

UNIVERSIDAD NACIONAL AUTONOMA DE NICARAGUA-LEON

FACULTAD DE CIENCIAS

DEPARTAMENTO DE COMPUTACIÓN



**DATAWAREHOUSING Y DATAMINNING
APLICACIONES EN VISUAL BASIC 6.0**

**Trabajo de monografía previo para optar al título de Licenciado
en Computación.**

Presentado por:

Br. Dudley Antonio Santamaría Jirón.

Tutor:

Msc. Martín E. Ibarra Padilla

León, Marzo, 2005

INDICE.

Tema.	Número de Página.
Introducción.	3
Antecedentes y Evolución.	5
Justificación.	7
Objetivos Generales.	8
Objetivos Específicos.	8
Diseño Metodológico.	9
Marco teórico.	10
Resultados.	34
Conclusiones.	35
Recomendaciones.	36
Bibliografía.	37

INTRODUCCIÓN

En la sociedad actual es comúnmente aceptado que la información constituye un activo esencial para proporcionar beneficios significativos a cualquier organización y que por lo tanto representa una ventaja competitiva en el mundo de los negocios.

El concepto de data warehousing se ha desarrollado rápidamente en una clase única y popular de aplicaciones para negocios a partir de la necesidad de disponer de un acceso sencillo e inmediato a determinada información de negocio estructurada y de calidad, que pueda ser empleada para la toma de decisiones.

Cada constructor de datawarehousing considera su sistema como un componente clave en la estrategia y arquitectura de su Tecnología de Información.

Hacer accesible la información de ayuda a la toma de decisiones, de manera instantánea a los responsables de la gestión de la empresa, constituye en definitiva, el objetivo intrínseco del data warehousing.

Podríamos definirlo informalmente como una copia de datos de transacciones específicamente estructurada para consultas y reportes. También se podría decir que un datawarehouse es la manipulación de datos estructurados después y fuera de un sistema operacional.

A estos conceptos aclararemos dos aspectos:

1. a veces se guardan datos no transaccionales en un almacén de datos aunque probablemente 95-99 % de los datos normalmente son datos de transacción
2. las preguntas e informes generados de datos guardados en un almacén de datos pueden o no usarse para el análisis. Finalmente algo que no se dice en la definición expuesta anteriormente es que la forma en que los datos son almacenados no tiene nada que ver con el hecho de que estos datos constituyan un datawarehouse.

Un datawarehousing puede normalizarse o denormalizarse, puede ser un banco de datos correlativo, un banco de datos multidimensional, un archivo llano, un banco de datos jerárquico, un banco de datos de objetos, etc. Los datawarehouses se enfocan a menudo en una actividad específica o entidad.

El datawarehousing no es un producto y no puede ser comprado; debe construirse paso a paso, es temático porque se organiza alrededor de conceptos de negocio de la empresa.

El **Data Mining**, extracción de información oculta y predecible de grandes bases de datos, es una poderosa tecnología nueva con gran potencial para ayudar a las compañías a concentrarse en la información más importante de sus Bases de información (Datawarehouse).

Las herramientas de Data Mining predicen futuras tendencias y comportamientos permitiendo en los negocios tomar decisiones proactivas y conducidas por un conocimiento acabado de la información (knowledge-driven). Los análisis prospectivos automatizados ofrecidos por un producto así van más allá de los eventos pasados provistos por herramientas retrospectivas típicas de sistemas de soporte de decisión. Las herramientas de Data Mining pueden responder a preguntas de negocios que tradicionalmente consumen demasiado tiempo para poder ser resueltas y a los cuales los usuarios de esta información no están dispuestos a aceptar. Estas herramientas exploran las bases de datos en busca de patrones ocultos, encontrando información predecible que un experto no puede llegar a encontrar porque se encuentra fuera de sus expectativas.

Muchas compañías ya colectan y refinan cantidades masivas de datos. Las técnicas de Data Mining pueden ser implementadas rápidamente en plataformas ya existentes de software y hardware para acrecentar el valor de las fuentes de información existentes y pueden ser integradas con nuevos productos y sistemas pues son traídas en línea (on line). Una vez que las herramientas de Data Mining fueron implementadas en computadoras cliente servidor de alta performance o de procesamiento paralelo, pueden analizar bases de datos masivas para brindar respuesta a preguntas tales como ¿Cuáles clientes tienen mas probabilidad de responder al próximo mailing promocional y por que? Y presentar los resultados en forma de tablas, con gráficos, reportes, texto, hipertexto, etc.

ANTECEDENTES Y EVOLUCIÓN

Por un lado tenemos el legado de sistemas de información que fueron desarrollados desde los años 70 sobre la plataforma IBM utilizando herramientas tales como Cobol, CICS, IMS; DB2, etc; durante los años 80 utilizando AS/400 y VAX/VMS y a finales de los 80 utilizando UNÍS, plataforma que introduce la arquitectura cliente/servidor.

Por otro lado, otra categoría popular de sistemas de apoyo en la toma de decisiones y los sistemas de información ejecutiva. Los sistemas de apoyo en la toma de decisiones tienden a enfocar mas los detalles de los datos y están enfocados a los encargados de las tomas de decisiones del nivel medio. Los sistemas de información ejecutiva generalmente proveen un alto nivel de consolidación y una vista multidimensional de los datos necesarios para los altos ejecutivos que parten y juegan con la información sin tener que detenerse en los detalles de los datos.

Estos dos tipos de sistemas son los más cercanos antepasados del datawarehouse. Hoy en día los sistemas basados en datawarehouse proveen todas las herramientas analíticas que exponen sus precursores.

Existen también otros antecedentes que favorecieron al desarrollo de sistemas basados en datawarehouse. Primero el incremento en las capacidades de procesamiento y almacenamiento del hardware moderno a la par del decremento de los precios de los mismos; segundo durante la década pasada se dio el incremento en la popularidad de las computadoras personales las que han introducido un sinnúmero de oportunidades y nuevas opciones para el análisis de los negocios. El puente entre los programadores y los usuarios finales ha iniciado ha cerrarse a medida de los analistas de los negocios tienen a sus posibilidades muchas de las herramientas requeridas para mejorar la eficiencia en el uso de las hojas de cálculo para la representación y análisis grafico de la información. Las capacidades modernas de las computadoras personales han permitido desarrollar sistemas basados en la arquitectura cliente / servidor de tal forma que la mayoría de los datawarehouses son accedidas mediante las herramientas para computadoras personales; tercero, desarrollo de sistemas operativos servidores tales como Windows NT y Unix han sido de vital importancia en el desarrollo de sistemas distribuidos cuyos costos han permanecido estables o en disminución.

Todo esto combinado con los conceptos de memoria virtual, multitarea y multiprocesamiento simétrico han ayudado a obtener plataformas de desarrollo para datawarehouses relativamente a bajo costo. Finalmente, el fenómeno más importante en el desarrollo de la computación ha sido la explosión en el desarrollo de sistemas basados en Internet y la Web.

Uno de los campos de mayor desarrollo en las industrias son las llamadas Intranets que no son más que redes privadas basadas en los estándares de Internet y que son diseñadas para uso interno de una empresa o negocio.

