



CENTRO DE CIENCIAS BÁSICAS  
DEPARTAMENTO DE SISTEMAS DE INFORMACIÓN

“Diseño de un modelo multidimensional de data mart del área de  
capacitación en el INEGI”

TRABAJO PRÁCTICO

QUE PARA OBTENER EL GRADO DE MAESTRÍA EN  
INFORMÁTICA Y TECNOLOGÍAS COMPUTACIONALES

PRESENTA

L.I. Jorge Armando Domínguez Martínez

ASESOR:

Dra. Laura A. Garza González

CO-ASESORES:

M.C. Jorge Eduardo Macías Luévano

M.C. César Eduardo Velázquez Amador

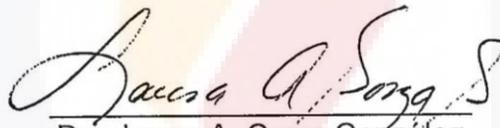
Aguascalientes, Ags., Mayo de 2008.

Por este conducto, autorizamos al tesista:

L.I. Jorge Armando Domínguez Martínez

La impresión de su documento final de Tesis, ya que cumple con los requisitos de contenido y forma exigidos en la Universidad Autónoma de Aguascalientes.

Asesor



---

Dra. Laura A. Garza González  
Asesor principal



---

M.C. Jorge Eduardo Macías Luévano  
Revisor



---

M.C. César Eduardo Velázquez Amador  
Revisor

**Contenido**

Lista de tablas ..... 1

Lista de figuras ..... 1

Resumen ..... 2

Capítulo 1. Formulación del problema ..... 3

    1.1 Introducción ..... 3

    1.2 Antecedentes ..... 4

    1.3 Planteamiento del problema ..... 6

    1.4 Objetivo General ..... 9

        1.4.1 Objetivos específicos ..... 9

    1.5 Limitaciones ..... 10

Capítulo 2. Marco teórico ..... 11

    2.1 Sistemas inteligentes y de soporte a la decisión (DSS) ..... 11

        2.1.1 Definición ..... 11

        2.1.2 Características y capacidades de un DSS ..... 13

        2.1.3 Componentes de un DSS ..... 14

        2.1.4 Proceso de un DSS ..... 15

    2.2 Business Intelligence (BI) ..... 16

        2.2.1 Definición ..... 16

    2.3 Data Warehouse (DW) ..... 19

        2.3.1 Definición ..... 19

        2.3.2 Data Warehousing ..... 21

        2.3.3 Características ..... 22

        2.3.4 Componentes del Data Warehouse ..... 23

        2.3.5 Arquitectura de referencia ..... 24

        2.3.6 A quién va dirigido el data warehouse ..... 26

        2.3.7 Beneficios de contar con un Data Warehouse ..... 27

    2.4 Definición data mart (DM) ..... 27

    2.5 Data warehouse vs. data mart ..... 31

    2.6 Diferencias entre la base de datos operacional y el data warehouse ..... 32

- 2.7 Modelado Multidimensional (MMD) ..... 33
  - 2.7.1 Componentes del Modelo Multidimensional..... 34
  - 2.7.2 Procesamiento Analítico en Línea (OLAP)..... 35
  - 2.7.3 Pasos básicos del Modelado Multidimensional ..... 39
- Capítulo 3. Metodología del desarrollo..... 40
  - 3.1 Introducción ..... 40
    - 3.1.1 Metodología Kimball – Ciclo de Vida..... 40
  - 3.2 Definición de requerimientos del negocio ..... 41
  - 3.3 Modelo conceptual ..... 41
  - 3.4 Construcción del modelo dimensional ..... 42
    - 3.4.1 Matriz de Data Marts ..... 42
    - 3.4.2 Descripción de los data marts ..... 43
    - 3.4.3 Descripción de las dimensiones ..... 43
    - 3.4.4 Modelo de los data mart's ..... 44
    - 3.4.5 Diseño lógico de la base de datos ..... 45
  - 3.5 Herramienta de desarrollo ..... 46
  - 3.6 Poblando el Data Mart ..... 48
  - 3.7 Implementación ..... 49
  - 3.8 Consulta de cubos por Intranet..... 49
- Capítulo 4. Conclusiones. .... 51
- Capítulo 5. Recomendaciones. .... 53
- Glosario..... 55
- Bibliografía ..... 60
- Anexos ..... 64
  - Anexo A. Misión, visión, valores y proceso..... 64
  - Anexo B. Proceso de registro en el sistema transaccional ..... 66
  - Anexo C. Entrevistas ..... 68
  - Anexo D. Reportes varios de capacitación ..... 72
  - Anexo E. Portafolio de análisis requeridos ..... 75
  - Anexo F. Código de la aplicación del ETL ..... 76
  - Anexo G. Código MDX ..... 79

**Lista de tablas**

Tabla 1. Capacidades de un DSS ..... 13  
Tabla 2. OLTP vs Data warehouse. .... 32  
Tabla 3. Comparación de bases de datos MOLAP y ROLAP. .... 38  
Tabla 4. Matriz de data marts..... 42  
Tabla 5. Descripción de los data marts ..... 43  
Tabla 6. Descripción de las dimensiones..... 43

**Lista de figuras**

Figura 1. El DSS y su ambiente computacional. .... 15  
Figura 2. Componentes del Datawarehouse. .... 24  
Figura 3. Arquitectura de un data warehouse, adaptado de Inmon, 1996..... 25  
Figura 4. Arquitectura de un esquema de data marts independientes. .... 25  
Figura 5. Arquitectura de un esquema de data marts dependientes..... 26  
Figura 6. Data Mart. Fuente: Inmon (1999). .... 29  
Figura 7. Ciclo de vida según Kimball. .... 40  
Figura 8. Modelo conceptual del data mart de capacitación. .... 42  
Figura 9. Modelo del data mart de cursos..... 44  
Figura 10. Modelo del data mart de participantes. .... 44  
Figura 11. Diseño lógico del data mart cursos. .... 45  
Figura 12. Diseño lógico del data mart participantes. .... 45  
Figura 13. Arquitectura del sistema Microsoft DW/BI..... 46  
Figura 14. SQL Server Management Studio ..... 47  
Figura 15. BI Studio. .... 48  
Figura 16. Procesar los cubos..... 49  
Figura 17. Consulta de cubos por intranet. .... 50

## **Resumen**

Este trabajo de grado tiene por objetivo el desarrollar un modelo que permita implementar un data mart en el área de Capacitación para el INEGI, con el propósito de apoyar la toma de decisiones estratégicas y tácticas de dicha área. Este proyecto surge de la necesidad imperante de proporcionar a los administradores de la capacitación del INEGI información que les permita realizar análisis profundo de ella a fin de evaluar la efectividad y el desempeño del programa de capacitación en el Instituto y tomar los correctivos que permitan alcanzar los objetivos que los directivos se plantean.

Cabe mencionar que en el presente trabajo se desarrolló el data mart de Capacitación usando la metodología Kimball. La metodología de Kimball se enfoca principalmente en el diseño de la base de datos que almacenará la información para la toma de decisiones. El diseño se basa en la creación de tablas de hechos (FACTS) que son tablas que contienen la información numérica de los indicadores a analizar, es decir la parte cuantitativa de la información.

# TESIS TESIS TESIS TESIS TESIS

## **Capítulo 1. Formulación del problema**

### ***1.1 Introducción***

Actualmente las organizaciones se están enfrentando a nuevas reglas de negocios producto de la globalización, así mismo los clientes son cada vez más exigentes debido a que sus necesidades son más cambiantes y complejas. Ante esto, se han desarrollado herramientas que han aprovechado las ventajas y beneficios que ofrece el internet y las tecnologías de información, dichas herramientas han ayudado a satisfacer las necesidades del cliente y a las empresas les ha permitido responder al cambiante ambiente de negocios.

En este nuevo panorama, la información es el recurso que ha ayudado a las organizaciones a hacer frente al ambiente de negocios, siendo el conocimiento un elemento muy cuidado, debido a que marca la diferencia entre ellos y su competencia.

En la actualidad toda empresa deposita mucha confianza en el proceso de toma de decisiones, el problema que ahora enfrentan es la posesión de mucha información, pero una escasez de conocimiento que les sea útil para tomar decisiones acertadas. Por ese motivo cada vez son más elementales las tecnologías que ayudan a minimizar el tiempo de análisis de información, hacerlo a mayor velocidad y precisión, de manera que les permita reaccionar a las necesidades del mercado antes que la competencia lo haga.

Este panorama y la necesidad de obtener utilidades, han obligado a las empresas a requerir más y mejor información, así como a diseñar sistemas que les permitan conseguirla y administrarla de la mejor manera, al respecto Buksard, Mollot y Richards (2000) opinan que esta necesidad de nuevas herramientas de acceso y reporte de información, para diversos tipos de usuarios, ha impulsado la creación

de nuevas tecnologías, colectivamente conocidas como Business Intelligence o Inteligencia de Negocios.

Entre los elementos que han contribuido al análisis de la información está el data warehouse, que forma parte del enfoque llamado Inteligencia de Negocios y a través del cual se pretende hacer de una empresa un negocio inteligente capaz de obtener el máximo provecho al cúmulo de información.

Un data warehouse es un proceso, no un producto, para la elaboración y administración de datos provenientes de diferentes fuentes con el propósito de obtener una vista particular, detallada de todo el ambiente de la organización. (Gardner, 1998).

En el presente trabajo, se propone el diseño de un data mart para el área de Capacitación del Instituto Nacional de Estadística, Geografía e Informática (INEGI), el cual responde a la necesidad de contar con información que permita analizar el desarrollo, la situación actual, la evolución de la capacitación y el desempeño de la gestión de la misma, para vincularla con tópicos relacionados de las áreas técnicas del Instituto, con el propósito de apoyar la toma de decisiones estratégicas y tácticas del Instituto.

## ***1.2 Antecedentes***

El Instituto Nacional de Estadística, Geografía e Informática emprende las acciones de capacitación como actividad estratégica y permanente del proceso de modernización. En el año 1989 se crea el Programa Integral de Capacitación, Formación e Investigación (PICFI), el cual modifica su nombre por el de Programa de Capacitación Institucional (PCI) en el año 2002. Desde ese año a la fecha, el crecimiento y evolución del programa ha sido considerable, alcanzando el propósito fundamental de desarrollar en el personal los conocimientos, habilidades

TESIS TESIS TESIS TESIS TESIS

y aptitudes que le faciliten el desempeño de sus labores y lo formen para realizar las actividades que los cambios institucionales y de avance tecnológico le requieran.

No obstante, el PCI, de 1995 al 2000, había tenido no sólo una trascendencia y diversificación importante en las áreas temáticas del conocimiento que lo integran, ya que a la postre también se crearon nuevos programas sustanciales como el de Educación para la Calidad y el de Apoyo a la Continuidad Académica, situación que ha permitido hasta ahora que el personal se perfile hacia mayores y mejores niveles de desempeño, y por ende, su ingreso a los beneficios de la Calidad Total.

En consecuencia, durante los últimos años la capacitación ha adquirido también una mayor fuerza como acción que procede al cambio institucional, articulándose en mayor grado a la dinámica y necesidades específicas de las áreas de trabajo.

Estas acciones del PCI tuvieron su marco de referencia en las siguientes líneas de orientación:

- Continuidad en la diversificación temática del campo del conocimiento institucional.
- Diversificación de servicios, aprovechando los recursos existentes y la tecnología educativa.
- Fortalecimiento de la formación, actualización y especialización del personal para el presente y futuro desarrollo tecnológico institucional.
- Capacitación orientada a la atención de las necesidades prioritarias de las áreas laborales.
- Contribución a la creación de una cultura de calidad organizacional.
- Fortalecimiento de los vínculos de colaboración e intercambio con organismos de América Latina y el Caribe a fines al INEGI.
- Impulso y apoyo a la continuidad de la formación académica y cultural para el desarrollo integral y la calidad de vida.

TESIS TESIS TESIS TESIS TESIS

A su vez, en el año 1995 se creó el Consejo Directivo de Calidad y en 1996 se creó la Secretaría Ejecutiva del Consejo Directivo de Calidad, integrándose a este Consejo en el 2002 las Direcciones de Calidad y Capacitación.

Posteriormente, en 2005, la Dirección de Capacitación se integró a la Dirección General de Innovación y Tecnologías de la Información, en tanto que en el año 2006 se logró la creación del Consejo Directivo de Capacitación, el cual se creó con el propósito de autorizar y priorizar las necesidades de capacitación y dictar las políticas y las líneas estratégicas para la consecución de las metas institucionales.

### **1.3 Planteamiento del problema**

Actualmente las organizaciones no pueden quejarse de tener poca información, *hoy el verdadero problema es que hay demasiada información*. Navarrete (2001) opina que *la gran mayoría de las organizaciones tienen una abundancia de datos, pero una escasez de conocimiento*. Las métricas del funcionamiento y los recursos de información más importantes siguen estando perdidos en un mar de números y de sistemas desconectados. La gran mayoría de las organizaciones tienen muchos sistemas dispersos, cada uno con sus propias fuentes de datos y mecanismos de representación. Esto hace que el mantenimiento de información actualizada a través de los departamentos y unidades de negocios sea extremadamente difícil. En contraparte, mientras más integrada sea una organización, es más fácil para cualquier integrante de ésta obtener la información que necesite y estar habilitado para tomar una mejor decisión.

En estos últimos años la información ha pasado de ser el proceso de automatización de tareas rutinarias a ser un proceso de verdadera “información” de las organizaciones. Ahora la información se convierte en un aliado estratégico de gran importancia para la supervivencia de las empresas.

TESIS TESIS TESIS TESIS TESIS

Para mantenerse competitiva una organización necesita una buena gestión de datos que minimice las duplicidades en su tratamiento y que aseguren la calidad de los mismos, de manera que puedan servir como fuente para la toma de decisiones estratégicas y tácticas.

