	Unidad_productiva	Es la clave primaria de la tabla Unidad_Productiva y representa una Unidad Productiva en particular.
tipo_unidad	Unidad_productiva, Nom_tipo_unidad_pro ductiva	Indica el tipo de unidad.
actividad_productiva	Unidad_productiva, Nom_actividad_produ ctiva	Actividad productiva que se realiza en la unidad.
nombre_brigada	Unidad_productiva, Brigada	Nombre de la brigada que trabaja en la unidad.
capacidad_máxima	Unidad_productiva	Capacidad máxima de internos a



		trabajar en la unidad.
fecha_inicio	Unidad_productiva	Fecha de inicio de la actividad productiva.
fecha_fin	Unidad_productiva	Fecha de fin de la actividad productiva.

**Tabla 4:** Indicadores y Perspectivas Proceso Trabajo.

La Tabla 5 recoge la información relacionada con el Proceso Características Socio - Antropológicas, los indicadores y perspectivas definidas para el mismo, así como las correspondencias establecidas a cada una de sus perspectivas.

Proceso	Características Socio - Antropológicas	
Indicadores	Descripción	
Total de internos por edad.	Representa la cantidad de internos según el rango de edad.	
Por ciento de internos edad.	Representa el Por ciento de internos por rango de edad que existen.	
Total de internos por sexo.	Representa la cantidad de internos según el sexo.	
Por ciento de internos por sexo.	Representa el Por ciento de internos según el sexo que existen.	
Total de internos por color de la piel.	Representa la cantidad de internos según el color de la piel.	
Por ciento de internos por color de la piel.	Representa el Por ciento de internos según el color de piel que existen.	
Total de internos por oficio.	Representa la cantidad de internos según un oficio en particular.	
Por ciento de internos por oficio.	Representa el Por ciento de internos por oficio que existen.	
Total de internos por nivel de	Representa la cantidad de internos según el nivel de	



escolaridad.	escolaridad.
Por ciento de internos por nivel de escolaridad.	Representa el Por ciento de internos por este concepto que existen.
Total de internos por estado civil.	Representa la cantidad de internos según el estado civil.
Por ciento de internos por estado civil.	Representa el Por ciento de internos por estado civil que existen.
Total de internos por religión.	Representa la cantidad de internos según la religión.
Por ciento de internos por religión.	Representa el Por ciento de internos por religión que existen.
Total de internos por personas a su abrigo.	Representa el total de internos por la cantidad de personas que tienen a su abrigo.
Por ciento de internos por personas a su abrigo.	Representa el Por ciento de internos por personas a su abrigo que existen.
Total de internos por número de hijos.	Representa la sumatoria de todos los internos según el número de hijos que tienen.
Por ciento de internos por número de hijos.	Representa el Por ciento de internos según el número de hijos que tienen.

**Análisis OLTP: Perspectiva Centro**

Atributo destino	Tablas origen	Descripción
id_centro	Centro_penitenciario	Es la clave primaria de la tabla "Centro_Penitenciario" y representa un centro penitenciario en particular.
nombre_centro	Estructura	Nombre del centro.
municipio	Estructura, Direcciones, Nom_municipio	Nombre del municipio donde se encuentra el centro.



provincia	Estructura, Direcciones, Nom_municipio, Nom_provincia	Nombre de la provincia donde se encuentra el centro.
tipo_centro	Centro, Tipo_centro	Indica tipo de centro.
tipo_estructura	Estructura, Nom_tipo_estructura	Tipo de estructura a la cual pertenece el centro.

**Tabla 5:** Indicadores y Perspectivas Proceso Características Socio - Antropológicas.

La Tabla 6 recoge la información relacionada con el Proceso Actividades Educativas, los indicadores y perspectivas definidas para el mismo, así como las correspondencias establecidas a cada una de sus perspectivas.

<b>Proceso</b>	Actividades Educativas	
<b>Indicadores</b>	<b>Descripción</b>	
Total de internos incorporados.	Representa la cantidad de internos que están incorporados a las actividades educativas.	
Por ciento de internos incorporados.	Representa el Por ciento de internos que están incorporados a las actividades educativas.	
Total de internos no incorporados.	Representa la cantidad de internos que no están incorporados a las actividades educativas.	
Por ciento de internos no incorporados.	Representa el Por ciento de internos que no están incorporados a las actividades educativas.	
<b>Análisis OLTP: Perspectiva Nivel</b>		
<b>Atributo destino</b>	<b>Tablas origen</b>	<b>Descripción</b>
id_nivel	Nom_niv_act	Es la clave primaria de la tabla "NOM_NIV_ACT" y representa un nivel en particular.
nivel	Nom_niv_act	Nivel donde se realiza la actividad.
<b>Análisis OLTP: Perspectiva Actividad</b>		



Atributo destino	Tablas origen	Descripción
id_actividad	Actividad	Es la clave primaria de la tabla Actividad y representa una actividad en particular.
nombre_brigada	Actividad	Nombre de la actividad.
lugar	Actividad	Lugar donde se efectúa la actividad.
responsable	Actividad, Plan_act_edu, Oficial	Responsable de la actividad.
tipo_actividad	Actividad, Nom_tipo_act_ed	Indica el tipo de actividad que se realiza.
clasificación	Actividad, Nom_clasif_act	Indica la clasificación de la actividad.
estado_actividad	Actividad, Nom_est_act	Indica el estado de cumplimiento de la actividad.
fecha_inicio	Actividad	Fecha de inicio de la actividad educativa.
fecha_fin	Actividad	Fecha de fin de la actividad educativa.

**Tabla 6:** Indicadores y Perspectivas Proceso Actividades Educativas.

A partir de los indicadores y perspectivas obtenidas se construye el modelo conceptual.

### 2.2.2 Modelo Conceptual.

El Modelo conceptual es una descripción de alto nivel de la estructura de la base de datos, en la cual la información es representada a través de objetos, relaciones y atributos.

A través de este modelo, se podrá observar con claridad cuáles son los alcances del proyecto para luego poder trabajar sobre ellos, quedando especificado los indicadores del proceso y las perspectivas mediante las cuales estos se analizarán.

### Modelo Conceptual Proceso Trabajo



Figura 4: Modelo Conceptual Proceso Trabajo.

**Modelo Conceptual Características Socio - Antropológicas**

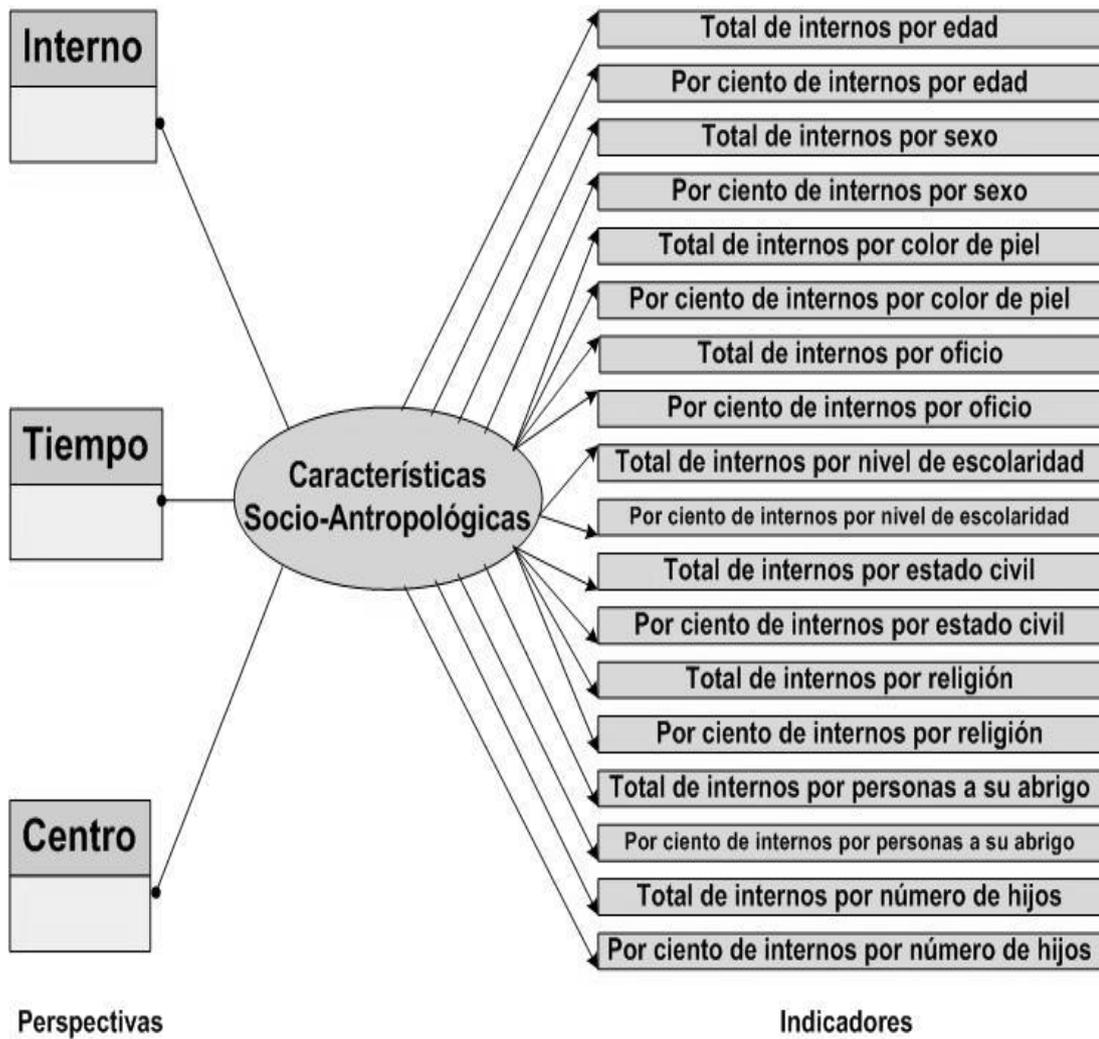


Figura 5: Modelo Conceptual Proceso Características Socio – Antropológicas.

**Modelo Conceptual Proceso Actividades Educativas**

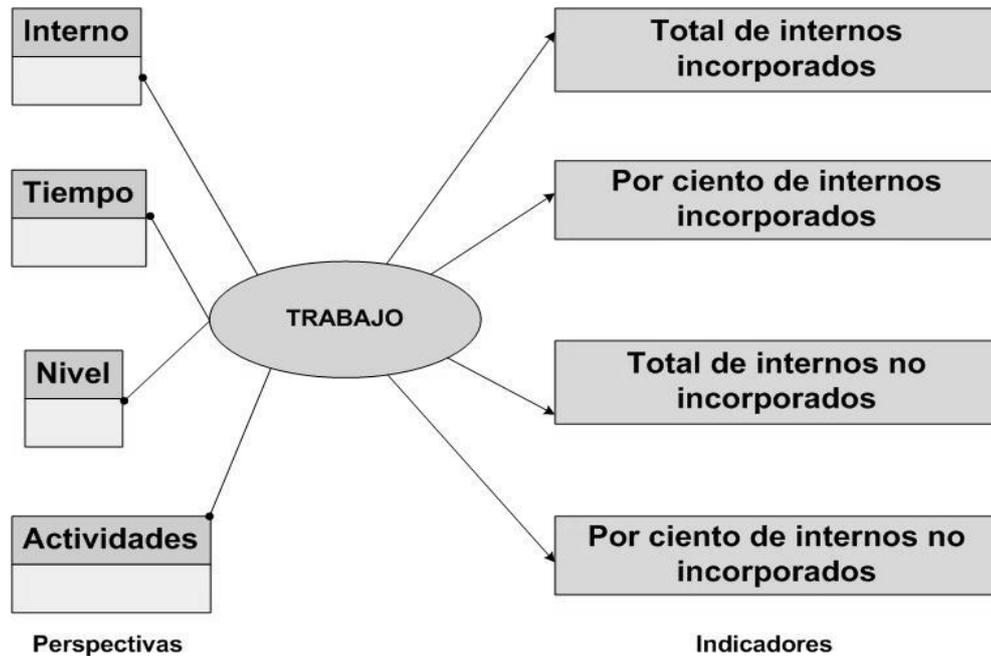


Figura 6: Modelo Conceptual Proceso Actividades Educativas.

