

**Universidad Nacional de La Matanza****Departamento de Ingeniería e Investigaciones Tecnológicas****Código:** C150**Título del Proyecto:** Implementación de un Data Warehouse para la toma de decisiones en el Área Académica.**Programa de Investigación:** *PROINCE***Director del proyecto:** Ryckeboer, Hugo Ludovico.**Integrantes Del Proyecto:** Sposito, Osvaldo Mario; Castro, Hugo Martín; Matteo, Lorena Romina; Barone, Miriam Andrea Teresa; Etcheverry, Martín Esteban; Gargano, Cecilia Victoria; Bossero, Julio Cesar.**Fecha de Iniciación del Proyecto:** 2013/01/01**Fecha de Finalización del Proyecto:** 2014/12/31**Resumen:**

Este proyecto implementó el primer Data Warehouse (DW cuya mejor traducción al español es Almacén de Datos) de la UNLaM. El mismo, se construyó a partir de la información que generan las aplicaciones que se encuentran en el entorno operativo de la gestión universitaria. Esta información se integra cuando pasa al almacén de datos a través de un proceso conocido como Extraer, Transformar y Cargar (Extract, Transform and Load, frecuentemente abreviado *ETL*). Este acto de integración es siempre una tarea compleja y tediosa que utiliza técnicas para consolidar y administrar información de variadas fuentes, eliminando gran cantidad de datos inútiles o no deseados, con el objetivo de responder consultas a las autoridades departamentales, y tomar decisiones de una forma que antes no era posible.

En este segundo año de trabajo, en principio se explica el desarrollo para la construcción del esquema lógico y físico del DW, el cual se llevó a cabo mediante transformaciones aplicadas al esquema lógico de la base de datos del Sistema de Gestión de Académica SIU-Guaraní. Se explican además, las rutinas empleadas en los proceso ETL, con el objetivo de poder concretar finalmente, el esquema conceptual OLAP del DW. Este esquema representa la información como matrices multidimensionales o cuadros de múltiples entradas denominados *cubos*. A los ejes de la matriz se los llama *dimensiones* y representan los criterios de análisis, y a los datos almacenados en la matriz se los denomina *medidas* y representan los indicadores o valores a analizar. Por último, se muestra como se realizaron los diferentes cubos o Data Marts.

Data Warehouse, Data Mart, ETL, OLAP, Multidimensional.



Área de conocimiento: Ingeniería de Comunicaciones, Electrónica y Control

Código de Área de conocimiento: 1800

Disciplina de conocimiento: Computación

Código Disciplina de conocimiento: 1802

Campo de Aplicación: Computación

Código Campo de Aplicación: 1802

Otras dependencias de la UNLaM que intervienen en el Proyecto: No aplica.

Otras instituciones externas a la UNLaM intervinientes: No aplica.

Otros proyectos con los que se relaciona: No aplica.

**Título del Proyecto:**

Implementación de un Data Warehouse para la toma de decisiones en el Área Académica.

Resumen del Proyecto:

Este proyecto continúa con línea de investigación propuesta por las autoridades del DIIT, para poder obtener indicadores de gestión para la toma de decisiones estratégicas tanto en relación con los estudiantes como con los docentes. Al reconocer la importancia del manejo de información en la gestión de toma de decisiones, fue necesario construir un repositorio que permita de una manera práctica el análisis de información para que de esta manera permita agilizar, facilitar y personalizar las consultas para dar soluciones a los requerimientos de los usuarios finales.

En esta última etapa del proyecto nos concentramos en el diseño lógico y físico de los distintos Cubos OLAP y en la construcción de rutinas que conformen el proceso ETL¹, este proceso permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, Data Mart, o Data Warehouse, para apoyar un proceso de negocio. Estos Cubos pretenden brindar un servicio realizando un nuevo enfoque en la reutilización de la información histórica, esto es muy importante para saber dónde y cómo solventar alguna falla, o mejorar las decisiones tomadas.

Palabra clave:

Cubo OLAP, Data Warehouse, Data Mart, Proceso ETL.

¹Del inglés **Extract, Transform and Load** (“extraer, transformar y cargar”)



Memoria Descriptiva



Etapa 3. El Manejo los datos

Introducción al Modelo Multidimensional

Como ya se mencionó, el término Data Warehouse se traduce literalmente como almacén o repositorio de datos. La ventaja principal de este tipo de sistemas se basa en su concepto fundamental, **la estructura de la información**. Este concepto significa el almacenamiento de información homogénea y fiable, en una estructura basada en la consulta y el tratamiento jerarquizado de la misma, y en un entorno diferenciado de los sistemas operacionales. Es una base de datos orientada al análisis que se caracteriza por integrar y depurar información de uno o más orígenes, para luego procesarla permitiendo su análisis desde una infinidad de perspectivas y con mejores velocidades de respuesta respecto a los sistemas transaccionales. En este tipo de repositorios el almacenamiento de la información se realiza generalmente en bases de datos del tipo OLAP², éste tipo de persistencia de la información permite al usuario obtener una visión multidimensional de la información, para cada actividad que es objeto de análisis, permitiendo la consulta y el tratamiento jerarquizado de la misma, siempre en un entorno diferente a los sistemas operacionales del tipo OLTP³. [1]

Las bases de datos multidimensionales son una variación del modelo relacional que utiliza cubos OLAP para organizar los datos y expresar las relaciones entre ellos, las principales ventajas de este tipo de bases de datos son la versatilidad para cruzar información y la alta velocidad de respuesta. [Kim98].

El modelado multidimensional es una técnica de diseño lógico que busca organizar los datos en un marco estándar, que sea intuitivo y de acceso rápido. [Kim02], estos modelos brindan la habilidad de visualizar datos en una forma que facilita la comprensión de los mismos. Este es el modelo utilizado para este proyecto de Data Warehousing⁴, en donde se organizan y presentan los datos definiendo dimensiones, líneas o áreas temáticas del negocio, de forma que permitan analizar la información a distintos niveles de agregación o sumalización para cada análisis, a estos niveles de granularidad⁵ se los caracteriza con el nombre de atributos. [Mic06] En el modelo multidimensional, los datos se pueden percibir como un cubo, donde cada celda contiene una medida o un conjunto de medidas de interés (probablemente agregadas) que se analizan de acuerdo a un conjunto de dimensiones de interés, ver figura nro. 1. Estas dimensiones normalmente están organizadas en jerarquías que favorecen el proceso de agregación de la información, por ejemplo, si definimos un punto en el espacio por la intersección de sus coordenadas X, Y, Z. Si le asignamos

² En inglés On-Line Analytical Processing (Procesamiento analítico en línea)

³Procesamiento de Transacciones en Línea (On-Line Transaction Processing, OLTP)

⁴ Es el proceso de extraer y filtrar los datos de las operaciones comunes a la organización, procedentes de los distintos sistemas de información y/o sistemas externos, para transformarlos, integrarlos y almacenarlos en un depósito o almacén de datos o DW.

⁵ Se refiere a la especificidad a la que se define un nivel de detalle en una tabla, es decir, si hablamos de una jerarquía la granularidad empieza por la parte más alta de la jerarquía, siendo la granularidad mínima, el nivel más bajo.



valores a esas coordenadas, digamos X representa Alumno, Y representa el tiempo y Z representa la asignatura, luego consideramos la siguiente combinación X = Alumno (Pedro Gómez); Y = Fecha (5/12/2013); Z = Materia (Base de Datos), obtendremos un punto en el espacio el cual podríamos definir como la Nota obtenida en el examen de ese día. A esta estructura también se la denomina Hipercono o Cubos OLAP.

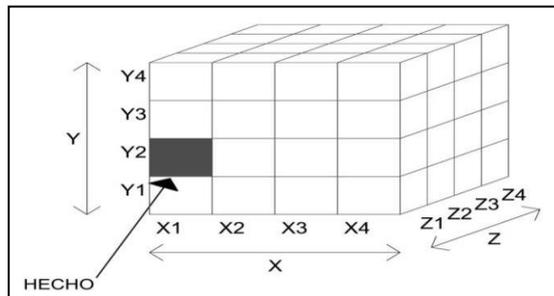


Figura 1. Ejemplo de un Cubo genérico

Componentes del Modelo Multidimensional

A continuación se detallan los componentes utilizados para este proyecto del Modelo Multidimensional y se explica cómo interactúan entre sí, para ofrecer a los estrategas de negocios la información necesaria para una toma de decisiones acertada. En este punto, se confeccionará el modelo lógico de la estructura del DW, teniendo como base el modelo conceptual que se obtuvo de las Etapas I y II de este proyecto. Para ello, primero se definirán que son las tablas de hechos y de dimensiones.

Tabla de Hechos

Es el núcleo del modelo multidimensional. Según el caso, los hechos pueden ser atómicos, es decir, almacenados con el máximo nivel de detalle del sistema OLTP, por ejemplo cada una de las notas de los exámenes finales de los alumnos o bien se pueden guardar los datos ya agregados⁶, como los totales de alumnos aprobados, desaprobado, alumnos ausentes, etc. como resultados de los distintos exámenes finales. Estas agregaciones son las que se forman al combinar uno o más hechos con alguna operación matemática o lógica y que también residen en una tabla de hechos. Estos datos poseen la ventaja de almacenarse previamente calculados, por lo cual pueden ser accedidos a través de consultas SQL sencillas y devolver resultados rápidamente, pero requieren más espacio físico en el DW, además de necesitar más tiempo de proceso en los ETL⁷ que los calculan. Estos se conocen como hechos básicos y derivados respectivamente. Resumiendo, las tablas de hechos contienen la información que será utilizada por los analistas de negocio para apoyarse en el proceso de toma de decisiones, en estas tablas es donde se almacenan las

⁶ Un registro agregado de la tabla de hechos representa totalización (sumarización) de registros de una tabla de hechos de nivel básico. Este está siempre asociado con uno o más registros de tablas de dimensión agregadas

⁷ En inglés Extract, Transform and Load (Extraer, Transformar y Cargar), es el proceso que permite mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otro repositorio.



mediciones numéricas del negocio (métricas o medidas). Contienen datos cuantitativos. [1]

Una tabla de Hechos está compuesta principalmente por dos tipos de campos: las claves foráneas de las dimensiones y las métricas. Estas medidas se hacen sobre el grano o unidad básica de la tabla. El grano o la granularidad de la tabla queda determinada por el nivel de detalle que se almacenará en la tabla. Mientras mayor sea el nivel de detalle de la información, se tendrán mayores posibilidades analíticas, ya que los mismos podrán ser resumidos o sumariados. Es decir, los datos que posean granularidad fina (nivel de detalle) podrán ser resumidos hasta obtener una granularidad media o gruesa. No sucede lo mismo en sentido contrario, ya que por ejemplo, los datos almacenados con granularidad media podrán resumirse, pero no tendrán la facultad de ser analizados a nivel de detalle. O sea, si la granularidad con que se guardan los registros es a nivel de día, estos datos podrán resumirse por semana, mes, semestre y año, en cambio, si estos registros se almacenan a nivel de mes, podrán sumariarse por semestre y año, pero no lo podrán hacer por día y semana.

Cada medida es interpretada como la intersección de todas las dimensiones que la definen. Toda tabla de hechos representa una relación muchos a muchos entre las dimensiones y contiene un conjunto de dos o más claves foráneas que la vinculan a sus respectivas tablas de dimensión.

Tabla de Dimensiones:

El modelo multidimensional organiza y presenta los datos definiendo dimensiones. Estas son líneas o áreas temáticas del negocio. Cada elemento de una dimensión tiene una clave primaria⁸ que lo identifica en el mayor nivel de detalle y un conjunto de atributos. Llevado al ámbito universitario y para el análisis de información académica de alumnos se podrían considerar ejemplo de dimensiones a las carreras, las materias que se cursan, los colegios del que provienen los alumnos, etc.

Las tablas de dimensión son un conjunto de tablas compañeras o guías para una Tabla de Hechos. Cada dimensión es definida por su clave primaria que sirve como base para la integridad referencial con cualquier Tabla de Hechos a la cual se une. La mayoría de las tablas de dimensión contienen muchos atributos o campos de texto que sirven para restringir y agrupar las consultas del DW y determinan los niveles de análisis. El objetivo de los atributos de estas tablas es servir como filtros o encabezamientos de las consultas de usuarios. Por ejemplo al consultar cantidad de alumnos de un determinado año discriminados por carreras, la dimensión año el nombre se está usando como restricción, para filtrar un valor, y la dimensión carreras como encabezamiento del reporte. Cada dimensión puede referirse a conceptos como '*tiempo*', '*cursos*', '*materias*', '*zona geográfica*', etc. Ahora bien, cada dimensión puede estar medida de diferentes maneras según la granularidad deseada, por ejemplo, para la dimensión **procedencia del alumno** podríamos considerar '*localidades*'

⁸ Se llama clave primaria a un campo o a una combinación de campos que identifica de forma única a cada fila de una tabla. No puede haber dos filas en una tabla que tengan la misma clave primaria.



'departamento', 'provincias', 'regiones', 'países' o 'continentes'. Las dimensiones se agrupan en jerarquías mediante relaciones uno-a-muchos. Una localidad agrupa a muchos alumnos. Un partido agrupa a muchas localidades. Una provincia está formada por varios partidos. Etcétera. En la figura 2 se muestra un ejemplo.