Esto conlleva a afirmar que los aspectos relativos a la explotación, el control y la auditoría de los almacenes de datos cobran gran importancia tanto para los especialistas informáticos como para los directivos de las empresas. Todo esto, con base en un plan de seguridad específico, con el fin de garantizar la continuidad de la operación y la disponibilidad del sistema.

Atendiendo a lo planteado anteriormente, las organizaciones necesitan sistematizar la información para la toma de decisiones, principalmente aquellas en las que se identifican grandes volúmenes de datos. En este tipo de organizaciones entran las que se dedican a la de capacitación, tal es el caso de la Dirección de Capacitación del INEGI, la cual tiene como misión de administrar el programa de capacitación institucional y promover la profesionalización del capital humano del Instituto; así como establecer estrategias de intercambio en materia de capacitación con organismos nacionales e internacionales, con la finalidad de fortalecer y desarrollar las competencias laborales del personal.

La Dirección de Capacitación maneja muchos datos. Esta sobreabundancia tiene como consecuencia directa la necesidad de aprovechar esos valiosos activos. El principal reto consiste por una parte en seleccionar la información justa y útil y, por otra, referenciar esta información y almacenarla correctamente para ser capaz de recuperarla al momento que se necesite.

De igual forma todos los responsables de la capacitación a nivel nacional requieren un acceso más rápido a los datos, más reportes y a mayor velocidad, análisis y reacciones oportunas para administración dicha capacitación en su ámbito, controlar la calidad del servicio, incluso con costos adicionales en

tecnología de la información, para crear y generar reportes oportunos y de esta manera que la información no pierda novedad.

Otro factor que incide en la necesidad de implementar una herramienta de soporte de decisión (DSS) es la excesiva complejidad de los modelos relacionales que soportan los sistemas operacionales (la base de datos se encuentra en SQL server 2000 e interactúa con Oracle para la actualización del personal vigente), lo cual dificulta en extremo la extracción de información para llevarlo a cabo labores de gestión de la capacitación.

La razón por la cual el director de capacitación considera necesario desarrollar el data mart para complementar el actual sistema de administración de la capacitación (SAC), es que éste último, no contempla opciones para generar reportes para la alta gerencia o al Consejo Directivo de Capacitación, que le permita realizar análisis profundos y seguros. Los requerimientos textuales de la Dirección de Capacitación son los siguientes:

- Poder consultar los avances del programa de capacitación en cualquier periodo y oficina donde se imparte capacitación. La consulta de los avances es a nivel de cursos o eventos de capacitación programados contra los realizados, así como el número de participantes programados contra los que se realmente se han capacitado.
- La forma en que se presentan los datos que no sea rígido, permitiendo la posibilidad de exportar éstos datos a otras herramientas que permitan mayor flexibilidad para hacer análisis complejos de la situación.
- Las variables que se tomen en cuenta sean suficientemente representativa de la categorización de la información.
- Comparativos entre diferentes años de lo realizado.

A través de los años, esta Dirección de Capacitación ha acumulado grandes cantidades de datos relacionadas con el personal del instituto y su capacitación,

los cuales son de tipo operacional. Estos depósitos contienen una mina de oro en cuanto a información histórica relacionada, de la cual pudiera derivarse información con elementos tales como, programas de capacitación, los participantes, los instructores, los cursos o eventos de capacitación, los cotos, los presupuestos y la experiencia. Es por esto que los encargados de la capacitación en las diferentes oficinas del INEGI necesitan de un sistema basado en el acceso interactivo e inmediato a la información estratégica del área de capacitación. Este acercamiento de información al usuarios final permite la toma de decisiones rápida y basada en datos objetivos a partir de las bases de datos de la organización.

Satisfacer no solo a los administradores de la capacitación sino a la alta dirección del INEGI será el enfoque, para lo cual será fundamental implementar una nueva tecnología informática de decisión para obtener una mayor comprensión del valor de la información disponible, definir indicadores de la capacitación pertinentes para facilitar la toma de decisiones operativas que apoyen los procesos sustantivos y el cumplimiento de la misión y visión del Instituto.

#### **1.4 Objetivo General**

Desarrollar un modelo que permita implementar un data mart en el área de Capacitación para el INEGI.

##### **1.4.1 Objetivos específicos**

- Conceptualizar un modelo funcional/operacional de data mart para el área de capacitación del INEGI.
- Estudiar, evaluar y elegir las herramientas (software) para los usuarios finales de data mart.
- Diseñar y desarrollar un data mart para el área de capacitación.

### **1.5 Limitaciones**

Este caso práctico estará enfocado al aspecto técnico-tecnológico y no analizará los aspectos económicos y social que intervienen en el proceso de diseño e implementación de un data mart. Es decir, se propondrá un modelo de data mart para el área de capacitación, se implementará bajo la infraestructura actual del INEGI (equipamiento y personal experto) y se utilizará las herramientas de análisis disponibles en el Instituto, por lo tanto no se requerirá hacer una inversión adicional.



## **Capítulo 2. Marco teórico**

### ***2.1 Sistemas inteligentes y de soporte a la decisión (DSS)***

Actualmente los sistemas de soporte a las decisiones han causado gran impacto en las organizaciones, las cuales buscan un sistema que les ayude y oriente para tomar mejores decisiones. La finalidad principal de los sistemas de información es dar soporte a los procesos sustantivos de la organización (ventas, producción, personal...).

#### **2.1.1 Definición**

El concepto de sistema de soporte a las decisiones (DSS) es muy amplio debido a que hay muchos acercamientos a la toma de decisiones, así como la amplia gama de los dominios en los cuales se toman las decisiones. Un DSS puede tomar diferentes formas. En general, un DSS es un sistema computarizado que ayuda en la toma de decisiones. Una decisión es una opción entre varias alternativas basados en las estimaciones de los valores de dichas alternativas. El soporte a las decisiones significa ayudar a la gente a trabajar solo o en grupo, generando alternativas y haciendo selección de las mismas. El soporte del proceso de selección implica el apoyo de la estimación, evaluación y/o comparación de las alternativas. En la práctica, cuando se habla de un DSS, generalmente se refiere a aplicaciones computacionales que realizan el papel de soporte (Alter, S. L. 1980).

El término de sistema de soporte a las decisiones ha sido utilizado de diversas maneras (Alter 1980, Power, 2002) y ha sido definido de varias formas dependiendo del punto de vista del autor (Druzdzal, M. J. and R. R. Flynn 1999). Finlay (1994) y otros autores definen un DSS como “un sistema computarizado que ayuda el proceso de toma de decisiones.” Turban (1995) lo define más específicamente como “un sistema de información computarizado interactivo,

flexible y adaptable, especialmente desarrollado para apoyar la solución de problemas no estructurados de la gerencia, mejorando la toma de decisiones. El sistema utiliza datos, proporciona una interface fácil de usar y permite a los responsables de tomar decisiones su propia interpretación."

Otras definiciones caen dentro de esos dos extremos. Para Little (1970) un DSS es un modelo basado en un conjunto de procedimientos para procesar datos y juicios que ayudan a los administradores a tomar decisiones. Para Keen (1978), un DSS acopla los recursos intelectuales de los individuos con las capacidades de las computadoras para mejorar la calidad de las decisiones. Es un sistema basado en computadora para los tomadores de decisión que tratan con problemas semiestructurados. Moore and Chang (1980) define un DSS como un sistema extendible, capaz de ayudar en el análisis de datos ad hoc y modelo de decisiones, orientado hacia la futura planeación, y a usarlo en intervalos irregulares y no planeados. Para Sprague and Carlson (1982), un DSS es sistema interactivo computarizado, que ayuda en la toma de decisión utilizando datos y modelos para encontrar la solución de un problema no estructurado. En contraste, Keen (1980) demanda que es imposible dar una definición exacta incluyendo todas las facetas del DSS ("no puede haber una definición de los sistemas para la toma de decisiones, solamente al soporte a las decisiones"). Sin embargo, según Power (1997), un DSS es un término útil y de carácter general que hace referencia a diversos tipos de sistemas de información que sirven de apoyo en la toma de decisiones. Son herramientas específicamente diseñadas para permitirle al usuario interpretar análisis de datos generados por sistemas interactivos en computadora. Por otra parte Bonczek et al (1980) define un DSS como un sistema basado en computadoras con tres componentes clave: un sistema de lenguaje (un mecanismo para proveer comunicación entre el usuario y otros componentes del DSS), un sistema de conocimiento (repositorio de conocimiento sobre problemas, datos o procesos) y el sistema procesador de problemas (la unión entre los otros dos componentes).

## 2.1.2 Características y capacidades de un DSS

Turban, E. (2006), debido a que no hay consenso en exactamente que constituye un DSS, por lo tanto no hay un acuerdo en las características y capacidades de estos. Sin embargo, la mayoría de los DSS tienen al menos algunos de los atributos mostrados en la siguiente tabla.

Tabla 1. Capacidades de un DSS

Un DSS provee ayuda para los que toman decisiones en todos los niveles de la administración, ya sea individuales o en grupos, principalmente en situaciones semiestructurados y no estructurado, trayendo el juicio humano e información objetiva.
Un DSS apoya varias decisiones secuenciales y/o interdependientes.
Un DSS ayuda a todas las fases del proceso de la toma de decisiones- inteligencia, diseño, elección e implementación – así como también una variedad de procesos para la toma de decisiones y estilos.
Un DSS es adaptable por el usuario con el paso del tiempo de acuerdo con las condiciones cambiantes.
Un DSS es fácil de construir y de usar en muchos casos.
Un DSS promueve el aprendizaje, lo cual conduce a nuevas demandas y refinamientos de la aplicación actual, lo que provoca a un aprendizaje adicional y así sucesivamente.
Un DSS usualmente utiliza modelos cuantitativos (estándar y/o hechos a la medida)
Los DSSs avanzados están equipados con un componente de administración del conocimiento que permite una solución eficiente y efectiva para muchos problemas complejos.
Un DSS puede ser usado vía Web.
Un DSS permite la fácil ejecución de análisis sensitivo.

Fuente: Turban, Efraim; Aronson, Leidner, Dorothy; Mclean, Ephraim; Wetherbe, James. "Information Technology for Management". John Wiley & Sons, Inc. 5th Edition. Cap. 11. Pag. 467.

### 2.1.3 Componentes de un DSS

Turban, E. (2006), cada DSS consiste al menos de un manejador de datos, la interfase del usuario, un manejador del modelo de componentes y el usuario final. En poco también contendrá un componente de administración del conocimiento.

1. Un subsistema de administración de datos de DSS contiene todos los datos que fluyen desde varias fuentes y que usualmente son previamente extraídos para su entrada en una base de datos del DSS o un almacén de los datos (data warehouse). E algunos DSSs, no hay una base de datos separada y los datos son metidos dentro del modelo DSS como se vaya necesitando.
2. Un subsistema de administración del modelo contiene modelos completos (financiero, estadístico, ciencia de la administración y otros modelos cuantitativos), y las rutinas para desarrollar aplicaciones DSSs. Un ejemplo es Excel, con todas sus funciones matemáticas y estadísticas.
3. La interface del usuario cubre todos los aspectos de la comunicación entre el usuario y el DSS. Algunos expertos en DSS consideran que la interface del usuario es el componente más importante de un DSS, debido a que mucho del poder, flexibilidad y facilidad de uso son derivados de este componente. La mayoría de las interfaces son vía Web y algunas son complementadas con voz.
4. El usuario. La persona que enfrenta los problemas o las decisiones con un DSS puede ser un usuario, un administrador o él que toma las decisiones. El usuario es considerado como parte del sistema.
  - a. DSS intermediarios. Los administradores pueden utilizar un DSS vía un intermediario quién realiza el análisis y reporta los resultados.

5. Un subsistema inteligente o basado en el conocimiento provee el expertis para resolver algunos aspectos del problema o el conocimiento que puede incrementar la operación de los otros componentes del DSS.

### 2.1.4 Proceso de un DSS

Según Hackney (2000) los componentes de un DSS es todo el software que corre en una computadora estándar y puede ser facilidad con multimedia. Herramientas como Excel incluye algunos de los componentes y puede ser utilizado por los usuarios finales.

Cuando el usuario tiene un problema lo evalúan usando este proceso.

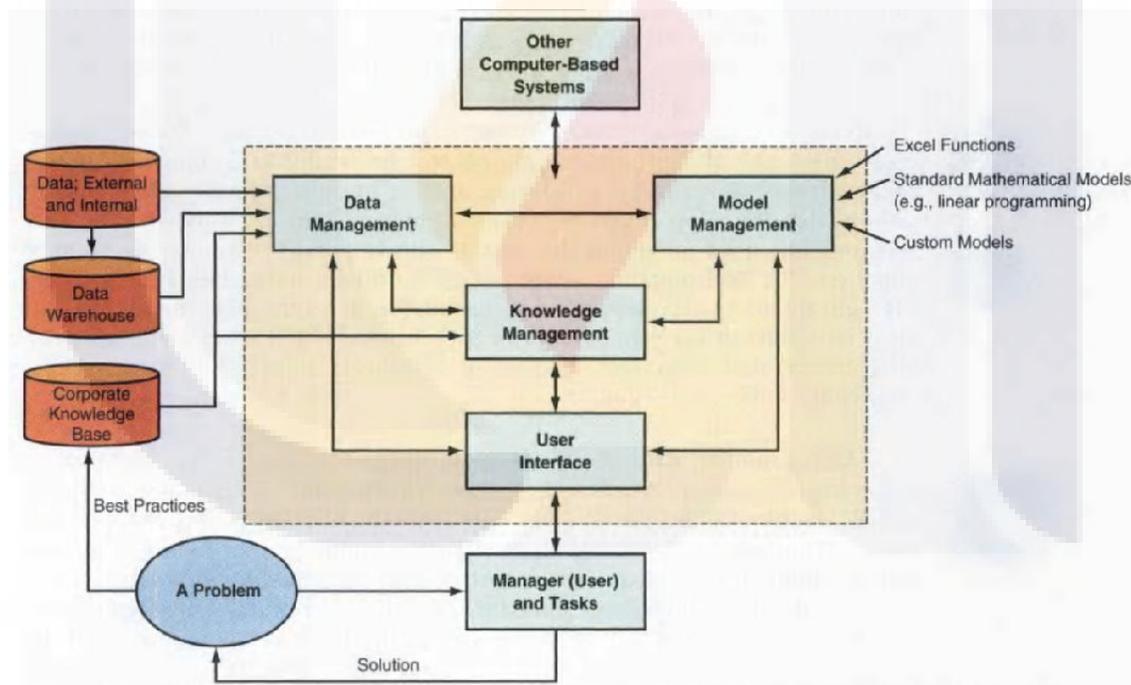


Figura 1. El DSS y su ambiente computacional.