Una vez diseñado el modelo conceptual se deben definir cómo serán calculados los indicadores.

### 2.2.3 Conformación de indicadores

Los **indicadores** del **Proceso Trabajo** se calculan de la siguiente manera:

- **“Total de internos incorporados al Trabajo”.**

Hechos: Internos incorporados al trabajo.

Función de sumarización: COUNT.

- **“Por ciento de internos incorporados al Trabajo”.**

Tiet: Total de internos en trabajo.

Tgi: Total general de internos.

Hechos:  $(Tiet / Tgi) * 100$ .

Función de sumarización: Por ciento.

- **“Total de internos no incorporados al Trabajo”:**

Tiet: Total de internos en trabajo.

Tgi: Total general de internos.

Hechos:  $Tgi - Tiet$ .

Función de sumarización: COUNT.



- **“Por ciento de internos no incorporados al trabajo”:**

Tiet: Total de internos en instrucción.

Tgi: Total general de internos.

Hechos:  $((Tgi - Tiet) / Tgi) * 100$ .

Función de sumarización: Por ciento.

Los **indicadores** del **Proceso Características Socio-Antropológicas** se calculan de la siguiente manera:

- **“Total internos por edad”:**

Hechos: edad.

Función de sumarización: COUNT.

- **“Por ciento internos por edad”:**

Tipe: Total internos por edad.

Tgi: Total general de internos.

Hechos:  $(Tipe / Tgi) * 100$ .

Función de sumarización: Por ciento.

- **“Total de internos por sexo”:**

Hechos: sexo.

Función de sumarización: COUNT.

- **“Por ciento de internos por sexo”:**

Tips: Total internos por sexo.

Tgi: Total general de internos.

Hechos:  $(Tips / Tgi) * 100$ .

Función de sumarización: Por ciento.

- **“Total de internos por color de la piel”:**

Hechos: color de la piel.

Función de sumarización: COUNT.

- **“Por ciento de internos por color de la piel”:**

Tipcp: Total internos por color de la piel.

Tgi: Total general de internos.

Hechos:  $(Tipcp / Tgi) * 100$ .



Función de sumarización: Por ciento.

- **“Total de internos por oficio”:**

Hechos: oficio.

Función de sumarización: COUNT.

- **“Por ciento de internos por oficio”:**

Tipo: Total internos por oficio.

Tgi: Total general de internos.

Hechos:  $(\text{Tipo} / \text{Tgi}) * 100$ .

Función de sumarización: Por ciento.

- **“Total de internos por nivel de escolaridad”:**

Hechos: nivel de escolaridad.

Función de sumarización: COUNT.

- **“Por ciento de internos por nivel de escolaridad”:**

Tipne: Total internos por nivel de escolaridad.

Tgi: Total general de internos.

Hechos:  $(\text{Tipne} / \text{Tgi}) * 100$ .

Función de sumarización: Por ciento.

- **“Total de internos por estado civil”:**

Hechos: estado civil.

Función de sumarización: COUNT.

- **“Por ciento de internos por estado civil”:**

Tipec: Total internos por estado civil.

Tgi: Total general de internos.

Hechos:  $(\text{Tipec} / \text{Tgi}) * 100$ .

Función de sumarización: Por ciento.

- **“Total de internos por religión”:**

Hechos: religión.

Función de sumarización: COUNT.

- **“Por ciento de internos por religión”:**

Tipr: Total internos por religión.



Tgi: Total general de internos.

Hechos:  $(\text{Tipr} / \text{Tgi}) * 100$ .

Función de sumarización: Por ciento.

- **“Total de internos por personas a su abrigo”:**

Hechos: personas a su abrigo.

Función de sumarización: COUNT.

- **“Por ciento de internos por personas a su abrigo”:**

Tipa: Total internos por personas a su abrigo.

Tgi: Total general de internos.

Hechos:  $(\text{Tipa} / \text{Tgi}) * 100$ .

Función de sumarización: Por ciento.

- **“Total de internos por número de hijos”:**

Hechos: número de hijos.

Función de sumarización: COUNT.

- **“Por ciento de internos por número de hijos”:**

Tipnh: Total internos por número de hijos.

Tgi: Total general de internos.

Hechos:  $(\text{Tipnh} / \text{Tgi}) * 100$ .

Función de sumarización: Por ciento.

Los **indicadores del Proceso Actividades Educativas** se calculan de la siguiente manera:

- **“Total de internos incorporados a las actividades”.**

Hechos: Internos incorporados a las actividades.

Función de sumarización: COUNT.

- **“Por ciento de internos incorporados a las actividades”.**

Tiea: Total de internos en actividades.

Tgi: Total general de internos.

Hechos:  $(\text{Tiea} / \text{Tgi}) * 100$ .

Función de sumarización: Por ciento.

- **“Total de internos no incorporados a las actividades”:**

Tiea: Total de internos en actividades.



Tgi: Total general de internos.

Hechos: Tgi – Tiea.

Función de sumarización: COUNT.

- **“Por ciento de internos no incorporados a las actividades”:**

Tiea: Total de internos en actividades.

Tgi: Total general de internos.

Hechos:  $((Tgi - Tiea) / Tgi) * 100$ .

Función de sumarización: Por ciento.

### **2.3 Conclusiones**

Se analizaron los procesos Trabajo, Características Socio-Antropológicas y Actividades Educativas que se desarrollan en el área de Sistema Educativo del Sistema Penitenciario Cubano y a partir de ellos se identificaron los indicadores y perspectivas a tener en cuenta para el diseño del Data Mart. A su vez se realizó el diseño conceptual del Data Mart y se establecieron las correspondencias con las fuentes de datos para determinar los atributos que contendrán cada una de las perspectivas, quedando todo preparado para el diseño del almacén.



### CAPÍTULO 3: DISEÑO DEL DATA MART

En este capítulo se realiza el diseño lógico del Data Mart, así como el proceso de carga inicial de los datos, teniendo en cuenta la arquitectura definida.

#### 3.1 Modelo Lógico del Data Mart

Basándose en el Modelo Conceptual creado se confecciona el modelo lógico del Data Mart. Se define primeramente el modelo que se utilizará y luego se llevan a cabo las acciones propias para diseñar las tablas de dimensiones y de hechos. Al finalizar, se realizan las uniones pertinentes.

##### 3.1.1 Tipo de modelo lógico del Data Mart

Existen tres tipos de esquemas para contener la estructura del depósito de datos: Esquema en estrella, Esquema copo de nieve y Esquema constelación.

Se propone para el diseño del modelo lógico del Data Mart utilizar el Esquema constelación, ya que el Data Mart va a tener varias tablas de hechos que se van a relacionar con tablas de dimensiones comunes y con otras nuevas tablas.

##### 3.1.2 Tablas de dimensiones

Se diseñan las tablas de dimensiones que forman parte del Data Mart, donde cada perspectiva definida en el modelo conceptual constituye una tabla de dimensión, como se muestra en la Figura 7.

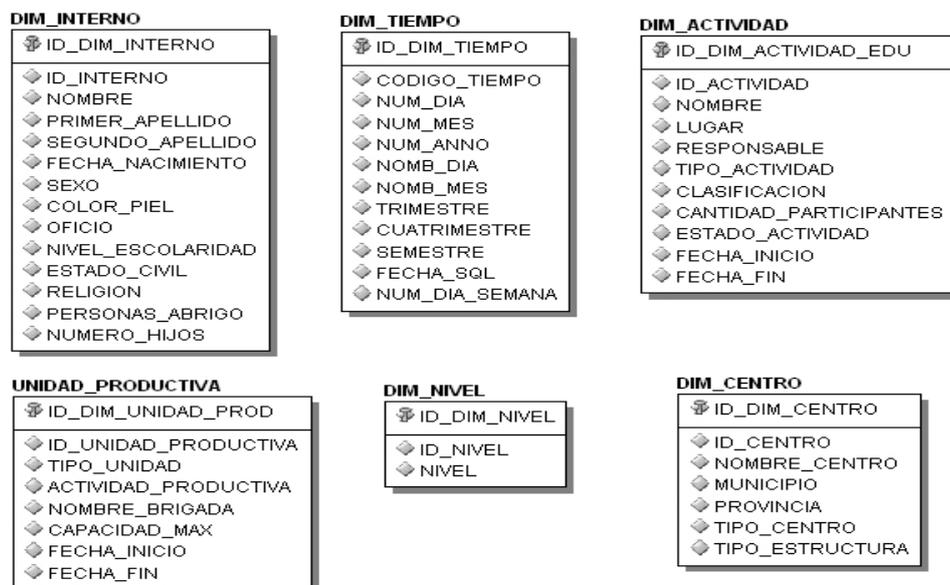


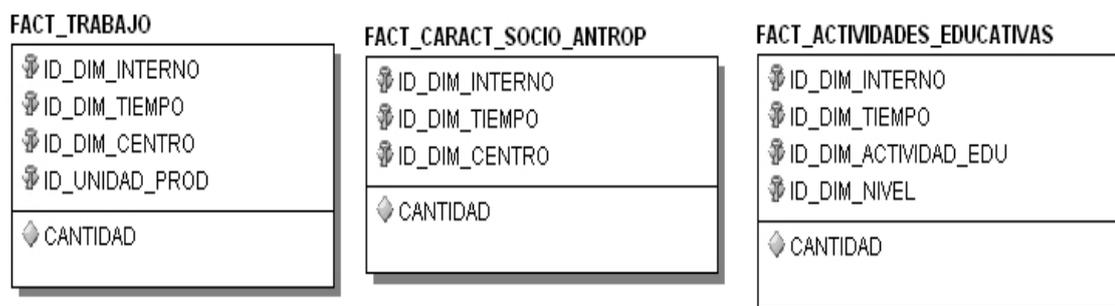
Figura 7: Tablas de dimensiones del Data Mart.



Luego de definir las tablas de dimensiones, se diseñan las tablas de hechos.

### 3.1.3 Tablas de hechos

Las tablas de hechos son las que contendrán los hechos a través de los cuales se construirán los indicadores a analizar. En el Esquema constelación se tienen en cuenta los indicadores y perspectivas analizados anteriormente. Los identificadores de cada tabla de hechos son las llaves primarias de las dimensiones con las que está asociada. Ver Figuras 8.



**Figura 8:** Tablas de hechos del Data Mart.

Luego de diseñar las tablas de hechos y de dimensiones se realizan las uniones correspondientes como se muestra en las siguientes figuras:

La Figura 9 describe la Tabla de Hechos FACT\_TRABAJO que representa el proceso Trabajo.

La Figura 10 describe la Tabla de Hechos FACT\_CARACT\_SOCIO\_ANTROP que representa el proceso Características Socio – Antropológicas.

La Figura 11 describe la Tabla de Hechos FACT\_ACTIVIDAD que representa al proceso Actividades Educativas.

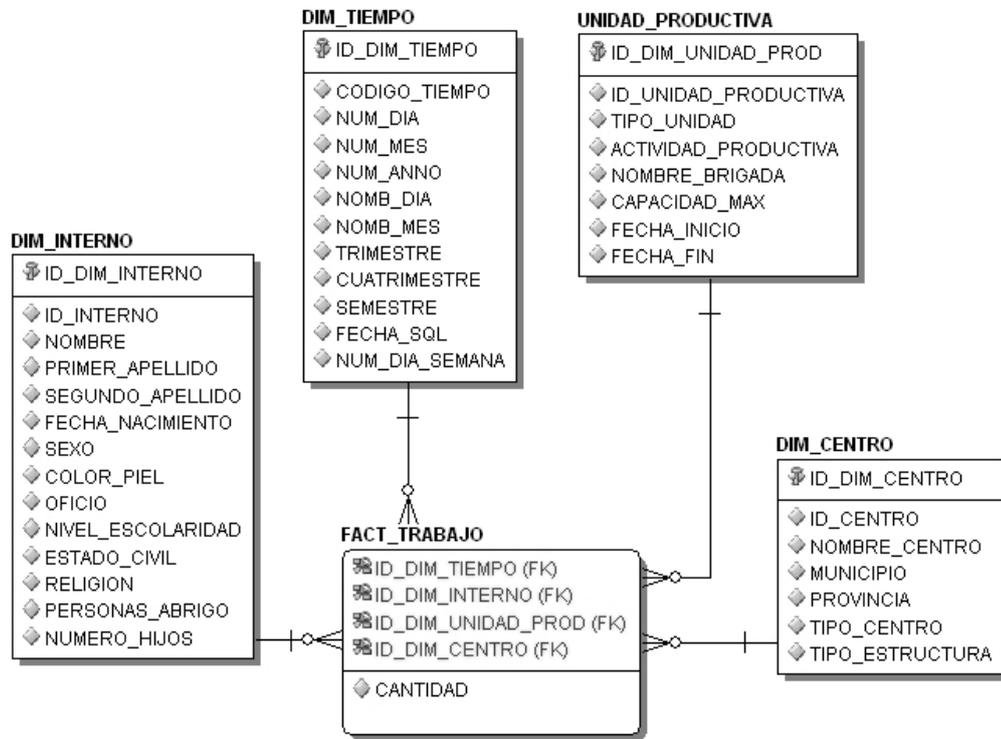


Figura 9: Tabla de hechos FACT\_TRABAJO.