Figura 2. Jerarquía geográfica

Tipos de Modelos de un DW

Un modelo de datos es un conjunto de conceptos, reglas y convenciones que permiten describir y manipular la estructura de una base los datos que queremos almacenar en una base de datos. En ellas se pueden usar distintos modelos de datos para describir la información con que operan. Un Modelo de Datos permite describir:

- La estructura de datos de la base: el tipo de los datos que hay en la base y la forma en que se relacionan.
- Las restricciones de integridad: Un conjunto de condiciones que deben cumplir los datos para reflejar correctamente la realidad deseada. [Elm97]

Los modelos estrella y copo de nieve son estructuras bajo las cuales se guía la creación de un Data Warehouse.

Esquema en Estrella

El esquema en estrella, consta de una tabla de hechos central y de varias tablas de dimensiones relacionadas a esta, a través de sus respectivas claves. En la siguiente figura se puede apreciar un esquema en estrella estándar:

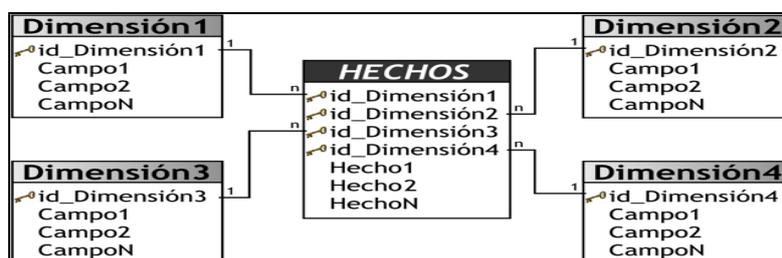


Figura 3. Esquema Estrella

Esquema Copo de Nieve

Como se puede apreciar en la figura 4, existe una tabla de hechos central que está relacionada con una o más tablas de dimensiones, quienes a su vez pueden estar relacionadas o no con una o más tablas de dimensiones. Uno de los motivos principales de utilizar este tipo de modelo, es la posibilidad de segregar los datos de



las tablas de dimensiones y proveer un esquema que sustente los requerimientos de diseño. Otra razón es que es muy flexible y puede implementarse después de que se haya desarrollado un esquema en estrella.

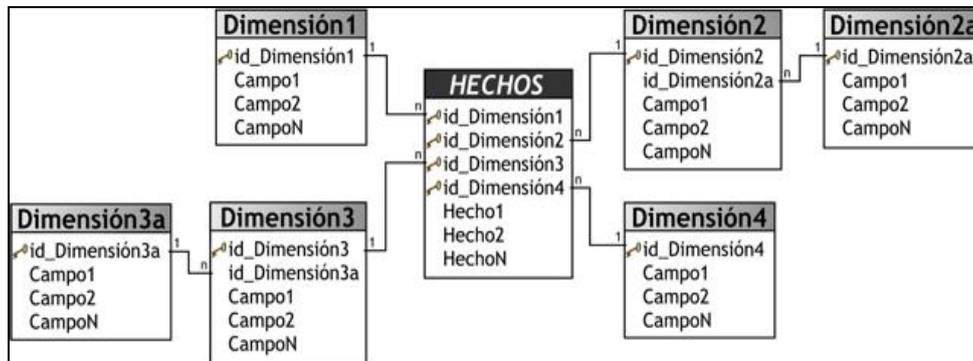


Figura 4. Esquema Copo de nueve.

Tipos de implementación de un DW

Un DW y las herramientas OLAP se pueden basar físicamente en varias tipos de organizaciones físicas: [2]

ROLAP

Este tipo de organización física se implementa sobre tecnología relacional, pero disponen de algunas facilidades para mejorar el rendimiento. ROLAP (Relational On Line Analytic Processing - Procesamiento Analítico en línea Relacional) cuenta con todos los beneficios de una SGBD⁹ Relacional a los cuales se les provee extensiones y herramientas para poder utilizarlo como un Sistema Gestor de DW. En los sistemas ROLAP, los cubos multidimensionales se generan dinámicamente al instante de realizar las diferentes consultas, haciendo de esta manera el manejo de cubos transparente los usuarios.

MOLAP

MOLAP "Multidimensional Online Analytical Processing – Procesamiento Analítico Multidimensional en Línea" es un proceso analítico en línea OLAP que indexa directamente en una base de datos multidimensional. Los procesos MOLAP usan información que ya están almacenados en un arreglo multidimensional, en el cual todas las posibles combinaciones de los datos se reflejan, cada uno en una celda que puede ser accedido directamente.

⁹ Sistema de gestión de Base de Datos) o en inglés Database management system (DBMS).



HOLAP

HOLAP (Hybrid On Line Analytic Processing o Procesamiento Analítico en Línea Híbrido) constituye un sistema que surge de la unión o mezcla de dos sistemas distintos como lo son MOLAP y ROLAP, que combina estas dos implementaciones para almacenar algunos datos en un motor relacional y otros en una base de datos multidimensional.

Los datos agregados y pre-calculados se almacenan en estructuras multidimensionales y los de menor nivel de detalle en estructuras relacionales. Es decir, se utilizará ROLAP para navegar y explorar los datos, y se empleará MOLAP para la realización de tableros. Como contrapartida, hay que realizar un buen análisis para identificar los diferentes tipos de datos.

ROLAP vs. MOLAP (Comparativa)

Cuando se comparan las dos arquitecturas, se pueden realizar las siguientes observaciones: [3]

- El ROLAP delega la negociación entre tiempo de respuesta y el proceso batch al diseño del sistema. Mientras, el MOLAP, suele requerir que sus bases de datos se pre-compilen para conseguir un rendimiento aceptable en las consultas, incrementando, por tanto los requerimientos batch.
- Los sistemas con alta volatilidad de los datos (aquellos en los que cambian las reglas de agregación y consolidación), requieren una arquitectura que pueda realizar esta consolidación ad-hoc. Los sistemas ROLAP soportan bien esta consolidación dinámica, mientras que los MOLAP están más orientados hacia consolidaciones batch.
- Los ROLAP pueden crecer hasta un gran número de dimensiones, mientras que los MOLAP generalmente son adecuados para diez o menos dimensiones.
- Los ROLAP soportan análisis OLAP contra grandes volúmenes de datos elementales, mientras que los MOLAP se comportan razonablemente en volúmenes más reducidos (menos de 5 Gb)

Por ello, y resumiendo, el ROLAP es una arquitectura flexible y general, que crece para dar soporte a amplios requerimientos OLAP. El MOLAP es una solución particular, adecuada para soluciones departamentales con unos volúmenes de información y número de dimensiones más modestos.



Modelo Lógico

En la arquitectura propuesta para nuestro Data Warehouse, en la Etapa II de este proyecto, se eligió la metodología propuesta por Kimball, quien define a un Data Mart como “...una porción de la torta completa que representaría el DW...”. Esta metodología denominada “**Bottom-Up**” comienza por el desarrollo de un Data Mart para construir luego un DW como unión de estos. [Kim02]

Los Data Marts que son pequeños Data Warehouses orientados a áreas específicas de una organización. Se puede entender un Data Mart como un subconjunto de los datos del Data Warehouse Corporativo con el objetivo de responder a un determinado análisis, función o necesidad y con una población de usuarios específica. Al igual que en un DW, los datos están estructurados en modelos de estrella o copo de nieve y un Data Mart puede ser dependiente o independiente de un Data Warehouse, es decir, la diferencia que existe entre un Data Mart y un DW es el alcance. Un Data Mart está pensado para cubrir las necesidades de un grupo de trabajo o de un determinado departamento dentro de la organización, es el almacén natural para los datos departamentales, en cambio, el ámbito del DW es la organización en su conjunto. Es el almacén natural para los datos corporativos comunes. Tanto a partir de los Data Marts como del Data Warehouse, se construyen reportes y gráficos en base a herramientas de extracción de información. Estas herramientas de extracción de información pueden ser de diversos tipos, pero típicamente permiten manipular la información de forma similar a una planilla de cálculo.

En este apartado se detallará el modelo lógico de los Data Marts propuesto para este trabajo, en la figura 5 ilustra la arquitectura propuesta de diseño para la construcción del esquema lógico, tomando como entrada un esquema conceptual multidimensional y una base de datos fuente previamente integrada. Para un estudio más detallado de las técnicas de modelado dimensional puede encontrarse información en [Kim98] y [Kim02].

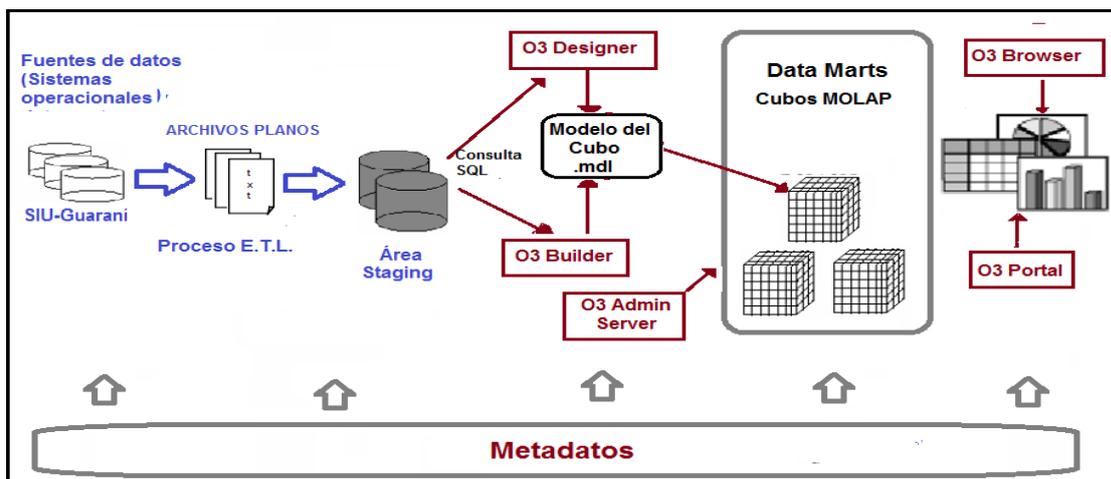


Figura 5. Arquitectura del DW propuesta.



Descripción de la Tabla de Hechos

Los hechos se usan para definir qué es lo que se quiere medir, para Kimball, todos los hechos candidatos en un diseño deben ser fieles a la granularidad definida en el modelado dimensional y los hechos que pertenezcan claramente a otra granularidad deben separarse en una tabla de hechos diferente. En forma similar, a la elección de las dimensiones, se especifican los hechos posibles que el DM debe contener para satisfacer las necesidades de información de los analistas de negocio. Los scripts de creación de las tablas se describen en el Anexo II.

DWH001_FT_RendimientoAcademico. En la figura 6 se muestra los atributos que contiene la tabla de hecho La idea principal de esta Tabla es registrar la actividad académica de los alumnos en la universidad. Para esto se consideran datos en tres niveles diferentes de análisis:

- Calificaciones del alumno en cada uno de los años académicos en que ha tenido actividad dentro de la carrera.
- A qué tipo de cursada corresponde dicha nota, si pertenece a la cursada actual o a una materia de otro periodo
- y por último el tipo de nota, si es la nota de una cursada, de un recuperatorio o de un examen final.

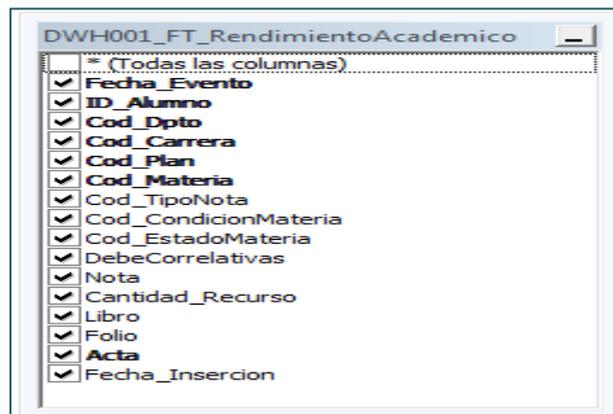


Figura 6. Tabla de Hecho Rendimiento Académico.

DWH002_FT_HistorialAlumno. En la siguiente figura se observa los atributos que contiene esta tabla con el objeto de tener los datos de la última materia aprobada por el alumno, y los demás atributos requeridos por la etapa de relevamiento:



Figura 7. Tabla de Hecho Historial Alumno



DWH002_FT_OfertaDemandaCarrera. En esta tabla se almacena los datos estadísticos de la oferta/demanda de vacantes para cada carrera del DIIT.:

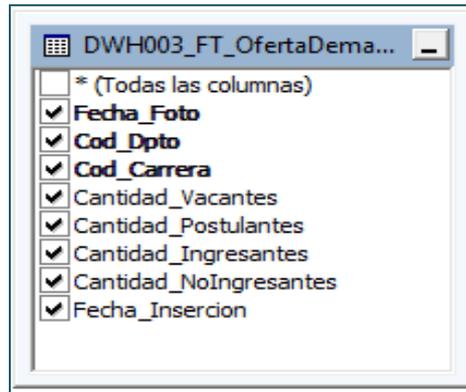


Figura 8. Tabla de Hecho Oferta Demanda Carrera.

DWH002_FT_HistorialPaseAlumno. En esta tabla se almacena los datos históricos del movimiento de los alumnos respecto a las carreras que elige.