También se dan factores tales como la globalización basada en la demanda de información, el cambio en la naturaleza de los negocios, la caída del comunismo y la liberación de las economías de Asia y América del Sur y el surgimiento de aplicaciones estandarizadas para negocios han tenido un efecto directo en el uso de datawarehouses como mecanismo para reemplazar los sistemas anteriores y para facilitar las transacciones sobre grandes bancos de datos que son accedidos a través de Intranet o de Internet a grandes servidores multitareas.

Justificación.

Debido al crecimiento de información en las empresas y al mismo tiempo el lanzamiento de nuevas tecnologías para contrarrestar esta problemática es que se realizó este documento para dar a conocer como podemos manipular más información en poco tiempo y así desarrollar software con mas rendimiento y eficaces, con tecnologías nuevas.

Un Data Warehouse con éxito es hacer que la información que un negocio ha reunido, tanto de los sistemas operacionales como de algunos sistemas externos, sea accesible, de forma que se pueda realizar la toma de decisiones.

Al utilizar el Data Warehouse estamos aumentando los conocimientos de los datos almacenados de la mega empresa corporativa que antes no tenían conocimiento y/o acceso a estos.

Con los cubos OLAP, ahora se pueden tomar decisiones basadas en la información en tiempo real a la que posee acceso, tanto en la sucursal local como departamental e internacional.

OBJETIVOS:

OBJETIVO GENERAL:

Definir los conceptos y documentar minuciosamente las recientes técnicas de consulta a las bases de datos (Data Warehousing y Data Mining).

OBJETIVOS ESPECÍFICOS:

1. Documentar detalladamente en que consisten DataWarehouse y DataMining.
2. Explicar los requerimientos de hardware y software para la implementación automatizada del DataWarehouse y el DataMining.
3. Indicar los pasos para el desarrollo de una aplicación que haga uso práctico de los conceptos expuestos.

DISEÑO METODOLÓGICO

El trabajo consiste en una investigación documental sobre el tema y la realización de una aplicación del concepto de datawarehouse. La información se adquirió de referencias bibliográficas y fuentes obtenidas a través de Internet ya que no se encontró información disponible en ninguna biblioteca ni centros de investigación informática.

La búsqueda de información consistirá en el proceso de documentar todo lo relacionado con el DataWarehouse y el DataMining. Esta información constituirá el marco teórico del trabajo. Para la realización de la aplicación se tendrá en cuenta el ciclo de vida de desarrollo del software, pero el trabajo no pretende exponer con detalles la realización del sistema en sí sino que la aplicación del datawarehouse dentro de la aplicación.

MARCO TEÓRICO

Datawarehouse

Un Datawarehouse (DW), es un ambiente estructurado y extensible diseñado para el análisis de datos no volátiles, lógica y físicamente transformados de múltiples aplicaciones fuentes para alinearlos con la estructura del negocio, actualizados y mantenidos por un largo periodo de tiempo y sumarizados para permitir un análisis rápido.

La característica no volátil de la información hace referencia a la imposibilidad no física, sino conceptual de realizar datos. Cada registro del DW tiene el tiempo indisolublemente grabado en él. Visto de otra forma, cada registro es una fotografía de información tomada en un momento concreto del tiempo; si existiera la necesidad de realizar cambios en el entorno del DW se realizaría una nueva fotografía en un momento posterior, el resultado neto es que el DW representa un registro histórico de fotografías de información.

Una vez creado de manera fiable un registro en el DW este no debe ser modificado.

Arquitectura DataWarehouse

Poder transformar los datos en conocimientos en un proceso complejo. Un DW es mucho más que simplemente copiar datos de un lugar a otro, de los sistemas operacionales a una base de datos informacional independiente. Un DW es, en primer lugar, una arquitectura que debe servir como infraestructura para proporcionar una solución completa.

Los elementos principales de un sistema DW son:

- El componente de diseño, para el diseño de la base de datos del DW.
- El componente de adquisición de datos, para la captura de datos (procedentes de los sistemas fuente, limpieza, transporte y carga) en la base de datos del DW.
- El componente de gestor de datos, para la creación, gestión y acceso a los datos del DW.
- El componente de gestión, para la administración global de las operaciones.

- El componente de directorio de información, para proporcionar información a los usuarios sobre contenidos y significados de la información incluida en el DW.
- El componente de acceso a datos, proporciona a los usuarios finales herramientas necesarias para acceder y analizar la información del DW.
- El componente Middleware, proporciona a las herramientas de usuario final acceso a la base de datos del DW.
- El componente de distribución de datos, para distribuir la información del DW a otros sistemas.

El Componente de Diseño.

Es utilizado por los administradores y diseñadores para proyectar la base de datos del DW. En su diseño se han de considerar factores adicionales en comparación con el diseño de la bases de datos operacionales como el tipo de modelado (tercera forma normal o dimensional). La desnormalización de tablas, la inclusión de información agregada o consolidada o el tratamiento temporal. Para el diseño del DW habitualmente se emplean herramientas CASE.

El Componente de Adquisición de Datos.

El objetivo de este componente, también denominado a veces componente de integración y transformación, es desarrollar y ejecutar, en primera instancia, las aplicaciones de adquisición de datos que capturan la información procedente de los sistemas fuentes. Generalmente este proceso es conocido por sus siglas en inglés, ETL (Extraction, transformation and load) y es fundamental para el éxito del DW. Las aplicaciones de adquisición de datos se desarrollan siguiendo reglas identificadas en la fase de levantamiento de requisitos de a interfaz de entrada. Estas reglas definen, principalmente, las fuentes de datos a partir de las cuales se obtendrá la información para el DW, así como la periodicidad de los procesos de adquisición, los formatos esperados de los datos y la calidad de los mismo. Existen diversas maneras de implementarlo. Desde el desarrollo manual de programas específicos hasta el empleo de herramientas especializadas en esta actividades. Independientemente del método, hay ciertos aspectos fundamentales que deben ser considerados. La transformación es algo más que simplemente modificar las estructuras de datos mientras se pasan de los sistemas fuentes al DW, un buen proceso de transformación realmente verifica y mejorar la calidad de los datos y, por tanto, el uso de la información que finalmente se introduce en el DW. Uno de los principales errores en muchos abordajes de DW es subestimar el esfuerzo necesario para establecer una

correcta política de transformación de los datos. Habitualmente se considera que estas actividades representan entre el 70 por ciento y el 80 por ciento de los recursos de desarrollo en el diseño y construcción de un DW. Entre los aspectos a considerar en la fase de adquisición, y que son responsables de la complejidad del proceso, podemos citar:

- cambio de tecnología entre las bases de datos y entre las plataformas origen y destino, formatos de almacenamiento diferentes, por ejemplo, EBCDIC Y ASCII, modos de operación distintos.
- Actualización y no actualización.
- Necesidad de especificar valores por defecto.
- Secuencia distinta de los datos.
- Selección de datos a extraer.
- No todos los datos deben ser extraídos.
- Elección del momento en el tiempo para la extracción.
- Conversión de los datos.
- Limpieza de los datos.
- Existencia de múltiples fuentes de datos.
- Inclusión de procesos de auditoria de los propios procesos de adquisición.
- Inserción de la información temporal en los datos y creación de agregaciones y consolidaciones, y todo ello sin olvidar mantener un tiempo de proceso aceptable.