Modelo conceptual de un DSS mostrando cuatro componentes principales de software y sus relaciones con otros sistemas.

## **2.2 Business Intelligence (BI)**

La gran mayoría de las organizaciones tienen una abundancia de datos, pero una penuria de conocimiento. Las métricas del funcionamiento y los recursos de información más importantes siguen estando perdidos en un mar de números y de sistemas desconectados. La gran mayoría de las organizaciones tienen muchos sistemas dispersos, cada uno de los cuales tiene sus propias fuentes de datos y mecanismos de representación. Esto hace que el mantenimiento de información actualizada a través de los departamentos y unidades de negocios sea extremadamente difícil. En contraparte, mientras más integrada sea una organización, es más fácil para cualquier integrante de esta obtener la información que se necesite, así cualquier individuo se encuentra más habilitado para tomar una mejor decisión.

La inteligencia de negocio (Business Intelligence, BI) lucha por eliminar las suposiciones y la ignorancia en las empresas al potenciar las montañas de datos cuantitativos que las empresas reúnen cada día en una variedad de aplicaciones corporativas, Dresnes (citado en Romero, 2002).

### **2.2.1 Definición**

El cambiante entorno económico y la problemática de sistemas descrita anteriormente impulsaron el surgimiento del llamado Business Intelligence, el cual es un concepto que trata de englobar todos los Sistemas de Información de una organización para obtener de ellos no solo información o conocimiento, si no una verdadera inteligencia que le confiera a la organización una ventaja competitiva por sobre sus competidores.

Hackney (2001) nos dice que el Business Intelligence se compone de todas las actividades relacionadas a la organización y entrega de información así como el análisis del negocio. Esto incluye minería de datos, administración del conocimiento, aplicaciones analíticas, sistemas de reportes y principalmente data warehousing.

TESIS TESIS TESIS TESIS TESIS

Michel (2000) afirma que el conjunto de tecnologías que usan Data Warehousing y OLAP (On-line Analytic Processing, procesamiento analítico en línea), combinado con herramientas de reporte, son referidas como Business Intelligence, porque ayudan a las compañías a ganar inteligencia en operaciones y desempeño.

Buksard, Mollot y Richards (2000) comentan en su artículo que la necesidad de nuevas herramientas de acceso y reporte de información, para diversos tipos de usuarios, ha impulsado la creación de nuevas herramientas, colectivamente conocidas como Business Intelligence. Business Intelligence no es una sola tecnología o aplicación. No es una "cosa", sino que se trata de un "suite" de productos que trabajan de manera conjunta para proveer datos, información y reportes analíticos que satisfagan las necesidades de una gran variedad de usuarios finales.

Business Intelligence es la habilidad de consolidar información y analizarla con la suficiente velocidad y precisión para descubrir ventajas y tomar mejores decisiones de negocios. Definición compatible con la necesidad actual de los negocios que ante la presión de ser cada día más competitivos, para mantenerse tienen la doble tarea no sólo de permanecer sino de ser lucrativos (Cano, 1999).

Por último, el vicepresidente y director de investigaciones del Grupo Gartner, Howard Dresner, citado en Hilson (2001) coincide con las anteriores definiciones, y él agrega: "Business Intelligence es simplemente la habilidad de los usuarios finales para acceder y analizar tipos cuantitativos de información y ser capaz de actuar en consecuencia".

Business Intelligence ha tomado la delantera en los últimos dos años, los proveedores de soluciones cuentan con tecnología más amigable y presentan datos más fáciles de analizar, apostando hacia la adopción masiva de dichos sistemas por parte de las organizaciones (Sullivan, 2001).

TESIS TESIS TESIS TESIS TESIS

Las compañías actualmente usan una amplia gama de tecnologías y productos para saber qué es lo que está pasando en la organización. Las herramientas más comunes (simple consulta y reporte de datos, procesamiento analítico en línea, análisis estadístico, predicciones y minería de datos) pueden ser usadas de una gran variedad de formas. El objetivo de todo esto es transformar las montañas de datos en información útil para la empresa (McGeever, 2000).

El radical crecimiento de nuevas formas de "inteligencia" generada por computadora es una de las dos revoluciones en materia de TI que se están dando en la actualidad. La segunda es el Internet, del cual ya todo mundo conoce (Martín, 2001). Las dos revoluciones van a soportar y amplificar una a la otra. El conjunto de ambas resultará en una radical reinención de los negocios.

Los sistemas actuales de Business Intelligence están contruidos en una moderna infraestructura, que consisten de una arquitectura federada (también conocida como modular) que acomoda todos los componentes en un moderno sistema de inteligencia del negocio (Hackney, 2000). Estos sistemas incluyen:

- Data warehousing y data marts, sistemas de almacén de datos.
- Aplicaciones analíticas.
- Data mining, herramientas para minería de datos.
- OLAP, herramientas de procesamiento analítico de datos.
- Herramientas de consulta y reporte de datos.
- Herramientas de producción de reportes personalizados.
- ETL, herramientas de extracción, transformación y carga de datos.
- Herramientas de administración de sistemas.
- Portales de información empresarial.
- Sistemas de base de datos.
- Sistemas de administración del conocimiento.

Desde luego, una organización puede implementar por separado cada una de éstas herramientas y alcanzar un buen nivel de inteligencia, o bien, implementar una solución completa de Business Intelligence que muchos proveedores ofrecen actualmente.

## **2.3 Data Warehouse (DW)**

### **2.3.1 Definición**

El concepto de data warehousing es un concepto definido desde 1981 por Bill Inmon, considerado el padre del tema. En esencia busca proporcionar un modelo arquitectónico para el flujo de datos de los sistemas operacionales hacia los ambientes de soporte a la decisión. Busca dar con varios problemas asociados a este flujo así como los altos costos asociados con él (Manning, 1999).

A través de su historia han ido surgiendo muchas definiciones, para Gardner (1998) un data warehouse es un proceso, no un producto, para la elaboración y administración de datos provenientes de diferentes fuentes con el propósito de obtener una vista particular, detallada de todo el ambiente de la organización.

Johnson (1999) da una definición sencilla pero completa, un data warehouse es una base de datos que recolecta la información del negocio de muchas fuentes de la empresa, cubriendo todos los aspectos de los procesos de la compañía, productos y consumidores. Provee a los usuarios del negocio de una vista multidimensional de los datos que ellos necesitan para analizar las condiciones del negocio.

Para Foote y Krishnamurthi (2001) es un software para el soporte a la toma de decisiones, la elaboración de reportes corporativos y análisis de datos.

Por su parte, Jones (1998) lo define como la optimización de hardware y software para Sistemas de Información Ejecutiva (EIS) así como Sistemas de Soporte a la Decisión (DSS) y están integrados para soportar procesamientos en línea más que procesamientos de transacciones. Es un ambiente de sistemas de información más que un producto, que consiste de:

- Programas de descarga
  - Transformación de datos extraídos de sistemas de producción.
  - Combinación de mapeo y de datos.
  - Inserta datos.
- Metadata y su administración
  - Estructura de los datos.
  - Algoritmos utilizados.
  - Mapeo de la estructura de datos.
- Datos actuales y detallados
- Información histórica
  - Por tema.
  - Integrada.
  - No volátil.
  - Variante en el tiempo.

Jones (1998) menciona que un data warehouse almacena los datos de la organización en repositorios únicos pero integrados en una base de datos relacional. Una estrategia de una base de datos distribuido es una combinación de base de datos relacionales y multidimensionales.

Con este tipo de tecnologías, los usuarios pueden obtener la respuesta a varias preguntas que antes no podían responderse fácilmente. Así mismo, con este tipo de datos inter-funcionales, es posible ver que está pasando con el negocio así como conocer la razón que lo ocasionó. (Gardner, 1998). Así mismo, los administradores pueden identificar las principales tendencias, realizar predicciones

TESIS TESIS TESIS TESIS TESIS

y entender completamente el ambiente organizacional así como tomar una foto de la situación en determinado momento. Permite al administrador entender sistemáticamente los eventos presentados, redefinir los procesos de negocios de una organización y lograr una ventaja competitiva. (Furlow, 2001).

### **2.3.2 Data Warehousing**

Es el proceso de extraer y filtrar datos de las bases de datos de las operaciones comunes de la empresa, procedentes de los distintos subsistemas operacionales, para transformarlos, integrarlos, sumarizarlos y almacenarlos en un depósito o repositorio, para poder acceder a ellos cada vez que se necesite (Dale, 2002).

Por otro lado, el Data Warehousing o almacenamiento de datos, es el proceso que facilita la creación y explotación de un data warehouse, haciendo real la gestión del conocimiento. Para conseguirlo es necesario la aplicación de una metodología y la implantación de una arquitectura tecnológica (Treviño, 2002).

Una de las novedades que aporta el data warehousing como sistema de análisis de información, es la creación de la Meta Información (metadata). Se trata de un archivo al que se le considera como diccionario de estructura de datos que el administrador del sistema define con el objetivo de asistir en los procesos de consulta a la base de datos. La metadata se adaptará a las definiciones que el usuario utilizará posteriormente en sus interrogaciones al sistema. De esta manera se conseguirá minimizar los complejos procedimientos de definición de nombres de campos, jerarquías y relaciones entre necesidades de información a la que desea acceder cada compañía.

### 2.3.3 Características

Según Turban, E. (2006), son 9 las características de los data warehouse:

1. Organización. Los datos están organizados por tema (ejemplo; cliente, vendedor, producto, precio y región), y contiene únicamente información relevante para la toma de decisiones.
2. Consistencia. En el almacén los datos serán codificados de una manera consistente.
3. Tiempo Variante. Los datos son conservados durante varios años y pueden ser usados para tendencias, pronósticos y comparación en el tiempo.
4. No volatilidad. Una vez que los datos entran al data warehouse no son actualizados.
5. Relacional. Típicamente los data warehouse usan estructuras relacionales.
6. Cliente/servidor. Usan principalmente arquitectura cliente/servidor para proveer fácil acceso a los datos a usuarios finales.
7. Basados en Web. Están diseñados para proveer un eficiente ambiente para aplicaciones Web.
8. Integración. Se integran los datos desde diferentes fuentes. Actualmente los Web services son utilizados para realizar la integración (Vaughan, 2003).
9. Tiempo real. Aunque la mayoría de las aplicaciones de data warehousing no están en tiempo real, es posible habilitar esta característica.

Según Bill Inmon fue uno de los primeros autores en escribir sobre el tema de los almacenes de datos, define un data warehouse en términos de las características del repositorio de datos:

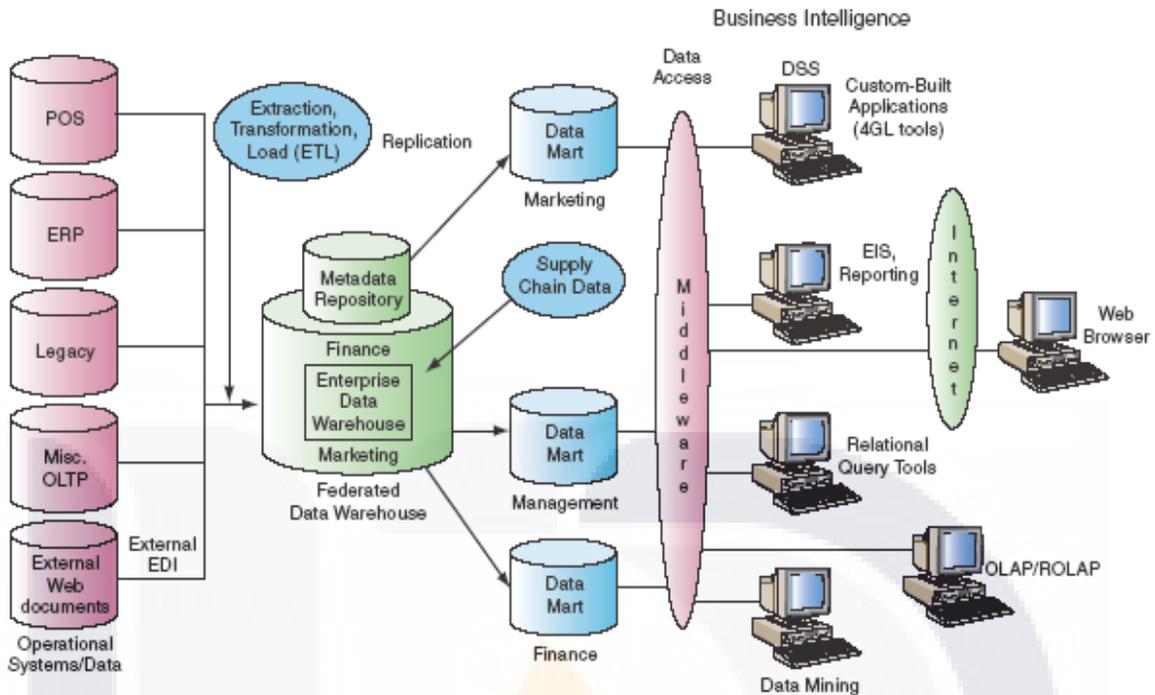
- **Orientado a temas.** Los datos en la base de datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real están unidos entre sí.