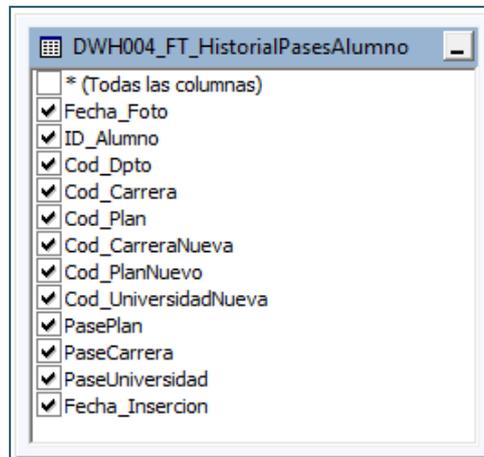


Figura 9. Tabla de Hecho Historial Pases Alumnos

DWH002_FT_HistorialLaboralAlumno. En esta tabla se almacena los datos relacionados a la actividad laboral de los alumnos.

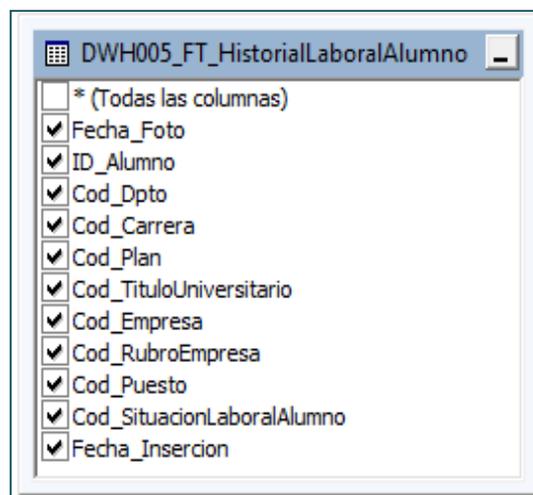


Figura 10. Tabla de Hecho



Tablas de Dimensiones

A continuación se explican las dimensiones más significativas del modelo, La descripción de la totalidad de las estructuras de datos se detallan también en el Anexo II junto con los scripts de creación de las tablas.

Dimensión DWH006_LK_Alumno

Esta dimensión contiene los datos de identidad de cada uno de los alumnos que estudian en el DIIT. Como se ve en la figura 11, en este modelo cada alumno vendrá identificado de manera inequívoca por el atributo ID_ALUMNO, que es un atributo compuesto por el tipo de documento más el número del mismo. Cada alumno tendrá una fila en la tabla de hechos, cada vez que curse una asignatura o se presente a rendir un examen.

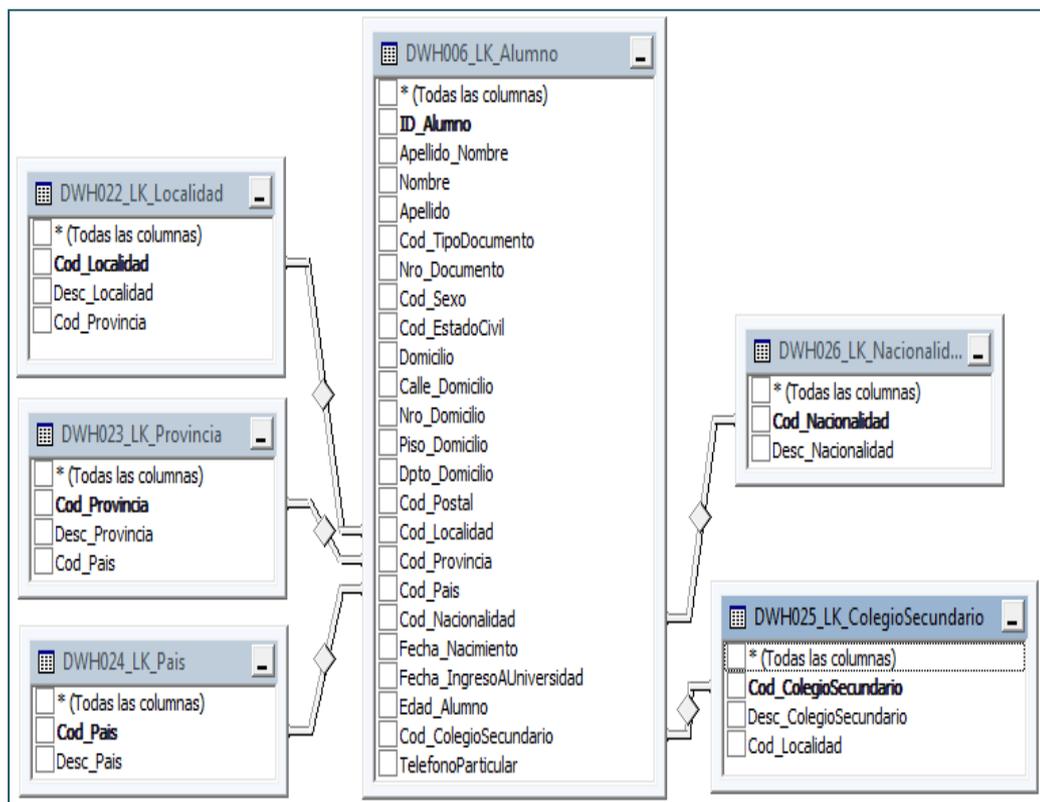


Figura 11. Esquema Estrella de la Dimensión DWH006_LK_Alumno.

El modelo permite a través de las demás dimensiones mostradas en la figura anterior, poder realizar análisis comparativos, por ejemplo, el sexo y la fecha de nacimiento, permitirán al usuario sacar resultados en ciertas asignaturas o carreras por el género del individuo o la edad que tiene en el momento de la realización del examen. Los atributos colegio secundario dará la posibilidad de realizar interesante estudios relacionados con la procedencia de los mismos. La *nacionalidad* permitirá conocer el grado de alumnos extranjeros que asisten a la universidad.



Dimensión DWH010_LK_MateriaPlan

Esta Dimensión junto con las tablas de dimensiones DWH0020_LK_Plan, DWH009_LK_Materia y DWH008_LK_Carrera, permiten identificar de que Carrera y de que Plan de Estudios¹⁰ es la Materia que está cursando/rindiendo en el momento del registro. Casi todas las carreras de la UNLaM, han sufrido modificaciones, en algún año, en su plan de estudio original.

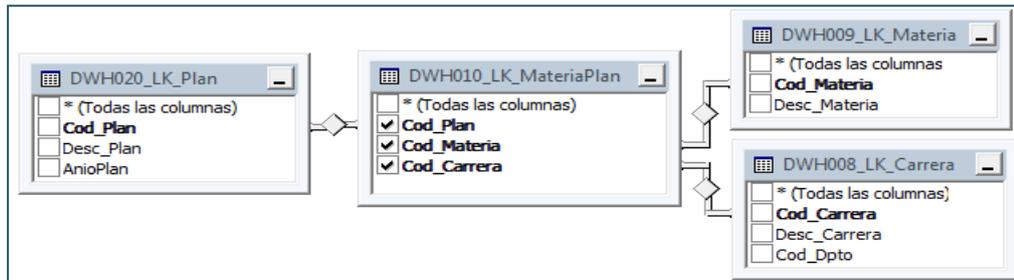


Figura 12. Modelo de datos lógico de la Dimensión DWH010_LK_MateriaPlan.

Dimensión DWH007_LK_Departamento y Dimensión DWH014_LK_Universidad

La Dimensión Departamento y Universidad sirven para poder identificar a que Departamento pertenece dicha asignatura.

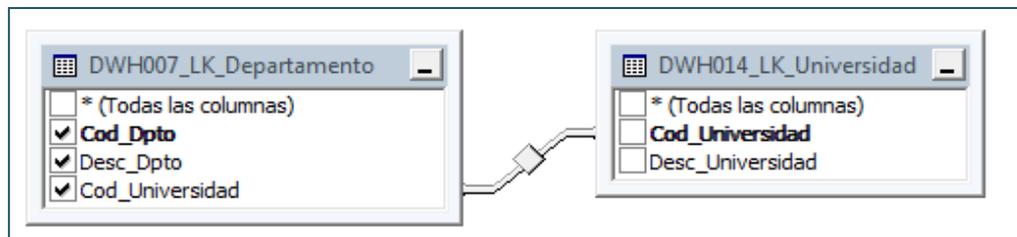


Figura 13. Dimensiones Departamento y Universidad.

Dimensión DWH013_LK_TipoNota

Como se mencionó anteriormente, se ingresa una fila cada vez que se registra una nota del alumno. Los posibles valores para nuestro desarrollo son:

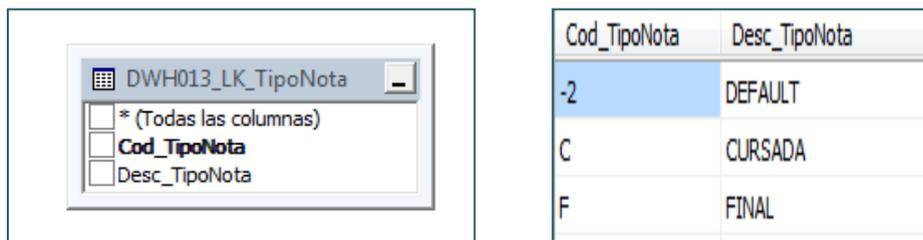


Figura 14. Dimensión DWH013_LK_TipoNota y los Posibles valores.

¹⁰ El plan de estudios es el esquema estructurado de las áreas obligatorias y fundamentales y de áreas optativas con sus respectivas asignaturas que forman parte del currículo de los establecimientos educativos.



Dimensión DWH013_LK_CondicionMateria

Al igual que en el caso anterior, la materia que se rinde puede tener 2 condiciones posibles al momento de registrarse la nota

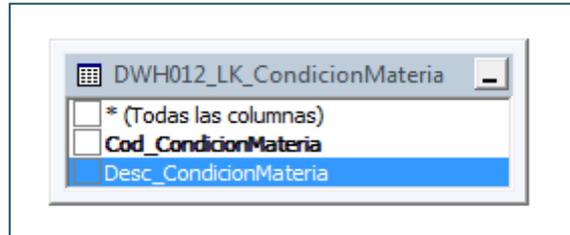


Figura 15. Modelo de datos lógico de la Dimensión DWH013_LK_CondicionMateria.

Los posibles valores para nuestro desarrollo son:

Cod_Condicion...	Desc_CondicionMateria
-2	DEFAULT
L	LIBRE
R	REGULAR

Figura 16. Posibles valores de la Dimensión DWH013_LK_CondicionMateria.

Dimensión DWH011_LK_EstadoMateria

Cada Nota va a corresponder a un Estado en particular.

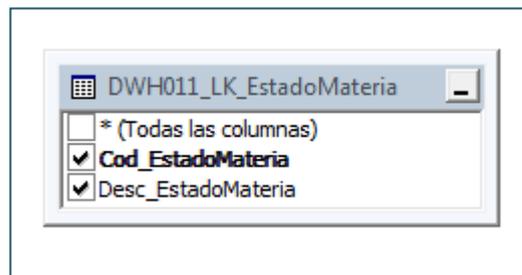


Figura 17. Modelo de datos lógico de la Dimensión DWH011_LK_EstadoMateria.

Los posibles valores para nuestro desarrollo son:

Cod_EstadoMa...	Desc_EstadoMateria
0	DEFAULT
2	CURSADA APROBADA
3	CURSADA REPROBADA
4	CURSADA AUSENTE
5	CURSADA PROMOCIONADA
6	FINAL APROBADA
7	FINAL REPROBADA
8	FINAL AUSENTE

Figura 18. Posibles valores de la Dimensión DWH011_LK_EstadoMateria.



Establecer Relaciones.

Como hemos visto en el punto anterior, estamos utilizando la metodología desarrollada por Kimball (y su enfoque dimensional). Ralph Kimball define a un Data Warehouse como un conglomerado de todos los Data Marts (DM) dentro de una organización, siendo una copia de los datos transaccionales estructurados de una forma especial para el análisis, que incluye las dimensiones de análisis y sus atributos, su organización jerárquica, así como los diferentes hechos de negocio que se quieren analizar. Por un lado tenemos tablas para las representar las dimensiones y por otro lado tablas para los hechos (Facts Tables).

Para facilitar el análisis, el data mart organiza los datos en una estructura que está compuesta por una tabla central - tabla de hechos - y un conjunto de tablas organizadas alrededor de ésta - tablas de dimensiones. En las tablas de dimensión que contienen los atributos de las aperturas que interesan al negocio que se pueden utilizar como criterios de filtro y son relativamente pequeñas. Cada tabla de dimensión se vincula con la tabla de hechos por un identificador

Las dimensiones pueden ser:

- ✓ Locales
- ✓ Compartidas

Las **dimensiones locales** son las que se definen y se utilizan dentro de un mismo cubo.

Las **dimensiones compartidas** son aquellas dimensiones que se definen independientes de los cubos y pueden ser utilizadas por varios de ellos.

Ventajas de las dimensiones compartidas

- ✓ Evitamos duplicar dimensiones locales
- ✓ Aseguramos que los datos analizados estén organizados de la misma forma en todos los cubos, lo que implica un menor costo de mantenimiento.
- ✓ Desventajas de las dimensiones compartidas
- ✓ Deben emplearse del mismo modo en los cubos que las usen.
- ✓ Un cambio implica que la dimensión deberá ser modificada en todos los cubos

Los diferentes Data Marts están conectados entre sí por la llamada bus structure, que contiene los elementos anteriormente citados a través de las dimensiones conformadas (que permiten que los usuarios puedan realizar queries conjuntos sobre los diferentes DM, pues este bus contiene los elementos en común que los comunican). Una dimensión conformada puede ser, por ejemplo, la dimensión cliente, que incluye todos los atributos o elementos de análisis referentes a los clientes y que puede ser compartida por diferentes data marts (ventas, pedidos, gestión de cobros, etc.).