Cabe destacar que en esta fase se incluyen actividades como la limpieza y mejora de los datos antes de integrarlos en la base de datos. La limpieza de datos puede incluir la reestructuración de los registros o campos, la eliminación de información específica de los sistemas fuente, la generación de valores artificiales y la verificación de la integridad y consistencia de los datos. La parte de mejora puede involucrar la decodificación y traducción de valores, la unificación de los nombres correspondientes a áreas distintas, la inclusión de atributos tiempo, la síntesis de datos o el cálculo de los valores derivados.

El Componente de Gestor de Datos.

Este componente es utilizado por otros de la arquitectura para crear, gestionar y acceder a los datos y meta datos en ocasiones residentes en el DW. El gestor de datos utilizado habitualmente es una base de datos relacional (RDBMS) o una base de datos multidimensionales (MDBMS).

La base de datos relacionales es empleada para la construcción de grandes DWs corporativos o pequeños DWs departamentales; mientras que la base de datos multidimensionales se suele utilizar para DWs departamentales. Por otra parte la base de datos de los DWs tienen requerimientos por encima de los exigidos a las aplicaciones de los sistemas operacionales; los factores claves a considerar son la escalabilidad (tamaño de la base de datos, complejidad de las consultas y número de usuarios) y el rendimiento (aplicaciones de administración y procesamiento de consultas complejas)

A medida que el tamaño de la base de datos y la complejidad de las consultas se incrementa, es necesario considerar la utilización de arquitecturas de hardware y sistemas de gestión de base de datos paralelas para lograr un rendimiento satisfactorio.

El Componente de gestión

Está formado por el conjunto de servicios de administración de los sistemas para el mantenimiento del entorno del DW. Estos servicios incluyen la gestión de las operaciones de adquisición y carga de datos en el DW, copias de seguridad y recuperación de la información, seguridad y autorizaciones de acceso, así como las tareas de administración y ajuste asociadas, también puede incluirse la creación de metadatos sobre los procesos de sistema y la actividad de auditoria del uso que hace del DW.

El Componente de directorio de información.

Tiene como función ayudar a los usuarios del negocio a acceder y explotar las capacidades del sistema DW. Para ello se dispone de un conjunto de herramientas para el mantenimiento y visualización del metadata del DW, este metadata puede ser creado por los desarrolladores y administradores del sistema durante el proceso de diseño y construcción, y/o puede ser importado de productos externos como el catálogo de sistema de base de datos, las herramientas CASE (Computer Aided Software Engineering) o DSS. El metadata también puede crearse durante la fase de adquisición de datos y cuando se accede al DW; entre otros proporciona información sobre los orígenes y destinos de los datos, las reglas de limpieza y mejora, así como las relaciones entre las fuentes de datos y la base de datos del DW. El directorio debe incluir información sobre el volumen de datos disponible en el sistema y la fecha de su creación y actualización.

El Componente de acceso a datos.

Proporciona las herramientas de acceso de datos que permiten a los usuarios finales consultar y analizar la información contenida en el DW. Su organización esta alineada de acuerdo a los términos de negocio y las áreas temáticas definidas en DW, de manera que el usuario final puede trabajar en un entorno de conceptos que le son familiares-en el que se oculta la complejidad interna del sistema.

Existen distintas categorías de componentes, dependiendo del tipo de usuario al que esté enfocado:

- En primer lugar, para los usuarios básicos se encuentran las aplicaciones que permiten realizar informes predefinidos que se ejecutaran de forma más o menos periódica.
- En segundo lugar están las aplicaciones que facilitan la ejecución de análisis elemental de los datos contenidos en el DW.
- En un tercer apartado se encuentran las aplicaciones que posibilitan la realización de análisis avanzados.

El Componente de “Middleware”

El componente Middleware proporciona la conectividad necesaria entre la base de datos del DW y las herramientas de acceso de usuario final. Existen dos tipos de Middleware especializado:

1. El Middleware inteligente, que proporciona una visión de negocio del DW a los usuarios finales y supervisa y registra el acceso a la información del DW.
2. Los servidores analíticos, que mejoran las prestaciones de las operaciones de análisis multidimensional en un entorno de base de datos relacional.

El Componente de Distribución de Datos.

Es el responsable de distribuir la información del DW a otras aplicaciones, como pueden ser Data Marking, Hojas de Cálculo, Bases de Datos Locales, aplicaciones DBM o de CRM, etc.

El Ciclo de Vida de Desarrollo de un DataWarehouse.

Se recomiendan los siguientes pasos para llevar a cabo el ciclo de vida de desarrollo de un Datawarehouse.

1. planeación.
 - A. seleccionar la estrategia de implementación.

- i. metodología Top Down (de arriba hacia Abajo).
 - ii. Metodología Bottom Up (de abajo hacia arriba).
 - iii. Combinación de las metodologías.
- B. Seleccionar la estrategia de desarrollo.
 - i. Análisis estructurado y método de diseño en cascada.
 - ii. Método de desarrollo en espiral.
- C. Desarrollar los objetivos del negocio.
 - i. ¿cuál es el mercado final para el Datawarehouse?
 - ii. ¿cuáles son las plataformas actuales en uso o planificadas a utilizar?
 - iii. ¿cuáles son las diferentes bases de datos fuentes que pueden y/o deben ser integradas en el Datawarehouse?
- D. Seleccionar una meta o alcance de implementación inicial.
 - i. Dimensiones típicas para las restricciones del alcance.
 - 1. Número y tipos de departamentos servidos.
 - 2. número de datos fuentes seleccionados.
 - 3. subconjunto del modelo de empresa seleccionado.
 - 4. localización del presupuesto.
 - 5. localización del tiempo para el 1^{er} proyecto.
 - ii. Las dimensiones pueden ser ampliamente divididas en dos categorías:
 - 1. alcance de terminado desde la perspectiva de los usuario del Datawarehouse del negocio.
 - a. ¿Cuáles son los departamentos que necesitan utilizar inicialmente el Datawarehouse y con qué propósitos?
 - 2. ¿Cuál es el rango de las consultas del negocio que deben ser inicialmente respondidas por el Datawarehouse?
- 2. Determinación del alcance basado en consideraciones tecnológicas.
 - a. ¿Cuál es el tamaño del metamodelo del Datawarehouse;
 - b. ¿Cuáles son los tamaños de los datos dentro del Datawarehouse?
 - c. ¿Cuáles y cuántas son las fuentes de entrada de datos?
 - d. ¿Cuán útiles son los datos de las fuentes de datos?
 - e. ¿Cuál es el nivel de Manejo de facilidades Integradas?

- f. ¿Cuál es la disponibilidad de modelos lógico y herramientas CASE?
- g. ¿Existe un modelo de datos de empresa disponible?
- h. ¿Pueden los recursos y habilidades humanas existentes ser utilizados?
- i. ¿El Datawarehouse será implementado en una nueva plataforma o en alguna plataforma similar a las existentes?

E. Seleccionar una aproximación de la arquitectura.

- i. Almacenaje operacional versus usando copias de datos operacionales.
- ii. Solamente Datawarehouse.
- iii. Solamente data marts (centro de datos)
- iv. Datawarehouse y data marts.
- v. Particionamiento de la plataforma y la infraestructura.
- vi. Arquitectura cliente/servidor de dos niveles.
- vii. Arquitectura cliente/servidor de tres niveles.

