Título del trabajo: "OLAP Server - FOMEC 10011"

Autores:

BARLATTO, Martín mmarlatto@yahoo.com.ar
SUPPO, Mónica msuppo@frsf.utn.edu.ar
TORRES LUYO, Diego dtluyo@yahoo.com.ar

Director: Dr. Aldo Vecchietti aldovec@crdsf2.ceride.gov.ar

Materia: Proyecto *Departamento:* Ingeniería en Sistemas de Información

Universidad: Universidad Tecnológica Nacional – Facultad Regional Santa Fe

Resumen:

El Proyecto FOMEC-10011 consiste en implementar un Sistema de Soporte de Decisión (DSS) para brindar apoyo en la gestión académica. Define una serie de indicadores calculados a partir de un DataWarehouse que concentra información de: Alumnado, Ciencia y Técnica, RRHH, Extensión Universitaria-Asuntos Estudiantiles, y, Administración-Finanzas. El proyecto "OLAP Server – FOMEC-10011", extiende el proyecto original, ya que implementa un OLAP Server sobre el DW, para brindar a los tomadores de decisión los beneficios que brinda OLAP.

Objetivos:

- Investigar sobre Servidores OLAP, Modelado dimensional, Base de Datos Multidimensionales (BDM).
- Generar una guía de uso de productos Oracle Express utilizados, para que otros grupos, puedan continuar con la implementación, incluyendo cuestiones de diseño, creación, carga, mantenimiento y administración de BDM, desarrollo aplicaciones para usuarios finales.
- Generar un DSS prototipo en Oracle Express que satisfaga requerimientos definidos en Proyecto FOMEC 10011.

Etapas:

a. Análisis de las tecnologías OLAP, MOLAP y Modelo de Datos Multidimensional.

b. Arquitectura del Proyecto basado en Oracle Express.

c. Diseño y Administración de una BD Multidimensional.

d. Carga y Mantenimiento de la BD Express.

e. Desarrollo de Aplicaciones OLAP para usuarios finales.

Además, se presentan ejemplos de consultas ad-hoc, conclusiones del trabajo y proyectos futuros.

Palabras Claves: OLAP, DSS (Decision Support System), MDB (Multidimensional Data Base), DataWarehouse, Add-hoc

Categoría: Otras: OLAP y DSS (Sistemas de Soporte de Decisión)

Carta de Recomendación: Archivo "Aval_JAIIO_Tesina_OLAP Server_FOMEC 10011.zip"

Copia de Tesina: archivo adjunto "Tesina_OLAP Server_FOMEC 10011.zip", que incluye la documentación presentada a la Cátedra Proyecto para la aprobación de la tesina. (Este archivo no incluye los códigos fuentes desarrollados)

OLAP SERVER - FOMEC 10011

RESUMEN

El Proyecto FOMEC-10011 consiste en la implementación de un Sistema de Soporte de Decisión (DSS) para brindar apoyo en la gestión académica. Define una serie de indicadores calculados a partir de un DataWarehouse que concentra información de: Alumnado, Ciencia y Técnica, RRHH, Extensión Universitaria y Asuntos Estudiantiles, y, Administración y Finanzas. El proyecto "OLAP Server – FOMEC-10011", que se presenta en este trabajo, extiende el proyecto original, ya que implementa un OLAP Server sobre el DW, para brindar a los tomadores de decisión los beneficios que brinda ese tipo de tecnología.Los objetivos del mismo son:

- Investigar sobre Servidores OLAP (On-line Analytical processing), Modelado dimensional, Base de Datos Multidimensionales (MOLAP, Multidimensional OLAP) con el fin de forjar conocimientos sobre estas nuevas tecnologías y ponerlos al servicio de la Facultad.
- Generar una guía de uso de los diversos productos de Oracle Express utilizados, para que otros grupos, en el futuro, puedan continuar con la implementación de otros indicadores. Dicha guía incluirá cuestiones de diseño, creación, carga y mantenimiento de la Base de Datos Multidimensional, diseño y desarrollo de aplicaciones para usuarios finales y administración de Base de Datos (DB) Express.
- Generar un prototipo de sistema de toma de decisión en Oracle Express que satisfaga un conjunto de requerimientos definidos en torno al Proyecto FOMEC 10011.

Para alcanzar dichos objetivos el proyecto se organizó en 5 etapas, las que se describen a lo largo de este trabajo. Dichas etapas son:

1. Análisis de las tecnologías OLAP, MOLAP y Modelo de Datos Multidimensional.
2. Arquitectura del Proyecto basado en Oracle Express.
3. Diseño y Administración de una BD Multidimensional.
4. Carga y Mantenimiento de la BD Express.
5. Desarrollo de Aplicaciones OLAP para usuarios finales.

Al final de este documento, se presenta ejemplos de consultas ad-hoc, las conclusiones del trabajo y los proyectos y trabajos futuros.

1. INTRODUCCIÓN

Debido a los cambios que se produjeron en los últimos años en lo que respecta a Tecnología de la Información, el surgimiento de nuevas lógicas de negocios, las crecientes necesidades de analizar información visualizada en un formato adecuado a los usuarios finales, y de poder realizar análisis Ad-hoc, surgen nuevas tecnologías, como es el caso de OLAP.