Esquema de Relación del Cubo Rendimiento Académico

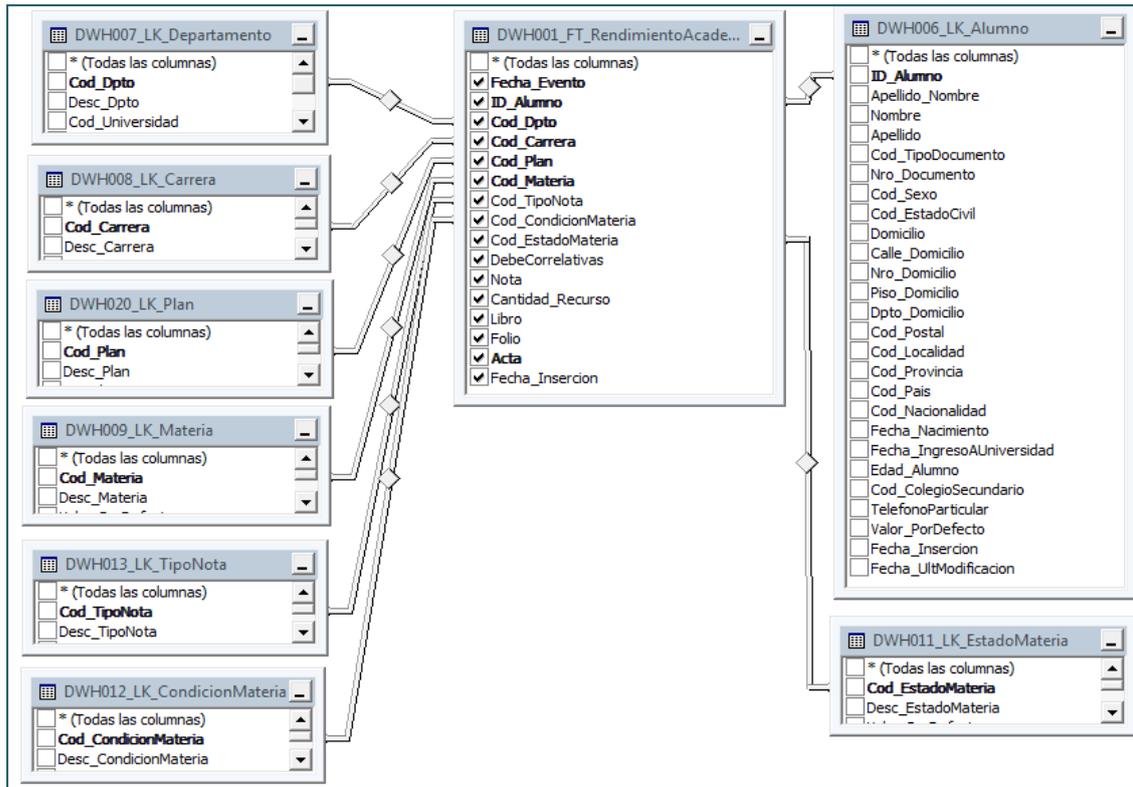


Figura 19. Relación Rendimiento Académico

Esquema de Relación del Cubo HistorialAlumno

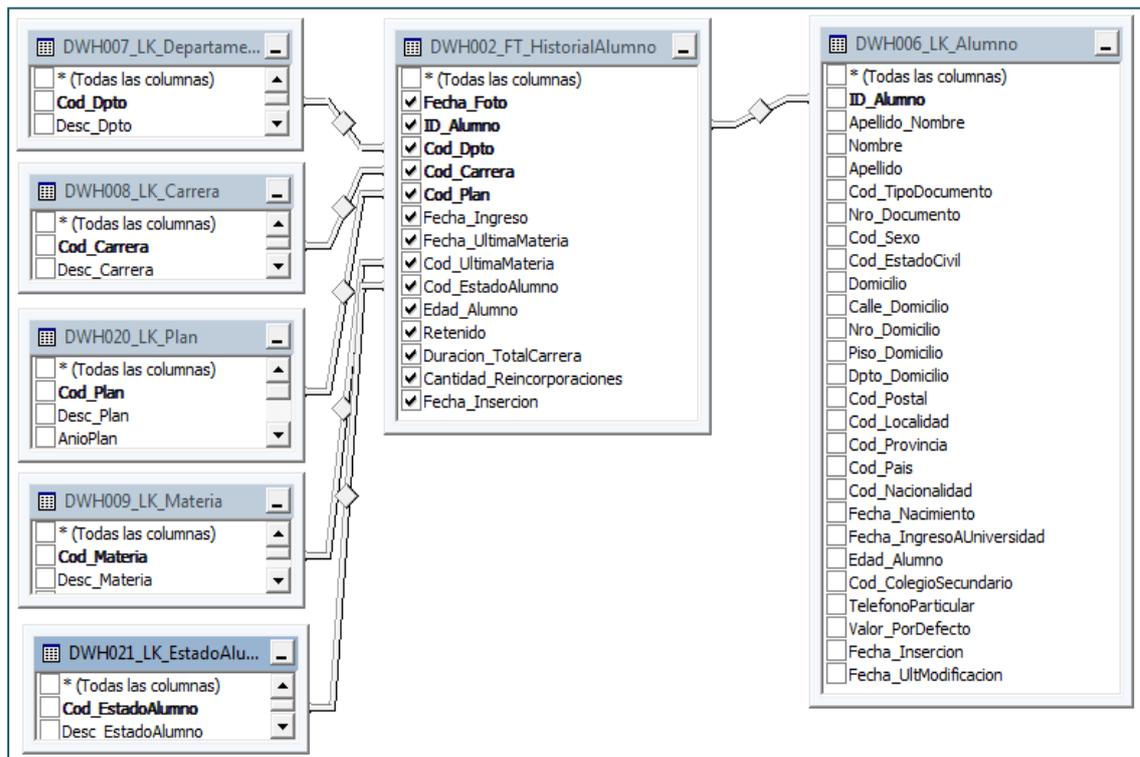


Figura 20. Relación Historial Alumno



Esquema de Relacione del Cubo HistorialLaboralAlumno

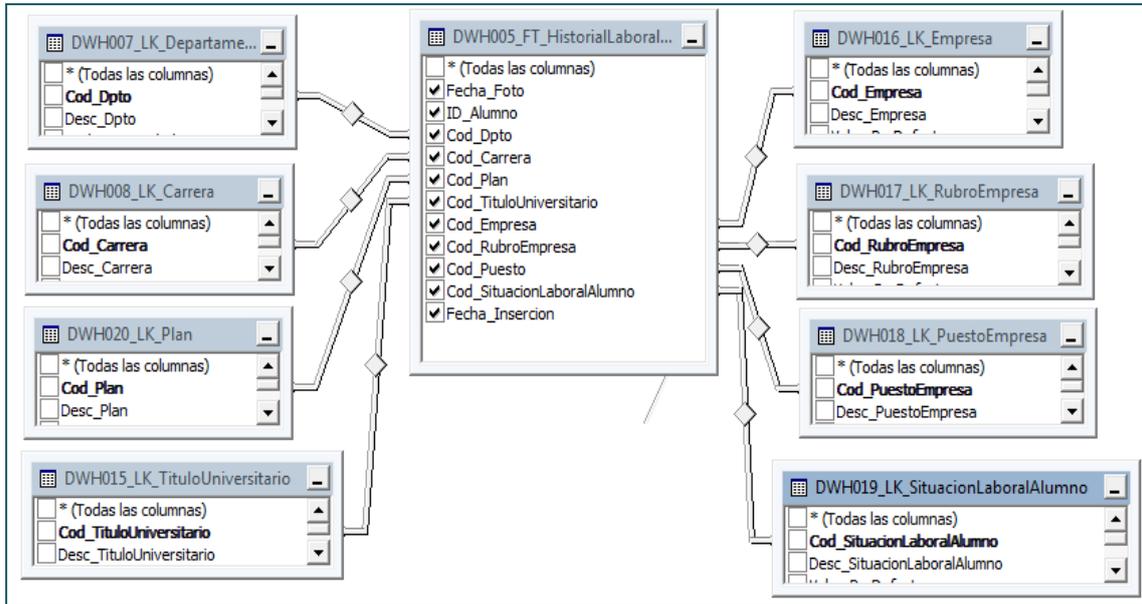


Figura 21. Relación HistorialLaboralAlumno

Esquema de Relacione del Cubo Oferta Demanda Carrera

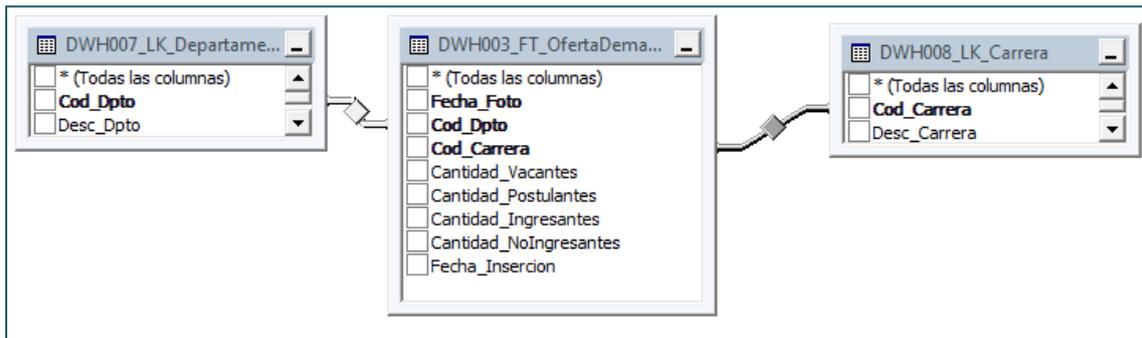


Figura 22. Relación Oferta Demanda Carrera

Esquema de Relacione del Cubo Historial Pase

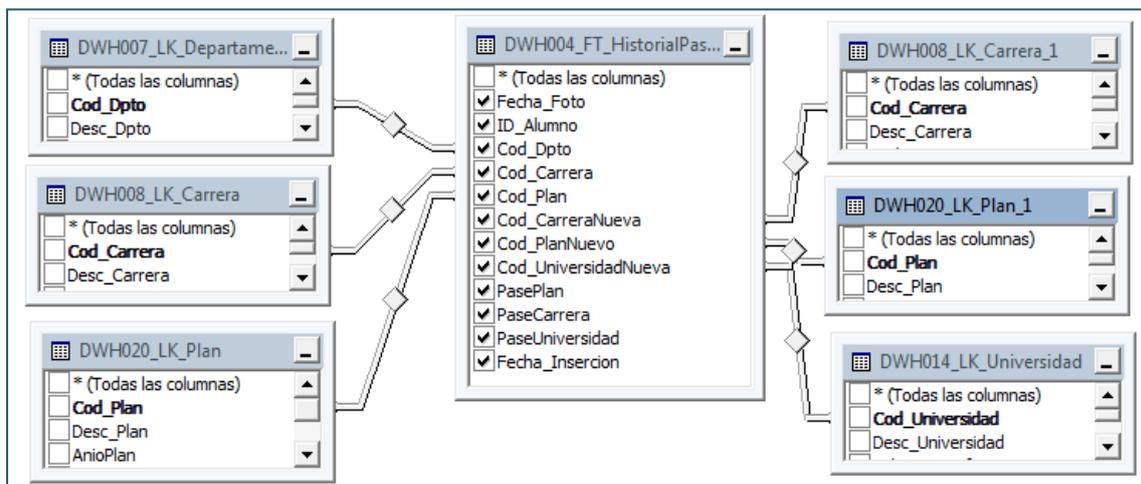


Figura 23. Relación Historial Pase.



Diseño Físico del Data Warehouse.

El modelo dimensional desarrollado en los puntos anteriores, debe convertirse ahora en un diseño físico. El diseño físico detalla cómo deben ser elaboradas las distintas entidades (tablas) que deberán contener una base de datos relacional para poder mantener la información accesible desde cualquier aplicación, también se indica el tipo de dato, tamaño de dato, la clave primaria, índices, el nombre físico y una breve descripción de cada campo, los cuales conforman la parte física desde donde será desplegada la información presentada por el cubo.

➤ Creación de las Tablas y Relaciones

Para este proyecto se utilizó la herramienta **Microsoft® SQL Server® 2008 R2**¹¹ que contiene cuatro componentes principales:

- ✓ **Database Engine:** es el servicio principal para almacenar, procesar y proteger datos. El Database Engine (Motor de base de datos) proporciona acceso controlado y procesamiento de transacciones rápido para cumplir con los requisitos de las aplicaciones consumidoras de datos más exigentes de su empresa.
- ✓ **SSAS:** SQL Server Analysis Services, Es el servidor de bases de datos multidimensionales que ofrece funciones OLAP, de procesamiento analítico en línea y minería de datos para aplicaciones de Business Intelligence. Permite diseñar, crear y administrar estructuras multidimensionales que contienen datos agregados desde otros orígenes de datos, como bases de datos relacionales. En el caso de las aplicaciones de minería de datos, Analysis Services permite diseñar, crear y visualizar modelos de minería de datos que se construyen a partir de otros orígenes de datos mediante el uso de una gran variedad de algoritmos de minería de datos estándar del sector
- ✓ **SSRS:** SQL Server Reporting Services, el servidor de informes. Permite a su vez leer la base de datos SSAS y producir informes integrando las principales demandas del reporting estadístico de Business Intelligence: tablas, gráficos, indicadores de nivel, indicadores visuales y cartografía.
- ✓ **SSIS:** SQL Server Integration Services, los servicios ETL. Reemplaza a los antiguos DTS (Data Transformation Services). Sirve para alimentar los almacenes de datos y automatizar tareas de administración.