F. Construir los presupuestos para el programa y el plan del proyecto.

- i. Articular un plan del programa y un conjunto de planes para el proyecto.
- ii. Reservar y adecuar el presupuesto para el programa mientras se consignan los gastos para proyectos específicos.
 - 1. Estimación de costos basada en el historial de la organización con respecto al desarrollo de software.
 - 2. Estimación de costos basada en la referencia arquitectónica.
- iii. Proveer medidas para la estimación de la compensación por el uso e implementación del Datawarehouse.
 - 1. Medidas de recuperación de costos y ahorro de costos.
 - 2. Medidas de oportunidades de creación.
 - 3. Medidas de creación de ingresos.
 - 4. Medidas del crecimiento del mercado.
 - 5. Medidas de ventajas competitivas.
 - 6. Medidas de satisfacción de los clientes.

G. Desarrollar los casos de uso y los escenarios para clarificar las expectativas de los usuarios.

i. Un escenario consistente de:

1. Usuarios del negocio claramente identificados con funciones del negocio claramente identificadas.
2. Un área funcional que patrocina el Datawarehouse o el data mart y que será utilizada cuando esta sea entregada. Una o más consultas del negocio de crucial interés para el área funcional que no está actualmente satisfecha con los sistemas de información existentes.

ii. Son útiles en las siguientes instancias:

1. Construir los criterios de aceptación para el Datawarehouse.
2. Identificar el metamodelado del Datawarehouse.
3. Identificar el monto de información histórica necesaria.
4. Identificar las dimensiones de interés de los usuarios finales.
5. Identificar las necesidades del DataWarehouse o del datamart.

H. Colectar los metadatos.

2. Requerimientos.

A. Definir los requerimientos del propietario.

i. Potenciales requerimientos por áreas de interés.

1. Objetivos del negocio.
2. Alcance, objetivos y usuarios del Datawarehouse.
3. Requerimientos de los clientes.
4. Fuentes de datos.
5. Planes, presupuestos, agenda y recursos.
6. Impacto de la inversión actual tales como gente, tecnología y entrenamiento.

ii. Parte de los requerimientos del negocio son también especificaciones del alcance para el Datawarehouse en términos de lo siguiente:

1. Áreas de estudio, tópicos de interés para varias funciones del negocio.

- a Análisis de tópicos seleccionados procurando áreas específicas de interés para el departamento de mercadeo.
 - 2. Granularidad, a menor granularidad, mayor el monto de detalles.
 - 3. Dimensiones.
- B. Definir los requerimientos arquitectónicos.
- C. Definir los requerimientos de los desarrolladores.
- D. Definir los requerimientos de los usuarios finales.
- 3. Análisis.
 - 4. Diseño.
 - 5. Construcción.
 - 6. Instalación.

Diseño y modelización

Los requerimientos de información identificados durante la anterior fase proporcionarán las bases para realizar el diseño y la modelización del Data Warehouse.

En esta fase se identificarán las fuentes de los datos (sistema operacional, fuentes externas,..) y las transformaciones necesarias para, a partir de dichas fuentes, obtener el modelo lógico de datos del Data Warehouse. Este modelo estará formado por entidades y relaciones que permitirán resolver las necesidades de negocio de la organización.

El modelo lógico se traducirá posteriormente en el modelo físico de datos que se almacenará en el Data Warehouse y que definirá la arquitectura de almacenamiento del Data Warehouse adaptándose al tipo de explotación que se realice del mismo.

La mayor parte estas definiciones de los datos del Data Warehouse estarán almacenadas en los metadatos y formarán parte del mismo.

Implementación

La implantación de un Data Warehouse lleva implícitos los siguientes pasos:

- Extracción de los datos del sistema operacional y transformación de los mismos.
- Carga de los datos validados en el Data Warehouse. Esta carga deberá ser planificada con una periodicidad que se adaptará a las necesidades de refresco detectadas durante las fases de diseño del nuevo sistema.
- Explotación del Data Warehouse mediante diversas técnicas dependiendo del tipo de aplicación que se de a los datos:

Query & Reporting

On-line analytical processing (OLAP)

Executive Information System (EIS) ó Información de gestión

Decision Support Systems (DSS)

Visualización de la información

Data Mining ó Minería de Datos, etc.

La información necesaria para mantener el control sobre los datos se almacena en los metadatos técnicos (cuando describen las características físicas de los datos) y de negocio (cuando describen cómo se usan esos datos). Dichos metadatos deberán ser accesibles por los usuarios finales que permitirán en todo momento tanto al usuario, como al administrador que deberá además tener la facultad de modificarlos según varíen las necesidades de información.

Con la finalización de esta fase se obtendrá un Data Warehouse disponible para su uso por parte de los usuarios finales y el departamento de informática.

Revisión

La construcción del Data Warehouse no finaliza con la implantación del mismo, sino que es una tarea iterativa en la que se trata de incrementar su alcance aprendiendo de las experiencias anteriores.

Después de implantarse, debería realizarse una revisión del Data Warehouse planteando preguntas que permitan, después de los seis o nueve meses posteriores a su puesta en marcha, definir cuáles serían los aspectos a mejorar o potenciar en función de la utilización que se haga del nuevo sistema.

Diseño de la estructura de cursos de formación

Con la información obtenida de reuniones con los distintos usuarios se diseñarán una serie de cursos a medida, que tendrán como objetivo el proporcionar la formación estadística necesaria para el mejor aprovechamiento de la funcionalidad incluida en la aplicación. Se realizarán prácticas sobre el desarrollo realizado, las cuales permitirán fijar los conceptos adquiridos y servirán como formación a los usuarios.

Estrategias de Implantación

Algunas consideraciones que recomendamos que deben seguirse a la hora de abordar un proyecto de este tipo:

"La Base de Datos de Riesgos debe estar separada de las Bases de Datos Operacionales" con objeto de no interferir en la actividad del día a día, disponiendo de la información necesaria para Riesgos (interna y externa) y en un entorno orientado hacia la consulta y el análisis (Data Warehouse).

"Concepción del sistema como un conjunto de herramientas de análisis", debido a que las actividades de Análisis de Riesgos no se pueden automatizar completamente, puesto que requieren análisis y decisiones del usuario.

"Diseño del sistema no orientado a procesos"; se debe disponer de un conjunto abierto de herramientas que se utilizan con propósitos determinados no relacionados con las necesidades operativas.

"Abordar el sistema con un enfoque de desarrollo gradual", se debe comenzar con un esqueleto básico de funcionalidad y datos que produzcan resultados a corto plazo y permita aprender en la práctica, y a continuación ir configurando progresivamente nuevas funcionalidades conforme la experiencia lo vaya requiriendo.

Los sistemas de soporte a la decisión usando tecnologías de Data Warehouse, se llaman sistemas OLAP (siglas de On Line Analytical Processing). En general, estos sistemas OLAP deben:

- Soportar requerimientos complejos de análisis
- Analizar datos desde diferentes perspectivas
- Soportar análisis complejos contra un volumen ingente de datos

La funcionalidad de los sistemas OLAP se caracteriza por ser un análisis multidimensional de datos corporativos, que soportan los análisis del usuario y unas posibilidades de navegación, seleccionando la información a obtener.