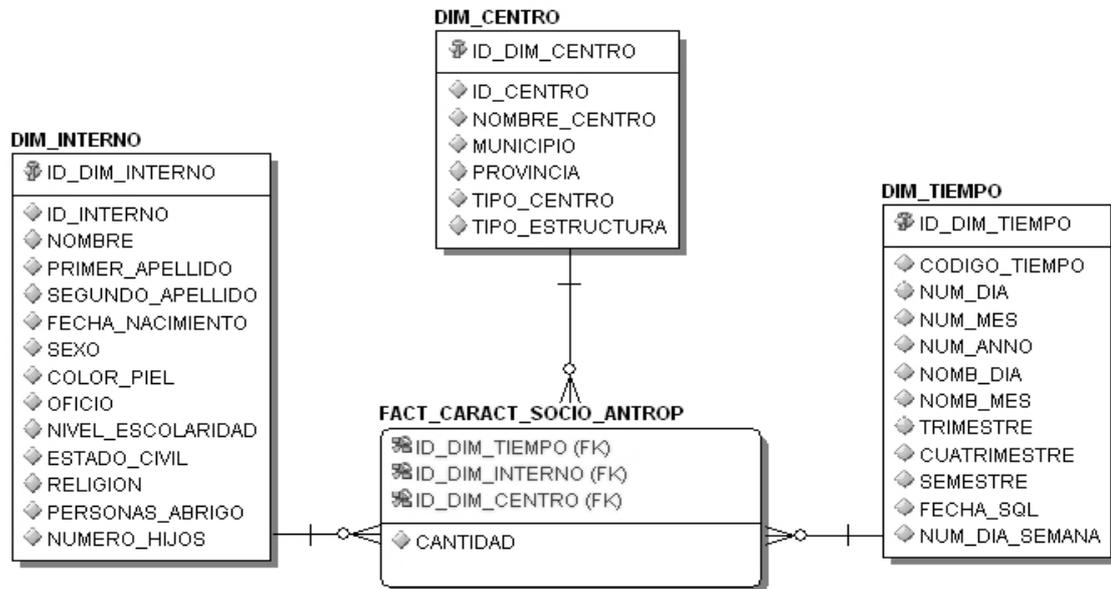


Figura 10: Tabla de hechos FACT\_CARACT\_SOCIO-ANTROP.

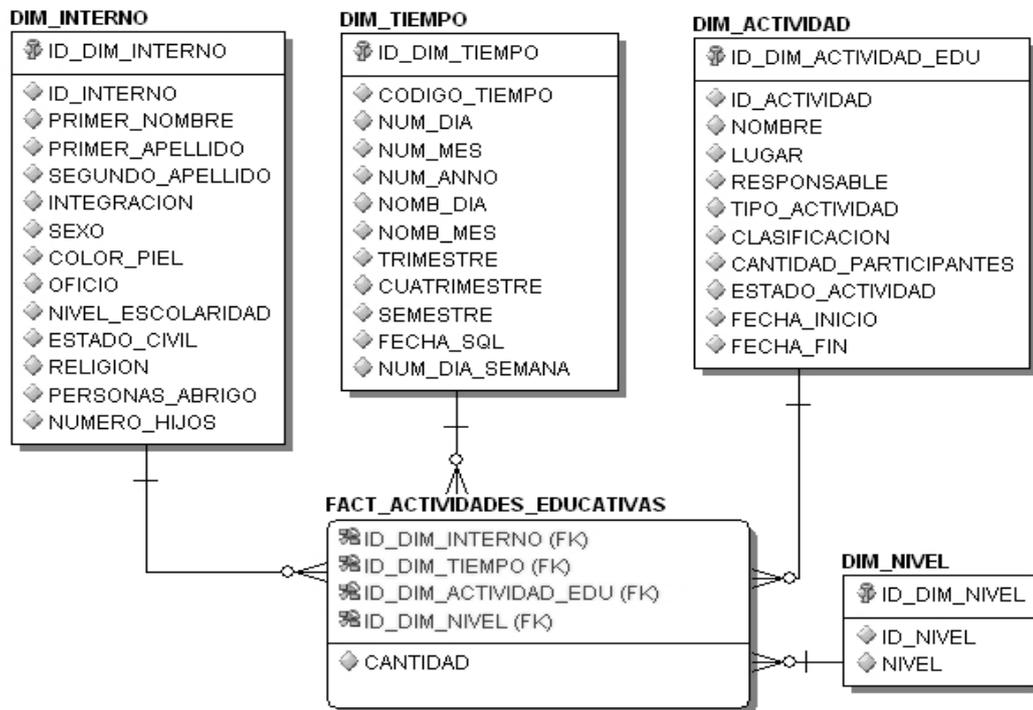


Figura 11: Tabla de hechos FACT\_ACTIVIDADES\_EDUCATIVAS.



### **3.2 Diseño de los procesos de Extracción, Transformación y Carga (ETL)**

Una vez diseñado el Almacén de Datos, se procederá a la carga de los datos mediante los procesos de Extracción, Transformación y Carga (ETL).

La realización de los procesos 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 este proceso y en correspondencia con la arquitectura de tres capas se crearon las Staging Área (Área de almacenamiento temporal) (16), las cuales constituyen una línea divisoria entre el sistema operacional y el depósito de datos, las mismas van a almacenar las tablas de dimensiones que ya han sido cargadas antes de integrarlas finalmente a las tablas de hechos.

#### **3.2.1 Procesos ETL**

Los procesos ETL como se muestra en la Figura 12, se basan en tres pasos fundamentales:(16)

- **Extracción**

Se extraen los datos relevantes desde las distintas fuentes de datos, generalmente son bases de datos o ficheros planos, pero pueden encontrarse también en otras estructuras. Los datos deben de analizarse antes de extraerse para verificar que cumplen con la estructura que se requiere.

- **Transformación**

Los datos se transforman para evitar las inconsistencias, ya que pueden existir diferentes fuentes de información, y los datos del depósito deben de estar en un mismo formato e integrados. La integración se realiza principalmente para establecer: una codificación íntegra de los datos, convenciones de nombramiento, medidas de atributo y fuentes múltiples, todo ello debido a que los datos pueden estar codificados de distintas maneras, tener diversos nombres y unidades de medidas o encontrarse repetido en varias fuentes. En este proceso también se pueden limpiar los datos erróneos o irrelevantes.

- **Carga**

Se realiza la primera carga de los datos desde el almacenamiento intermedio hasta el almacén. La frecuencia de actualización o mantenimiento está en dependencia de los requerimientos de los



usuarios. Serán cargados o actualizados en el almacén aquellos datos que han sido transformados y residen en el almacenamiento intermedio, y los datos de las fuentes que tienen correspondencia directa con los del depósito.

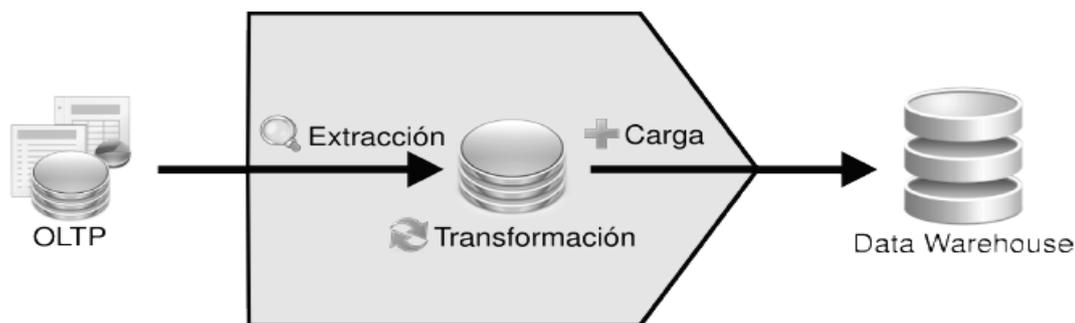


Figura 12: Proceso ETL. (10 p. 26)

### 3.2.2 Procesos ETL para cargar las dimensiones.

El diseño de los procesos ETL se realiza en el Oracle Warehouse Builder (OWB).

- **ETL Dimensión Interno.**

Las transformaciones para cargar la dimensión Interno (DIM\_INTERNO) inician seleccionando las fuentes de datos origen correspondiente (tablas que se examinaron cuando se establecieron las correspondencias entre los atributos de las perspectivas y los OLTP).

Posteriormente se utilizan los componentes Joiner para establecer las relaciones entre las tablas (similar al JOIN en SQL) y Aggregator para agrupar los datos necesarios (similar al GROUPBY en SQL) y luego se declara la Condición de Unión de Joiner (similar a una consulta SQL).

A algunos datos extraídos se les hicieron transformaciones para integrarlos correctamente al Data Mart. Estas transformaciones fueron:

- **CONCAT\_NOMBRES:** Unifica el primer y segundo nombre del interno en un solo atributo.
- **DECOD\_SEXOS:** Muestra el sexo del interno como una cadena (el dato se extraía codificado en 0 y 1, ahora se integra como “Femenino” si es 0 y “Masculino” si es 1).
- **NO\_OFICIOS:** Muestra la cadena “No tiene” en caso de que el interno no posea un oficio, evitando que se carguen en el Data Mart valores nulos.
- **NO\_HIJOS:** Muestra 0 si el interno no tiene hijos, evitando que se carguen en el Data Mart valores nulos.



El identificador de la dimensión es un valor generado por la Secuencia SEQ\_DIM\_INTERNO, una secuencia es un Operador del OWB que genera valores aleatorios continuos (Ej. 1, 2,3... o 74, 75,76).

Una vez obtenido los datos se proceden a mapearlos con los de la tabla DIM\_INTERNO, la Figura 13 muestra todo el proceso ETL realizado.