En la figura 24 se muestra el diagrama de entidades resultantes para este proyecto.

¹¹<http://www.microsoft.com/es-ar/download/details.aspx?id=30438>

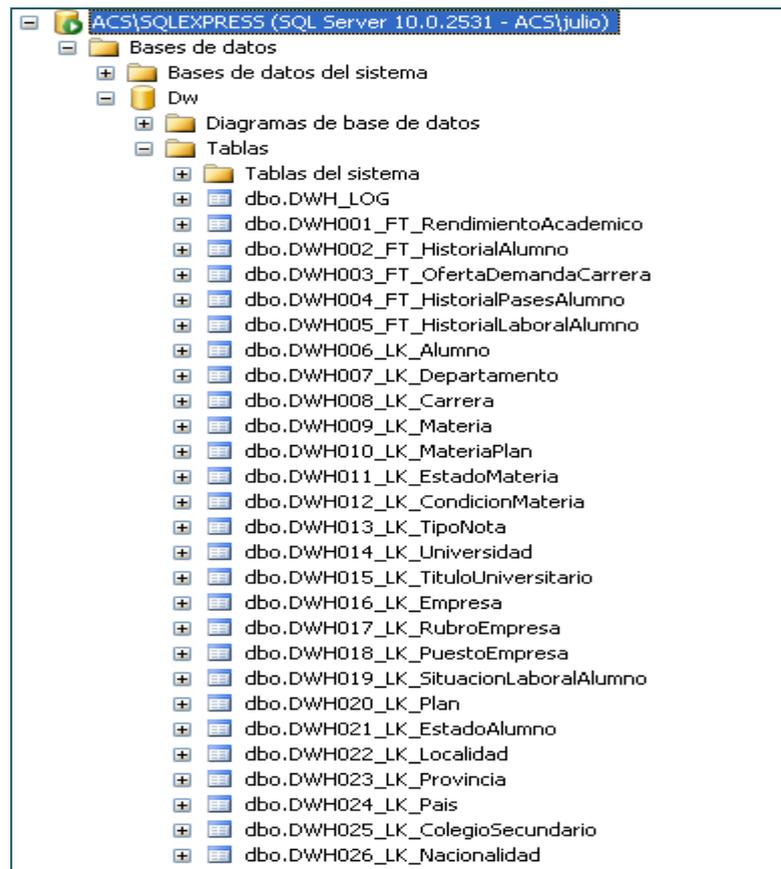


Figura 24. Esquema de tablas en SQL Server 2008 R2.

➤ **Establecer los permisos necesarios.**

Los usuarios finales y sus requerimientos impactan siempre en la implementación de un DWH. Según la perspectiva de Kimball, los requerimientos del negocio se posicionan en el centro del “universo del Data Warehouse” [Rod10].

De la misma forma que hay una gran cantidad de maneras para organizar un data warehouse, es importante notar que también hay una gama cada vez más amplia de usuarios finales. No todos los usuarios del DWH necesitan el mismo nivel de análisis. Es por ello que es necesario identificar los roles o perfiles de usuarios para los diferentes tipos de aplicaciones necesarias en base al alcance de los perfiles detectados. Para este proyecto, se consideraron las siguientes categorías:

- Autoridades del Departamento
- Usuarios de consultas.
- Administradores de bases de datos.
- Diseñadores y responsables del proceso ETL.
- Soporte técnico.

Cada una de estas categorías diferentes de usuario tienen su propio conjunto de requerimientos para los datos, acceso, flexibilidad y facilidad de uso.



Diseño del ETL

En esta etapa es donde se realiza los procesos que se conocen como sistema ETL (Extracción, Transformación y Carga) del inglés “Extract, Transform and Load”. Este proceso de extracción de datos consiste en la búsqueda y recuperación de datos que son relevantes para el DW, usando para ello los sistemas operacionales de la organización o fuentes de datos externas. Este proceso exige frecuentemente, un trabajo de programación ya que no existe un método estándar que permita hacerlo de manera automática debido a que las fuentes de datos suelen estar soportadas por distintas tecnologías. En este apartado se explican algunos de los procesos del ETL.

➤ Control de la extracción de los datos y su automatización

Esta etapa consiste en el diseño y desarrollo del proceso ETL, aquí es donde se definen como “*Procesos de Transformación*” los procesos para convertir o recodificar los datos fuente a fin de poder efectuar la carga efectiva del Modelo Físico.

Todas las actividades de esta etapa son altamente críticas pues tienen que ver con la materia prima: los datos. La calidad de los datos es un factor determinante en el éxito de cualquier proyecto de análisis de información. Es en esta etapa donde deben sanearse todos los inconvenientes relacionados con la calidad de los datos fuente. La desconfianza y pérdida de credibilidad del DW serán resultados inevitables si el usuario encuentra información inconsistente. La alimentación del DW a partir de los orígenes de datos se divide generalmente en dos etapas.

La primera consiste en alimentar una base de datos intermedia y temporal, comúnmente llamada Staging Area (SA). Esta es un área temporal donde se recogen los datos que se necesitan de los sistemas origen. Se recogen los datos estrictamente necesarios para las cargas, y se aplica el mínimo de transformaciones a los mismos. No se aplican restricciones de integridad ni se utilizan claves, los datos se tratan como si las tablas fueran ficheros planos. De esta manera se minimiza la afectación a los sistemas origen, la carga es lo más rápida posible para minimizar la ventana horaria necesaria, y se reduce también al mínimo la posibilidad de error. Una vez que los datos están traspasados, el DW se independiza de los sistemas origen hasta la siguiente carga. Lo único que se suele añadir es algún campo que almacene la fecha de la carga. Obviamente estos datos no van a dar servicio a ninguna aplicación de reporting, son datos temporales que una vez hayan cumplido su función serán eliminados, de hecho en el esquema lógico de la arquitectura muchas veces no aparece, ya que su función es meramente operativa.

La segunda se usa para alimentar el DW a partir de la *Staging Area*. Las tablas que componen el SA se muestran en la figura 25.



La primera etapa efectúa controles sobre los tipos de datos y controles de campos obligatorios, de conversión y de limpieza de datos para garantizar que no habrá problemas en la siguiente etapa. La segunda etapa realiza la búsqueda de las claves técnicas en el DW, alimenta las dimensiones y, a continuación, las tablas de hechos.

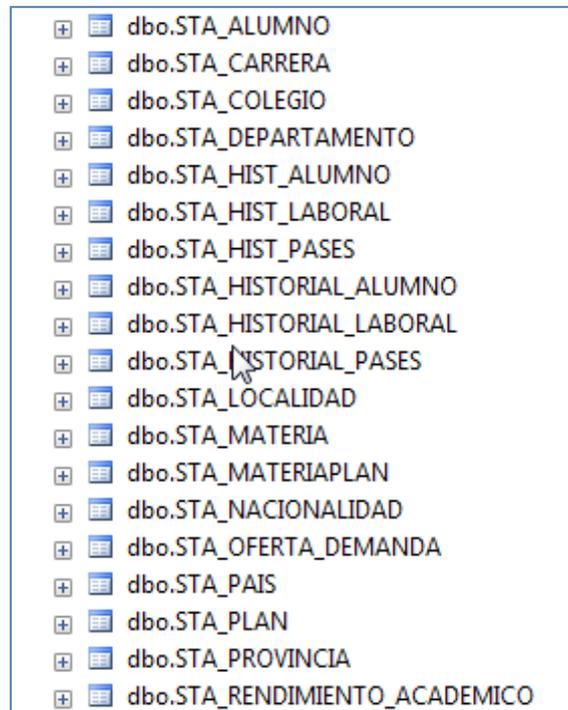


Figura 25. Esquema de tablas del Staging Área.

Definición de funciones sobre los datos extraídos para convertirlos en datos que serán cargados

Procesos ETL del Data Warehouse UNLaM utilizando SSIS

En esta sección se muestran los procesos ETL del Data Warehouse UNLaM, los cuales fueron desarrollados utilizando la suite de Microsoft 2008, particularmente SSIS (SQL Server Integration Services). Esta herramienta permite:

- generar uno o más paquetes dentro de un proyecto para trasladar datos de diferentes orígenes al destino.
- extraer datos de diferentes fuentes origen a través de conectores (.Net y/o proveedores OLE DB nativos).
- transformar datos mediante tareas de transformación, búsqueda, integración, combinación, entre otras.
- controlar la secuencia que los datos deben seguir mediante el control de eventos de éxito y fallo.
- cargar la información a diferentes fuentes de datos destino.

SSIS está constituido por varios elementos que permiten construir paquetes de ETL complejos, sus componentes son:

- **Data Source:** Identifica las conexiones hacia los diversos orígenes de datos.



- **Data Source View:** Definen vistas parciales o totales de los objetos de un Data Source.
- **Task:** Una tarea está asociada a una unidad de transformación, procesamiento de datos o transferencia de registros de un origen a un destino.
- **Control Flow:** Define la secuencia lógica de transferencia de información. Por ejemplo, a través de los flujos de control se permiten controlar la secuencia de transformación, limpieza, búsqueda del dato, entre otros.
- **Container:** Permite agrupar diferentes tareas de transformación para facilitar el entendimiento o la secuencia de las tareas.
- **Package:** Contiene las tareas del proyecto SSIS.

A continuación se muestra el entorno de desarrollo de los paquetes de ETL utilizando la herramienta:

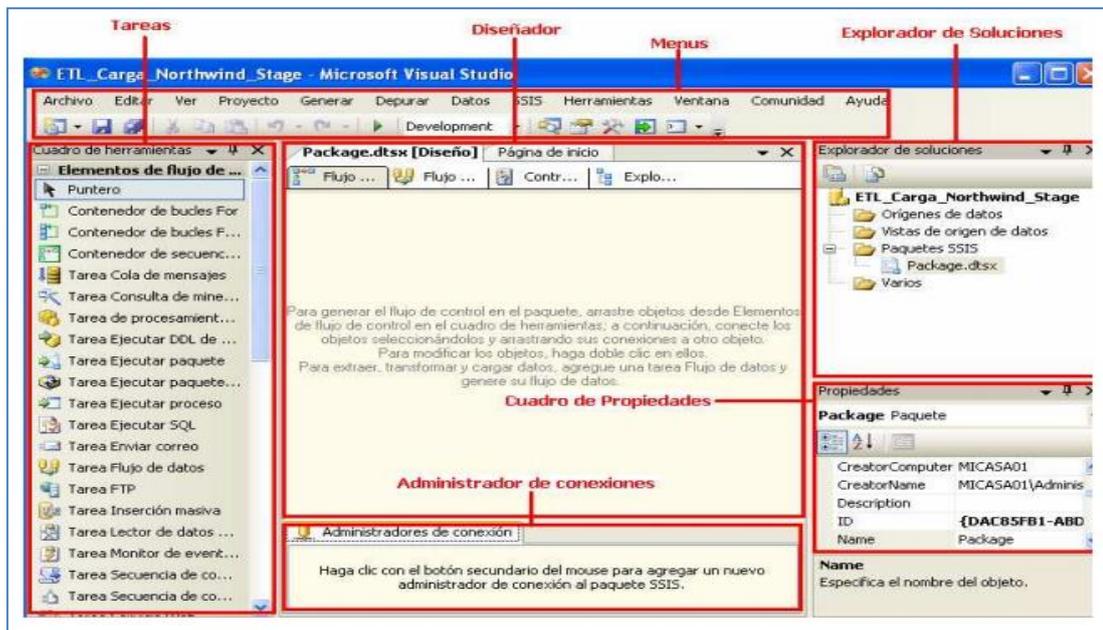


Figura 26. Pantalla principal de SSIS.