- **Variante en el tiempo.** Los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones.
- **No volátil.** La información no se modifica ni se elimina, una vez almacenado un dato, éste se convierte en información de sólo lectura, y se mantiene para futuras consultas.
- **Integrado.** La base de datos contiene los datos de todos los sistemas operacionales de la organización, y dichos datos deben ser consistentes.

#### 2.3.4 Componentes del Data Warehouse

De acuerdo con Turban, E. (2006), los datos de las organizaciones están almacenados en los sistemas operacionales (a la izquierda de la figura 2). Se utiliza software especial llamada ETL (extracción, transformación y carga) para procesar los datos que van a estar en el data warehouse. No todos los datos son necesariamente transferidos al data warehouse y frecuentemente solo un resumen de ellos son transferidos. Los datos transferidos al data warehouse son organizados de tal forma que sean acezados y localizados por parte de los usuarios finales. Asimismo, los datos son estandarizados y organizados por tema.

Los data warehouse tiene un almacén para los metadatos, los cuales son datos acerca de los datos (Sen, 2004). Los Metadatos incluyen programas acerca de los datos, reglas para organizarlos y datos sumariados que son más fáciles de indexar y buscar, especialmente con herramientas Web. Finalmente las herramientas de middleware tools las cuales permiten el acceso al data warehouse.



**Figura 2. Componentes del Datawarehouse.**

**Fuente: Information Technology For Management 5th Edition. Turban, Leidner, McLean, Wetherbe.**

### 2.3.5 Arquitectura de referencia

Inmon define la arquitectura de un data warehouse con cuatro componentes:

1. Los sistemas fuente, donde se gestiona la información relevante de la operación de la organización.
2. El área intermedia (o staging area), en la cual se hace la integración, unificación y limpieza de los datos que vienen de los diferentes sistemas fuente.
3. El área de almacenamiento, conformada por dos elementos: el repositorio y los metadatos.
4. El área de acceso a los datos a través de diferentes herramientas de consulta, tales como publicación en la web, generadores de reportes dinámicos y predefinidos, herramientas de minería de datos y OLAP. En la Figura 3 se presenta la arquitectura general de un data warehouse.

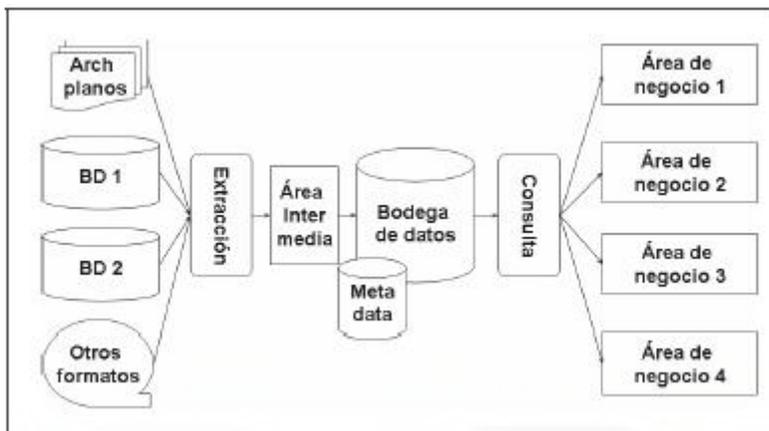


Figura 3. Arquitectura de un data warehouse, adaptado de Inmon, 1996.

Se han propuesto algunas variaciones a la arquitectura anterior (Barclay, Slutz y Gray, 2000) ; (Bédard, Merrett y Han, 2000); (Corey y Abey, 1997); (Inmon, 1996); (Kimball, 1996); (Kimball, et al., 1998), entre las que se incluye la creación de data marts, los cuales se pueden considerar pequeñas bodegas porque sólo contienen información de un tema o área de negocio en particular (a diferencia de la bodega corporativa, que abarca toda la empresa). Los data marts pueden ser de dos tipos: independientes, si cada uno, actuando como una bodega, obtiene su propia información, como los mostrados en la Figura 4, o dependientes, si obtienen su información de una bodega, como los mostrados en la Figura 5.

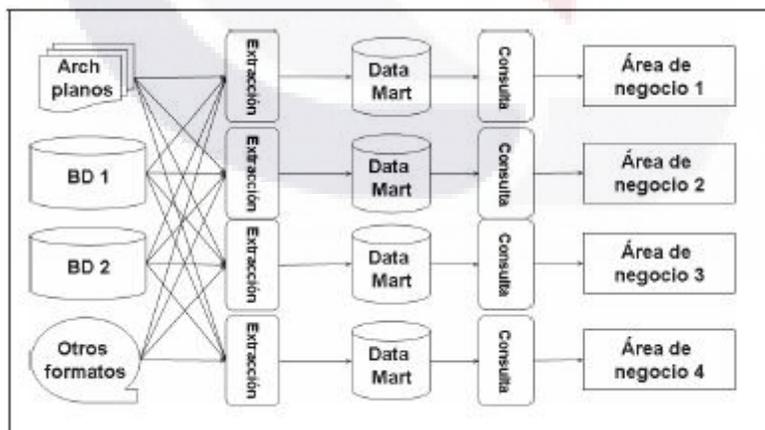


Figura 4. Arquitectura de un esquema de data marts independientes.

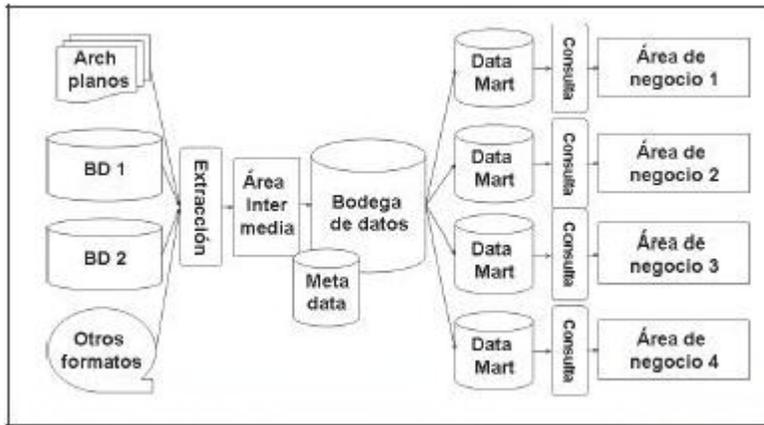


Figura 5. Arquitectura de un esquema de data marts dependientes.

### 2.3.6 A quién va dirigido el data warehouse

Dadas las características de un sistema de data warehouse, su aplicación puede tener varios fines, en una diversidad de industrias. No obstante, en términos generales, podemos decir que su aplicación más rica corresponde a entornos de empresas en los que se identifican grandes volúmenes de datos, asociados a: cantidad de clientes, variedad de productos y cantidad de transacciones (Dyché, 2001).

De acuerdo con Turban, E. (2006), el data warehousing es más apropiados para organizaciones que tienen algunas de los siguientes características:

1. Grandes cantidades de datos requieren ser acezados por los usuarios finales.
2. Los datos operacionales son almacenados en diferentes sistemas.
3. El enfoque de la administración está basada en información.
4. Tiene una gran cantidad de clientes.
5. Los mismos datos son representados de manera diferentes por los diferentes sistemas.
6. Los datos están almacenados con un formato técnicamente muy alto y dificulta su descriptación.
7. Extenso uso computacional por parte de los usuarios finales.

### **2.3.7 Beneficios de contar con un Data Warehouse**

Gray (1998) menciona que los beneficios que un data warehouse puede aportar son:

- Proporciona una herramienta para la toma de decisiones en cualquier área funcional, basándose en información integrada y global del negocio.
- Facilita la aplicación de técnicas estadísticas de análisis y modelización para encontrar relaciones ocultas entre los datos del almacén obteniendo un valor añadido para el negocio de dicha información.
- Proporciona la capacidad de aprender de los datos del pasado y de predecir situaciones futuras en diversos escenarios.
- Simplifica dentro de la empresa la implantación de sistemas de gestión integral de la relación con el cliente.
- Supone una optimización tecnológica y económica en entornos de centro de información, estadística o de generación de informes con retornos de la inversión espectaculares.

### **2.4 Definición data mart (DM)**

Un data mart es un conjunto de datos para un departamento en especial que se crea con la arquitectura base para el data warehouse (Demarest,1997).

De acuerdo con (Gupta, 1997), el DW puede ser utilizado por otras aplicaciones como la fuente de sistemas operacionales de datos, puede alimentar de datos a otros data warehouses o a data marts (también conocido como DW Departamental).

De acuerdo a Johnson (1999): “El data mart es un conjunto especializado de información de negocio enfocado a un aspecto particular de la empresa, como los departamentos o los procesos del negocio. La información en un data mart siempre viene de muchos sistemas de datos. Muchas compañías deciden

TESIS TESIS TESIS TESIS TESIS

alimentar a los data marts desde los data warehouses debido a que la información en el DW ya está consolidada y procesada desde la misma fuente.

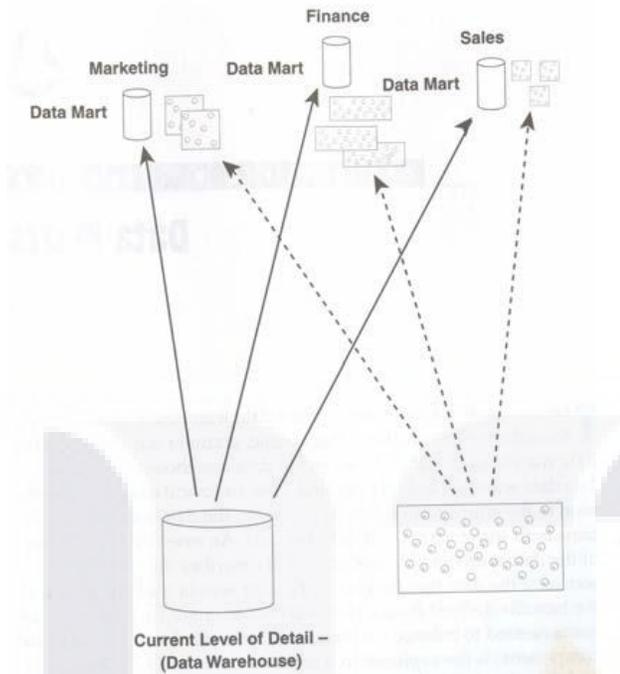
La estrategia de data mart apareció para ser más popular y fácil de entender. La creación de un data mart orientado a un área para resolver problemas particulares representa una solución más simple.

Las bases de datos departamentales, que pueden llegar a crear un data warehouse, son llamadas data marts. Un data mart es un conjunto de datos para un departamento en especial que se crea con la arquitectura base para el data warehouse. El dato que contiene el data warehouse está a un nivel muy exacto, detallado, mientras que el dato que contiene un data mart se encuentra a un nivel resumido o genérico. Cuando el data mart que contiene información de varios departamentos entonces contiene diferentes combinaciones y selecciones de datos al mismo nivel detallado encontrado en el data warehouse (Demarest,1997).

Generalmente, cada departamento operacional quiere su propio data mart. En este sentido, los departamentos que suelen tener sus propios data marts son finanzas, mercadotecnia, ventas y contabilidad. Aunque, ciertamente un mismo data mart puede servir a más de un departamento.

De acuerdo con Devlin (1997), el usuario de un data mart suele ser algún analista departamental que mayormente es una persona del negocio, no un técnico. El tipo de decisiones que suelen tomar son a mediano plazo y a nivel estratégico. Suelen ser recolectores más que exploradores.

En la figura 6 se ilustra la relación que existe entre un data mart y el data warehouse.



**Figura 6. Data Mart. Fuente: Inmon (1999).**

La información de un data mart siempre viene de muchos sistemas de datos. Muchas compañías deciden alimentar a los data marts desde los DWH debido a que la información del DWH ya está consolidada y procesada desde la misma fuente.

El data mart es el antecedente del data warehouse, incluso puede no haber data warehouse, sin embargo el data warehouse mejora el rendimiento del data mart (Russell, 1998).

Poe, Klauer y Brobst (1998) señalan que las características de un data mart son:

- De implementación rápida, sencilla y de bajo costo.
- Cubre las necesidades de una unidad de negocio.
- Proporciona protección a la información sensible.
- Tiempos de respuesta más cortos ya que los volúmenes de datos son menores.
- Está construido de forma ascendente.

Inmon (1998) señala que los data marts son conocidos con varios nombres como:

- Base de datos reporteadoras.
- Base de datos departamentales.
- Bases de datos OLAP.
- Base de datos multidimensionales.

Gray (1998) sostiene que los data marts ofrecen una serie de ventajas sobre el uso de data warehouse, ventajas como:

- Los data marts son más baratos.
- Requieren de un menor tiempo de implementación.
- Son controlados localmente y no centralmente con los DWH.
- Almacenan menos información que el DWH, por lo cual tiene un tiempo de respuesta menor, así mismo es más sencillo entender y navegar en él.
- Permite a las unidades de negocio construir su propio sistema de soporte a la decisión.