Normalmente este tipo de selecciones se ve reflejada en la visualización de la estructura multidimensional. En unos campos de selección que nos permitan elegir el nivel de agregación (jerarquía) de la dimensión, y/o la elección de un dato en concreto, la visualización de los atributos del sujeto, frente a unas(s) dimensiones en modo tabla, pudiendo con ello realizar, entre otras las siguientes acciones:

Proceso de Diseño:

Básicamente, el desarrollo de un DW consta de los siguientes pasos:

1. identificar los requerimientos de usuario y delimitar el ámbito del proyecto.
2. desarrollar el modelo de datos lógico del DW, correspondiente a los requerimientos del punto anterior. Normalmente, este modelo cubre aspectos particulares del negocio, como información.
3. implementar la arquitectura que constituirá la infraestructura del DW.
4. implementar físicamente el modelo lógico de datos, es decir, crear la base de datos.
5. dar a conocer el metadata del DW al usuario final. El propósito es familiarizar al usuario desde el primer momento con la visión analítica que propone el sistema.
6. identificar de manera detallada las fuentes de datos, sistemas operacionales y/o externos del DW, así como describir los procesos de conversión que los datos origen sufrirán hasta llegar a ser incorporados.
7. desarrollar los programas manual o automáticamente de extracción, limpieza, conversión e integración de los datos fuente.
8. cargar la base de datos del DW utilizando los programas anteriores.
9. verificar la satisfacción del usuario respecto al DW, incluyendo la calidad de los datos, así como su disponibilidad y el rendimiento.

Las arquitecturas base de una implantación de Data Warehouse en Internet, pueden tener las siguientes alternativas:

1. Usar el Servidor Internet como router (encaminador), y ejecutar la petición desde el cliente al servidor directamente.
2. Hacer uso del navegador para visualizar una página Internet residente en el servidor de Internet. Esta página contendría información que se actualizaría en el servidor Internet, desde el servidor DW, a petición del usuario haciendo uso de CGI's.
3. El cliente podría lanzar su consulta directamente al servidor de DW, con "applets" de Java, haciendo el servidor Internet únicamente de encaminamiento (router).
4. El cliente podría ejecutar la aplicación DW desde el navegador, pero con un plug-in, que haría que se tuvieran las mismas opciones que la aplicación DW.

Seguridad de Acceso y Manipulación de la Información en el DW

A continuación tratare las consideraciones a contemplar en cuanto a seguridad de accesos y seguridad de datos (backup), puesto que si bien la seguridad de accesos (al nivel de datos y de aplicación) debe ser tratada de la misma manera que en los sistemas operacionales, los procedimientos de copias de seguridad merecen un especial tratamiento.

Tal y como ocurre en los sistemas operacionales, un sistema Data Warehouse debe poder realizar procedimientos de recuperación de la información desde cualquier momento en el que los datos estaban validados. Un Data Warehouse, debe poder contar con procedimientos de recuperación, que permitan recuperar los datos ante cualquier situación de catástrofe.

No obstante, es preciso tener en cuenta otras consideraciones, así por ejemplo dependiendo del tamaño de un Data Mart, se puede elegir no realizar un backup, sino realizar un refresco especial desde los datos operacionales, dependiendo de la periodicidad estándar de carga.

En cuanto a la seguridad de acceso, se cumple en los sistemas de Data Warehouse, que es preciso el implantar niveles de acceso a la información, realizando un plan completo de seguridad que contemple:

- Acceso a recursos de la red (local o intranet)
- Asignación de usuarios a grupos con perfiles de seguridad diferenciados

- Asignación de niveles de autorización de aplicación a grupos de usuarios
- Seguridad a nivel de Base de Datos, mediante los procedimientos provistos por las mismas.
- Etc.

Entorno Hardware:

Host: (Tipo Máquina, Sistema Operativo y Base de Datos Operacional)

Servidor de la aplicación de DW (Tipo Máquina, Número de procesadores, Memoria total y por procesador, Sistema Operativo, Capacidad de Disco)

Clientes (Tipo de Máquina, Memoria, Sistema Operativo, Capacidad de Disco)

Red local (Topología, Protocolos, Software de Cliente y Número de Usuarios soportados)

Entorno Software:

Gestor de Base de Datos para el Data Warehouse

Volumen estimado de la Base de Datos

Data Mining.

El DataMining es un proceso que, a través del descubrimiento y la cuantificación de relaciones predictivas en los datos, permite transformar la información disponible en conocimiento útil de negocio. Esto es debido a que no es suficiente “navegar” por los datos para resolver los problemas de negocio, sino que se hace necesario seguir una metodología ordenada que permita obtener rendimientos tangibles de este conjunto de herramientas y técnicas de las que dispone el usuario.

Constituye por tanto una de las vías clave de explotación del Data Warehouse, dado que es este su entorno natural de trabajo.

Se trata de un concepto de explotación de naturaleza radicalmente distinta a la de los sistemas de información de gestión, dado que no se basa en coeficientes de gestión o en información altamente agregada, sino en la información de detalle contenida en el almacén. Adicionalmente, el usuario no se conforma con la mera visualización de datos, sino que trata de

obtener una relación entre los mimos que tenga repercusiones en su negocio.

Técnicas de Data Mining.

Para soportar el proceso de Data Mining, el usuario disponible de una extensa gama de técnicas que le pueden ayudar en cada una de las fases de dicho proceso, las cuales pasamos a describir:

1. ANOVA: o Análisis de la Varianza contrasta si existen diferencias significativas entre las medidas de una o más variables continuas en grupo de población distintos.
2. Regresión: define la relación entre una o más variables y un conjunto de variables predictorias de las primeras.
3. Ji Cuadrado: contrasta la hipótesis de independencia entre variables.
4. Componentes Principales: permite reducir el número de variables observadas a un menor número de variables artificiales, conservando la mayor parte de la información sobre la varianza de las variables.
5. Análisis Cluster: Permite clasificar una población en un número determinado de grupos, en base a semejanzas y desemejanzas de perfiles existentes entre los diferentes componentes de dicha población.
6. Análisis Discriminante: Método de clasificación de individuos en grupos que previamente se han establecido, y que permite encontrar la regla de clasificación de los elementos de estos grupos y por tanto identificar cuáles son las variables que mejor definan la pertenencia al grupo.

Metodología de aplicación:

Para utilizar estas técnicas de forma eficiente y ordenada es preciso aplicar una metodología estructurada, al proceso de Data Mining. A este respecto proponemos la siguiente metodología, siempre adaptable a la situación de negocio particular a la que se aplique:

Muestreo

Extracción de la población muestral sobre la que se va a aplicar el análisis. En ocasiones se trata de una muestra aleatoria, pero puede ser también un

subconjunto de datos del Data Warehouse que cumplan unas condiciones determinadas. El objeto de trabajar con una muestra de la población en lugar de toda ella, es la simplificación del estudio y la disminución de la carga de proceso. La muestra más óptima será aquella que teniendo un error asumible contenga el número mínimo de observaciones.

En el caso de que se recurra a un muestreo aleatorio, se debería tener la opción de elegir

- El nivel de confianza de la muestra (usualmente el 95% o el 99%).
- El tamaño máximo de la muestra (número máximo de registros), en cuyo caso el sistema deberá informar del el error cometido y la representatividad de la muestra sobre la población original.
- El error muestral que está dispuesto a cometer, en cuyo caso el sistema informará del número de observaciones que debe contener la muestra y su representatividad sobre la población original.

Para facilitar este paso se debe disponer de herramientas de extracción dinámica de información con o sin muestreo (simple o estratificado). En el caso del muestreo, dichas herramientas deben tener la opción de, dado un nivel de confianza, fijar el tamaño de la muestra y obtener el error o bien fijar el error y obtener el tamaño mínimo de la muestra que nos proporcione este grado de error.