Control Flow

El proyecto SSIS para cargar el DWH_UNLaM se ha organizado en una serie de secuencias que se ejecutan en el siguiente orden:

- A. Carga STAsLKs
- B. Carga Defaults LKs y Valores Iniciales
- C. Carga LKs
- D. Carga STAsFTs
- E. Inserta Códigos faltantes en LKs provenientes de FTs
- F. Carga FTs

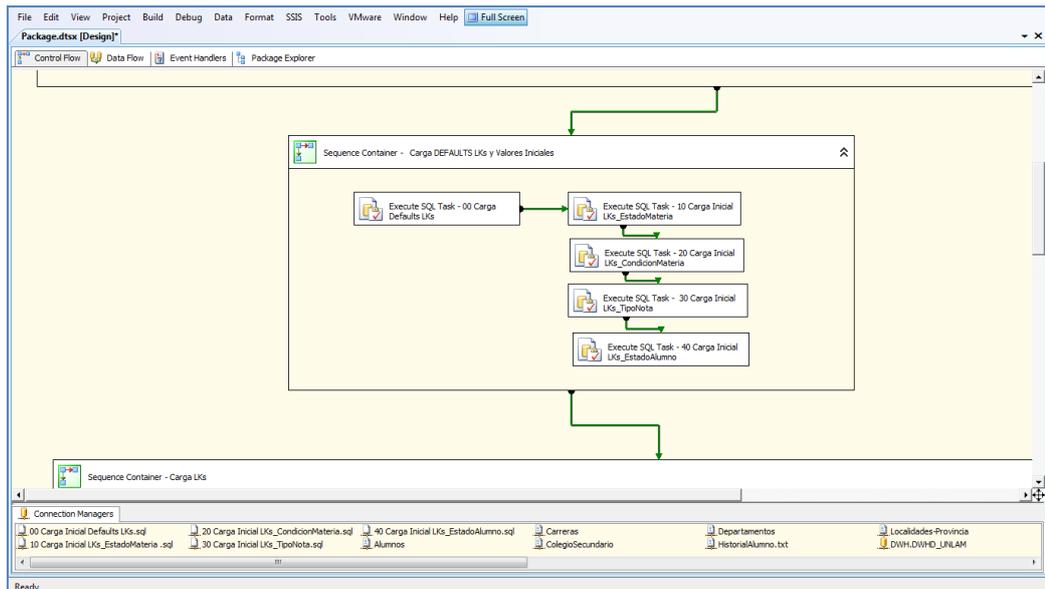


Figura 28. Ejemplo de la Carga Defaults LKs y Valores Iniciales.

Carga LKs

En este Container se agrupan las tareas que se ocupan de cargar las tablas de Dimensión (LKs = Lookups) del DWH y que anteriormente fueron cargadas en tablas intermedias (STAs = StaginAreas). Durante este proceso pueden aplicarse transformaciones de tipos de datos, reagrupaciones y/o normalización de códigos, creación de campos derivados, entre otras tareas.

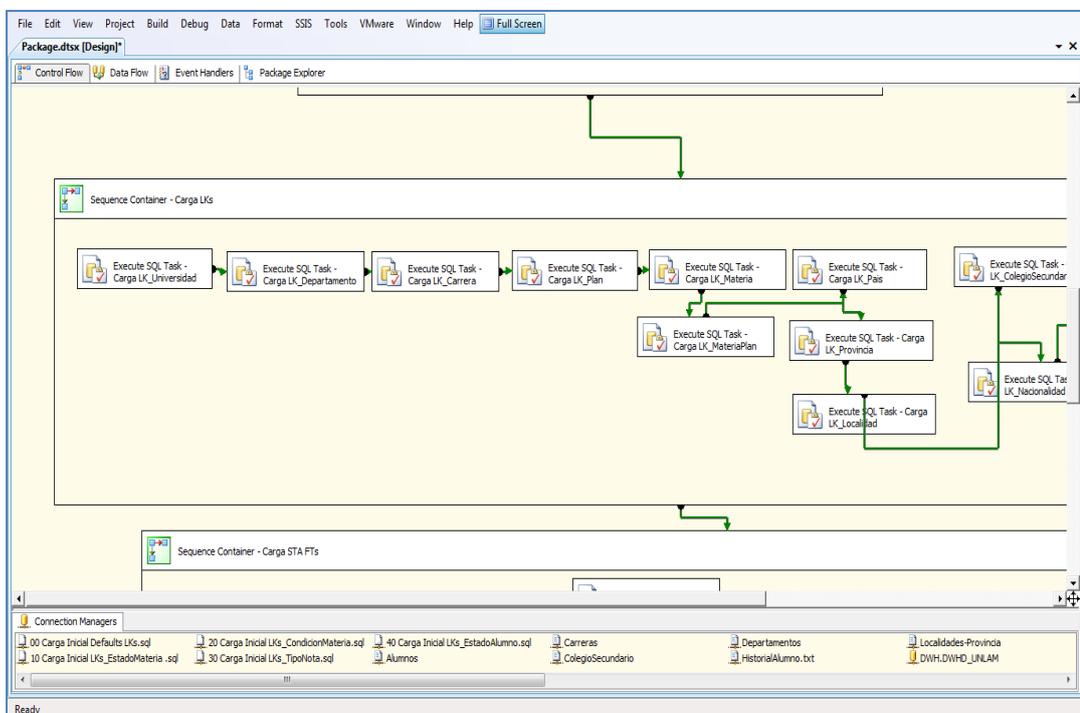


Figura 29. Ejemplo de la Cargar LKs.



Carga STAsFTs

En este Container se agrupan las tareas que se ocupan de cargar los archivos de texto recibidos de los sistemas fuente y que contienen datos de Tablas de Hechos (FTs = FactTables) en las tablas intermedias (STAs = StaginAreas) del DWH.

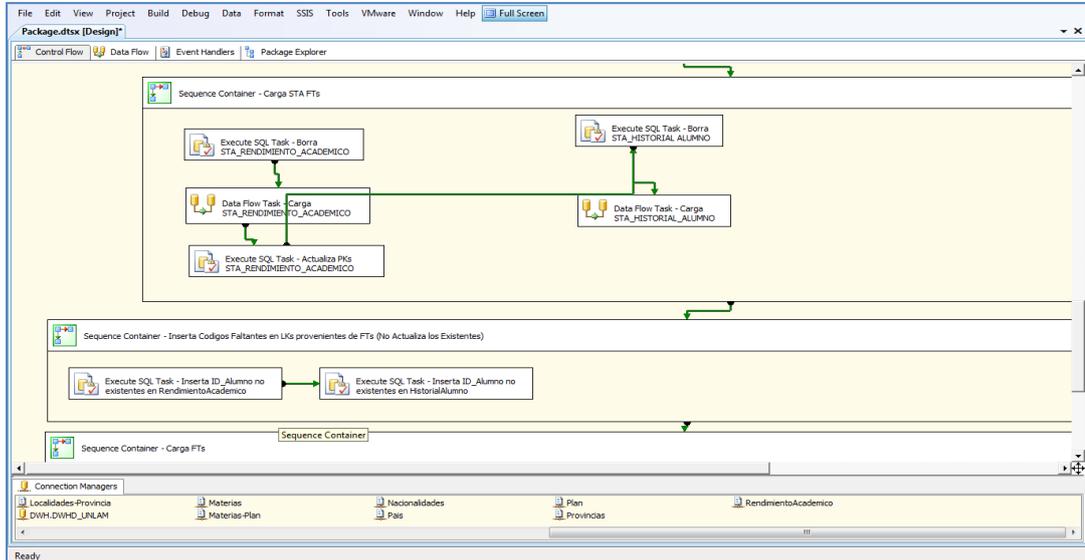


Figura 28. Ejemplo de la Carga STAsFTs

Inserta Códigos faltantes en LKs provenientes de FTs

En este Container se agrupan las tareas que se ocupan de insertar en las tablas de Dimensión aquellos códigos recibidos en los archivos de texto que contienen la información de los Hechos.

Carga FTs

En este Container se agrupan las tareas que se ocupan de cargar las Tablas de Hechos (FTs = FactTables) del DWH y que anteriormente fueron cargadas en tablas intermedias (STAs = StaginAreas). Durante este proceso pueden aplicarse transformaciones de tipos de datos, reagrupaciones y/o normalización de códigos, creación de campos derivados, entre otras tareas.

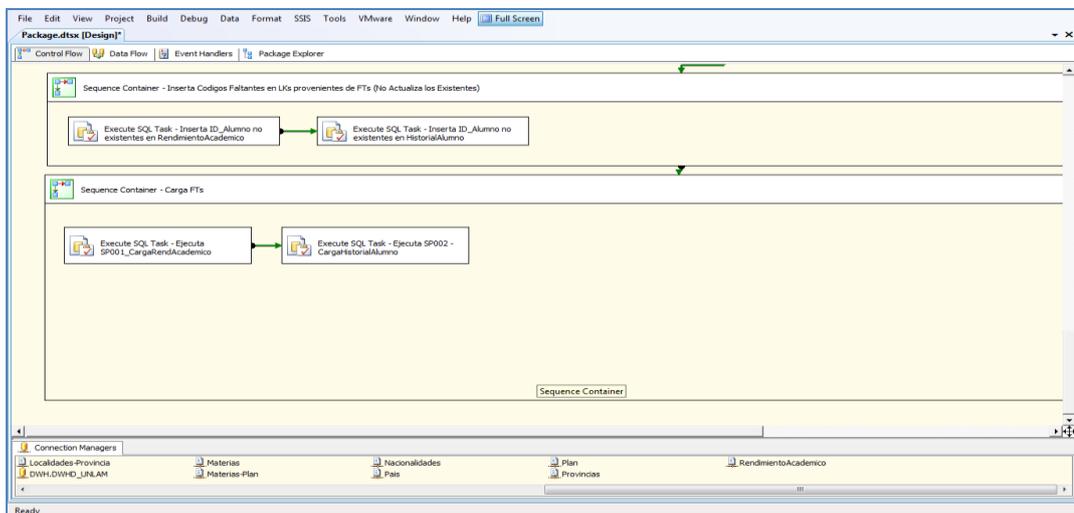


Figura 29. Ejemplo de la Carga FTs

ETAPA 4 Generación de los Cubos

En esta etapa se describirán como se construyen los Cubos OLAP agregando, según los requisitos de cada área o departamento, las dimensiones y los indicadores necesarios. El modo de creación, explotación y mantenimiento de los cubos OLAP es muy heterogéneo, en función de la herramienta final que se utilice, en nuestro proyecto utilizamos la herramienta O3 Designer¹³.

Uso del Diseñador

El O3 Designer es un instrumento completo y flexible utilizado para construir, cargar y modificar los cubos multidimensionales que contienen la información referente a la empresa. La estructura de los cubos permite a un usuario tener puntos de vista diferentes, revelando así dimensiones desde enfoques diversos. El O3 Designer ordena el proceso de diseño del siguiente modo:

- Define los atributos del cubo
- Define fuentes de datos (archivos de texto, bases de datos, campos virtuales, planillas, otros)
- Define las dimensiones y sus jerarquías
- Define medidas
- Verifica el diseño del modelo
- Carga el cubo con la información de las fuentes de datos

El O3 Designer permite la transformación de información dispersa en datos significativos para el negocio. A su vez, sostiene todas las funciones analíticas, inclusive las fechas relativas de informes y cálculos complejos.

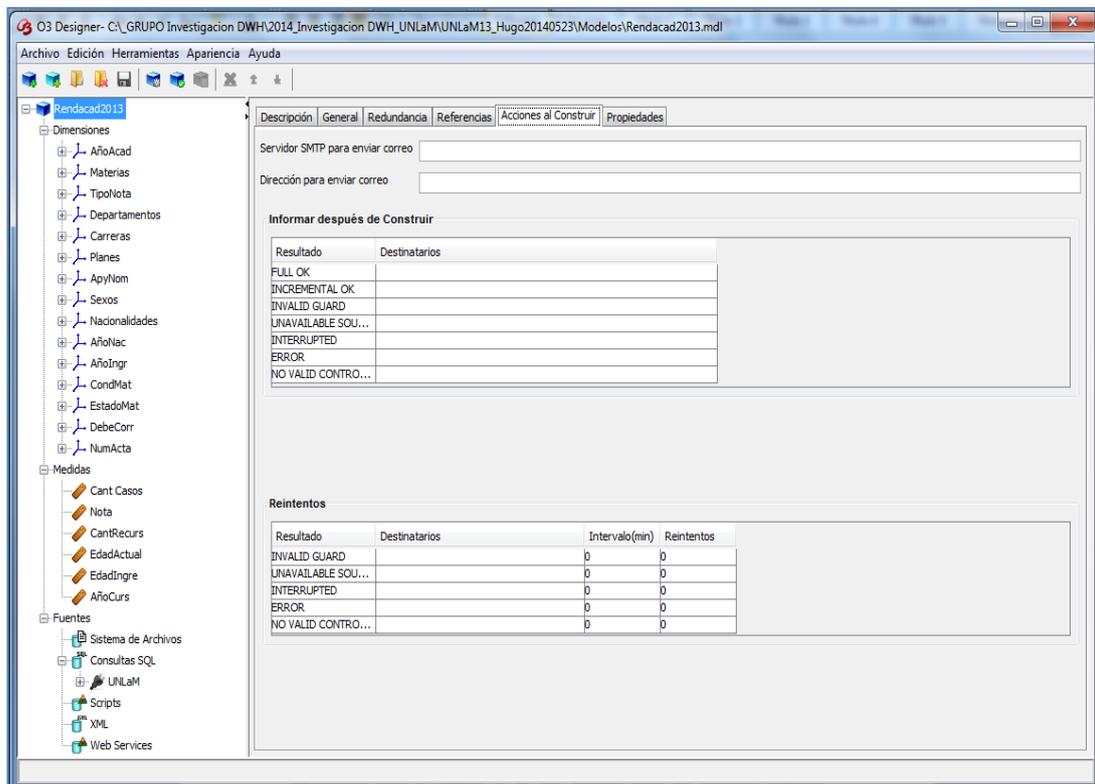


Figura 30. Vista del Cubo Rendimiento Académico desde el O3 Designer.