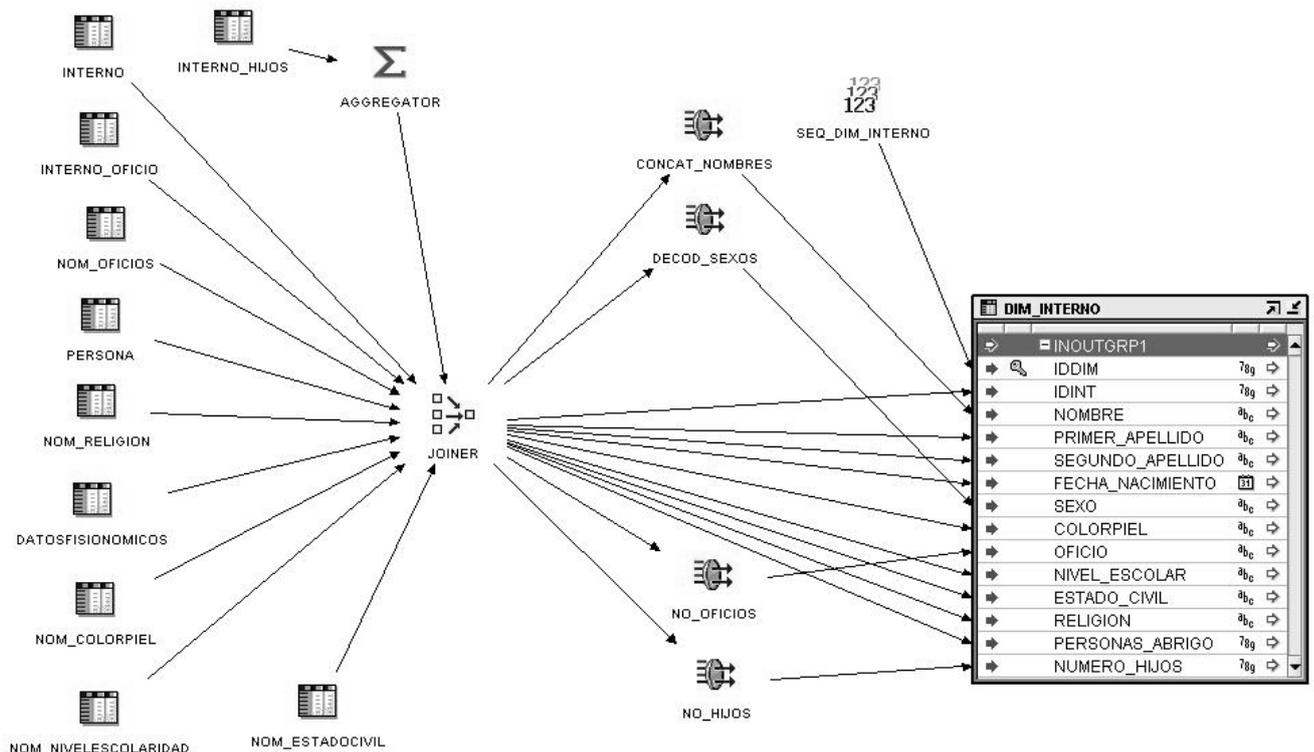


Figura 13: ETL DIM\_INTERNO.

De esta manera en la tabla DIM\_INTERNO del Data Mart quedan cargados los atributos del interno (sexo, estado civil, nombre, religión, etc.) seleccionados para filtrar los indicadores de los procesos con los que se relaciona la dimensión Interno.

- **ETL Dimensión Centro.**

Las transformaciones para cargar la dimensión Centro (DIM\_CENTRO), inician seleccionando las tablas correspondientes desde la fuente de datos, posteriormente se utiliza el componente Joiner, para obtener los datos necesarios a través de la Condición de Unión de Joiner.



El identificador de la dimensión es un valor generado por la Secuencia SEQ\_DIM\_CENTRO. Una vez obtenido todos los datos se proceden a mapearlos con los de la tabla DIM\_CENTRO como se muestra en la Figura 14.

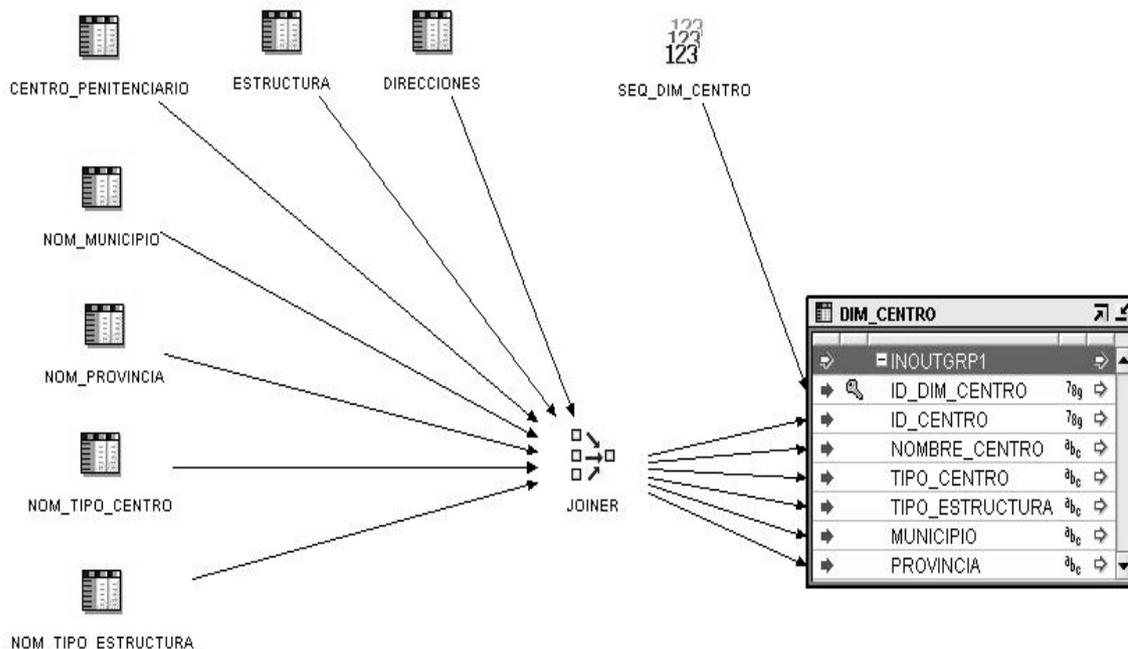


Figura 14: ETL DIM\_CENTRO.

Al concluir el ETL DIM\_CENTRO en el Data Mart quedan cargados los atributos del centro (nombre, tipo, municipio, provincia, etc.) seleccionados para filtrar los indicadores de los procesos con los que se relaciona la dimensión Centro.

- **ETL Dimensión Tiempo.**

Las transformaciones para cargar DIM\_TIEMPO, inician utilizando el Parámetro de Entrada de Correspondencia DATE\_INPUTS que se conecta al Operador de Función de Tabla Day Table Function, ambos son funciones desarrollados por OWB.

Posteriormente se utilizan Operadores de Transformación ya implementados por el OWB como:

- **WB\_CAL\_MONTH\_NAME:** Dada una fecha devuelve el nombre del mes correspondiente.
- **WB\_DAY\_NAME:** Dada una fecha devuelve el nombre del día correspondiente.
- **WB\_CAL\_YEAR:** Dada una fecha devuelve el número del año correspondiente.



- **WB\_CAL\_MONTH OF YEAR:** Dada una fecha devuelve el número del mes correspondiente.

Se realizan las transformaciones siguientes:

- **Function TRIMESTRE:** Dado el número de un mes devuelve el número del trimestre correspondiente.
- **Function CUATRIM:** Dado el número de un mes devuelve el número del cuatrimestre correspondiente.
- **Function SEMESTRE:** Dado el número de un mes devuelve el número del semestre correspondiente.

El identificador de la dimensión es un valor generado por la Secuencia SEQ\_DIM\_TIEMPO. Una vez obtenido todos los datos se proceden a mapearlos con los de la tabla DIM\_TIEMPO como se muestra en la Figura 15.

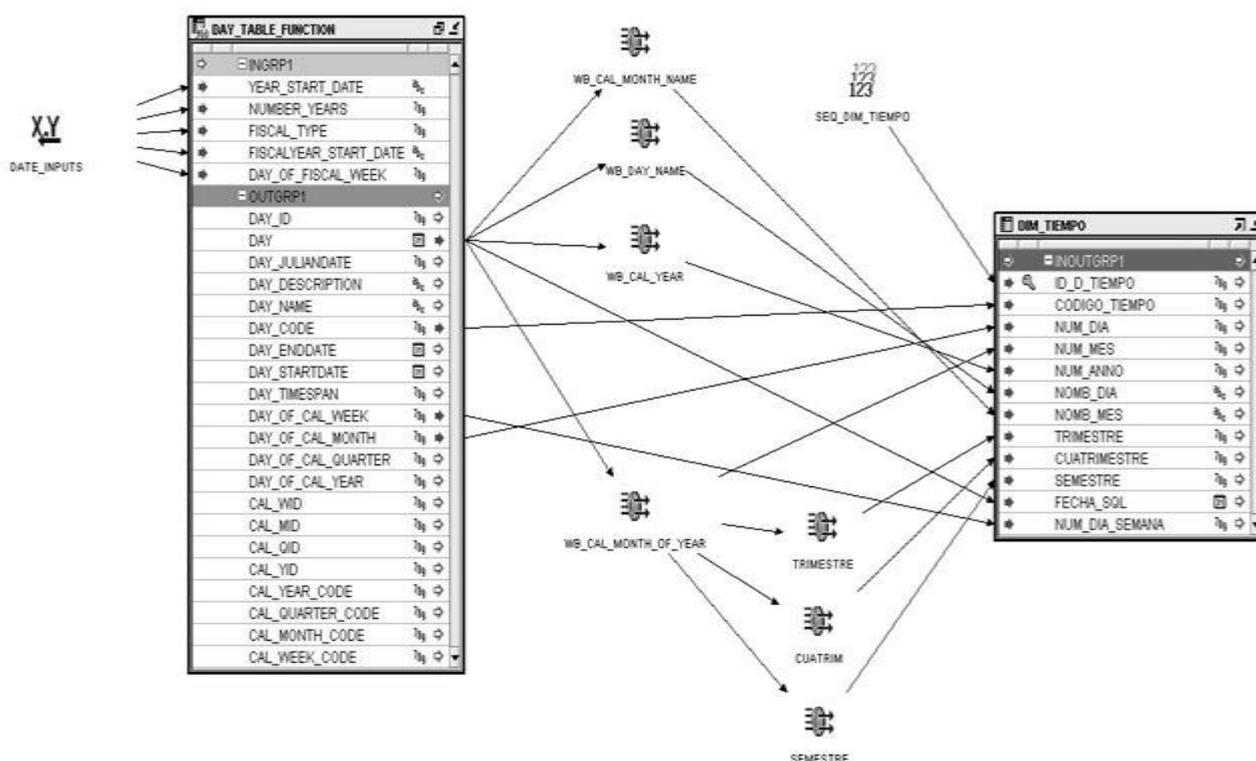


Figura 15: ETL DIM\_TIEMPO.

Finalizado el ETL DIM\_TIEMPO en el Data Mart queda la tabla DIM\_TIEMPO con los atributos (mes, año, día, etc.) que permitirán filtrar los indicadores de todos los procesos.

- **ETL Dimensión Unidad Productiva.**



Las transformaciones para cargar DIM\_UNIDAD\_PRODUCTIVA, inician seleccionando las tablas correspondientes desde la fuente de datos, posteriormente se utiliza el componente Joiner para obtener los datos necesarios a través de la Condición de Unión de Joiner.

El identificador de la dimensión es un valor generado por la secuencia SEQ\_DIM\_UNIDAD\_PRODUCTIVA. Una vez obtenido todos los datos se proceden a mapearlos con los de la tabla DIM\_UNIDAD\_PRODUCTIVA, quedando finalmente en el Data Mart la tabla DIM\_UNIDAD\_PRODUCTIVA con los atributos específicos para analizar los indicadores. (Ver Figura 16).