Exploración

Una vez determinada la población que sirve para la obtención del modelo se deberá determinar cuales son las variables explicativas que van a servir como "inputs" al modelo. Para ello es importante hacer una exploración por la información disponible de la población que nos permita eliminar variables que no influyen y agrupar aquellas que repercuten en la misma dirección.

Modelo en Cascada vs. Ciclo de Vida Data Warehouse.

Modelo en Cascada. Big Bang	Data Warehouse. Teoría de la Evolución.
1. Ingeniería de Sistema	1. Escuchar al Usuario
2. Análisis de Requisitos del Software.	2. Implementar una primera visión comercial del Warehouse.
3. Diseño	3. Escuchar al Usuario.

4. Codificación.	4. Desarrollar mecanismos de soporte analítico.
5. Prueba	5. Escuchar al Usuario.
6. Mantenimiento.	6. Ir al paso 3.

Archivo DSN (Data Source Name).

Almacena la información de conexión para una base de datos en un archivo que se guardó en su equipo. El archivo es un archivo de texto con la extensión. .sn. la información de conexión consta de parámetros y sus valores correspondientes que el administrador del controlador de ODBC utiliza para establecer una conexión.

ODBC (Conectividad Abierta de Base de Datos)

Es un protocolo estándar para servidores de base de datos que proporciona un lenguaje común para aplicaciones Windows®.

Características generales de ODBC

- Rendimiento muy eficaz.
- Dificultad de programación.
- Requisitos de memoria razonables.
- Compatibilidad con tecnologías existentes de base de datos.
- Portabilidad entre muchas plataformas de sistemas Operativos.
- Un modelo de conexión que admite diferentes redes, sistemas de seguridad y opciones de base de datos.

Control ADO:

El Control de datos ADO (Objetos de acceso a datos) es similar al control intrínseco Data y al Control de datos remotos (RDC). El Control de datos ADO permite crear rápidamente una conexión con una base de datos mediante Objetos de datos ActiveX de Microsoft (ADO).

Ejemplo de un Data Warehouse y Data Mining.

Este ejemplo tendremos la operaciones de consultas que se generan de forma estática, con esta nueva tecnología podremos hacerlas dinámicas, donde el usuario final podrá realizar las consultas cuales quiera según sus necesidades de generar reportes e informes.

Este sistema tiene por objetivo primordial llevar el control de asistencia mensual de niños y niñas; también el control de inventario del alimento que es donado a ellos como es el desayuno escolar, de las escuelas de los distintos municipios del departamento de Jinotega.

Pasos para realizar un Data Warehouse.

1. Realizamos una fuente da datos ODBC, esta se encuentra en Panel de Control, para acceder a esta damos clic en **Inicio**, luego damos clic en **Configuración**, **Panel de Control**, **Herramientas Administrativas**. Damos clic en **Orígenes de Datos**, luego nos aparecerá la siguiente ventana.



2. Donde seleccionaremos la pestaña DSN de archivo, donde escogeremos la opción agregar.



3. Luego seleccionamos el controlador del nuevo origen el

que será en este caso Microsoft Acces Driver, damos clic en el botón finalizar.



4. Nos aparecerá la siguiente ventana donde pondremos el nombre al controlador y seleccionaremos la base de datos que queremos enlazar, en este caso la Base de Datos se llama sistemaescolar.mdb.

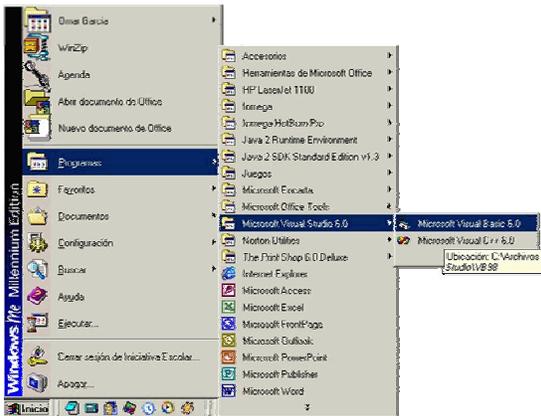


5. En esta ventana daremos la dirección de ubicación de la base de datos que se encuentra en C:\PC\sistemaescolar.MDB.



6. Damos clic en el botón de aceptar de todas las ventanas que vayan quedando.

7. Luego abrimos la aplicación en modo diseño en Visual Basic 6, que se encuentra en: Inicio, Programa, Microsoft Visual Studio 6.0, Microsoft Visual Basic 6.



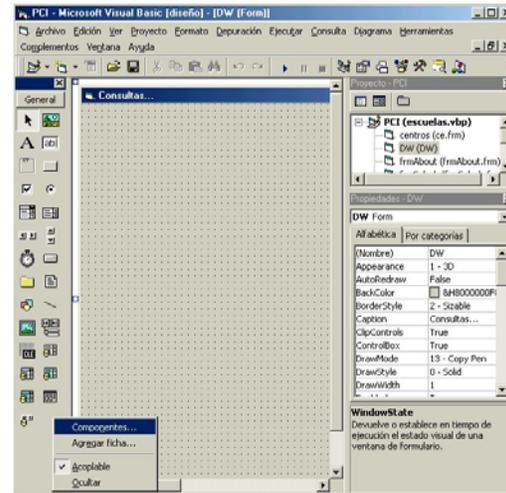
8. Abrimos nuestra aplicación dando clic en el icono "abrir", donde escogeremos la carpeta donde tenemos guardada nuestra aplicación en nuestro caso sistema escolar.

9. Una vez abierta nuestra Aplicación procedemos a insertar un formulario dando clic en Proyecto luego en Agregar

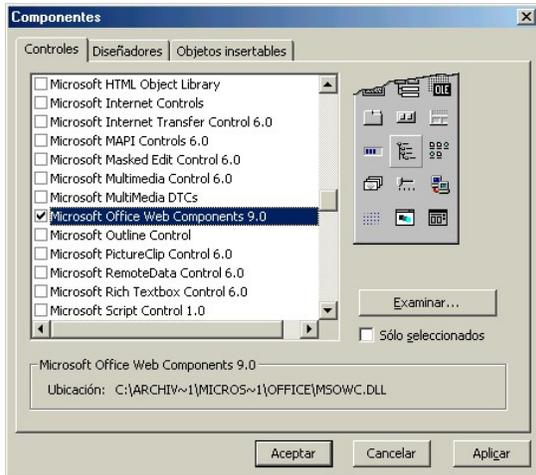
Formulario cuyas propiedades serán:

Objeto	Propiedad	Valor
Formulario	Nombre	DW
	WindowState	Maximizada
	Maxbutton	False
	Minbutton	False
	Caption	Consultas

10. Luego agregamos los componentes, haciendo clic derecho en la barra de herramienta, luego en Componentes...



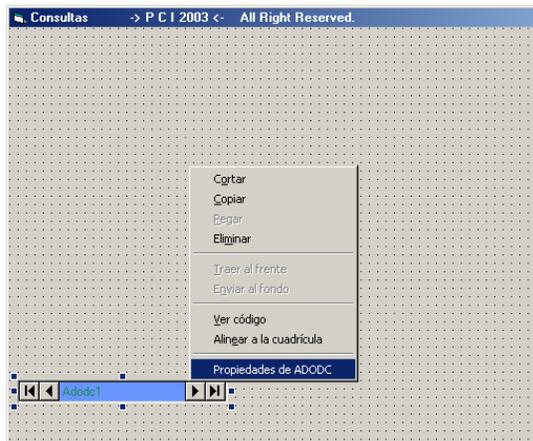
Aquí buscaremos la siguiente herramienta: Microsoft Office Web Components 9.0.