¹³ <https://www.ideasoft.biz/wiki/display/o3man/Componentes+de+O3+Designer>



Al comienzo de una sesión con el **O3 Designer**, se presentan los siguientes elementos:

- Árbol de Diseño
- Panel de Propiedades
- Barra de Menú
- Barra de Herramientas

Árbol de Diseño

El Árbol de Diseño aparece sobre la izquierda de la pantalla. Representa el modelo o la estructura del modelo. El árbol se divide en tres ramas principales:

- Dimensiones
- Medidas
- Fuentes de Datos

Definición de dimensiones

La definición de las dimensiones se realiza identificando los códigos únicos de los valores para cada nivel de estas y las descripciones correspondientes. En la figura 32 se muestra la definición de la dimensión “Carrera”, compuesta por un único nivel, pero pueden tener más de uno. Para cada nivel se debe indicar el campo que es la clave que identifica los distintos valores y las descripciones correspondientes.

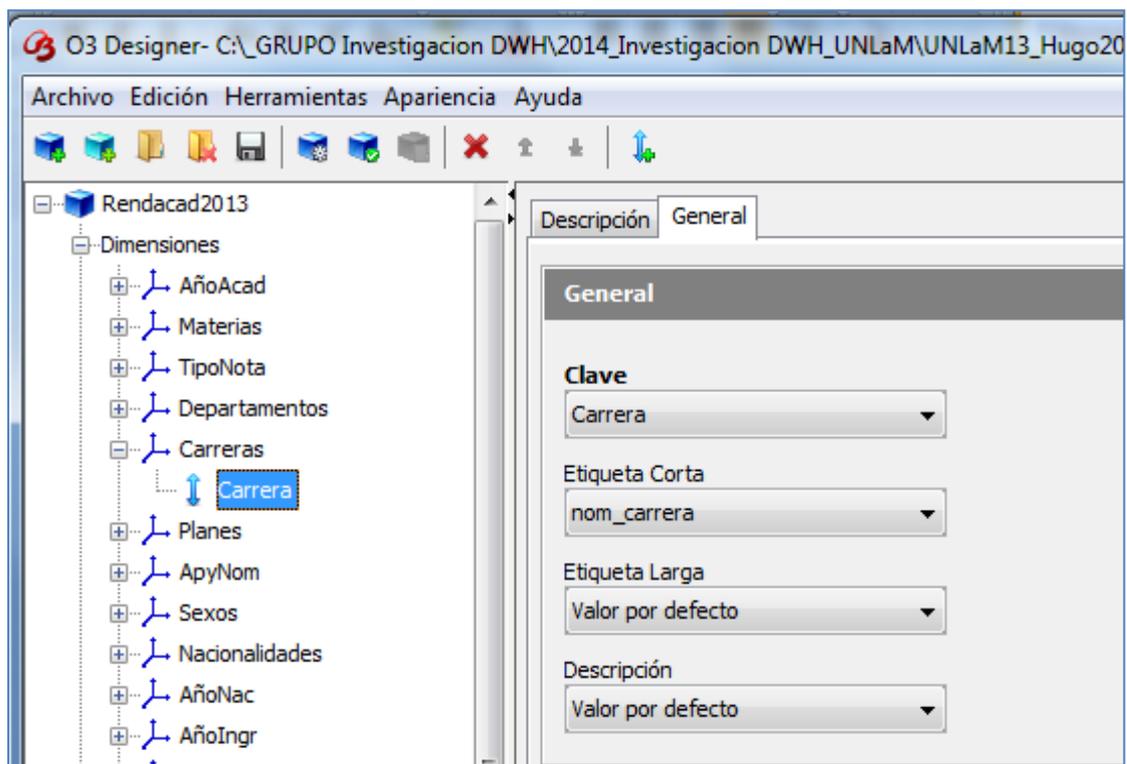


Figura 31. Ejemplo de la Carga FTs



Definición de medidas

Las **medidas** del modelo permiten al usuario analizar los valores almacenados en el datamart. Representan la cantidad de un elemento determinado de la Dimensión y es posible obtenerlas directamente de la Fuente de Datos o calcularlas desde la misma. Las medidas pueden representar conceptos tales como precios, unidades físicas vendidas, cantidad de facturas y costos, márgenes de ganancia, horas trabajadas, etc. También es posible definir Medidas que representen escalas de valores, o el número de veces que aparece un determinado valor. En la figura 33 vemos las medidas utilizadas para este cubo. Existen dos tipos de medidas:

- **Básicas:** Las medidas que se extraen directamente de un campo de la fuente de datos.
- **Derivadas:** Las medidas que son calculadas de otras medidas a través de la aplicación de expresiones de cálculo.

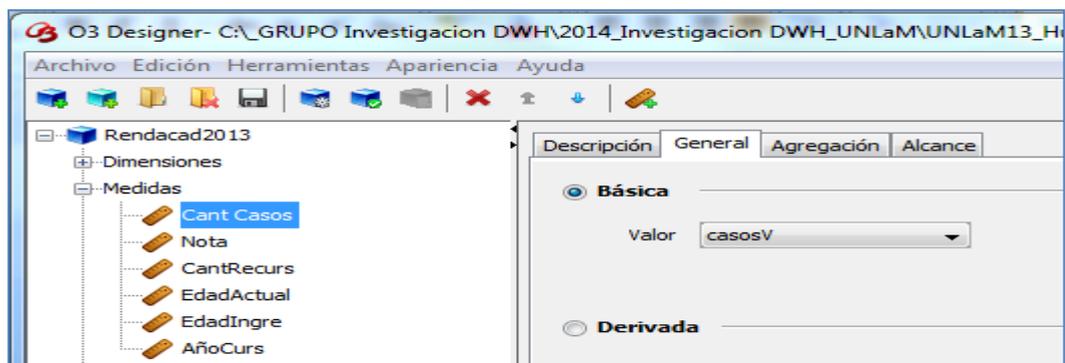


Figura 32. Medidas correspondiente

Definición de fuente de datos

Con las fuentes de datos que se incluyan en la definición del modelo, O3 obtiene los datos necesarios para la creación del datamart correspondiente al modelo. La información se usa para crear las Dimensiones y las jerarquías así como también para calcular las Medidas. En la figura 34 se muestra los campos que componen la consulta SQL, que es de donde se obtienen los datos del cubo.

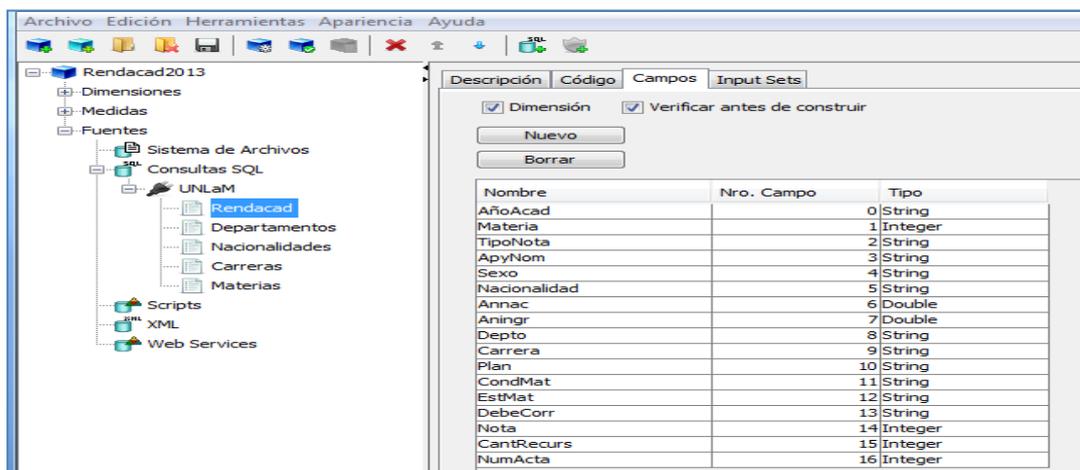


Figura 33. Vista de una de las fuentes de datos.

Actualización de la Información

Para que el contenido del almacén de datos esté sincronizado con el contenido de las fuentes (sistemas operacionales), cualquier actualización en una fuente, que sea relevante para el almacén, debe ser transmitida al almacén para que éste sea actualizado. El desajuste temporal entre estos dos procesos de actualización, puede hacer que el almacén no esté sincronizado con las fuentes, es decir el almacén no contenga datos que serían derivables en un instante de tiempo, del contenido de las fuentes.

En la figura 35 se muestran los componentes básicos del proceso para la actualización de los cubos, este proceso se prevé realizar al final de cada cuatrimestre. A continuación se describen que se hace en cada etapa:

- *Origen de los datos*: Proviene de diversos sistemas OLTP, en este desarrollo los datos fueron exportados en sistema SIU-GUARANI y convertidos en archivos planos.
- *Staging Area* es un área temporal donde se van a almacenar los datos necesarios de los sistemas origen. En este caso, se van a recoger los datos de las tablas alumno, asignatura y titulación de la base de datos operacional del entorno universitario que son los estrictamente necesarios para las cargas de datos en el DM.
- *Data Marts*: son las bases de datos ya modeladas de forma multidimensional que se alimentan periódicamente mediante procesos ETL.
- *Herramientas analíticas y de presentación*: que permiten al usuario del negocio poder acceder a la información, compartirla y analizarla.

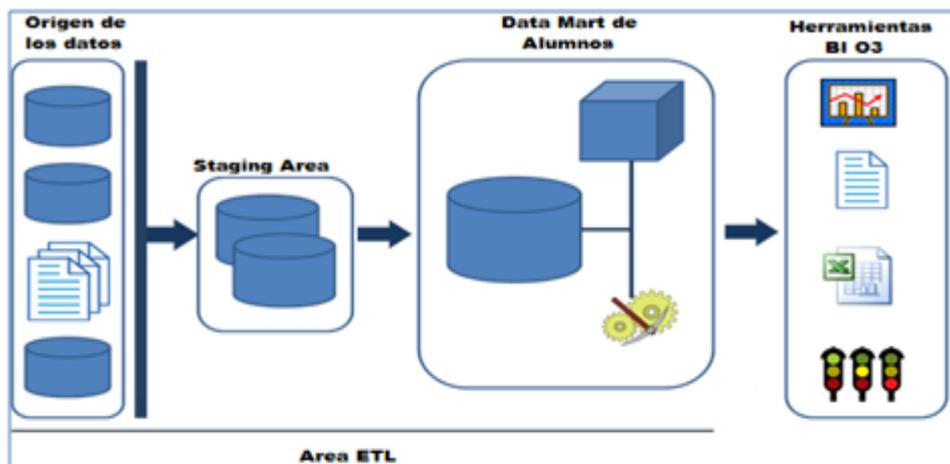


Figura 34. Arquitectura empleada en la propuesta DM de Alumnos.



ETAPA 5 Puesta en Marcha

Instalación del DW en un Servido designado.

Como el DIIT tiene licencias de Microsoft, este proyecto se realizó bajo la suite SQL Server para el desarrollo de soluciones de BI, además se adquirió las licencias para el uso del producto Ideasoft O3 Business Intelligence¹⁴ (BI) versión 5.4.0 que está compuesto por los siguientes componentes:

- O3 AdmServer.
- O3 Browser.
- O3 Designer.
- O3 Builder.
- O3 Portal.
- O3 JBoss.
- O3 Reporting.

Determinación de Usuarios y contraseñas

El módulo de O3 Server Administrator, posee una opción para administrar usuarios y contraseñas, crear y asignar perfiles a los usuarios que podrán utilizar el sistema, conceder o denegar permisos a las consultas, vistas y reportes, etc. En la figura 40, se puede observar el menú dentro de las opciones de O3 Server Administrador, que permite administrar los Roles. En la figura 41, se muestra la pantalla para ingresar nuevos Usuarios.

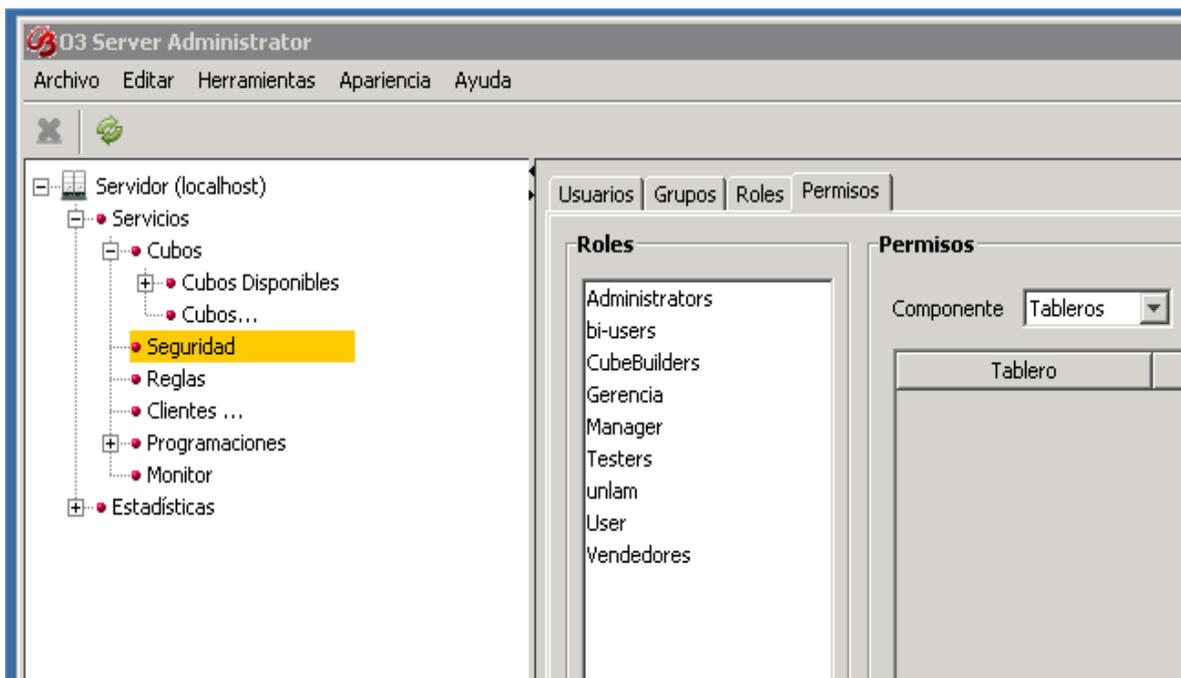


Figura 40. Modulo de Roles de O3.