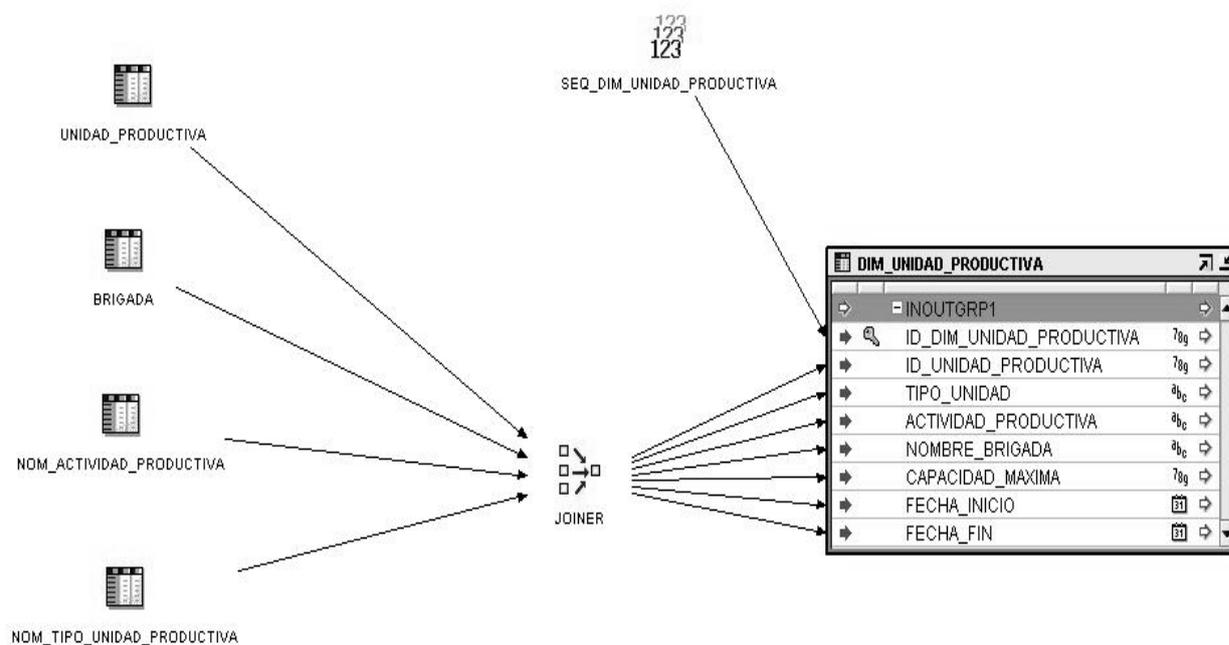


Figura 16: ETL DIM\_UNIDAD\_PRODUCTIVA.

Concluido el ETL DIM\_UNIDAD\_PRODUCTIVA en el Data Mart queda la tabla DIM\_UNIDAD\_PRODUCTIVA con los atributos (nombre, brigada, capacidad, etc.) que permitirán filtrar los indicadores de todos los procesos.

- **ETL Dimensión Actividad.**

Las transformaciones para cargar DIM\_ACTIVIDAD, inician seleccionando las tablas correspondientes desde la fuente de datos, posteriormente se utiliza el componente Joiner para obtener los datos necesarios a través de la Condición de Unión de Joiner.



El identificador de la dimensión es un valor generado por la secuencia SEQ\_DIM\_ACTIVIDAD. Una vez obtenido todos los datos se proceden a mapearlos con los de la tabla DIM\_ACTIVIDAD como muestra la Figura 17.

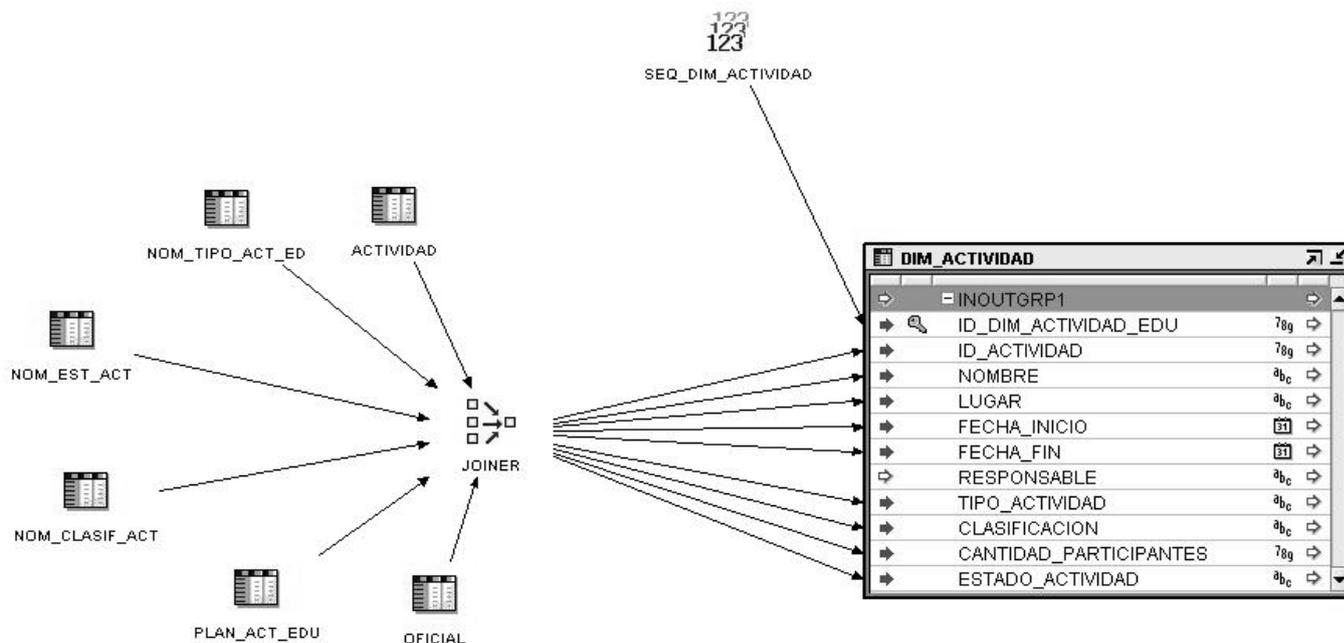


Figura 17: ETL DIM\_ACTIVIDAD.

Al terminar el ETL queda almacenado en el Data Mart la tabla DIM\_ACTIVIDAD, la cual recoge los atributos necesarios(lugar, responsable, clasificacion, etc.) para analizar los indicadores propuestos del proceso con el que se relaciona la dimensión.

Luego de cargar las tablas de dimensiones se procede a cargar las tablas de hechos.

### 3.2.3 Procesos ETL para las tablas de hechos

- ETL FACT\_TRABAJO

Antes de realizar el ETL de la tabla FACT\_TRABAJO se realiza el de la Staging Área diseñada para este proceso que es la STG\_TRABAJO, donde se almacenan de manera intermedia los datos que luego contendrá la FACT\_TRABAJO.

- ETL STG\_TRABAJO



Las transformaciones para cargar STG\_TRABAJO, inician seleccionando las tablas correspondientes, posteriormente se utilizan los componente Joiner y Agregator para obtener los datos necesarios a través de la Condición de Unión de Joiner. Una vez obtenido todos los datos se proceden a mapearlos con los de la tabla STG\_TRABAJO, como muestra la Figura 18.

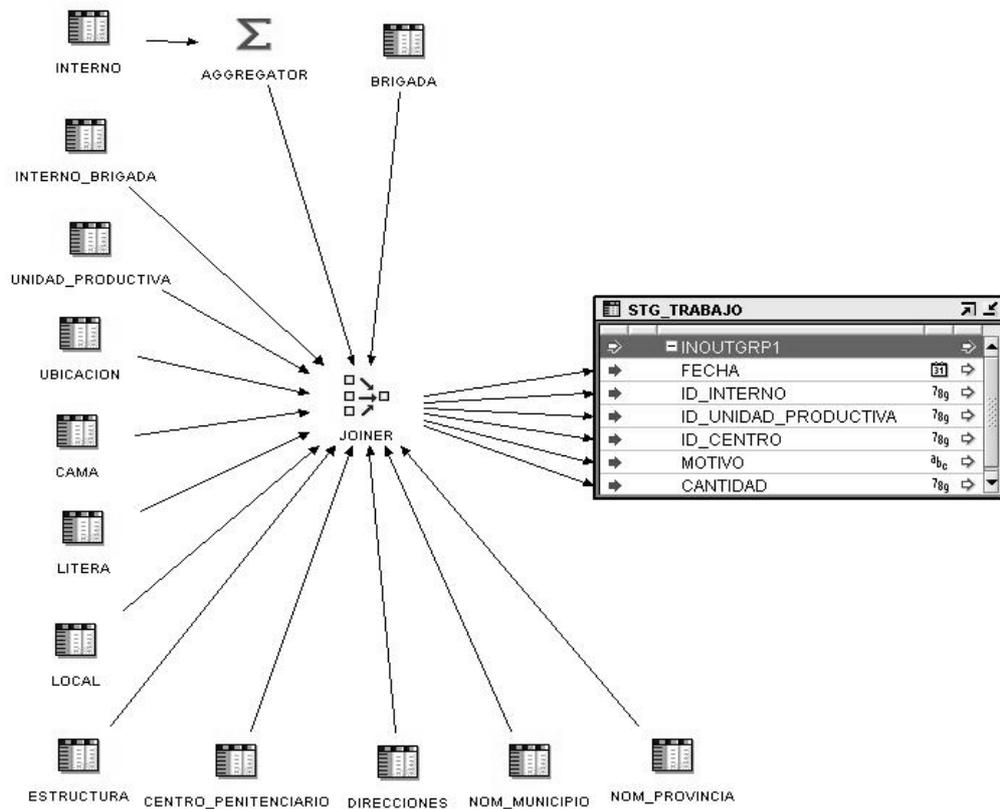


Figura 18: ETL STG\_TRABAJO.

Diseñada la STG\_TRABAJO se procede a realizar el ETL de la tabla FACT\_TRABAJO como muestra la Figura 19.

Para realizar el mapeo correspondiente entre los datos que tiene almacenado la Staging Área y la tabla de hechos es necesario utilizar el Operador de Consulta de Claves KEY\_LOOKUP, el cual permite a través de los datos que tiene el STG\_TRABAJO filtrar cada atributo en sus correspondientes tablas de dimensiones, para luego pasarlo a la FACT\_TRABAJO que almacena las llaves de todas las dimensiones con las que se relaciona.

Para esto se implementaron los operadores:



- **KEY\_LOOKUP\_TIEMPO:** Recibe una fecha como dato de entrada y filtra ese atributo en la tabla DIM\_TIEMPO.
- **KEY\_LOOKUP\_CENTRO:** Recibe el ID del Centro como dato de entrada y filtra ese atributo en la tabla DIM\_CENTRO.
- **KEY\_LOOKUP\_INTERNO:** Recibe el ID del Interno como dato de entrada y filtra ese atributo en la tabla DIM\_INTERNO.
- **KEY\_LOOKUP\_UNIDAD:** Recibe el ID de la Unidad Productiva como dato de entrada y filtra ese atributo en la tabla DIM\_UNIDAD\_PRODUCTIVA.

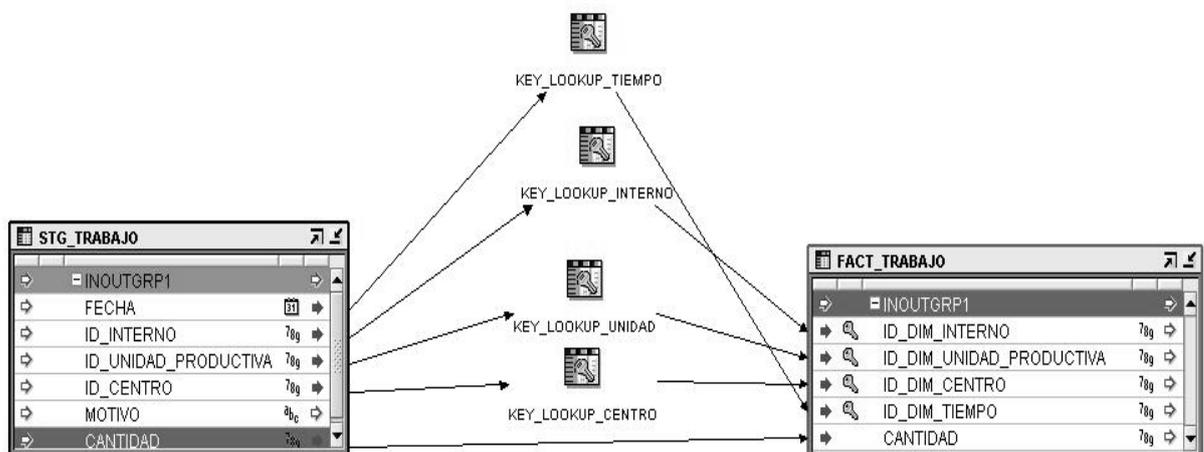


Figura 19: ETL FACT\_TRABAJO.

Al concluir el ETL FACT\_TRABAJO queda en el Data Mart cargada la Tabla FACT\_TRABAJO con todos los identificadores de las dimensiones relacionadas a esta tabla de hechos, permitiendo el análisis de los indicadores propuestos para el Proceso Trabajo.