Es importante mencionar que así como ofrece ventajas, este tipo de tecnología tiene sus limitaciones o desventajas, entre ellas están (Gray, 1998):

- Degradación del desempeño a través del tiempo.
- La administración de múltiples data marts puede volverse compleja, sobre todo si el número de data marts aumenta. Es necesario mantener una estandarización, seguridad y consistencia en los datos, de manera que esta arquitectura siga siendo funcional para la organización.
- El desarrollo e implementación. Aunque la construcción es más sencilla, aspectos como los proveedores, la integración de software, la descarga, pueden hacer difícil la implementación. Los data marts como solución no resuelven todos los problemas, las organizaciones deben tomar decisiones

en cuanto a la lógica de la extracción de datos y la definición consistente de los datos.

- Tienen un acceso remoto.

## **2.5 Data warehouse vs. data mart**

Las razones para desarrollar en una empresa data warehouse antes de data mart o viceversa, dependen de factores tales como la naturaleza de la organización, tamaño, requerimientos, estructuración y distribución geográfica de la misma. Por este motivo no existe un planteamiento estandarizado, que permita aplicarlo, indistintamente del tipo de empresa que lo requiera [Inmon1].

El diseñador de los data marts o data warehouse debe tener presente algunos aspectos que hacen la diferencia entre estos dos planteamientos:

- Los data marts son creados para satisfacer las necesidades específicas de un departamento de acuerdo a los objetivos de los mismos. El Data warehouse se crean para satisfacer las necesidades globales de una corporación, pero no se puede satisfacer las necesidades de ambos.
- La granularidad entre un planteamiento y otro es muy diferente; los data marts están constituidos por datos resumidos o agregados, por otro lado los datos en el data warehouse mantienen más detalle sobre los datos de la empresa, por esto resulta muy complejo pasar datos de los data marts hacia la data warehouse.
- Los datos contenidos en los data marts son más recientes históricamente que los contenidos en la data warehouse.
- Las relaciones en cuanto a temas en los data marts no son las mismas que las que se pueden encontrar en el data warehouse.
- Los tipos de consultas realizadas en los data marts son muy diferentes a las realizadas en la data warehouse.

- En los data marts los usuarios son recolectores de información, mientras que en la data warehouse los usuarios son exploradores de información.

## **2.6 Diferencias entre la base de datos operacional y el data warehouse**

Los sistemas tradicionales de transacciones y las aplicaciones de data warehousing son polos opuestos en cuanto a sus requerimientos de diseño y sus características de operación. Es importante comprender estas diferencias para evitar caer en el diseño de un data warehouse como si fuera una aplicación de transacciones en línea (OLTP). Las principales diferencias son:

- Las aplicaciones OLTP administran las transacciones (operaciones) mientras que un data warehouse administra áreas o conceptos. Tal cual se señala en la Tabla 2.

**Tabla 2. OLTP vs Data warehouse.**

<b>Enfoque OLTP</b>	<b>Enfoque data warehouse</b>
Merma de productos	Inventario
Adquisición de material de oficina	Compras
Cobro de una factura	Clientes
Agregar un producto a la línea de producción	Producción

- Un data warehouse es de mayor tamaño y puede estar conformado por varios sistemas OLTP's.
- En cuanto al diseño, un OLTP está normalizado y el DWH está desnormalizado.

- El OLTP está formado de más tablas con pocas columnas y un DWH tiene un menor número de tablas pero con un mayor número de columnas.
- Los OLTP se actualizan día con día mientras que el DWH lo hace de manera periódica.
- Las estructuras de los OLTP son muy estables, con pocos cambios, mientras que las estructuras del DWH sufren cambios constantes derivados de su evolución. Esto se debe a que los tipos de consultas a los cuales están sujetos son muy variados y es imposible preverlos todos de antemano.
- Para finalizar tenemos que los sistemas OLTP tienen un mayor número de usuarios con referencia al data warehouse, los sistemas OLTP realizan cientos de transacciones por segundo mientras que una consulta al data warehouse puede tomar minutos.

Una vez analizadas las diferencias entre ambos conceptos, tenemos que el concepto data warehouse es más estratégico mientras que OLTP es más operacional.

## ***2.7 Modelado Multidimensional (MMD)***

Para el diseño de un DW, en general, se emplea la representación de un modelo multidimensional. El modelo mutidimensional es una técnica de diseño lógico que busca presentar los datos en un estándar, que permita una recuperación adecuada de estos. La idea fundamental es que el usuario visualice fácilmente la relación que existe entre los distintos componentes del modelo (Wolf, 1999).

El modelo multidimensional es un modelo adecuado que provee un camino viable para agregar hechos a los largo de múltiples atributos, llamados dimensiones. Los datos son almacenados como hechos y dimensiones en un modelo de datos relacional.

## 2.7.1 Componentes del Modelo Multidimensional

### 2.7.1.1 Tabla Fact o de Hechos

Es la tabla central en un esquema dimensional. Es en ella donde se almacenan las mediciones numéricas del negocio. 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. Por ejemplo, para el caso de producto, mercado y tiempo antes visto, el grano puede ser la cantidad de madera vendida 'mensualmente'. El grano revierte las unidades atómicas en el esquema dimensional.

Cada medida es tomada de la intersección de las dimensiones que la definen. Idealmente está compuesta por valores numéricos, continuamente evaluados y aditivos. La razón de estas características es que así se facilita que los miles de registros que involucran una consulta sean comprimidos en unas pocas líneas en un set de respuesta.

La clave de la tabla fact recibe el nombre de clave compuesta o concatenada debido a que se forma de la composición (o concatenación) de las llaves primarias de las tablas dimensionales a las que está unida.

Así entonces, se distinguen dos tipos de columnas en una tabla fact: columnas fact y columnas key. Donde la columna fact es la que almacena alguna medida de negocio y una columna key forma parte de la clave compuesta de la tabla.

### **2.7.1.2 Tablas Look-up o Dimensionales**

Estas tablas son las que se conectan a la tabla fact, son las que alimentan a la tabla fact. Una tabla lock\_up almacena un conjunto de valores que están relacionados a una dimensión particular. Tablas lock\_up no contienen hechos, en su lugar los valores en las tablas lock\_up son los elementos que determinan la estructura de las dimensiones. Así entonces, en ellas existe el detalle de los valores de la dimensión respectiva.

Una tabla lock\_up está compuesta de una primary key que identifica unívocamente una fila en la tabla junto con un conjunto de atributos, y dependiendo del diseño del modelo multidimensional puede existir una foreign key que determina su relación con otra tabla lock\_up.

Para decidir si un campo de datos es un atributo o un hecho se analiza la variación de la medida a través del tiempo. Si varía continuamente implicaría tomarlo como un hecho, caso contrario será un atributo.

Los atributos dimensionales son un rol determinante en un DDW. Ellos son la fuente de todas las necesidades que debieran cubrirse. Esto significa que la base de datos será tan buena como lo sean los atributos dimensionales, mientras más descriptivos, manejables y de buena calidad, mejor será el DDW.

### **2.7.2 Procesamiento Analítico en Línea (OLAP)**

El Procesamiento Analítico en Línea, OLAP por sus siglas en inglés On Line Analytical Processing, es un conjunto de herramientas de software que permite el análisis de los datos almacenados en una base de datos. Es una tecnología que le permite a los analistas, administradores y ejecutivos realizar consultas de forma rápida, consistente, así como un acceso interactivo a una gran variedad de vistas de la información, que ha sido transformada a la forma en que le es útil al usuario.

TESIS TESIS TESIS TESIS TESIS

Este tipo de tecnologías se caracteriza por un análisis multidimensional o MDDB's (Multidimensional Databases).

Un sistema OLAP se puede entender como la estandarización de un elaborador de informes debido a que las aplicaciones informáticas clásicas de consulta, orientadas a la toma de decisiones, deben ser programadas y estandarizadas, ya que atendiendo a las necesidades del usuario, se crea una y otra interfaz. Ante esto, muchos desarrolladores se dieron cuenta de que estas aplicaciones eran susceptibles de ser generalizadas y servir para casi cualquier necesidad de cualquier base de datos. Los sistemas OLAP evitan la necesidad de desarrollar interfaces de consulta, y ofrecen un entorno único válido para el análisis de información histórica orientado a la toma de decisiones. En cambio, es necesario definir dimensiones, jerarquías y variables, organizando de esta forma los datos. (De la Herrán, 2000).

Fernández (2005) señala que la funcionalidad de los sistemas OLAP se caracteriza por ser un análisis multidimensional que soportan los análisis del usuario, ofreciendo la oportunidad de navegación, a través de la selección de la información requerida.

La forma en que los sistemas OLAP organizan estos datos es a través de dos tipos de arquitecturas: ROLAP (Relational OLAP) y MOLAP (Multidimensional OLAP).

- ROLAP (Relational OLAP) es la arquitectura de base de datos multidimensional en la que los datos se encuentran almacenados en una base de datos relacional que tiene forma de estrella (también llamada copo de nieve o araña). En ROLAP, en principio la base de datos sólo almacena información relativa a los datos en detalle, evitando acumulados (evitando redundancia). Tanto los datos precalculados y agregados como los datos fuente residen en la misma base de datos relacional. Treviño (2002) sostiene que un sistema ROLAP utiliza una arquitectura de tres niveles, en

la cual la base de datos relacional maneja los requerimientos de almacenamiento de datos y el motor ROLAP proporciona la funcionalidad analítica.

- El nivel de base de datos utiliza una base de datos relacional para el manejo, acceso y obtención de los datos.
  - El nivel de aplicación, realiza las consultas multidimensionales de los usuarios.
  - El motor ROLAP se integra con niveles de presentación y es aquí donde los usuarios realizan los análisis ROLAP.
- MOLAP (Multidimensional OLAP), es un conjunto de interfases, aplicaciones y tecnologías de bases de datos, en los cuales los datos se encuentran almacenados en archivos con estructura multidimensional, en los que se reservan espacio para todas las combinaciones de todos los posibles valores de todas las dimensiones de cada una de las variables, incluyendo los valores de dimensión que representan acumulados. Es decir, un sistema MOLAP contiene precalculados (almacenados) los resultados de todas las posibles consultas a la base de datos. (De la Herrán, 2000). Optimiza las búsquedas, pero requiere más espacio de disco y diferente software. El primer punto está dejando ser un problema: el espacio de disco cada vez es más barato. Treviño (2002) menciona que un sistema MOLAP utiliza una arquitectura de dos niveles: la base de datos multidimensionales y el motor analítico.
    - La base de datos multidimensional es la encargada del manejo, acceso y obtención del dato.
    - El nivel de aplicación ejecuta los requerimientos OLAP. El nivel de presentación se integra con el de aplicación y proporciona una interfaz a través de la cual los usuarios finales visualizan los análisis

OLAP. A través de un arquitectura cliente/servidor, varios usuarios accedan a la misma base de datos multidimensional.

Gray (1998) realizó una comparativa de ROLAP y MOLAP, y señala que los sistemas ROLAP soportan los sistemas con alta volatilidad de los datos. Los ROLAP pueden crecer a un gran número de dimensiones y los MOLAP soportan 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 (menores a 5 GB).

En la tabla 3 se muestra un listado de funcionalidades importantes y qué tipo de arquitectura las soporta de la mejor forma.

**Tabla 3. Comparación de bases de datos MOLAP y ROLAP.**

<b>Funcionalidad</b>	<b>Arquitectura adecuada</b>
Complejidad de cálculos	MOLAP
Dispersión de datos	ROLAP
Cálculos a nivel de fila	MOLAP
Actualización de bases de datos	MOLAP
Volatilidad de datos	ROLAP
Mayor número de dimensiones	ROLAP
Mayor volumen de datos	ROLAP
Mejor tiempo de desarrollo, curva de aprendizaje	ROLAP
Estándares, interoperabilidad	ROLAP
Mejor tiempo de búsqueda	MOLAP
Consistencia, confiabilidad	MOLAP
Mejor tiempo de descarga	ROLAP
Seguridad	ROLAP
Impacto de red	MOLAP
Estabilidad del proveedor	ROLAP

*Fuente: Gray (1998).*

Analizado todo esto, es posible ver que la arquitectura ROLAP es más flexible y capaz de crecer acorde a las necesidades del usuario, mientras que el MOLAP es una opción más limitada.

### **2.7.3 Pasos básicos del Modelado Multidimensional**

1. Decidir cuáles serán los procesos de negocios a modelar, basándose en el conocimiento de éstos y de los datos disponibles. Ejemplo: Gastos realizados por cada mercado para cada ítem a nivel mensual. Productos vendidos por cada mercado según el precio en cada mes.
2. Decidir el Grano de la tabla Fact de cada proceso de negocio. Ejemplo: Producto x mercado x tiempo. En este punto se debe tener especial cuidado con la magnitud de la base de datos, con la información que se tiene y con las preguntas que se quiere responder. El grano decidirá las dimensiones del DDW. Cada dimensión debe tener el grano más pequeño que se pueda puesto que las preguntas que se realicen necesitan cortar la base en caminos precisos (aunque las preguntas no lo pidan explícitamente).
3. Decidir las dimensiones a través del grano. Las dimensiones presentes en la mayoría de los DDW son: tiempo, mercado, producto, cliente. Un grano bien elegido determina la dimensionalidad primaria de la tabla fact. Es posible usualmente agregar dimensiones adicionales al grano básico de la tabla fact, donde estas dimensiones adicionales toman un solo valor para cada combinación de las dimensiones primarias. Si se reconoce que una dimensión adicional deseada viola el grano por causar registros adicionales a los generados, entonces el grano debe ser revisado para acomodar esta dimensión adicional.
4. Elegir las mediciones del negocio para la tabla fact. Se deben establecer los ítems que quedarán determinados por la clave compuesta de la tabla fact.

### Capítulo 3. Metodología del desarrollo

#### 3.1 Introducción

Para el desarrollo del data mart se hará uso de la metodología Kimball. La metodología de Kimball se enfoca principalmente en el diseño de la base de datos que almacenará la información para la toma de decisiones. El diseño se basa en la creación de tablas de hechos (FACTS) que son tablas que contienen la información numérica de los indicadores a analizar, es decir la parte cuantitativa de la información.