¹⁴ <http://www.ideasoft.biz/web/?lang=es>

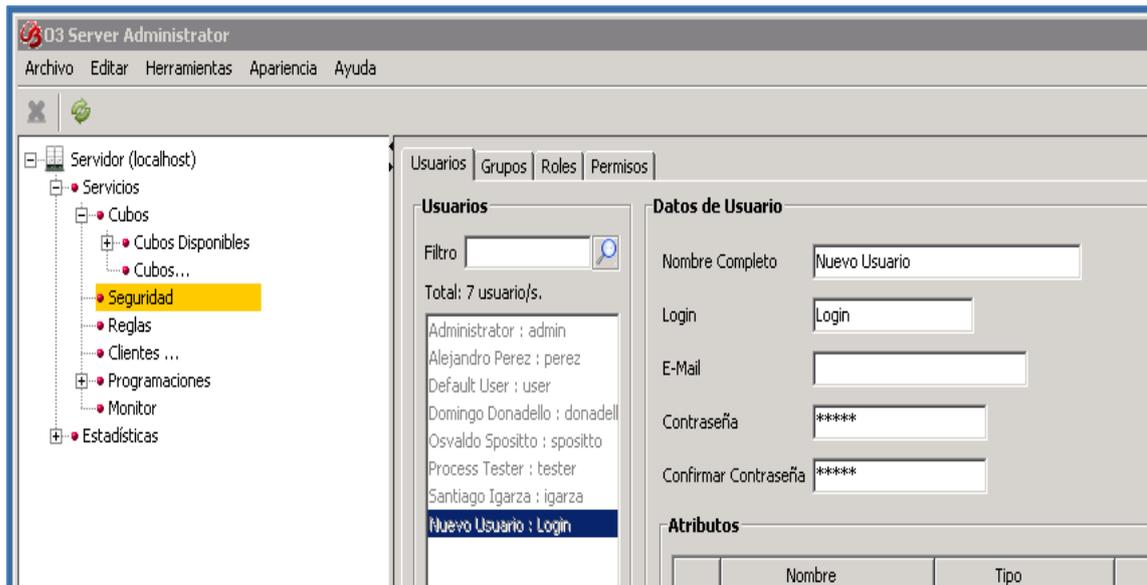


Figura 41. Creación de nuevos Usuarios.

Capacitación de los usuarios y asistencia en el uso.

La capacitación debe orientarse al conocimiento de la herramienta en su uso y la obtención e interpretación de la información obtenida mediante el cruce de variables, la capacitación se preparará según los grupos de usuarios, es necesario conocer las necesidades de información de cada grupo.

Para esta tarea se prepararon distintas metodologías de capacitación, preparando el material a distribuir que sirva de apoyo y seleccionado los ejercicios que serán desarrollados para que la práctica sea lo más didáctica posible y así además reducirá errores de captación.

Para la primera capacitación se desarrollo un PowerPoint, el cual se entrega junto a los demás anexos.

Plan de monitoreo y mantenimiento.

En esta etapa se debe tener un manejo adecuado de las peticiones de modificaciones del sistema, categorizándolas y priorizándolas adecuadamente, por tal motivo, en el momento en el que se registra la petición, se procede a diagnosticar de qué tipo de mantenimiento se trata. Atendiendo a los fines, podemos establecer los siguientes tipos de mantenimiento:

- **Correctivo:** son aquellos cambios precisos para corregir errores del producto software.
- **Evolutivo:** son las incorporaciones, modificaciones y eliminaciones necesarias en un producto software para cubrir la expansión o cambio en las necesidades del usuario.
- **Adaptativo:** son las modificaciones que afectan a los entornos en los que el sistema opera, por ejemplo, cambios de configuración del hardware, software de base, gestores de base de datos, comunicaciones, etc.



- **Perfectivo:** son las acciones llevadas a cabo para mejorar la calidad interna de los sistemas en cualquiera de sus aspectos: reestructuración del código, definición más clara del sistema y optimización del rendimiento y eficiencia.”

Relevamiento de nuevos requerimientos.

Este sistema apoyar a las autoridades que han de tomar decisiones, pues dan acceso a las bases de datos corporativas de la empresa y fuentes de datos externas, con lo que analistas y directivos disponen de un análisis flexible de la información. Los usuarios pueden dividir y aislar la información en múltiples formas, empezando en las de alto nivel y profundizando hasta niveles de detalle específicos, lo que proporciona un análisis robusto, combinado con una interfaz de usuario intuitiva que brinda un funcionamiento flexible.

El éxito final del proyecto dependerá en gran medida del uso y rendimiento que pueda ofrecer en el ámbito universitario alcanzando su máximo valor cuando esté sumamente explotado, permitiendo al usuario obtener beneficios materiales con su uso.

Conclusiones.

Actualmente la información se ha convertido en un elemento muy necesario para el análisis y seguimiento de cualquier actividad, sea cual fuese el ambiente en que se desarrolle, por esta razón los datos históricos y la información proveniente de las bases de datos operacionales se hacen más accesibles y que al ser procesada en un DW puede convertirse en una herramienta que colabore a nivel directivo con tomar sus decisiones.

En la realización de este trabajo, como primera etapa, se realizaron distintas actividades destinadas a analizar los requerimientos de información de las autoridades del DIIT, con el fin de identificar cuáles serán las bases de datos que serán tomadas en cuenta para la construcción del Data Warehouse Departamental o Data Mart.

En una segunda etapa se crearon lógica y físicamente las tablas donde se almacenaron los datos, luego de haber sido procesado por las rutinas ETL, para luego, ser convertidos al Modelo Multidimensional, que son los que en definitiva son usados por la herramienta O3. En este proyecto se pretende presentar la construcción de una solución utilizando la tecnología OLAP materializada mediante la implementación de un sistema de análisis multidimensional para el departamento de Ingeniería e Investigaciones Tecnológicas del la Universidad Nacional de La Matanza. Esta herramienta convierte los datos históricos en información estratégica que apoya la toma de decisiones.

La gran dificultad para el desarrollo de este trabajo fue la estructuración de los datos de manera que satisfaga las necesidades de los requerimientos.



Futuras líneas de trabajo.

Debido a la buena aceptación que se obtuvo de la herramienta dentro del DIIT, se recomienda, ya que ellos no poseen personal idóneo en el tema, extenderlo al resto de los Departamentos de la UNLaM.

Para esto es necesario realizar un nuevo análisis para obtener nuevos requerimientos que provengan de las restantes unidades académicas. Si bien también habría que realizar una promoción y motivación para lograr el uso real en la toma de decisiones, pasando por la definición de criterios sobre consideraciones de la información relevante.

Además, se podrían incorporar modificaciones a las soluciones actuales o desarrollar nuevos componentes dentro de la totalidad del sistema para análisis de información. También se podrían desarrollar modelos de análisis sobre nuevas temáticas.

Archivos adjuntos

Junto a este documento se entrega un CD que contiene:

- los procesos para generar los datos desde el SIU-Guaraní,
- el modelo y el ejecutable para poder generar el cubo,
- el cubo con datos de ejemplo usado para los ejemplos en este documento,
- los archivos de texto con datos de ejemplo para generar el cubo utilizado,
- y la versión digital de este documento.

**Bibliografía:**

- [Kim98] Kimball, Ralph, et al. “*The Data Warehouse Lifecycle Toolkit*”, John Wiley & Sons, 1998.
- [Kim02] Kimball & Ross, “*The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*”. Second Edition, New York, Wiley, 2002.
- [Mic06] “*Teoría sobre Business Intelligence*” Concurso MicroStrategy Experiencia Business Intelligence. 3ra. Edición. Año 2006.
Disponible e: <http://es.scribd.com/doc/86221564/Teoria-DW>
- [Elm97] Ramez Elmasri & Shamkant B. Navathe. Fundamentos de Sistemas de Bases de Datos, 5ta Edición –1997
- [Rod10] M. Rodríguez Sanz. “*Análisis y Diseño de un Data Mart para el seguimiento Académico de Alumnos en un entorno universitario*”.Universidad Carlos III de Madrid. 2010.
Disponible en : http://e-archivo.uc3m.es/bitstream/handle/10016/9856/PFC_Miguel_Rodriguez_Sanz.pdf?sequence=6.
- [1] <http://www.dataprix.com/datawarehouse-manager>
- [2] http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx
- [3] “*DataWarehouse*” Universidades Andaluzas y Consejería de Educación y Ciencia de la Junta de Andalucía. Disponible en http://www.gestion.uco.es/gestion/datawarehouse/doc/m_usuario/07000_olap.pdf



Tabla de contenido

Título del Proyecto:	3
Resumen del Proyecto:	3
Etapa 3. El Manejo los datos	5
Introducción al Modelo Multidimensional.....	5
Componentes del Modelo Multidimensional	6
Tabla de Hechos	6
Tabla de Dimensiones:	7
Tipos de Modelos de un DW	8
Esquema en Estrella	8
Esquema Copo de Nieve.....	8
Tipos de implementación de un DW	9
ROLAP	9
MOLAP.....	9
HOLAP.....	10
ROLAP vs. MOLAP (Comparativa).....	10
Modelo Lógico.....	11
Descripción de la Tabla de Hechos.....	12
Tablas de Dimensiones.....	14
Establecer Relaciones.....	17
Esquema de Relacione del Cubo Rendimiento Académico.....	18
Esquema de Relacione del Cubo HistorialAlumno	18
Esquema de Relacione del Cubo HistorialLaboralAlumno	19
Esquema de Relacione del Cubo Oferta Demanda Carrera	19
Esquema de Relacione del Cubo Historial Pase	19
Diseño Físico del Data Warehouse.....	20
➤ Creación de las Tablas y Relaciones	20
➤ Establecer los permisos necesarios.....	21
Diseño del ETL	22
➤ Control de la extracción de los datos y su automatización.....	22
Definición de funciones sobre los datos extraídos para convertirlos en datos que serán cargados	23



ETAPA 4 Generación de los Cubos	28
Uso del Diseñador	28
Árbol de Diseño	29
Actualización de la Información	31
ETAPA 5 Puesta en Marcha	32
Instalación del DW en un Servido designado.....	32
Determinación de Usuarios y contraseñas	32
Capacitación de los usuarios y asistencia en el uso.	33
Plan de monitoreo y mantenimiento.....	33
Relevamiento de nuevos requerimientos.....	34
Conclusiones.	34
Futuras líneas de trabajo.....	35
Archivos adjuntos.....	35
Bibliografía:	36

**Tabla de Ilustraciones**

Figura 1. Ejemplo de un Cubo genérico	6
Figura 2. Jerarquía geográfica	8
Figura 3. Esquema Estrella	8
Figura 4. Esquema Copo de nueve.....	9
Figura 5. Arquitectura del DW propuesta.....	11
Figura 6. Tabla de Hecho Rendimiento Académico.....	12
Figura 7. Tabla de Hecho Historial Alumno	12
Figura 8. Tabla de Hecho Oferta Demanda Carrera.....	13
Figura 9. Tabla de Hecho Historial Pases Alumnos	13
Figura 10. Tabla de Hecho.....	13
Figura 11. Esquema Estrella de la Dimensión DWH006_LK_Alumno.....	14
Figura 12. Modelo de datos lógico de la Dimensión DWH010_LK_MateriaPlan.....	15
Figura 13. Dimensiones Departamento y Universidad.....	15
Figura 14. Dimensión DWH013_LK_TipoNota y los Posibles valores.....	15
Figura 15. Modelo de datos lógico de la Dimensión DWH013_LK_CondicionMateria.....	16
Figura 16. Posibles valores de la Dimensión DWH013_LK_CondicionMateria.....	16
Figura 17. Modelo de datos lógico de la Dimensión DWH011_LK_EstadoMateria.....	16
Figura 18. Posibles valores de la Dimensión DWH011_LK_EstadoMateria.....	16
Figura 19. Relación Rendimiento Académico.....	18
Figura 20. Relación Historial Alumno	18
Figura 21. Relación Historial Laboral Alumno	19
Figura 22. Relación Oferta Demanda Carrera	19
Figura 23. Relación Historial Pase.....	19
Figura 24. Esquema de tablas en SQL Server 2008 R2.....	21
Figura 25. Esquema de tablas del Staging Área.....	23
Figura 26. Pantalla principal de SSIS.....	24
Figura 27. Ejemplo de la Carga STAsLKS.....	25
Figura 30. Ejemplo de la Carga STAsFTs.....	27
Figura 31. Ejemplo de la Carga FTs.....	27
Figura 32. Vista del Cubo Rendimiento Académico desde el O3 Designer.....	28
Figura 32. Ejemplo de la Carga FTs.....	29
Figura 33. Medidas correspondiente	30
Figura 34. Vista de una de las fuentes de datos.....	30
Figura 39. Arquitectura empleada en la propuesta DM de Alumnos.....	31