- **ETL FACT\_CARAC\_SOCIO\_ANTRO**

Antes de realizar el ETL de la tabla FACT\_CARAC\_SOCIO\_ANTRO se realiza el de la Staging Área diseñada para este proceso que es la STG\_CARACT\_SOCIO\_ANTRO, donde se almacenan de manera intermedia los datos que luego contendrá la tabla de hechos.

- **ETL STG\_CARACT\_SOCIO\_ANTRO**

Las transformaciones para cargar STG\_CARACT\_SOCIO\_ANTRO, inician seleccionando las tablas correspondientes, posteriormente se utilizan los componentes Joiner y Aggregator, para concatenar los



datos necesarios a través de la Condición de Unión de Joiner. Una vez obtenido todos los datos se proceden a mapearlos con los de la tabla STG\_CARACT\_SOCIO\_ANTRO como indica la Figura 20.

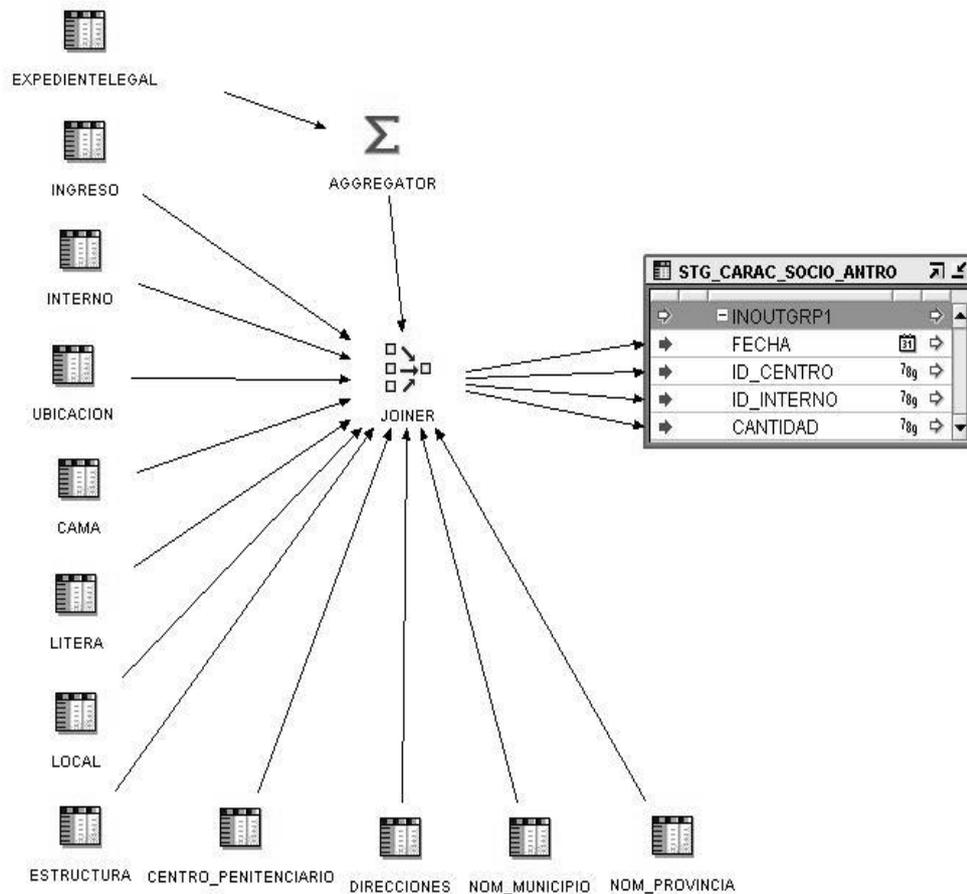


Figura 20: ETL STG\_CARACT\_SOCIO\_ANTRO.

Terminado el ETL STG\_CARACT\_SOCIO\_ANTRO se procede a cargar el ETL FACT\_CARACT\_SOCIO\_ANTRO detallado en la Figura 21.

Para realizar el mapeo correspondiente es necesario utilizar el Operador de Consulta de Claves KEY\_LOOKUP, y el proceso que se realiza es similar al explicado en el ETL FACT\_TRABAJO. Se implementaron los operadores KEY\_LOOKUP\_TIEMPO, KEY\_LOOKUP\_CENTRO y KEY\_LOOKUP\_INTERNO.

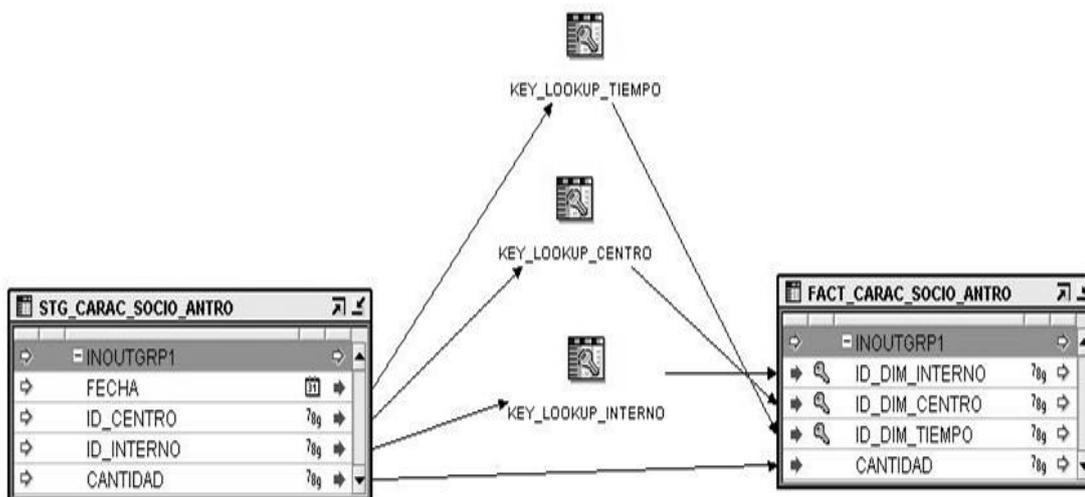


Figura 21: ETL FACT\_CARAC\_SOCIO\_ANTRO.

Al concluir el ETL FACT\_CARAC\_SOCIO\_ANTRO queda en el Data Mart cargada la Tabla FACT\_CARAC\_SOCIO\_ANTRO con todos los identificadores de las dimensiones relacionadas a esta tabla de hechos, permitiendo el análisis de los indicadores propuestos para el Proceso Trabajo.

- **ETL FACT\_ACTIVIDADES\_EDUCATIVAS**

Antes de realizar el ETL de la tabla FACT\_ACTIVIDADES\_EDUCATIVAS se realiza el de la Staging Área diseñada para este proceso que es la STG\_ACTIVIDADES\_EDUCATIVAS, donde se almacenan de manera intermedia los datos que luego contendrá la tabla de hechos.

- **ETL STG\_ACTIVIDADES\_EDUCATIVAS**

Las transformaciones para cargar STG\_ACTIVIDADES\_EDUCATIVAS, inician seleccionando las tablas correspondientes, posteriormente se utilizan los componentes Joiner y Aggregator, para concatenar los datos necesarios a través de la Condición de Union de Joiner. Una vez obtenido todos los datos se proceden a mapearlos con los de la tabla STG\_ACTIVIDADES\_EDUCATIVAS como muestra la Figura 22.

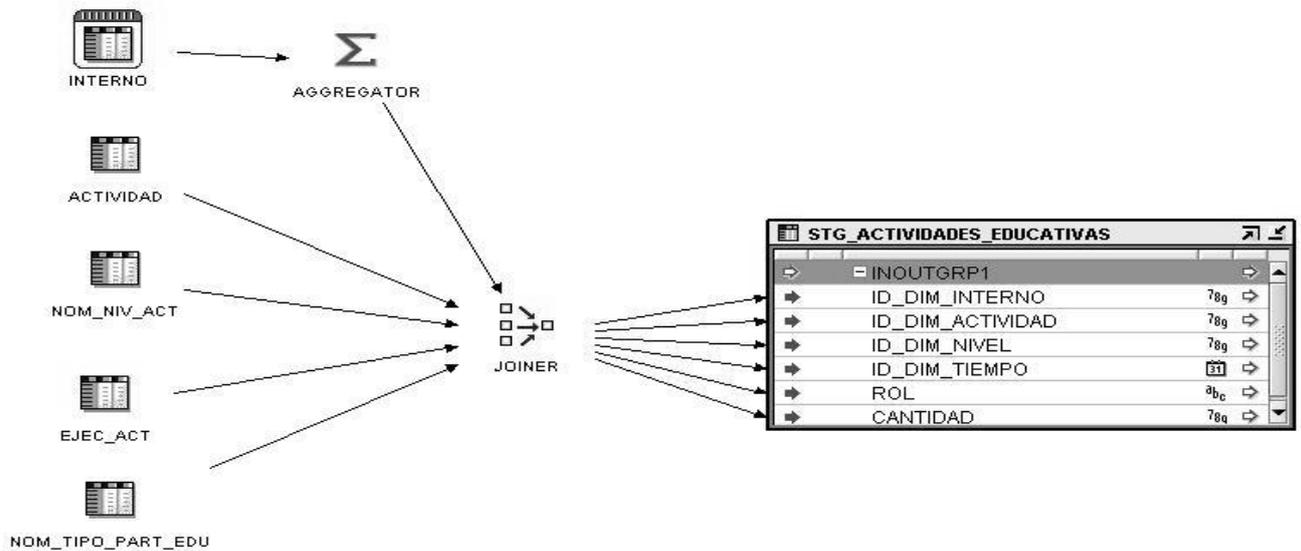


Figura 22: ETL STG\_ACTIVIDADES\_EDUCATIVAS.

Terminado el ETL STG\_ACTIVIDADES\_EDUCATIVAS se procede a cargar la tabla FACT\_ACTIVIDADES\_EDUCATIVAS, ver figura 23.

Para realizar el mapeo correspondiente es necesario utilizar el Operador de Consulta de Claves KEY\_LOOKUP, y el proceso que se realiza es similar a lo explicado en los anteriores. Implementándose los operadores: KEY\_LOOKUP\_TIEMPO, KEY\_LOOKUP\_INTERNO y:

- **KEY\_LOOKUP\_NIVEL:** Recibe el ID del Nivel como dato de entrada y filtra ese atributo en la tabla DIM\_NIVEL.
- **KEY\_LOOKUP\_ACTIVIDAD:** Recibe el ID de la ACTIVIDAD como dato de entrada y filtra ese atributo en la tabla DIM\_ACTIVIDAD.

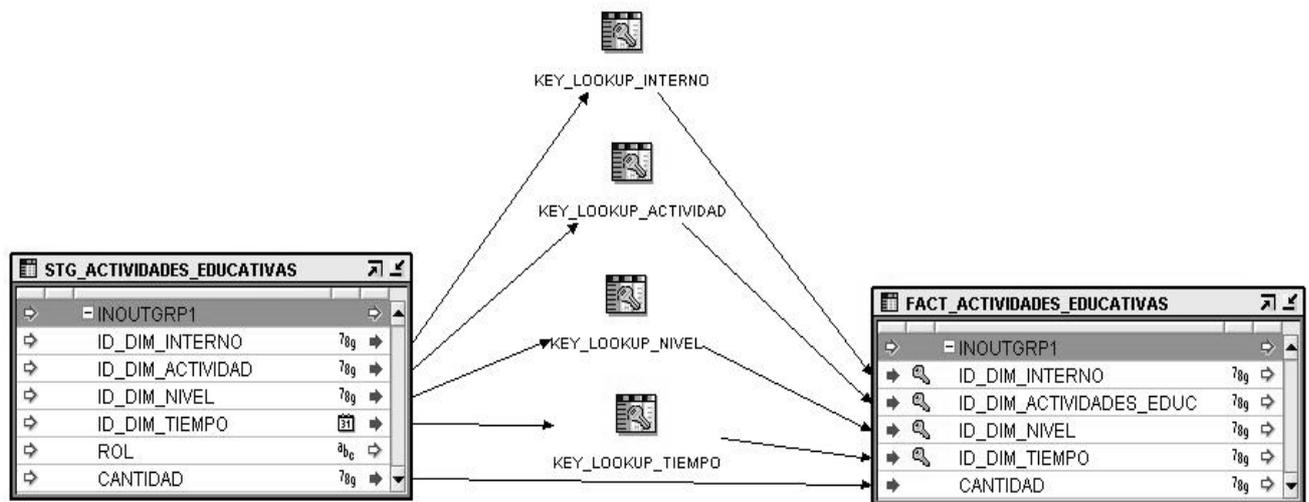


Figura 23: ETL FACT\_ACTIVIDADES\_EDUCATIVAS.