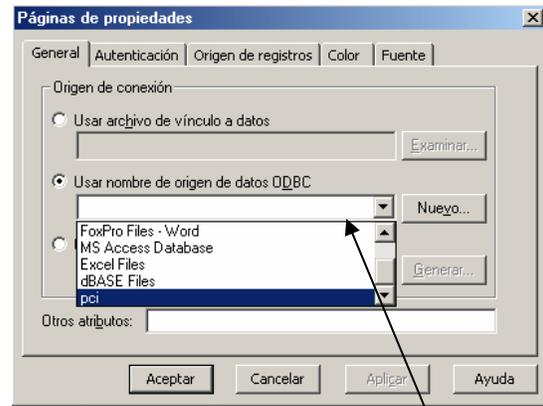
11. Damos clic en aceptar y se nos agregaran diferentes herramientas de las cuales ocuparemos primeramente el control ADODC.



Dibujamos el control, luego damos clic derecho sobre la imagen y escogemos la opción propiedades de ADODC



Seguidamente nos aparecerá la siguiente ventana donde seleccionaremos el botón de opción “Usar nombre de origen de Datos ADODC”



12. Luego buscamos en lista desplegable el nombre del control que realizamos anteriormente, en nuestro ejemplo usaremos el nombre del control PCI; pulsamos sobre la pestaña **origen de registro**, donde escogeremos el tipo de comando (adcmd table) y en tabla o nombre de procedimiento almacenado puede ser cualquiera de las tablas de nuestro sistema. Damos clic en aplicar, aceptar.



13. Luego Dibujamos el Pívor Table dando clic en el icono de la barra de Herramientas en nuestro formulario. Las propiedades que tendrá este son las siguientes:

Control	Propiedades
Conectionstring	DNS = pciiniciativaescolar
Datamember	Consulta1
Datasource	Adodc1

Pasos para realizar un cubo OLAP, utilizando Microsoft Query 2002

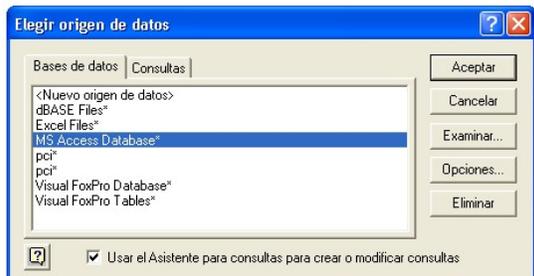
Este ejemplo consiste en una consulta para saber que Ordenes de Entrega fueron ya entregadas en el área del Sistema de Información para su procesamiento, de la cual tomaremos para crear el cubo OLAP que será implementado en nuestra aplicación.

1. Primero damos clic en inicio, todos los programas (en Microsoft Windows Xp), ejecutar y buscamos la siguiente ruta: "C:\Archivos de programa \ Microsoft Office\Office10\MSQRY32.EXE", ver imagen.



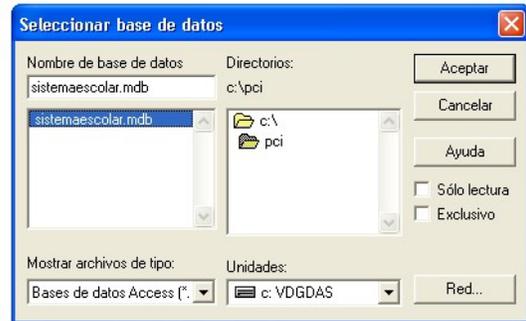
Seleccionamos el icono MSQRY32.EXE que es el nombre del archivo ejecutable de dicho programa y damos clic en abrir. Una vez abierta la aplicación procederemos a realizar el cubo OLAP, dando clic en archivo luego en nuevo.

2. Donde escogeremos el origen de los datos a procesar, siendo estos de MS Access Database.



Luego de escoger el origen de los datos damos clic en aceptar.

3. Aparecerá una nueva ventana donde tendremos que seleccionar la base de datos que queremos extraer los datos en nuestro caso la ruta era: c:\pci\sistemaescolar.mdb, y damos clic en aceptar.



5. Luego nos aparecerá una ventana donde escogeremos las tablas de las cuales queremos obtener los datos en este caso ocuparemos las siguientes tablas: mes, centros e inventario.

Resumen de los pasos del asistente.

Paso 1: Definir los campos de datos. En el primer paso del asistente se identifican los campos que se van a utilizar como campos de datos, se elige una función de resumen para calcular los valores de cada campo y se asigna un nombre a los campos.

Paso 2: Definir las dimensiones y los niveles. En el segundo paso se organiza el resto de los campos por dimensiones y niveles, y se asigna un nombre a las dimensiones y a los niveles.

Paso 3: Seleccionar el tipo de cubo. El asistente puede tener un tercer paso, en función del programa que estemos utilizando para obtener acceso a los datos y ejecutar el asistente. Aquí decidimos si se va a generar el cubo en memoria o se va a guardar como archivo de cubo.

Cuando termine de ejecutar el asistente, puede cambiar cualquiera de las opciones que haya elegido. Puede agregar o quitar campos, cambiar el modo en que se resumen los campos de datos y cambiar la organización de las dimensiones.

Opciones del paso 3 (¿Qué tipo de cubo desea crear?).

Opción 1: Recuperar los datos que se soliciten.

Acción: esta opción almacena las instrucciones para crear el cubo con la consulta o el informe, y crea el propio cubo sólo cuando se abre el informe. El cubo recupera inicialmente sólo los datos que se necesitan para mostrar el informe y va recuperando más datos a medida que se necesitan cuando el usuario cambia lo que muestra el informe.

Uso: Utilice esta opción en informes que usted o los usuarios del informe abran y vean pero que no cambien casi nunca, y cuando surjan problemas por falta de memoria o espacio en disco con las otras opciones.

Relación entre la velocidad y los recursos: Puesto que el cubo recupera sólo los datos que el informe muestra inicialmente, se reduce el periodo de tiempo que se tarda en abrir el informe.

Opción 2: Recuperar todos los datos a la vez.

Acción: Como en la primera opción, esta opción almacena las instrucciones para crear el cubo con la consulta o el informe, y crea el cubo sólo cuando se abre el informe. Sin embargo, en este caso el cubo recupera todos los datos del informe a la vez, cuando se abre el informe por primera vez, para que pueda modificar lo que el informe muestra sin tener que esperar a que el cubo recupera más datos.

Uso: Utilice esta opción en informes que vayan a ser objeto de un uso interactivo o que los usuarios vayan a cambiar repetidamente, si no desea asignar una cantidad importante de espacio en disco para los informes.

Relación entre la velocidad y los recursos: Puesto que esta opción hace que todos los datos estén disponibles inmediatamente, pueden cambiar la vista y mostrar datos diferentes rápidamente. Sin embargo, este tipo de cubo consume más memoria y espacio en disco temporal que un cubo que recupere los datos que se soliciten que no disponga de suficiente recursos si el cubo incluye muchos datos.