Hoy en día, la mayoría de las empresas cuentan con una/s Bases de Datos Relacional(es) en la/s que almacenan sus datos. La estructura completa del Sistema de Administración de Base de Datos Relacional (RDBMS) está específicamente diseñada para realizar transacciones en línea sobre los datos (OLTP: online transaction processing). El objetivo principal de los RDBMS es almacenar grandes cantidades de datos y realizar transacciones en la forma más eficientemente posible. El punto de vista con el que se diseñan este tipo de Bases de Datos está enfocado a la lógica de los procesos que forman parte del sistema y a la lógica operacional establecida. Las empresas que sólo cuentan con BD

Relacionales se enfrentan con diversos problemas a la hora analizar e integrar información que proviene de diversas áreas, los cuales van desde accesibilidad y visibilidad, por ejemplo, datos de ciertas áreas no se pueden "cruzar" con los de otra por cuestiones físicas, hasta formas distintas de representar un dato, cada área maneja la fecha en un formato determinado, utiliza diferentes unidades, identifica la misma entidad con diferentes claves.

El siguiente paso que dieron algunas empresas para resolver los inconvenientes antes mencionados, fue implementar un DataWarehouse (DW), que es un repositorio común e histórico con un alto grado de detalle de datos de toda la empresa. El modelo de datos que se implementa en el RDBMS de un DW tiene como objetivo facilitar el análisis sobre los datos de diversos sectores de la organización. El diseño de un DW está enfocado en los requerimientos de información del sector gerencial, quienes toman decisiones de gran impacto sobre la empresa, y en la lógica de negocios establecida, difiriendo mucho del tipo de estructura que se utiliza en un RDBMS de un sistema operacional.

Aún cuando un DW se diseña para soportar la recuperación de datos, su estructura es muy compleja y su uso complicado. Como una evolución de los DW surge la tecnología OLAP (online analytical processing), que tiene como clave principal brindar a un tomador de decisiones la capacidad de maximizar el análisis de datos, por medio de herramientas visuales de uso simple.

OLAP ayuda a los tomadores de decisiones a sintetizar la información de la organización, realizar comparaciones y personalizar vistas. Analizar información histórica, actual y proyectada en diversos escenarios (análisis "what-if") mediante herramientas que les permiten realizar sus propias consultas sobre los datos de la empresa sin tener que realizar complejas consultas como joins entre diferentes tablas de una RDBMS, ni conocer SQL, ni tener un sistema de información hecho a medida. El usuario puede crear sus propias consultas en cuestión de minutos.

Según el OLAP Council, que es una asociación de fabricantes y proveedores de esta tecnología, "On-Line Analytical Processing (OLAP) es una tecnología de software que permite a analistas, managers y ejecutivos compenetrarse dentro de los datos y acceder fácil, consistente y de forma rápida a una gran variedad de posibles vistas de la información, la cual fue transformada desde datos crudos, para reflejar la dimensionalidad real que el usuario tiene sobre su negocio".

Las aplicaciones OLAP visualizan y manipulan los datos usando un Modelo de Datos Multidimensional (MDM) específicamente diseñada para análisis de datos. A diferencia de la mayoría de las BD, una BD Multidimensional representa mejor la *vista de negocios* de los datos (en vez de la *vista del sistema*). El MDM refleja la forma en que los usuarios ven su negocio u organización.

OLAP es una tecnología multiusuario de gran capacidad, especialmente diseñada para soportar y operar sobre estructuras de datos multidimensionales, útil en ambientes de: Sistemas de Soporte de Decisión, Análisis Ad-hoc, DataWarehouse y Data Marts, etc. El diseño de un OLAP Server (como el caso de ORACLE EXPRESS) y de la estructura de datos Multidimensional es optimizado para la recuperación ad-hoc de información en cualquier orientación y nivel de agregación en forma rápida, simple y eficiente. Las consultas se ejecutan rápidamente debido a la organización de los datos. Se usan funciones aritméticas para acceder a las celdas y rotar los cubos OLAP. En la tabla 1, se presenta una comparación de las tecnologías que se presentaron.

El proyecto original, FOMEC 10011, consiste en la implementación de un Sistema de Soporte de Decisión (DSS) para brindar apoyo en la gestión académica. Dicho proyecto pertenece a dos Facultades de una conocida Universidad Argentina, y genera una serie de indicadores calculados a partir de un Data Warehouse que concentra información de los diversos Data Mart (Alumnado, Ciencia y Técnica, RRHH, Sec. Extensión Universitaria y Asuntos Estudiantiles, y, Administración y Finanzas). El proyecto "OLAP Server - FOMEC 10011", presentado en este trabajo, es una extensión

del proyecto original, ya que implementa un Servidor OLAP sobre el DW, para brindar a los tomadores de decisión los beneficios que brinda OLAP. (Véase punto 4)

	RDBMS OLTP	DataWarehouse	MDBMS OLAP
Operación Típica	Actualización	Reporte	Análisis
Cantidad de datos por transacción	Pequeña	Moderada	Grande
Nivel de detalle de los datos	Detallada	Detallada-Sumarizada	Mayormente Sumarizada
Edad de los datos	Actual	Histórica y actual	Histórica, actual y proyectada

Tabla 1: Comparación de las tecnologías para toma de decisión

2. MODELO DE DATOS MULTIDIMENSIONAL

El propósito de esta sección es describir los conceptos fundamentales del Modelo Multidimensional. Una gráfica de estos conceptos se puede ver en la Figura 1.

En el modelo de datos multidimensional, una *dimensión* es un objeto que contiene un conjunto de valores únicos que sirven para organizar los datos, con el propósito de analizarlos. En este proyecto definimos, por ejemplo, la dimensión Nivel Académico. Cada elemento en la lista que compone una dimensión, se denomina *dimension value;* por ejemplo, para nuestro caso: Paradigmas de Información, Sintaxis y Semántica de Lenguajes de Programación, Versión 1, Plan 95, Ing. en Sistemas de Información, Ing. Eléctrica. Una dimensión se asemeja a una clave primaria en el modelo de datos relacional (en un DW, los *dimension values* son almacenados en *dimension tables*).