Ralph Kimball, es reconocido como uno de los padres del concepto de Data Warehouse, se ha dedicado desde hace más de 10 años al desarrollo de su metodología para que éste concepto sea bien aplicado en las organizaciones y se asegure la calidad en el desarrollo de estos proyectos.

#### 3.1.1 Metodología Kimball – Ciclo de Vida

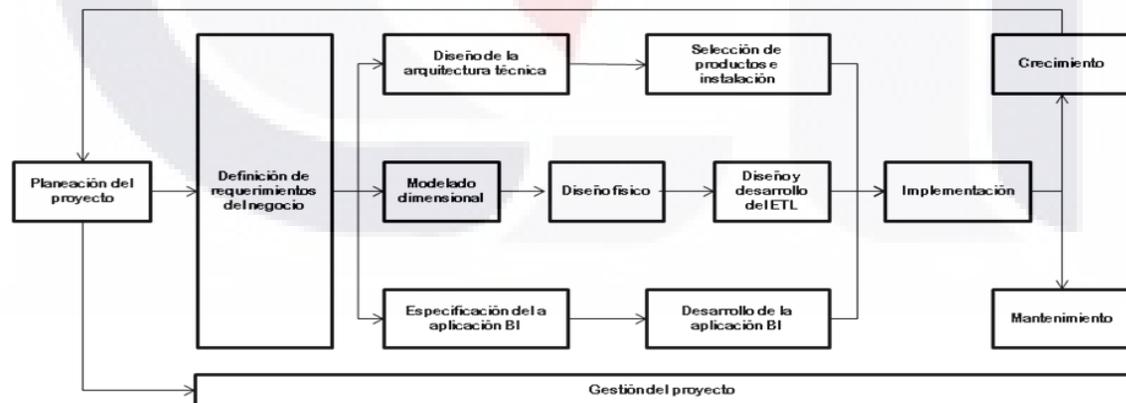


Figura 7. Ciclo de vida según Kimball.

En la figura anterior se presenta el ciclo de vida para el desarrollo de un proyecto data mart o data warehouse.

### **3.2 Definición de requerimientos del negocio**

Se realizó una serie de actividades para identificar y conocer el proceso de capacitación en el INEGI, así como los requerimientos del área para la inteligencia del negocio. Entre las actividades que se realizaron fueron las siguientes:

- Identificar la misión, visión, valores, objetivos, estrategias y el proceso de la dirección de Capacitación. Consultar el anexo A.
- Identificar el proceso de registro de información del sistema transaccional vigente, incluyendo los reportes que se generan actualmente. Consultar el anexo B.
- Entrevistas. Consultar el anexo C.
- Se identificó un portafolio de análisis de requerimientos. Consultar el anexo E.
- Se priorizó los requerimientos, concluyendo en los siguientes indicadores:
  - Número de cursos solicitados
  - Número de cursos realizados
  - Número de participantes programados
  - Número de participantes atendidos
- Se definió la granularidad a mensual.

### **3.3 Modelo conceptual**

Como resultado del análisis de requerimientos se definió el siguiente modelo conceptual del proyecto.

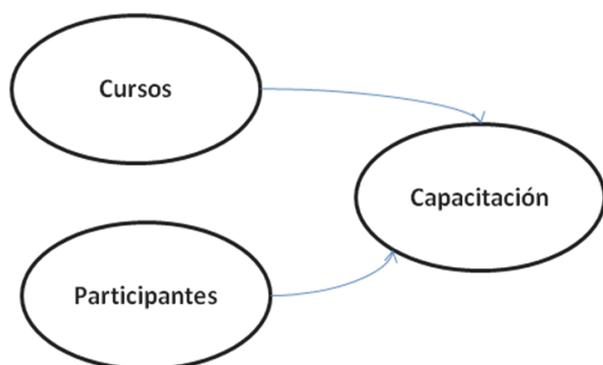


Figura 8. Modelo conceptual del data mart de capacitación.

Modelo propuesto.

### 3.4 Construcción del modelo dimensional

En base al modelo conceptual resultante se acordó definir dos data marts, uno para los cursos y otro para los participantes. De la misma manera se determinó las dimensiones por las cuales que se requiere realizar las consultas.

#### 3.4.1 Matriz de Data Marts

La siguiente matriz muestra la relación entre los data marts y las dimensiones resultantes.

Tabla 4. Matriz de data marts.

Data marts	Dimensiones									
	Oficina	Anio	Estado del participante	Evento	Mes	Tipo de evento	Programa	Estado del evento	Reconocimiento	Área
Cursos	x	x			x	x	x	x	x	x
Participantes	x	x	x	x	x					

### 3.4.2 Descripción de los data marts

En la tabla 5 se describe los data marts que se implementaron en este proyecto.

Tabla 5. Descripción de los data marts

Nombre	Descripción
Cursos	Los cursos de capacitación que se han programado.
Participantes	Los participantes que se han programado a cursos de capacitación.

### 3.4.3 Descripción de las dimensiones

En la tabla 6 se describen las dimensiones obtenidas y su contenido.

Tabla 6. Descripción de las dimensiones

Nombre	Descripción
Evento	Nombres de los cursos que se han impartido.
Año	Años en que se realizó un curso.
Área	La clasificación de los cursos por su área de conocimiento.
Programa	La clasificación de los cursos por su tipo.
Mes	Los meses del año.
Estado del participante	Los estados en que puede estar un participante en un curso de capacitación.
Oficina	Las oficinas donde se realizan eventos de capacitación.
Tipo de evento	Especifica si un curso es modular o no.
Estado del evento	Estado en que se encuentra un curso de capacitación.
Reconocimiento	Tipo de reconocimiento que se entrega al participante cuando concluye un curso.

### 3.4.4 Modelo de los data mart's

Se realizaron dos modelos para los siguientes data marts:

#### Cursos

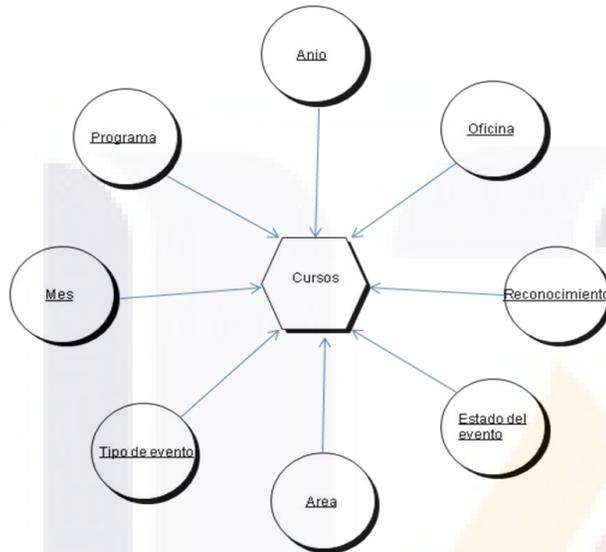


Figura 9. Modelo del data mart de cursos.

#### Participantes

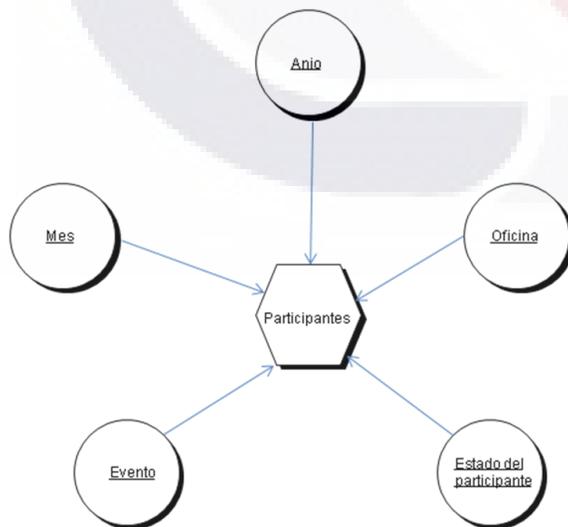


Figura 10. Modelo del data mart de participantes.

### 3.4.5 Diseño lógico de la base de datos

Utilizando la herramienta de SQL Server Management Studio, se implementó el diseño lógico de la base de datos, obteniéndose los siguientes diagramas:

#### Cursos

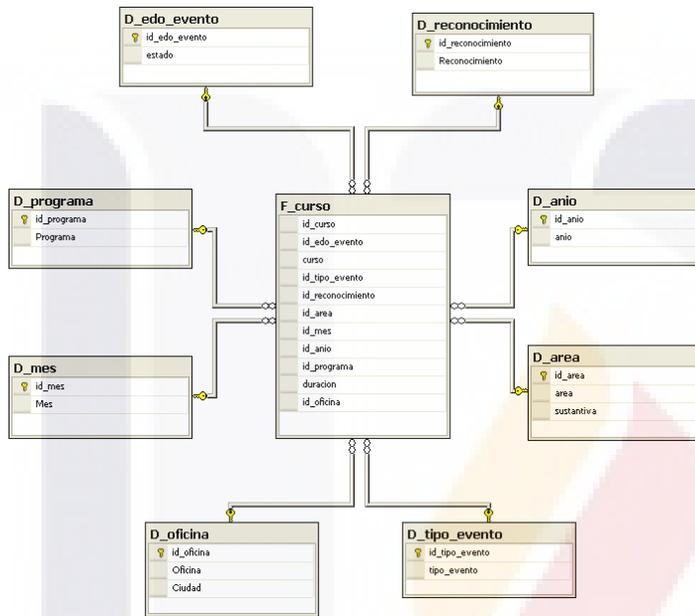


Figura 11. Diseño lógico del data mart cursos.

#### Participantes

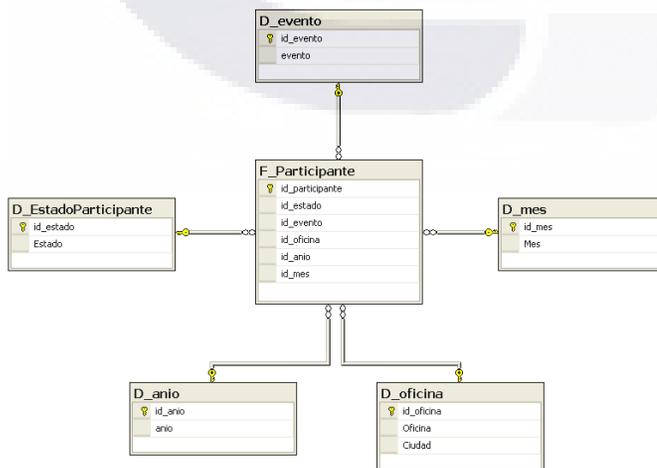


Figura 12. Diseño lógico del data mart participantes.

### 3.5 Herramienta de desarrollo

Para el desarrollo del proyecto se utilizó la solución de Microsoft DW/BI, la cual presenta la arquitectura mostrada en la siguiente figura.

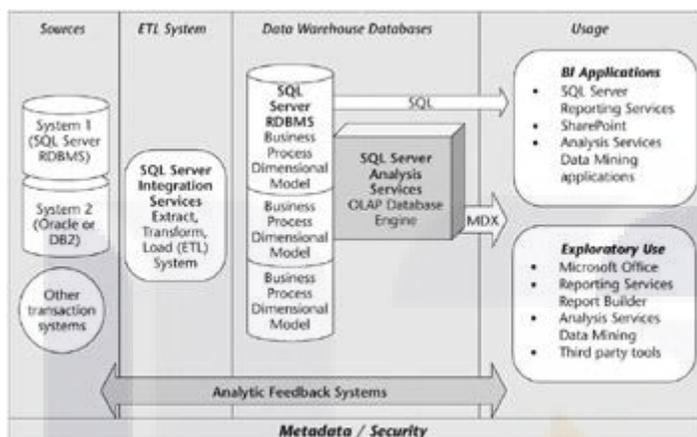


Figura 13. Arquitectura del sistema Microsoft DW/BI.

Cabe mencionar que el INEGI cuenta con el soporte técnico y tecnológico necesario para la implementación del data mart de capacitación, así como el licenciamiento correspondiente.

Las herramientas que se utilizaron en este proyecto son:

#### SQL Server Management Studio

Es la herramienta principal para los administradores de base de datos de SQL server 2005. Es una plataforma que permite el análisis y la administración de datos empresariales con herramientas de Inteligencia de Negocios, análisis, reporte, integración y notificación integradas. Provee de mayor escalabilidad, disponibilidad y seguridad al tiempo que simplifica la creación, implementación y gestión de aplicaciones altamente disponibles y de buen desempeño.

Con esta herramienta se realizó:

- La creación de las tablas de hechos y sus dimensiones.
- Implementación los data marts de cursos y participantes.

- Obtención de los obtuvo los diagramas de las base de datos.
- Programación de actividades en la consola del propio servidor de base de datos utilizando el SQL server agent.