Opción 3: Guardar un archivo de cubo.

Acción: Esta opción crea un archivo de cubo independiente en disco; recupera todos los datos para el cubo y los almacena en este archivo.

Uso: Utilice esta opción para informes interactivos que vaya a modificar con frecuencia, cuando no hay problemas relacionados con la cantidad de espacio en disco que utiliza el informe, o bien cuando desee almacenar el cubo en un servidor de red al que otros usuarios puedan obtener acceso para cerrar sus propios informes. Un archivo de cubo

puede proporcionar algunos de los datos de la base de datos de origen inicial a la vez que omite datos confidenciales que no desee que otros usuarios vean. Un archivo de cubo también permite trabajar con los informes sin conexión utilizando una instantánea de la base de datos de origen inicial.

Relación entre la velocidad y los recursos:

La creación inicial de un archivo de cubo puede tardar más tiempo que las otras opciones, pero la apertura y la modificación de informes se puede acelera. El tamaño del archivo de cubo afecta a la velocidad con la que se pueden abrir y modificar los informes. El tamaño del archivo está determinado por la cantidad de datos incluidos y el modo de organizar el cubo por campos de datos y dimensiones. Si sólo dispone de espacio en disco limitado, es posible que tenga que experimentar con diferís organizaciones y selecciones de datos para reducir el tamaño del archivo de cubo, o quizás sea necesario utilizar una de las otras opciones para crear un archivo de cubo.

Archivo de cubo y archivos de definiciones de cubo.

Guardar un archivo de cubo.

Cuando se elige la tercera opción, se crea un archivo de cubo con la extensión .cub en la ubicación que se especifique. Este archivo contiene todos los datos del cubo. No se crea inmediatamente al hacer clic en el comando **finalizar**, sino cuando se guarda la definición del cubo en un archivo o cuando se crea el informe que está basado en el cubo.

Guardar la definición en un archivo.

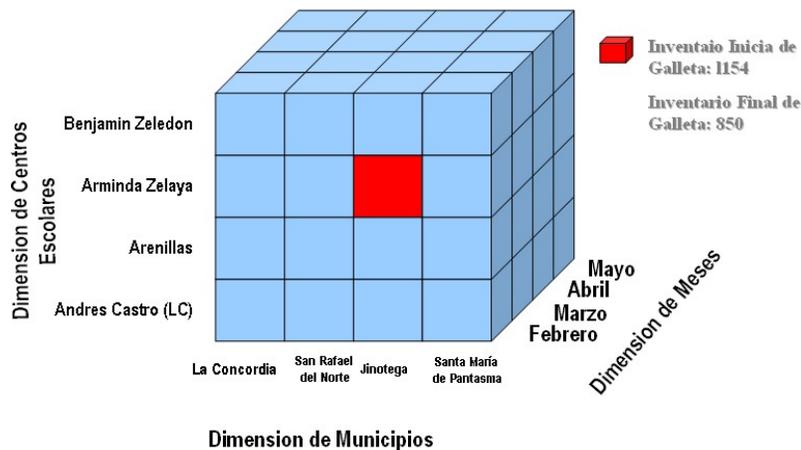
Esta sección se aplica sólo si se ejecuta el asistente para cubos OLAP desde Microsoft Query, en las tres opciones, Query pide que se guarde un archivo de definición de cubo (.oqy), que se puede abrir en Excel para crear informes que estén basados en el cubo. Está separado de cualquier archivo .cub que cree para almacenar los datos del cubo real. Si desea realizar cambios en el cubo, puede abrir el archivo .oqy en Query para volver a ejecutar el Asistente para cubos OLAP.

Resultados.

En la investigación se obtuvo los resultados esperados ya que pudimos definir los conceptos de DataWarehouse y DataMining, así como todas las metodologías que se recomiendan para su implementación y las demás herramientas que necesitamos para llegar a la aplicación, dando así una visión completa de todos los pasos y procedimientos que se requieren en un proyecto de este tipo.

Obtuvimos mayor cantidad de consultas sobre los datos que pueda manejar un sistema automatizado, implementando estas herramientas ya que la ventaja que tienen estas consultas dinámicas sobre las consultas estáticas (prediseñadas) es que el diseñador no tiene que realizar la codificación de una consulta y el diseño de un formulario o reporte para presentar los resultados de dicha consulta en el formato que desea el cliente.

Con la utilización de los cubos OLAP podemos observar con mayor rapidez las comparaciones de consumo de alimento entre los diferentes meses del año 2003, en los Centros Educativos como podemos observar en la gráfica.



Como podemos observar el recuadro rojo representa el mes de febrero enlazado con la dimensión de Municipio y con la dimensión de Centro Escolares para obtener el Inventario Inicial y Final de Consumo de Alimento del Centro Educativo: Arminda Zelaya de dicho mes.

Conclusiones.

Las herramientas de hoy en día hacen fácil la implementación, pero como siempre, si no existe un buen análisis de requerimientos no alcanzaremos el éxito esperado.

Uno de los propósitos logrado es demostrar una forma de soportar la toma de decisiones basada en el análisis oportuno de la información, no las operaciones de un negocio.

Los conceptos de DataWarehouse y DataMining son aplicables a sistemas de información que manejan gran volumen de información (varios gigabytes y hasta terabytes) brindando así mas documentación para el manejo de estas nuevas tecnologías de tercera generación.

Un DataWarehouse con éxito es hacer que la información que un negocio ha reunido, tanto de los sistemas operacionales como de algunos sistemas externos, sea accesible, de forma que se pueda realizar la toma de decisiones.

Podemos concluir que los conceptos de datawarehouse y data Mining son de gran utilidad a la hora de implementar la toma de decisiones basadas en el análisis de la información de las empresas corporativas.

Mediante los cubos, el servidor de bases de datos realizará un procesamiento previo de los resúmenes de los datos de forma que la creación y la actualización de informes sean mucho más rápidas.

Podemos observar que una de las grandes herramientas para la creación de cubos OLAP es Microsoft Query, ya que esta nos brinda un sinnúmero de tipos de consultas las cuales podemos acoplarlas en los cubos OLAP para posteriormente generar los reportes.

Con esta documentación di a conocer los requerimientos mínimos de Hardware y software para implementar un Data Warehouse y Data Mining.

Recomendaciones.

Se recomienda una vez instalada la aplicación en la máquina del usuario final configurar el ODBC, para el buen funcionamiento de los cubos y DataWarehouse.

Para utilizar estas técnicas de forma eficiente se recomienda realizar un muestreo al subconjunto de datos del Data Warehouse que cumpla unas condiciones determinadas.

Fuentes de Información Consultadas Sitios de Internet

- ☞ <http://www.map.es/csi/silice/DW31.html>
- ☞ <http://www.map.es/csi/silice/DW221.html>
- ☞ <http://www.map.es/csi/silice/DW223.html>
- ☞ <http://www.map.es/csi/silice/DW2251.html>
- ☞ <http://www.datawarehousing.com>.
- ☞ <http://www.map.es/csi/silice/DW2253.html#Introduccion#Introduccion>
- ☞ <http://www.map.es/csi/silice/DW2253.html#Tecnicas#Tecnicas>
- ☞ <http://www.map.es/csi/silice/DW2253.html#Metodologia#Metodologia>