Una *jerarquía* representa una forma particular de organizar los niveles dentro de una dimensión. Las jerarquías definen cómo van a ser agregados los datos desde niveles inferiores a niveles superiores, cómo los usuarios de las aplicaciones podrán navegar los datos usando herramientas de selección y drilling. Una dimensión puede tener múltiples jerarquías. Una jerarquía contiene múltiples niveles.

Un *nivel* es un objeto en el modelo de datos multidimensional que representa un nivel de agregación particular dentro de una dimensión. Los *dimension values* son organizados dentro de niveles. Cada nivel representa grupos de datos agregados (según una función de agregación) a partir del nivel inmediato inferior. Por ejemplo, dentro de la dimensión Nivel Académico, existe el nivel 1 correspondiente a Especialidad, cuyos valores pueden ser: Ing. en Sistemas de Información, Ing. Industrial, etc. Cada uno de estos valores tiene asociados hijos del Nivel 2. Por ejemplo Ingeniería en Sistemas de Información tiene asociada los Planes 95, 85, etc. Estos a su vez, tienen asociados las Versiones (nivel 3) de Plan, y cada Plan, tiene asociadas las Materias (nivel 4) que incluye.

Una *variable* es un objeto en el modelo de datos multidimensional que representa un hecho (fact) y contiene datos de un tipo particular. Una variable generalmente es numérica, también puede ser booleana o texto. Una variable se asemeja a una columna, que no pertenece a claves primarias ni foráneas en el modelo relacional, en la que se guarda un hecho en particular (en un DW, las variables son almacenadas en *fact tables*).

Cada combinación de valores de dimensiones identifica unívocamente cada dato dentro de la BD Multidimensional. Una variable es organizada o dimensionada por una o más dimensiones. Por ejemplo, la variable Nota (Figura 1) está dimensionada por: Alumno, Nivel Académico (mediante esta dimensión se identifica la materia, versión, plan y especialidad), Turno de Examen (orden en el calendario académico y año al que pertenece). Es decir, se generan todas las combinaciones posibles de estas tres dimensiones, y se almacena la Nota, en aquellas combinaciones para las que exista el dato.

Mediante *drilling down*, los usuarios pueden visualizar los valores "hijos" que están asociados con un valor "padre" en la jerarquía. Mediante *drilling up*, los usuarios colapsan la lista de valores descendientes asociados a un valor "padre". Por ejemplo, seleccionando "turnos normales 2000" se pueden desplegar "1er. turno normal del 2000", "2do turno normal del 2000", etc.

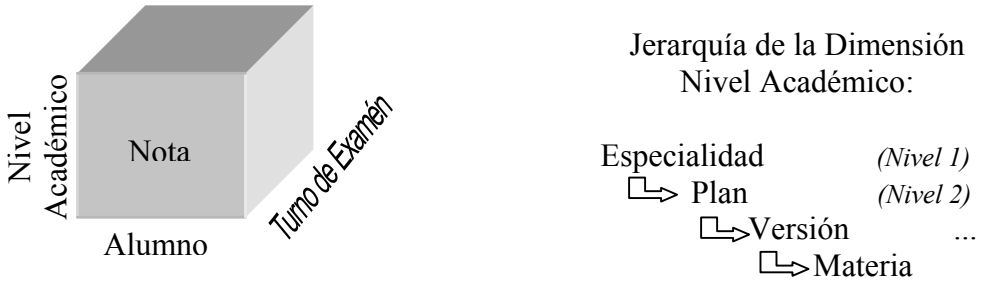

Figura 1: Conceptos del Modelo Dimensional

3. TECNOLOGÍAS OLAP

Existen tres tipos de tecnologías OLAP, MOLAP, ROLAP y HOLAP. En la primera (MOLAP - Multidimensional Online Analytical Processing) todos los datos están almacenados y precalculados en la BD Multidimensional, ó ciertos datos están almacenados y precalculados en la BD Multidimensional y el resto de los datos son calculados en el momento de la consulta según se requieran ("on the fly"). Existen pocos productos comerciales que implementan esta tecnología, a pesar que permiten aprovechar al máximo los beneficios de OLAP. En este proyecto se utilizó este tipo de tecnología.

En la segunda tecnología (ROLAP - Relational Online Analytical Processing) todos los datos están almacenados en una BD Relacional. Se define un Modelo de Datos Multidimensional al cual son "mapeados" los datos almacenados físicamente en tablas relacionales. Las aplicaciones trabajan con el modelo multidimensional y el acceso se realiza en forma transparente. A diferencia de MOLAP, no necesita almacenamiento en el MDBMS, pero necesita un mayor tiempo de respuesta en las consultas, ya que cada operación requiere el mapeo intermedio.

Por último, HOLAP (Hybrid Online Analytical Processing), es una combinación de MOLAP y ROLAP. Ciertos datos son almacenados en la BD Multidimensional y otros datos son recuperados desde la BD Relacional según sea necesario (en base a la demanda). Esta tecnología permite combinar la potencia de un MDBMS con la robustez de un RDBMS.

Oracle Express provee un módulo llamado Relational Access Manager que permite utilizar las dos últimas de estas tecnologías. Es decir, se define un modelo dimensional y se define qué datos se van a almacenar en el MDBMS y cuáles (o todos) se acceden desde el RDBMS mediante este módulo en forma transparente, sin necesitar de almacenamiento en el MDBMS. Se debe encontrar un equilibrio entre cantidad de almacenamiento y la velocidad de respuesta de las consultas.