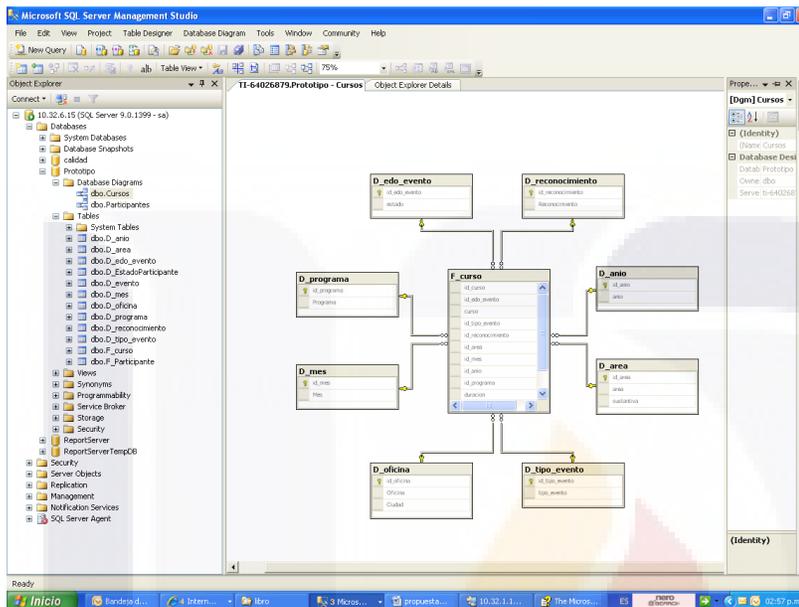


Figura 14. SQL Server Management Studio

## Business Intelligence Development Studio (BI Studio)

Business Intelligence Development Studio es Microsoft Visual Studio 2005 con tipos de proyecto adicionales que son específicos de Business Intelligence de SQL Server 2005. Business Intelligence Development Studio es el entorno principal que se utilizará para desarrollar soluciones empresariales que incluyan proyectos de Analysis Services, Integration Services y Reporting Services. Cada tipo de proyecto proporciona plantillas para crear los objetos necesarios para las soluciones de Business Intelligence y ofrece varios diseñadores, herramientas y asistentes para trabajar con los objetos.

Con esta herramienta se realizó:

- La creación del proyecto donde se definieron dos cubos.

- Creación de una base de datos de Analysis Services, conteniendo los cubos y sus dimensiones.

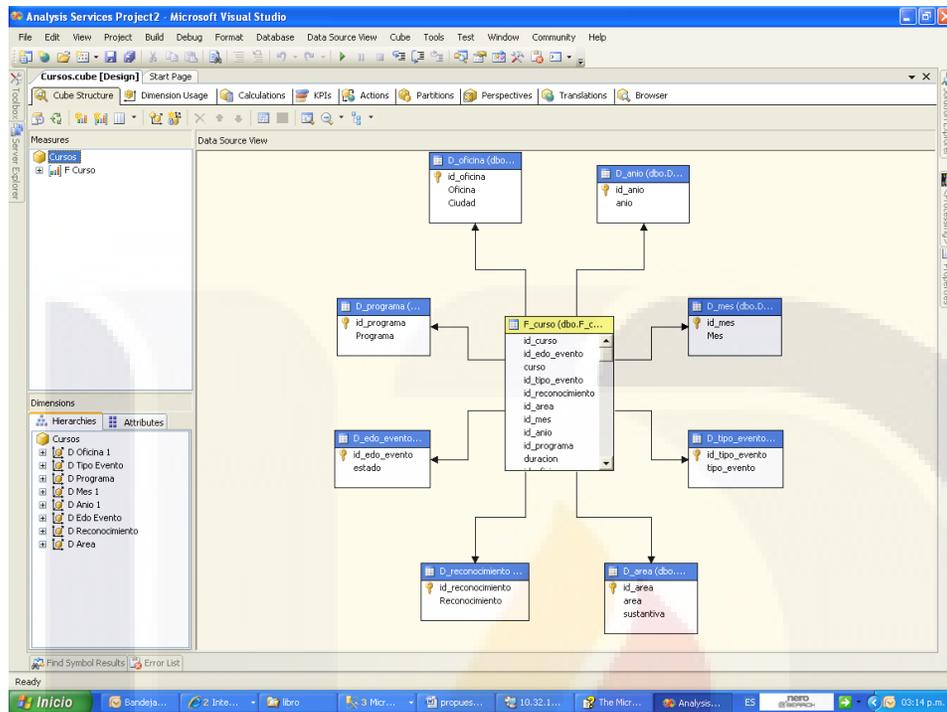


Figura 15. BI Studio.

### 3.6 Poblando el Data Mart

Se definió una aplicación utilizando tecnología .Net, la cual realiza la funcionalidad de extracción, transformación y carga de datos (ETL). Dicha aplicación permitirá actualizar los catálogos (dimensiones) y las tablas de hechos diariamente.

Las fuentes de los datos están en SQL server 2000 y mediante la aplicación desarrollada los carga a un servidor de SQL server 2005. Los cálculos que se realizan son para determinar el año y mes en que se realizó un curso o cuando un participante asistió a clases.

En el anexo F se puede consultar el código de la aplicación.

### 3.7 Implementación

Para la implementación del proyecto se realizaron las siguientes actividades:

1. Se automatizó el proceso de ETL, publicando la aplicación que se realizó para tal fin en un servidor de aplicaciones, de tal forma que diariamente a las 1:00 am se ejecute y se actualice la base de datos que contiene los data marts.
2. Se automatizó la actividad de reproceso de los cubos creados en este proyecto, mediante la SQL server agent, donde se programaron dos tareas, una para procesar el cubo de Cursos y la otra para procesar el cubo de Participantes. Lo anterior permitirá tener los cubos actualizados al día anterior. En el anexo G, se muestra el código en XMLA que se ejecuta para el reproceso de los cubos.

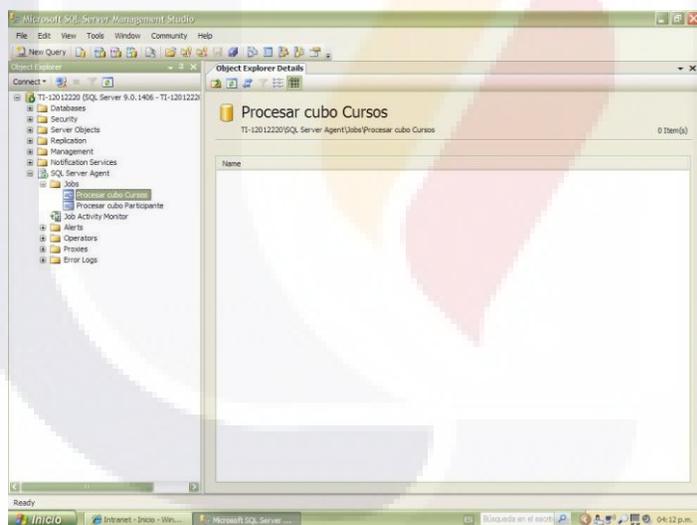


Figura 16. Procesar los cubos.

### 3.8 Consulta de cubos por Intranet

Para la consulta de los cubos por parte de los usuarios finales, se desarrolló una aplicación en .NET y utilizando un componente llamado CellSetGrid.

CellSetGrid es un control ASP.NET para examinar cubos de Analysis Services 2005. Este soporta navegación basada en menú, drill up & down, ordenación, graficación, entre otras. Es 100% basado en Web y no requiere la instalación de un cliente, sólo un navegador. Este control es freeware y es muy simple su configuración.

Finalmente se publicó la aplicación en el sitio de colaboración de la dirección de Capacitación, desde el cual todos los responsables de la misma a nivel nacional tienen acceso.

La siguiente figura muestra un ejemplo de una consulta a través de la herramienta desarrollada.

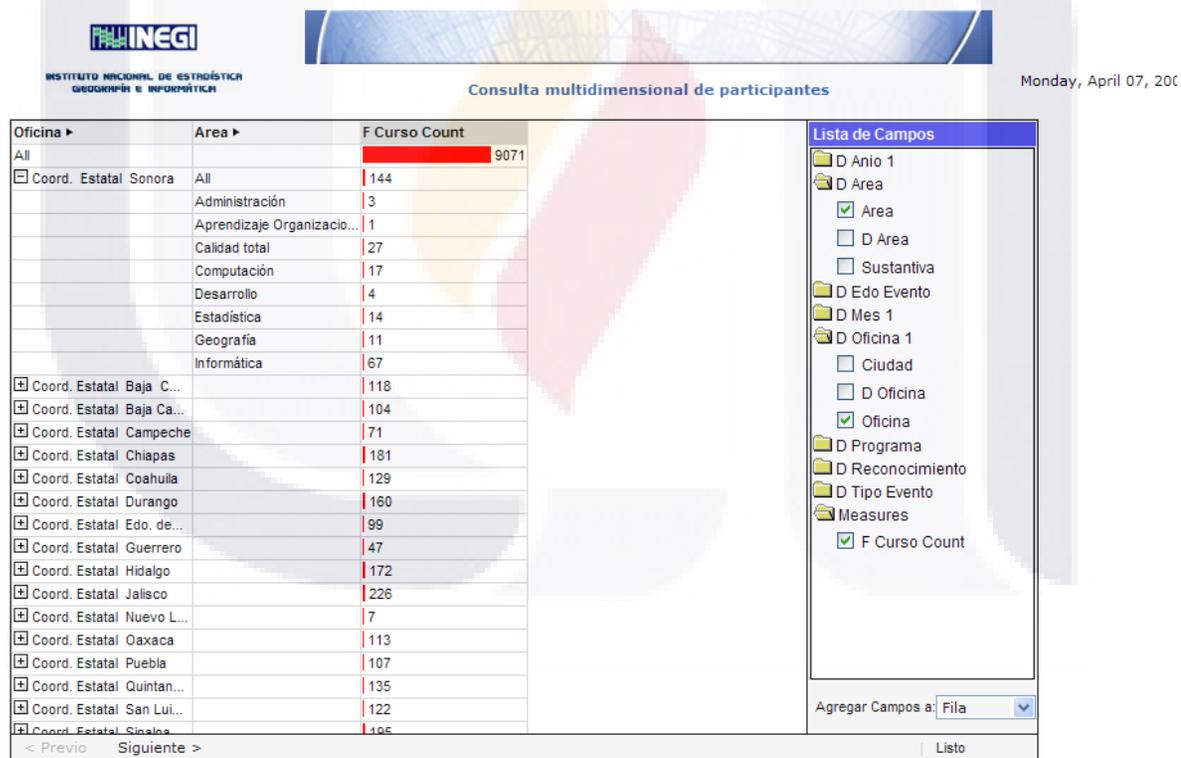


Figura 17. Consulta de cubos por intranet.

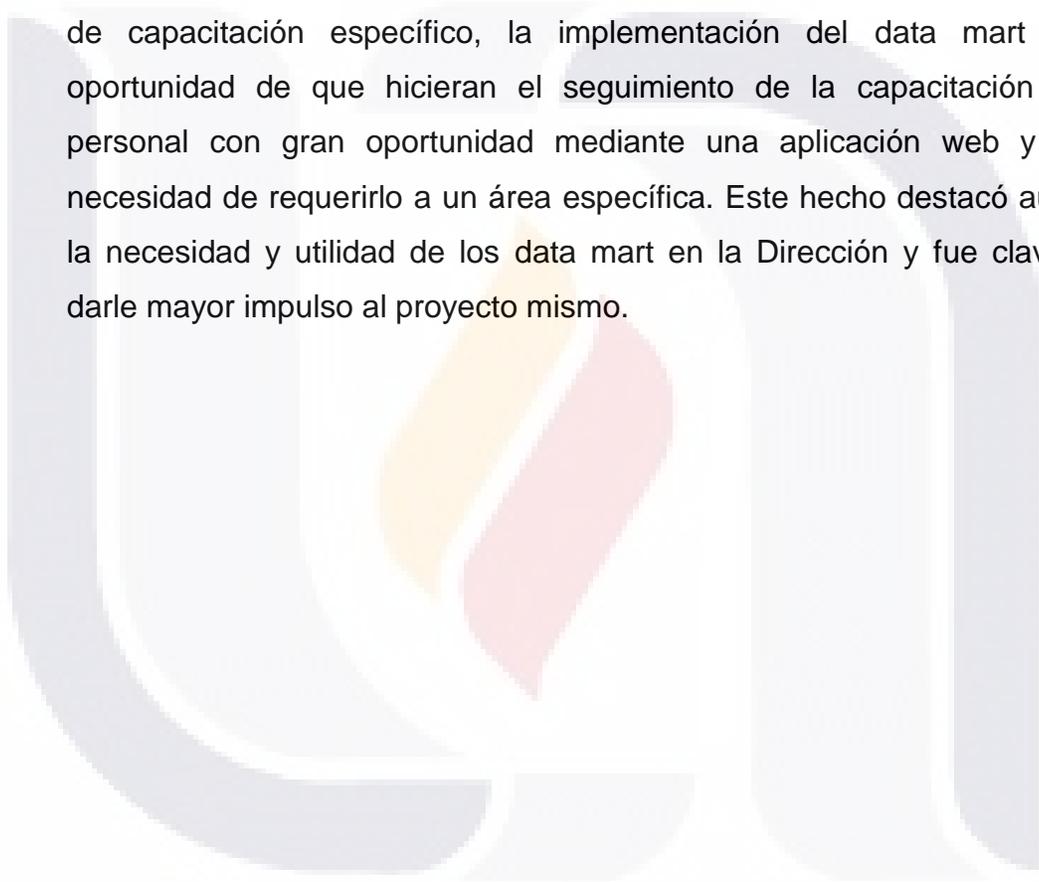
## Capítulo 4. Conclusiones.

Una vez finalizado la implementación de los data marts para capacitación, se tiene como conclusiones generales las siguientes:

- Se conceptualizó un modelo de data mart para el área de capacitación del INEGI, donde se identifican los elementos principales (insumos) para la capacitación en el Instituto.
- Se logró desarrollar un modelo conceptual de data mart para el área de capacitación, el cual en su primera fase contempló dos cubos, el primero de cursos y el segundo de participantes, estos a su vez, quedaron implementados y publicados para su respectiva explotación.
- Referente a la solución tecnológica y con base en el análisis de requerimientos se tomó la decisión de aprovechar la infraestructura con la que cuenta el Instituto. En lo que corresponde al software se utilizó la solución de BI de Microsoft, la cual consta de dos herramientas para el diseño y desarrollo de los data marts (Business intelligence development studio y SQL Management studio) y el motor de la base de datos SQL server 2005. En este punto se contó con la asesoría de los DBA y los responsables de la creación y administración de los cubos en el data warehouse del INEGI, aunque no se tiene el expertis en la versión de SQL server 2005 que se utilizó para este proyecto, por que se requirió la consulta de bibliografía sobre el tema.
- En relación con la herramienta desarrollada para la consulta de los data marts, permitió que los responsables de la capacitación desde su misma oficina puedan tener información oportuna y con esto mejorar en la toma de decisiones. No tendrán que recurrir a oficinas centrales en la ciudad de Aguascalientes para solicitar los reportes correspondientes a su área, ni

esperar a que sea fin de mes para conocer los avances en los indicadores establecidos.