Finalizado el ETL queda en el Data Mart cargada la Tabla FACT\_ACTIVIDADES\_EDUCATIVAS con todos los identificadores de las dimensiones asociadas a esta tabla de hechos, permitiendo el análisis de los indicadores propuestos para el Proceso Actividades Educativas.

El Oracle Warehouse Builder brinda la posibilidad de crear los Flujos de Procesos (20) los cuales permiten ejecutar en paralelo distintas actividades e iniciar otras después que las anteriores se terminen. Una vez diseñados todos los procesos ETL de las dimensiones, las Staging Área y las tablas de hechos del Data Mart, estos son cargados a través flujos que han sido diseñados y no de manera independiente.

### 3.2.4 Carga inicial

La carga inicial del depósito se ha diseñado a través de un Flujo de Procesos como se muestra en la Figura 24. Todos los flujos tienen un inicio (componente STAR1) y un fin (componente END\_SUCCESS) y la secuencia de las actividades que se realizarán.

Para cargar el Data Mart primeramente se ejecuta el flujo que carga las dimensiones comunes para al menos dos tablas de hechos y luego se define el orden en el que se cargarán las tablas de hechos, para las cuales se han diseñado también flujos independientes.

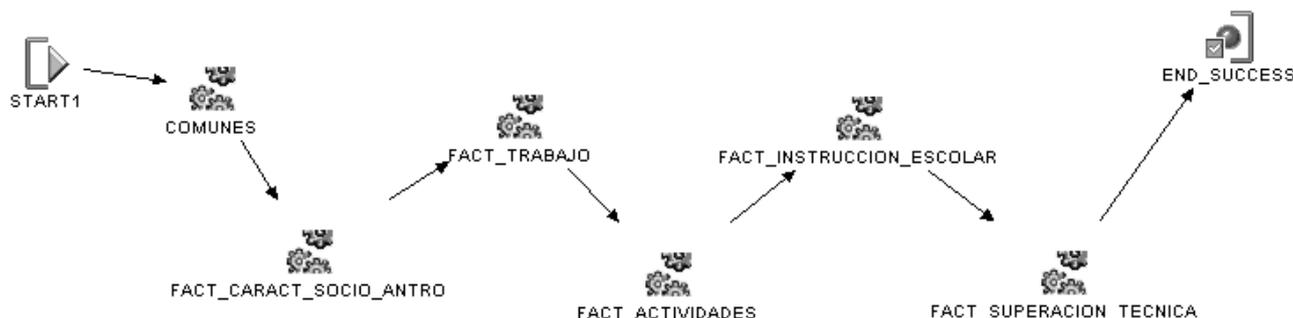


Figura 24: Flujo de Procesos para la Carga Inicial.

El flujo propuesto para cargar las dimensiones comunes como se muestra en la Figura 25, inicia, de ahí se ejecutan las correspondencias, que son algunos de los ETL explicados anteriormente y otros diseñados para los procesos restantes del área, y luego finaliza.

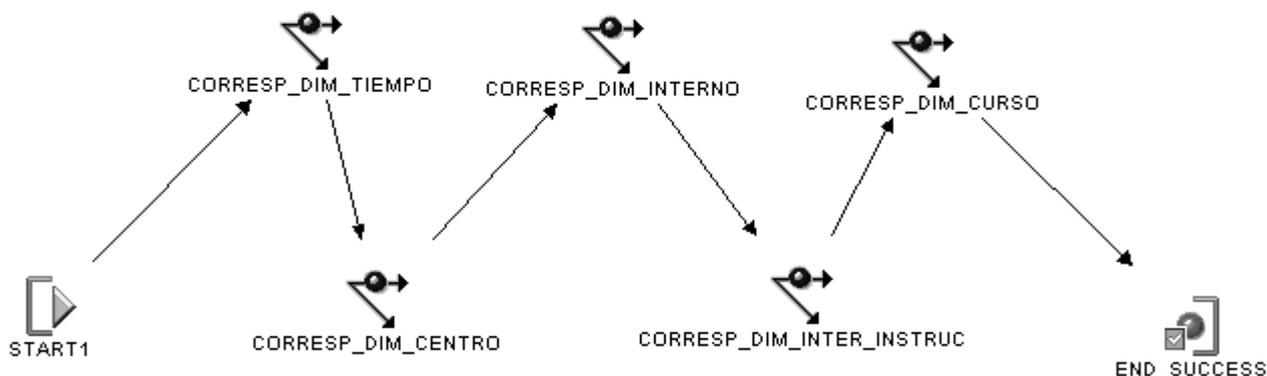


Figura 25: Flujo de procesos para cargar las dimensiones comunes.

A continuación comienzan a cargarse cada uno de los Flujos de las tablas de hechos según el orden definido en la carga inicial.

Estos flujos lo que hacen de manera general es ejecutar paralelamente los ETL de las dimensiones, Staging Área y tablas de hechos correspondientes explicados en los epígrafes 3.3.2 y 3.3.3.

Para diseñar los flujos se utilizan los siguientes operadores del OWB:

- **FORK:** Permite iniciar dos o más actividades, una vez terminada una actividad. Se utilizará para ejecutar varios ETL a la vez.



- **AND:** Especifica la terminación de todas las actividades entrantes antes de iniciar otra actividad.

La Figura 26 muestra el Flujo de Procesos para la tabla de hechos FACT\_TRABAJO: los ETL de las dimensiones Tiempo, Interno y Centro asociados a este proceso se ejecutaron como parte de las dimensiones comunes, se cargan entonces los ETL Unidad\_Productiva y STG\_TRABAJO y finalmente el ETL FACT\_TRABAJO.

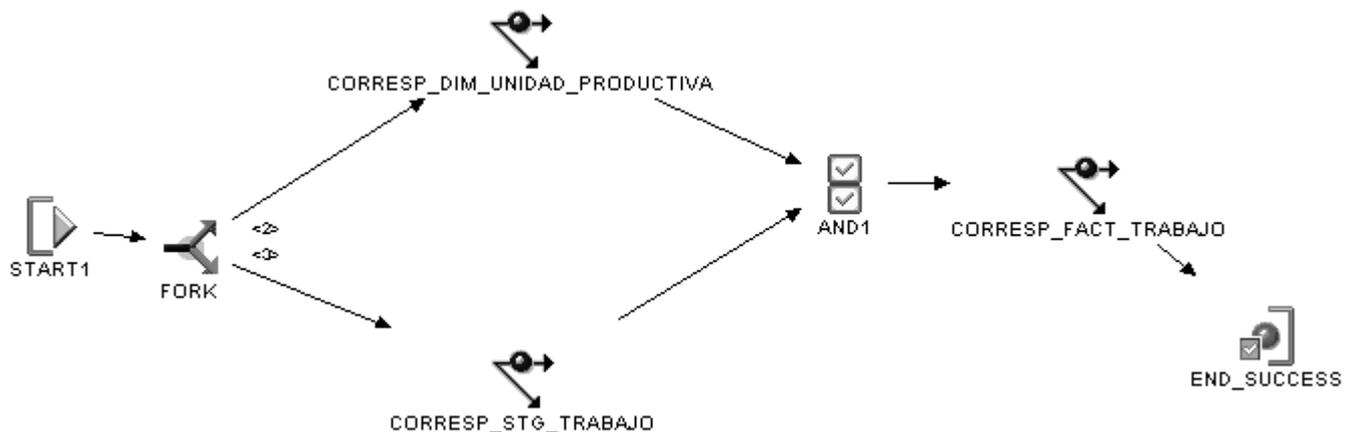


Figura 26: Flujo de Procesos para cargar FACT\_TRABAJO.

La Figura 27 muestra el Flujo de Procesos para cargar la tabla de hechos FACT\_CARAC\_SOCIO\_ANTRO: los ETL de las dimensiones Tiempo, Interno y Centro asociados a este proceso se ejecutaron como parte de las dimensiones comunes, se carga entonces el ETL STG\_CARACT\_SOCIO\_ANTRO; finalmente el ETL FACT\_CARACT\_SOCIO\_ANTRO.

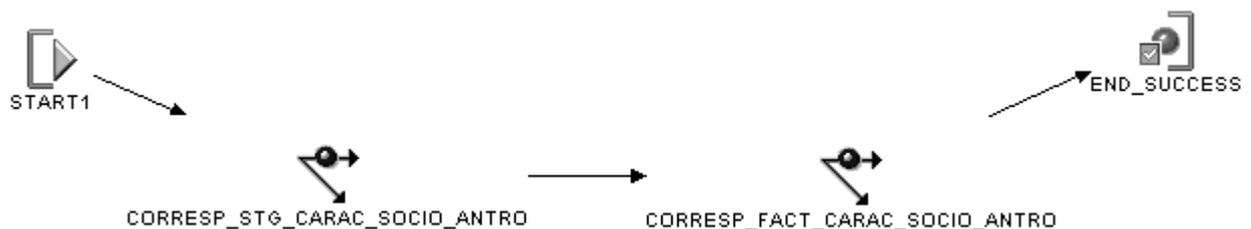
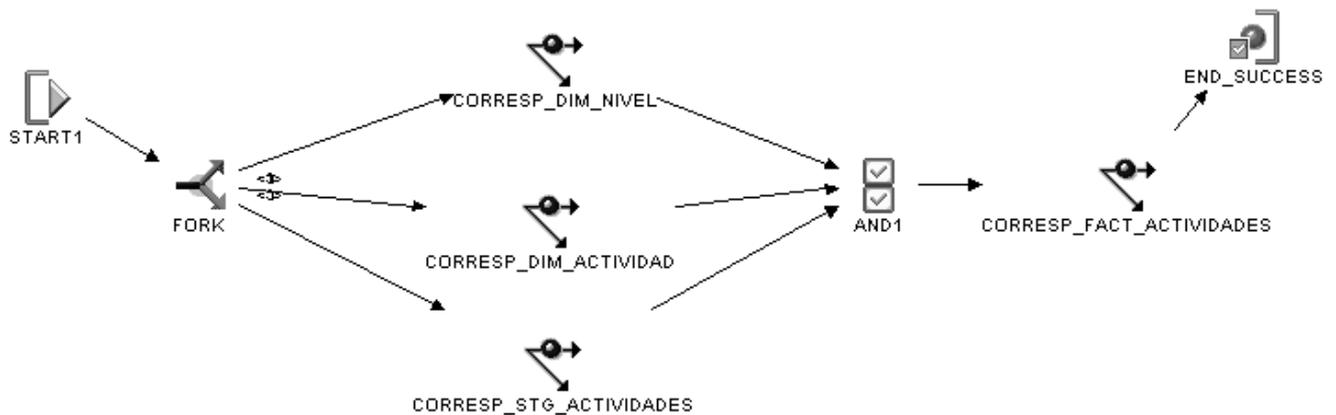


Figura 27: Flujo de Procesos para cargar FACT\_CARACT\_SOCIO\_ANTRO.



La Figura 28 muestra el Flujo de Procesos para cargar la tabla de hechos FACT\_ACTIVIDADES\_EDUCATIVAS: los ETL de las dimensiones Tiempo e Interno asociados a este proceso se ejecutaron como parte de las dimensiones comunes, se cargan entonces los ETL Nivel, Actividad y STG\_ACTIVIDADES; finalmente el ETL ACTIVIDADES.



**Figura 28:** Flujo de Procesos para cargar FACT\_ACTIVIDADES\_EDUCATIVAS.

La implementación de los Flujos de Procesos tiene como ventajas que permite cargar todos los ETL diseñados para las dimensiones, Staging Área y tablas de hechos una única vez y de manera simultánea desde el Flujo de Procesos Carga Inicial, lo cual contribuye al rendimiento del Data Mart y hace más rápido el proceso de carga de los datos.