4. ARQUITECTURA DEL PROYECTO – ORACLE EXPRESS

A continuación se describe la arquitectura del Proyecto FOMEC 10011, que brinda el marco para el proyecto que se presenta en este trabajo.

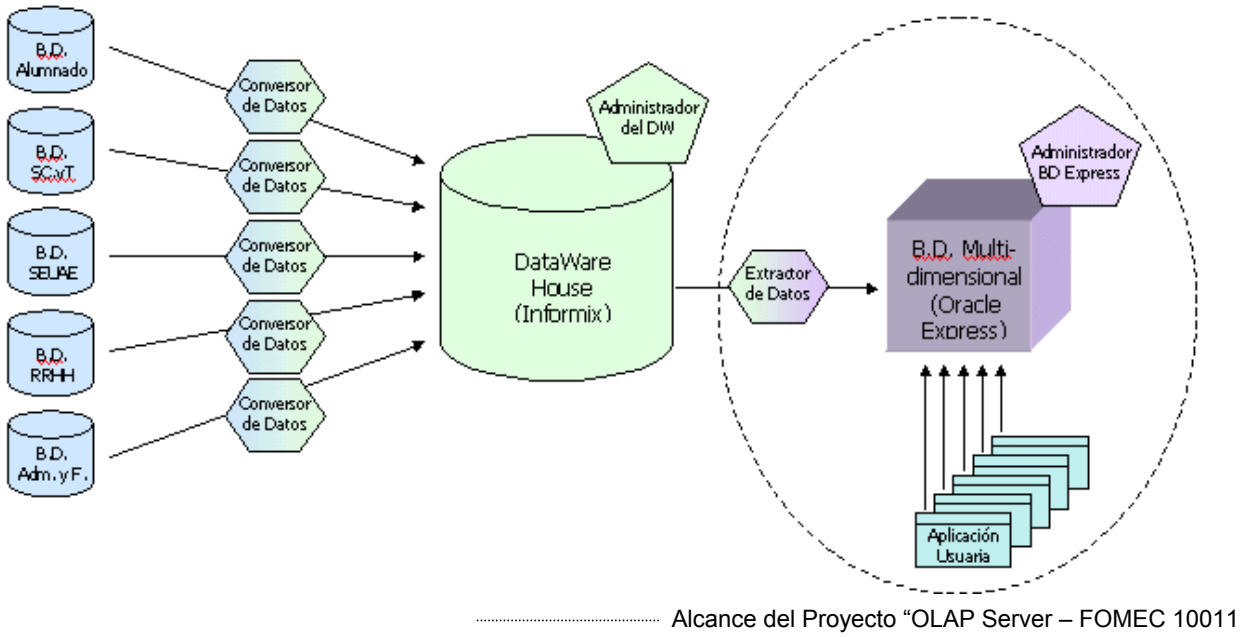

Figura 2: Arquitectura Proyecto FOMEC 10011

Para poder realizar un análisis sobre la problemática académica, primero, es necesario, localizar los datos en sus diversos orígenes, tales como los Sistemas Operacionales de Alumnado, Ciencia y Técnica, RRHH, Sec. Extensión Universitaria y Asuntos Estudiantiles, y, Administración y Finanzas. Cada uno de estos sistemas trabaja con su Base de Datos correspondiente. Para este proyecto fue necesario implementar un Data Warehouse para tener un repositorio común donde almacenar los datos de los distintos sistemas operacionales. Con esto se solucionan problemas de consistencia y así solucionar los problemas de accesibilidad, visibilidad y unificar las distintas formas de representar un dato. En el DW se guardan los datos históricos y con alto nivel de detalle. Cabe destacar que un OLAP Server puede tomar datos desde cualquier fuente homogénea y consistente, no necesariamente desde un DW.

Una vez que identificamos los datos en sus fuentes origen, se realiza la extracción, transformación y carga (ETL) de datos desde las BD Operacionales al DW mediante una serie de Conversores de Datos que integran y vuelcan la información. El DW está implementado sobre un RDBMS Informix, y requiere de un Administrador para su creación y mantenimiento. Es importante que el DW este bien creado y los datos se encuentren de manera consistente, dado que son la clave principal para el correcto funcionamiento del OLAP Server.

Luego, mediante un Extractor de Datos, se toman datos del DW, se efectúan las transformaciones correspondientes, y se cargan a la BD Multidimensional implementada en Oracle Express OLAP Server. Es indispensable contar con un Administrador para este servidor, ya que, además de realizar tareas de creación y mantenimiento, es responsable del *tuning* de datos, es decir, realizar las operaciones correspondientes para generar datos en el formato adecuado. Para esta tarea se utilizó O. E. Administrator y O. Relational Access Manager.

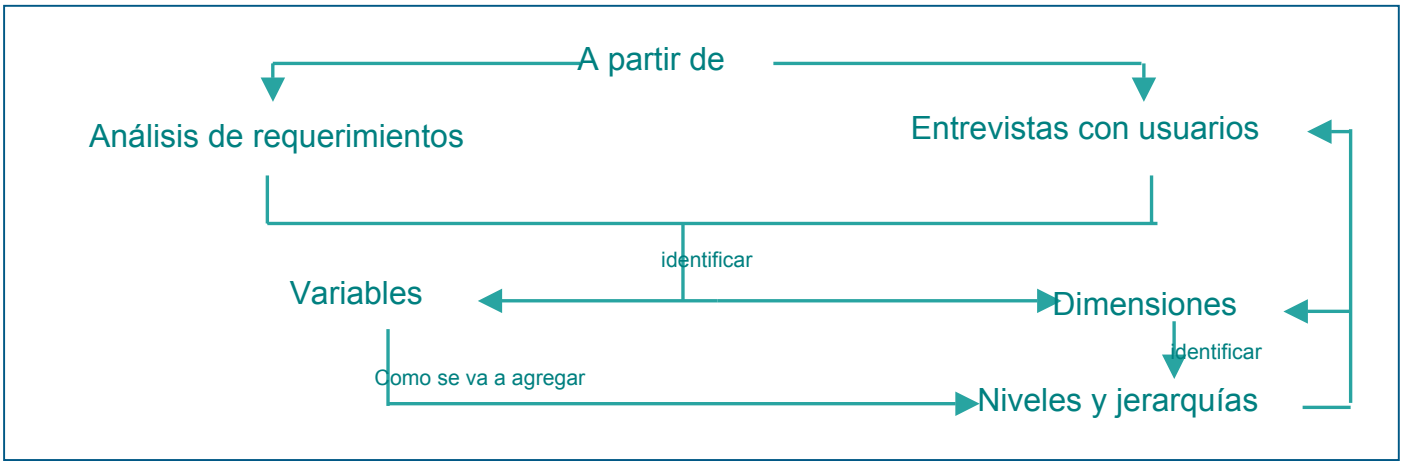

Figura 3: Diseño de la B.D. Multidimensional

Una vez cargada la BD Multidimensional y generados los datos correspondientes, se desarrollan las Aplicaciones para Usuarios Finales que consultan dichos datos. Para esta tarea se utilizó O.E. Object y O.E. Analyzer.

5. DISEÑO Y ADM. DE LA BD MULTIDIMENSIONAL

El primer paso para el diseño de la Base de Datos Multidimensional del proyecto, fue averiguar cuáles eran las necesidades de los usuarios, es decir, identificar cuáles eran los requerimientos que se debían satisfacer. Por ejemplo: ¿Qué datos se querían analizar?, ¿Qué tipos de consultas querían hacer sobre los datos? ¿A qué nivel de detalle necesitaban analizar dichos datos?

Cuando se planeó la estructura de los datos se debió pensar en términos de jerarquías y en el nivel de detalle debían tener. A partir de los niveles más bajos de detalle se formaron jerarquías para poder agrupar estos en conjuntos más generales, y así obtener niveles de datos agregados. En las Bases de Datos Multidimensionales la información de hechos (fact), como por ejemplo, *Nota de exámen,* (la nota que obtuvo un alumno en un determinado turno de examen para una materia en particular), son almacenados en variables. En forma paralela con la identificación de dimensiones, se identificaron las variables (fact) que almacenan los datos relacionados al proyecto, los cuales, son visualizados a través de dichas dimensiones.

Se recomienda planear primero y luego definir todos los objetos que se utilizarán en la Base de Datos multidimensional, ya que al definir las variables se debe indicar las dimensiones a las que está asociada, por lo que éstas últimas deben estar previamente definidas. En la figura 3 se esquematiza el proceso.

6. CARGA Y MANTENIMIENTO DE LA BD EXPRESS

Una vez definidas las dimensiones y variables, se procede a la carga de los datos a la BD Multidimensional del OLAP Server, para ello, se utilizaron dos métodos:

- Realizar una conexión vía ODBC mediante el módulo Oracle Relational Access Manager para acceder a un RDBMS (mediante esta módulo no sólo se cargan los datos al MOLAP, sino que se puede implementar un modelo ROLAP o HOLAP).
- Importar datos desde uno o más archivos de texto

En cuanto a la carga de datos, hay que tener en cuenta dos variantes, la carga masiva inicial y las de actualización. Estas últimas pueden ser utilizadas para agregar nuevos datos a las variables (notas de exámenes generadas después de la última carga) o para modificar algún objeto del modelo de datos. Luego de cargar los datos, se aplicaron las transformaciones y operaciones correspondientes, para que los mismos queden en el formato adecuado. Un *dato derivado* es información obtenida a partir de otros datos, ya sea mediante una fórmula, una expresión o a partir de la agregación de una variable, desde niveles inferiores de una jerarquía perteneciente a una dimensión dada. Los datos derivados pueden ser *precalculados* (se almacenan permanentemente en la BD multidimensional) o calculados en el momento de la consulta según se requieran ("on the fly").

En el caso del presente proyecto, las variables están agregadas hasta los primeros niveles de las jerarquías. De esta manera, los niveles más bajos están precalculados y almacenados en la B.D. Multidimensional, mientras que, para los niveles más altos, la agregación se realiza en el momento de la consulta según sea necesario. Mantener los datos de esta manera brinda un buen tiempo de respuesta para las consultas realizadas, ya que, en niveles bajos donde hay grandes cantidades de datos, los datos ya se encuentran agregados, mientras que, en los niveles superiores donde el volumen de los datos es pequeño, agregar datos en el momento de la consulta es muy rápido.

Cabe destacar que el diseño, creación y carga de la BD Multidimensional insumió un gran porcentaje del tiempo de este proyecto, principalmente porque se deben tener en cuenta varias cuestiones de performance para brindar flexibilidad y eficiencia. Si la BD Multidimensional no está perfectamente definida y con información consistente, las aplicaciones para usuarios finales no pueden accederla. Una de las razones por las que tuvimos que rever el diseño de la BD Multidimensional fue el esparcimiento de los datos, por ejemplo, sólo el 5 % de todas las posibles combinaciones entre las dimensiones asociadas a la variable *Nota* contenían datos, lo que afectaba drásticamente el espacio de almacenamiento.