### 3.3 Conclusiones

En este capítulo se describió todo el proceso de diseño del Data Mart guiado por la metodología Hefesto. Se identificaron y unieron las dimensiones y tablas de hechos, tal como propone el modelo lógico seleccionado de Esquema constelación. Se realizaron los procesos ETL para la carga inicial de las dimensiones, las Staging Área y las tablas de hechos. Se diseñaron además los Flujos de Procesos que permiten ejecutar los ETL realizados.

Con la carga inicial se logró que el Data Mart tenga almacenado las tablas de dimensiones y de hechos con los atributos que permiten analizar los indicadores propuestos para los procesos del área de Sistema Educativo del Sistema Penitenciario Cubano.



## **CONCLUSIONES GENERALES**

Al concluir la investigación se obtiene que:

- Se diseñó el Data Mart utilizando la metodología Hefesto para apoyar la toma de decisiones de los procesos Características Socio – Antropológicas, Trabajo, Actividades Educativas, Instrucción Escolar y Superación Técnica del departamento de Sistema Educativo del Sistema Penitenciario Cubano.
- Se diseñaron y ejecutaron los procesos de Extracción, Transformación y Carga (ETL).
- Se seleccionaron y utilizaron herramientas que responden a las posibilidades tecnológicas de los clientes y desarrolladores.



## **RECOMENDACIONES**

Se recomienda:

- Continuar el desarrollo de la solución, implementando el Data Mart y las herramientas de análisis de reportes.
- Extender la solución a los restantes departamentos del Departamento de Establecimientos Penitenciarios e implementar un Almacén de Datos central con los Data Mart creados.
- Tener en cuenta las nuevas necesidades de información que pueden presentar los usuarios.



## REFERENCIAS BIBLIOGRÁFICAS

1. Sistemas de Soporte a la Decisión (DSS). [En línea] 2007.  
[http://www.sinnexus.com/business\\_intelligence/sistemas\\_soporte\\_decisiones.aspx..](http://www.sinnexus.com/business_intelligence/sistemas_soporte_decisiones.aspx..)
2. **Del Risco Batista, Yanet.** *Sistema de Gestión Penitenciaria de la República de Cuba.* Infraestructura Productiva, Universidad de Ciencias Informáticas.
3. **Moral, Juan Carlos.** EPVNET . [En línea] 23 de 07 de 2007. <http://www.juancarlosmoral.es/epvnet>.
4. **Guerrero, José Iván.** *Evolución histórica del Sistema Penitenciario.* Panamá : s.n.
5. Portal Iberoamericano de la Justicia Electrónica. [En línea]  
<http://www.piaje.org/ES/Services/Paginas/statesservices.aspx?Country=Ecuador>.
6. **Martínez Bravet, René.** “Repositorio de Datos de la Sala Situacional del SIGEP, desde el punto de vista del Control Penal.”. Universidad de Ciencias Informáticas. Ciudad de la Habana,Cuba : s.n., 2010. Tesis.
7. **Blasco Domínguez, Antonio.** Data Warehousing. *Programación en Castellano.* [En línea] 2003.  
[http://www.programacion.com/articulo/data\\_warehousing\\_201/1](http://www.programacion.com/articulo/data_warehousing_201/1).
8. Bill Inmon - Ralph Kimball. [En línea] <http://kle.sisorg.com.mx/articulo04.html>.
9. **Inmon H, William.** *Building the Data Warehouse.* Indianapolis y Canadá : Wiley Publishing, 2005. Vol. 4ta Edición. Inc. 543.9780764599446..
10. **Ricardo Dario, Bernabeu.** *DATA WAREHOUSING: Investigación y Sistematización de Conceptos - HEFESTO: Metodología propia para la Construcción de un Data Warehouse.* V2.0. Córdoba,Argentina : s.n., 2010.
11. **Murillo Alfaro, Félix.** Manual para la Construcción de un Data Warehouse. *INSTITUTO NACIONAL DE ESTADISTICA E INFORMÁTICA.* [En línea] 09 de 1997.  
<http://www.inei.gob.pe/web/metodologias/attach/lib619/131.HTM>.
12. Data Mart. [En línea] 18 de 01 de 2007.  
[http://www.sinnexus.com/business\\_intelligence/datamart.aspx](http://www.sinnexus.com/business_intelligence/datamart.aspx).
13. **Kimball, Ralph.** *El Juego de Herramientas del Almacén de Datos.* New York : John Wiley & Sons, 1996. Inc. 388 p.0471153370.



14. **Martínez Orol, Alfredo.** OLAP y el diseño de cubos. [En línea] 03 de 2007.  
<http://www.gestiopolis.com/canales8/ger/olap-online-analytic-processing.htm>.
15. Sinneuxs. *Bases de datos OLTP y OLAP.* [En línea] Sinergia e Inteligencia de Negocios S.L, 2007.  
[http://www.sinnexus.com/business\\_intelligence/olap\\_vs\\_oltp.aspx](http://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx).
16. **Zorrilla, Marta.** *Data warehouse y OLAP.* Universidad de Cantabria : s.n., 2007/08.
17. **Ponniah, P.** *Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals.* New York, Chichester, Weinheim, Brisbane, Singapore y Toronto : John Wiley & Sons, 2001. Inc. 518 p.0471412546.
18. **Curto Díaz, Josep.** Arquitectura de un datawarehouse. [En línea] 03 de 06 de 2007.  
<http://informationmanagement.wordpress.com/2007/06/03/arquitectura-de-un-datawarehouse/>.
19. **Oracle.** [En línea] <http://www.oracle.com/index.html>.
20. —. *Oracle Warehouse Builder User's Guide 11g Release (11.1).* January 2009. B31278-06.
21. **Lujan Mora, Sergio.** *Data Warehouse Design with UML.* Universidad de Alicante, España : s.n., 2005.
22. **DEP.** *Procedimientos de Trabajo Sistema Educativo.* Departamento de Establecimientos Penitenciarios, MININT. Ciudad de la Habana,Cuba : s.n., 2009.



## BIBLIOGRAFÍA

1. Sistemas de Soporte a la Decisión (DSS). [En línea] 2007.  
[http://www.sinnexus.com/business\\_intelligence/sistemas\\_soporte\\_decisiones.aspx..](http://www.sinnexus.com/business_intelligence/sistemas_soporte_decisiones.aspx..)
2. **Del Risco Batista, Yanet.** *Sistema de Gestión Penitenciaria de la República de Cuba.* Infraestructura Productiva, Universidad de Ciencias Informáticas.
3. **Moral, Juan Carlos.** EPVNET . [En línea] 23 de 07 de 2007. <http://www.juancarlosmoral.es/epvnet>.
4. **Guerrero, José Iván.** *Evolución histórica del Sistema Penitenciario.* Panamá : s.n.
5. Portal Iberoamericano de la Justicia Electrónica. [En línea]  
<http://www.piaje.org/ES/Services/Paginas/statesservices.aspx?Country=Ecuador>.
6. **Martínez Bravet, René.** “Repositorio de Datos de la Sala Situacional del SIGEP, desde el punto de vista del Control Penal.”. Universidad de Ciencias Informáticas. Ciudad de la Habana,Cuba : s.n., 2010. Tesis.
7. **Blasco Domínguez, Antonio.** Data Warehousing. *Programación en Castellano.* [En línea] 2003.  
[http://www.programacion.com/articulo/data\\_warehousing\\_201/1](http://www.programacion.com/articulo/data_warehousing_201/1).
8. Bill Inmon - Ralph Kimball. [En línea] <http://kle.sisorg.com.mx/articulo04.html>.
9. **Inmon H, William.** *Building the Data Warehouse.* Indianapolis y Canadá : Wiley Publishing, 2005. Vol. 4ta Edición. Inc. 543.9780764599446..
10. **Ricardo Dario, Bernabeu.** *DATA WAREHOUSING: Investigación y Sistematización de Conceptos - HEFESTO: Metodología propia para la Construcción de un Data Warehouse.* V2.0. Córdoba,Argentina : s.n., 2010.
11. **Murillo Alfaro, Félix.** Manual para la Construcción de un Data Warehouse. *INSTITUTO NACIONAL DE ESTADISTICA E INFORMÁTICA.* [En línea] 09 de 1997.  
<http://www.inei.gob.pe/web/metodologias/attach/lib619/131.HTM>.
12. Data Mart. [En línea] 18 de 01 de 2007.  
[http://www.sinnexus.com/business\\_intelligence/datamart.aspx](http://www.sinnexus.com/business_intelligence/datamart.aspx).
13. **Kimball, Ralph.** *El Juego de Herramientas del Almacén de Datos.* New York : John Wiley & Sons, 1996. Inc. 388 p.0471153370.



14. **Martínez Orol, Alfredo.** OLAP y el diseño de cubos. [En línea] 03 de 2007.  
<http://www.gestiopolis.com/canales8/ger/olap-online-analytic-processing.htm>.
15. Sinneuxs. *Bases de datos OLTP y OLAP.* [En línea] Sinergia e Inteligencia de Negocios S.L, 2007.  
[http://www.sinnexus.com/business\\_intelligence/olap\\_vs\\_oltp.aspx](http://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx).
16. **Zorrilla, Marta.** *Data warehouse y OLAP.* Universidad de Cantabria : s.n., 2007/08.
17. **Ponniah, Paulraj.** *Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals.* New York, Chichester, Weinheim, Brisbane, Singapore y Toronto : John Wiley & Sons, 2001. Inc. 518 p.0471412546.
18. **Curto Díaz, Josep.** Arquitectura de un datawarehouse. [En línea] 03 de 06 de 2007.  
<http://informationmanagement.wordpress.com/2007/06/03/arquitectura-de-un-datawarehouse/>.
19. **Oracle.** [En línea] <http://www.oracle.com/index.html>.
20. —. *Oracle Warehouse Builder User's Guide 11g Release (11.1).* January 2009. B31278-06.
21. **Lujan Mora, Sergio.** *Data Warehouse Design with UML.* Universidad de Alicante, España : s.n., 2005.
22. **DEP.** *Procedimientos de Trabajo Sistema Educativo.* Departamento de Establecimientos Penitenciarios, MININT. Ciudad de la Habana, Cuba : s.n., 2009.
23. *Data Warehousing For Dummies.* Indianapolis, Indiana : Wiley Publishing. Vol. 2nd Copyright © 2009.
24. **Velasco Hernando, Roberto.** *Estructura de un almacén de datos.* [En línea]  
<http://www.rhernando.net/modules/tutorials/doc/bd/dw.html>.
25. **Varios.** *Importancia de la utilización de un Data Warehouse (DW) en las empresas.*  
<http://www.bibliociencias.cu/gsd/collect/libros/index/assoc/HASH0106/b6fac6b9.dir/doc.pdf>.
26. **Wolf, Carmen.** *Implementando un DataWarehouse.* Ingeniería Informática y Ciencias de la Computación, Universidad de Concepción. . Concepción, Chile : s.n., 2002. Tesis.
27. **Griesemer, Bob.** *Oracle Warehouse Builder 11g Getting Started.* Birmingham, Mumbai : Packt Publishing, 2009.



28. **Ricardo Dario, Bernabeu.** *DATA WAREHOUSING: Investigación y Sistematización de Conceptos - HEFESTO: Metodología propia para la Construcción de un Data Warehouse.* Córdoba, Argentina : s.n., 7 de noviembre de 2007. Vol. V 0.1.
29. **Hernández León, Rolando Alfredo y Coello González, Sayda.** *El proceso de investigación científica.* Ciudad de La Habana : Editorial Universitaria, 2011. ISBN.
30. **Rodríguez Lorenzo, Orestes.** *Almacén de datos para los subsistemas de Reclutamiento y Potencial Humano.* Universidad de Ciencias Informáticas. Ciudad de La Habana, Cuba : s.n., 2010. Tesis.
31. **Pedraza Hernández, Libisdey y Ramos Hernández, Alejandro.** *Diseño de un Almacén de Datos para apoyar la toma de decisiones en el Centro de Informática Médica.* Universidad de Ciencias Informáticas. Ciudad de La Habana, Cuba : s.n., 2010. Tesis.