7. DESARROLLO DE APLICACIONES OLAP PARA USUARIOS FINALES

En este proyecto se definieron dos tipos de interfaz. La primera a la que denominamos Interfaz de Usuario Simple, fue diseñada de forma estructurada. Es decir, en la interfaz se muestra el resultado de un indicador particular (sea éste simple o complejo) y el usuario puede ver los datos de la forma en que se le presentan, con dos o tres variantes preprogramadas. Esta interfaz está dirigida a tomadores de decisión con ningún conocimiento de informática. No brinda mucha flexibilidad, a cambio de su simplicidad.

El segundo tipo la llamamos Interfaz de Usuario Compleja, ya que no sólo brinda una variedad de posibles formas de navegar y explorar los datos, sino que además, mediante el "Selector" (Véase puntos 8 y 9) el tomador de decisiones puede definir y ejecutar sus

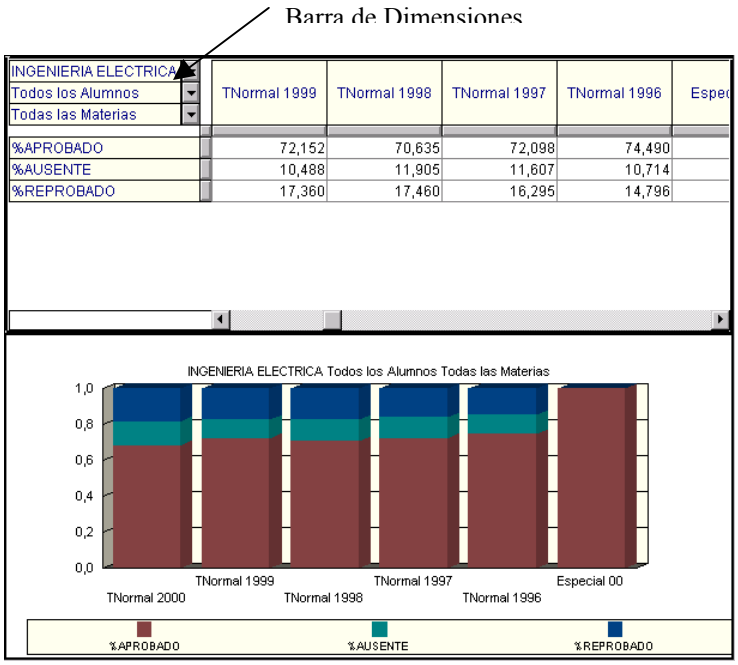

Figura 4: Barra de Dimensiones

propias consultas en cuestión de minutos. Esta interfaz está dirigida a usuarios con algún conocimiento de informática, o no lo tienen, pero han recibido una capacitación (1 o 2 días) en la herramienta. En este caso, se prioriza la flexibilidad para el usuario a costa de mayor complejidad.

Ambas interfaces le permiten a quienes la usan, visualizar los datos mediante tablas o gráficos, a los cuales se les puede configurar diversos permisos dependiendo el tipo de usuario. Las consultas pueden ser compartidas vía LANs (como se realizó en este proyecto) o publicadas en la Web para ser accedidas mediante cualquier browser (utilizando un módulo especial Oracle Express Web Publisher).

Existe una "Barra de Dimensiones", asociada a cada tabla o gráfico (Figura 4), la cual permite rotar los cubos OLAP (reorganizar los datos y visualizarlos en diversas perspectivas) con sólo hacer "drug and drop", además de poder navegar los distintos niveles de una jerarquía seleccionando simplemente las opciones los drill down/up correspondientes.

Para formular consultas de datos se utiliza el "*Selector*" (Figura 5), el cual incluye un conjunto de herramientas para desarrollar consultas multidimensionales, permite especificar criterios para incluir valores de dimensión en una selección, ordenar los valores de dimensión, buscar cierto valor de dimensión, seleccionar diversas variables a mostrar, filtrar ciertos niveles de agregación etc., esta posibilidad se explica en el siguiente item. El usuario final simplemente realiza las selecciones en el selector y los datos se actualizan inmediatamente a la forma correspondiente. Los usuarios finales pueden explorar y manipular los datos directamente sobre las tablas y gráficos. Se pueden visualizar valores agregados (sumarizados, promediados, extrayendo los mayores o menores, su desviación estándar) en vez de valores individuales en una tabla o gráfico. Para el caso de las notas, el tipo de agregación más utilizado fue el promedio.

8. ANÁLISIS AD HOC

Realizando análisis de manera Ad-hoc las posibilidades de analizar y visualizar la información a través de un servidor OLAP son muy amplias. A continuación se ejemplifica consultas que puede definir y ejecutar el Decano de la Facultad, o Secretario Académico o el Director de un Departamento, en cuestión de minutos, contando sólo con un cubo OLAP (utilizado como ejemplo en este trabajo, referidos a las notas de los exámenes).

En la figura 5 se muestra la ventana del selector. El botón (1) se utiliza para matchear valores de dimensión especificados. Permite seleccionar valores de cadenas de caracteres que "coincidan" con los valores que están seleccionados dentro de una dimensión, por ejemplo para la dimensión alumnos mostrar los datos de los alumnos que en el apellido se incluya la cadena "xyz".

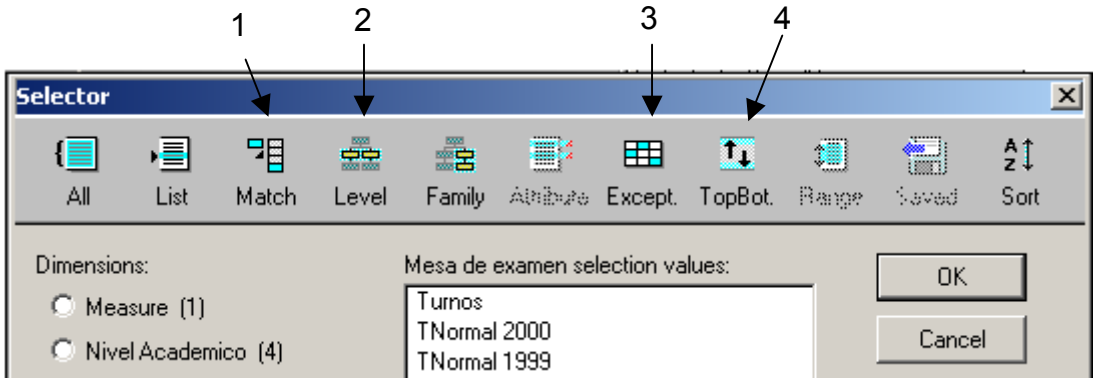

Figura 5: Selector

El botón (2) se utiliza para seleccionar un determinado nivel dentro de una jerarquía perteneciente a una dimensión. Por ejemplo, para la jerarquía de Nivel Académico antes mencionada, es posible seleccionar los valores pertenecientes al nivel 3 de la misma, o sea las versiones. El botón (3) de Excepciones se utiliza para seleccionar los valores que cumplan con una condición determinada. Por

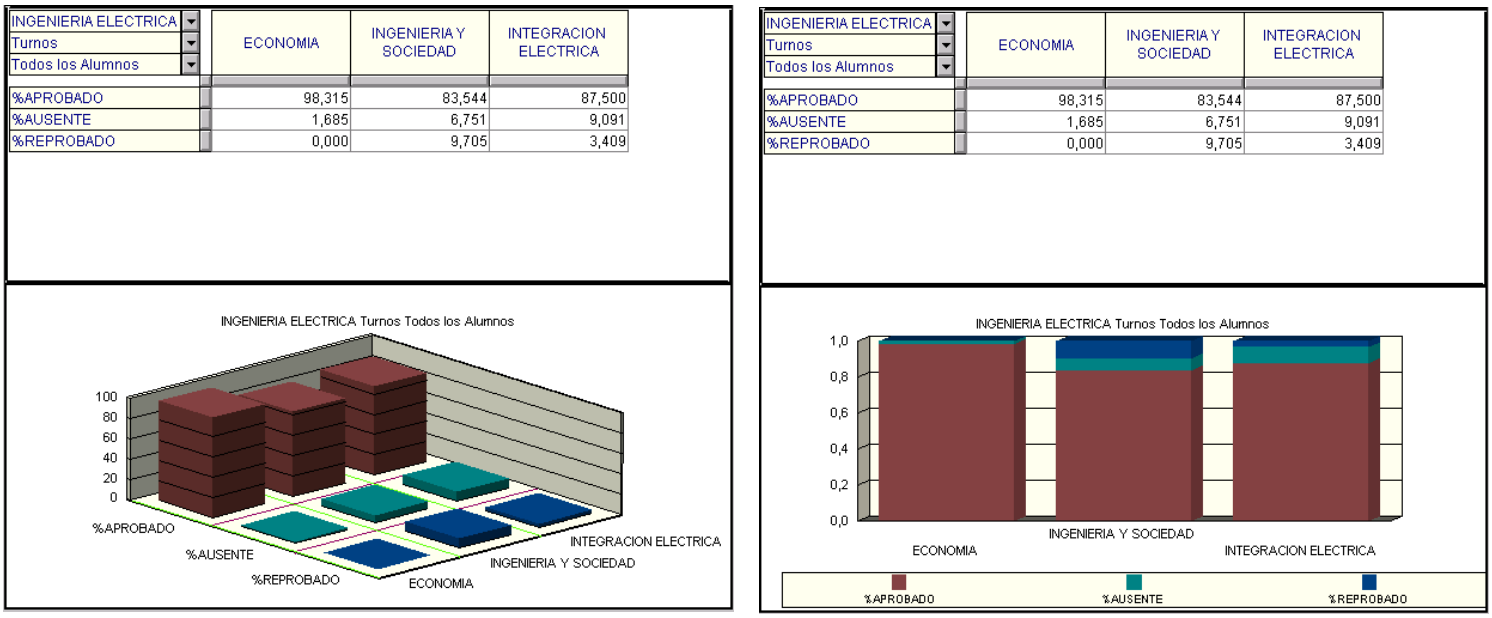

Figura 6: Ejemplo de consulta de Excepción

ejemplo visualizar el porcentaje de aprobados, no aprobados y ausentes, de aquellas materias que tengan un grado de aprobación mayor al 80 %, graficado en las figura 6.

Otro ejemplo de excepción, podría ser seleccionar los alumnos que pertenecen a un rango de una dimensión, como puede ser "Ing. en Sistemas" dentro de la dimensión "Nivel Académico" o a un período de tiempo determinado para la dimensión "Tiempo" y además que cumplen con una condición, como por ejemplo que tengan x cantidad de materias aprobadas. El botón (4), se utiliza para listar los "x" más altos o más bajos para una métrica dada. Por ejemplo, los "x" mejores o peores promedios (siendo x la cantidad de alumnos que se desee ver). Mediante otros botones, se pueden ordenar valores según diversos criterios tanto de los dimension value como de las variables, seleccionar un rango de tiempo determinado, listar ciertos valores a elegir, filtrar hechos que satisfagan algún atributo, etc.

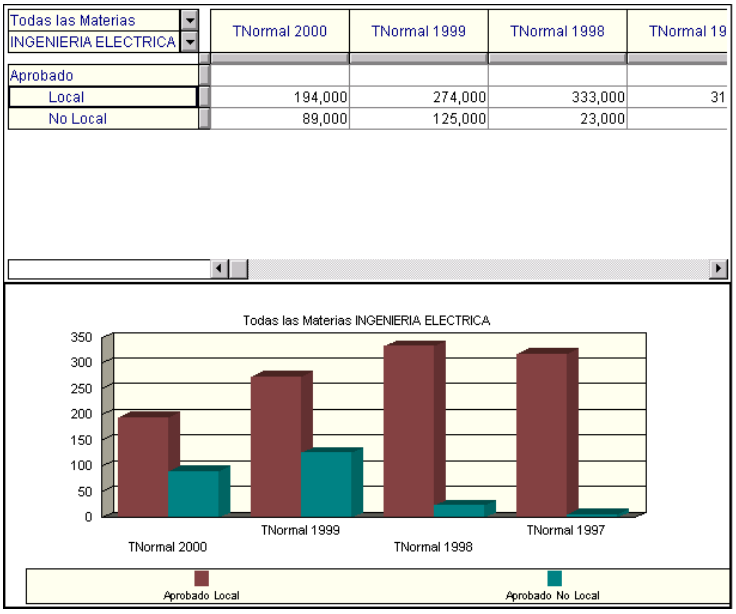

Figura 7: Visualizar datos según un atributo

En la figura 7 se grafica un ejemplo en el que se visualizan la cantidad de aprobados en turnos de exámenes de diversos años, en todas las materias de alumnos de Ing. Eléctrica, según el origen del alumno. El atributo tiene el valor local o no local, dependiendo si el alumno es de la ciudad en donde se encuentra la facultad o no.

9. CONCLUSIÓN Y TRABAJOS FUTUROS

Por medio de este proyecto, que implementa un OLAP Server en el ámbito académico, es posible demostrar que un tomador de decisión (como es el Decano, Secretario Académico o Director de Departamento), puede navegar y visualizar información en diversas perspectivas, en forma fácil, rápida y eficiente. Además de poder definir y ejecutar sus propias consultas ad-hoc sobre los datos, en cuestión de minutos, mediante una interfaz gráfica y amigable, sin la necesidad de tener conocimientos en bases de datos, ni SQL. También se demuestra la variedad de información que se puede generar a partir de un sólo cubo OLAP, es decir, a partir de sólo un fact, como es la *Nota* de los alumnos en sus respectivos exámenes finales.

Cabe destacar, que para que ello sea posible, es indispensable realizar un buen análisis y diseño de la base de datos multidimensional, ya que esta es la pieza fundamental que determina el grado de rapidez, flexibilidad y eficiencia que brinda un OLAP Server.

El proyecto que se presenta en este trabajo sirve como punto de partida para continuar trabajando con otras funcionalidades, como es el acceso web y data mining, así como también, en la construcción de indicadores más complejos, utilizando, por ejemplo, análisis de correlación.

En el caso del *acceso Web*, las características principales son la flexibilidad y la disponibilidad de los datos, ya que pueden ser accedidos desde cualquier parte y con las mismas funcionalidades que en el acceso local y no es necesario ningún aplicativo más que el browser.

Con respecto a las técnicas de *Data Mining* que se pueden aplicar a los datos residentes en el DW de una facultad o universidad, un futuro proyecto incluye la definición perfiles de alumnos, en base a actitudes, aptitudes y conocimientos, evaluación de una muestra de alumnos según estos perfiles mediante test y evaluaciones, seguimiento del rendimiento académico de los mismos a través de la carrera, para luego tratar de clasificar dichos alumnos en diversos grupos (alumnos excelentes que egresan en tiempo y forma, alumnos que abandonan en 1º o 2º año, etc.) y estudiar la relación entre cada grupo y el tipo de perfil que prevalece en los mismos. De esta manera se puede lograr un seguimiento más estrecho del alumno durante todo el período de estudio, y predecir, mediante una serie de test y evaluaciones que determinen el perfil del alumno, el futuro académico del mismo con cierta probabilidad.

BIBLIOGRAFIA

- R. Kimball, John Wiley & Sons, The Data Warehouse Lifecycle Toolkit, 1998.
- R. Kimball, Help for Dimensional Modeling, www.rkimball.com, 1998
- R. Kimball, Help for Hierarchies, www.rkimball.com, 1998
- W.H. Inmon, John Wiley & Sons, Building the Data Warehouse, 1996
- W.H. Inmon, What is a Data Mart?, www.informix.com, 1996
- W.H. Inmon, Building The Data Mart Or The Data Warehouse First?, www.informix.com, 1997
- W.H. Inmon, Data Marts and the Data Warehouse: Information Architecture for the Millenium, www.informix.com, 1999
- W.H. Inmon, The DSS Community: Tourists, Explorers and Farmers, www.informix.com, 1997
- Roger Geiwitz, Ebusiness Intelligence, www.billinmon.com, 2001
- Adding Value to the Data Warehouse, http://www.oracle.com/, 1998
- Delivering OLAP to the Enterprise, http://www.oracle.com/, 2000
- Productos Oracle, http://www.oracle.com/