- La consulta de los data marts es a través de la web y es muy intuitiva, por lo que no se requirió realizar una capacitación de uso de la misma.
- Al momento del desarrollo coincidió con un requerimiento especial por parte de la Dirección General de Estadística para dar seguimiento a un programa de capacitación específico, la implementación del data mart dio la oportunidad de que hicieran el seguimiento de la capacitación de su personal con gran oportunidad mediante una aplicación web y sin la necesidad de requerirlo a un área específica. Este hecho destacó aún más la necesidad y utilidad de los data mart en la Dirección y fue clave para darle mayor impulso al proyecto mismo.



## Capítulo 5. Recomendaciones.

En relación de las conclusiones obtenidas, se recomienda las siguientes actividades:

- Que se fomente el uso de soluciones de BI hacia el interior del Instituto, reconociéndolas como “buenas prácticas”, tales como los data marts y minería de datos. Durante el desarrollo de este proyecto, me he dado cuenta que el INEGI tiene un data warehouse bien soportado, pero sin embargo, cumple con un objetivo muy claro que es el de dar servicio a los usuarios internos y externos acerca de la información estadística y geográfica del país. Faltaría desde mi punto de vista el uso de los data marts en las áreas administrativas para saber en corto tiempo el estatus de la plantilla de personal, cómo se encuentra distribuido entre las diversas unidades administrativas, qué perfiles tiene el personal del Instituto, a cuánto asciende el gasto corriente, etc.
- En el caso particular de la Dirección de Capacitación, faltaría implementar más cubos para la consulta y análisis de la información, que cubran la totalidad del proceso de capacitación en el INEGI, por ejemplo, implementar los data marts correspondientes de instructores, aulas de cómputo, materiales didácticos, etc., permitiría realizar una planeación basadas en hechos.
- Realizar actualizaciones, cuando sea requerido, de los reportes gerenciales, para que los mismos se ajusten y se adapten a las necesidades y requerimientos del INEGI, los cuales son cambiantes con el transcurrir del tiempo.

- TESIS TESIS TESIS TESIS TESIS
- El uso de técnicas de minería de datos con la finalidad de encontrar patrones ocultos, ya que suele contener información valiosa oculta entre una gran cantidad de datos y que solo puede ser extraída sometiendo a la base de datos a procesos de análisis exhaustivos que busquen relaciones no obvias en los datos y revelen tendencias y parámetros que permitan identificar el origen de dichas tendencias.



## **Glosario.**

**Agregación:** Actividad de combinar datos desde m múltiples tablas para formar una unidad de información más compleja, necesitada frecuentemente para responder consultas del DataWarehouse en forma más rápida y fácil.

**Atributo:** Elemento de información, que describe una característica de una entidad.

**Base de datos (BD):** Conjunto de datos no redundantes, almacenados en un soporte informático, organizado de forma independiente de su utilización y accesible simultáneamente por distintos usuarios y aplicaciones.

**Base de datos operacional:** Una base de datos que apoya sistemas de software que están soportando normalmente a las operaciones empresariales de una organización. También llamada base de datos OLPT.

**Consulta:** Una petición formal y claramente especificada de información del Data Warehouse planteada por un usuario o una herramienta operada por un usuario.

**Datawarehouse:** Base de datos que almacena una gran cantidad de datos transaccionales integrados para ser usados para análisis gestionales por usuarios especializados (tomadores de decisión de la empresa).

**Data mart:** Conjunto de hechos y datos organizados para soporte de decisiones basados en la necesidad de un área o departamento específico. Los datos son orientados a satisfacer las necesidades particulares de un departamento dado teniendo sólo sentido para el personal de ese departamento y sus datos no tienen porque tener las mismas fuentes que los de otro data mart.

**Datamining:** Análisis de los datos para descubrir relaciones, patrones, o asociaciones desconocidas.

**Data Warehousing:** Proceso de implementación de un proyecto de Data Warehouse.

**Dato:** unidad básica de información.

**Diccionario de Datos:** Compendio de definiciones y especificaciones para las categorías de datos y sus relaciones.

**Dimensión:** Entidad independiente dentro del modelo multidimensional de una organización, que sirve como llave de búsqueda (actuando como índice), o como mecanismo de selección de datos.

**Drill Down:** Exponer progresivamente más detalle (dentro de un reporte o consulta), mediante selecciones de ítems sucesivos.

**Drill-Up:** Efecto contrario a drill -down. Significa ver menos nivel de detalle, sobre la jerarquía significa generalizar o sumarizar, es decir, subir en el árbol jerárquico.

**DSS:** Sistema de Soporte de Decisiones. Sistema de aplicaciones automatizadas que asiste a la organización en la toma de decisiones mediante un análisis estratégico de la información histórica.

**ETT (Extracción, Transformación y Transporte de datos):** Pasos por los que atraviesan los datos para ir desde el sistema OLTP ( o la fuente de datos utilizada) a la bodega dimensional. Extracción, se refiere al mecanismo por medio del cual los datos son leídos desde su fuente original. Transformación (también conocida como limpieza) es la etapa por la que puede atravesar una base de datos para

estandarizar los datos de las distintas fuentes, normalizando y fijando una estructura para los datos.

Finalmente está el Transporte, que consiste básicamente en llevar los datos leídos y estandarizados a la bodega dimensional (puede ser remota o localmente). Generalmente, para un Data Mart no es necesario atravesar por todos estos pasos, pues al ser información localizada, sus datos suelen estar naturalmente estandarizados (hay una sola fuente).

**Extracción:** La actividad relacionada con transferir datos de bases de datos operacionales (fuentes de datos) al Data Warehouse.

**Front-end:** Formato de pantalla. Dícese de la presentación que se le hace al usuario de la pantalla de trabajo.

**Granularidad:** Término que se usa en los Data warehouse para expresar el nivel de detalle. A más alto nivel de granularidad más bajo nivel de detalle (más alto nivel de abstracción).

**Hecho:** Dato numérico que sirve de base para la definición de los indicadores en un modelo multifuncional.

**Jerarquía:** Es un conjunto de atributos descriptivos que permite que a medida que se tenga una relación de muchos a uno se ascienda en la jerarquía. Por ejemplo : los Centros de Responsabilidad están asociados a un Tipo de Unidad, el cual pueden corresponder a una gerencia, subgerencia, superintendencia, etc.; por otro parte, cada CR está asociado a otro CR a nivel administrativo y, también existe una clasificación a nivel funcional.

**Manejador de base de datos:** Colección de numerosas rutinas de software interrelacionados, cada una de las cuales es responsable de alguna tarea

específica tales como: almacenar físicamente, garantizar consistencia, garantizar integridad, atomicidad transaccional y manejar vistas a la información.

**Metadata:** Información que describe un dato. En un contexto de Data Warehouse, califica un dato precisando por ejemplo su semántica, las reglas de gestión asociadas, su fuente, su formato, etc.

**Métrica:** indicador. Entre los indicadores pertinentes, algunos distinguen los hechos, almacenados físicamente en la base de decisión, de las métricas, derivadas de estos hechos.

**Modelo dimensional (o multidimensional):** Técnica de modelado consistente en modelar una base de decisión a partir de la identificación de los hechos a analizar y las dimensiones de análisis asociados.

**OLAP (On Line Analytical Processing):** Caracteriza la arquitectura necesaria para implementar un sistema de información de decisión. Se opone a OLTP. El término OLAP designa a menudo las herramientas de análisis basadas en base de datos multidimensionales.

**OLTP (On-line Transaction Processing):** Sistema transaccional diario (o en detalle) que mantiene los datos operacionales del negocio.

**SQL (Structure Query Language):** Lenguaje estructurado de consulta, usando con casi todas las herramientas de programación para gestionar bases de datos y extraer información de ellas.

**Sumarización:** Actividad de incremento de la granularidad de la información en una base de datos. La sumarización reduce el nivel de detalle, y es muy útil para presentar los datos para apoyar al proceso de Toma de Decisiones.

**Tabla Dimensional:** Dentro del esquema estrella, corresponde a las tablas que están unidas a la tabla central a través de sus respectivas llaves. La cantidad de estas tablas le otorgan la característica de multidimensionalidad a esta estrategia.



## **Bibliografía**

Alter, S. L. (1980). Decision support systems: current practice and continuing challenges. Reading, Mass., Addison-Wesley Pub.

Barclay, T., Gray, J. and Slutz, D., TerraServer: A Spatial Data Warehouse., Proceedings of the 2000 ACM-SIGMOD Conference, 2000.

Bédard, Y., Merrett, T. and Han, J., Fundamentals of spatial data warehousing for geographic knowledge discovery., In: H. Miller and J. Han (Editors), Geographic data mining and knowledge discovery, 2000.

Buskard, D., Mollot., & Mollot, M. (2000). Business Intelligence made easy. Insurance & Technology.

Buskard, D., Mollot., & Mollot, M. (2000). Business Intelligence made easy. Retomado el 10 de Mayo, 2001 de ProQuest Direct en el World Wide Web <http://www-cib.mty.itesm.mx>

Cano, C. (1999) Business Intelligence, decisiones de negocio basadas en tecnología: ruta crítica del negocio moderno. Retomado el 5 de Junio, 2001, de ProQuest Direct en el World Wide Web <http://www-cib.mty.itesm.mx>

Cognizant Technology Solutions (2005). "Consolidations and compílanse drive Data Warehouse growth". <http://www.datawarehouse.com/article/?articleid=5696>.

Cognos (2005), Cognos posicionado como líder en el último estudio de Gartner sobre el mercado del Business Intelligence.

Corey, M. J. and Abey, M., Oracle Data Warehousing., McGraw Hill, 1997.

Dale, T. (2002). Estrategías para la implementación de un Data Warehouse.

De la Herrán, Manuel. "Cómo diseñar grandes variables en bases de datos multidimensional". Revista Digital Universitaria. Junio 2000.

Demarest, M.; Building the data mart; DBMS, Julio 1997.

Devlin, Barry; Data Warehouse, from Architecture to Implementation; Addison –

Druzdel, M. J. and R. R. Flynn (1999). Decision Support Systems. Encyclopedia of Library and Information Science. A. Kent, Marcel Dekker, Inc.

Dyché. J. (2001). E-data, Transformando datos en información con Data Warehousing (1ra. Edición). Printice Hall.

Fernández, Luis Antonio. “Principios de Data Mining”. Tomada de <http://www.monografias.com/trabajos26/data-mining/data-mining.shtml>.

Finlay, P. N. (1994). Introducing decision support systems. Oxford, UK Cambridge, Mass., NCC Blackwell; Blackwell Publishers.

Furlow, Gerri. (2001). “The case for building a data warehouse”. IT Professional. Vol. 3, Issue 4.

Gardner, Stephen. “Building the data warehouse”. Communications of the ACM. Vol. 41, Issue 9. (1998).

Gray, Pau; Watson, Hugh J; Decision support in the DATA WAREHOUSE; the Data warehouseing Institute, Prentice Hall; 1998.

Gupta, 1997; Vivek R. Gupta, “An Introduction to Data Warehousing”, System Services Corporation Papers, Agosto 1997.

Hackney, D. (2000). Your Business Intelligence arsenal. Telephony. Retomado el 10 de Mayo, 2001, de ProQuest Direct en el World Wide Web <http://www-cib.mty.itesm.mx>

Hilson, G. (2001). BI market fraught with instability. Computing Canada. Retomado el 10 de Mayo, 2001, de ProQuest Direct en el World Wide Web <http://www-cib.mty.itesm.mx>

TESIS TESIS TESIS TESIS TESIS

Inmon, W. H., Using the Data Warehouse., John Wiley & Sons, 1996.

Johnson, 1999; Amy Helen Johnson, "Data Warehouse", Computerworld, 6 – Diciembre – 1999.

Keen, P. G. W. (1978). Decision support systems: an organizational perspective. Reading, Mass., Addison-Wesley Pub. Co. ISBN 0-201-03667-3

Keen, P. G. W. (1980). Decision support systems: a research perspective. Decision support systems: issues and challenges. G. Fick and R. H. Sprague. Oxford ; New York, Pergamon Press.

Kimball, R., Reeves, L., Ross, M. and Thornthwaite, W., The Data Warehouse Lifecycle Toolkit: expert methods for designing, developing and deploying data warehouses., John Wiley & Sons, Inc, 1998.

Kimball, R., The Data Warehouse Toolkit: practical techniques for building dimensional data warehouses., John Wiley & Sons, Inc, 1996.

Little, J.D.C.(1970, April). "Models and Managers:The Concept of a Decision Calculus." Management Science, Vol.16, NO.8.

Martín, J. (2001). Alien Intelligence. The Journal of Business Strategy. Retomado el 10 de Mayo, 2001 de ProQuest Direct en el World Wide Web <http://www-cib.mty.itesm.mx>

McGeever, C. (2000). Business Intelligence. Computer World. Retomado el 10 de Mayo, 2001, de ProQuest Direct en el World Wide Web <http://www-cib.mty.itesm.mx>

Michel, R. (2000). Business Intelligence solutions. Manufacturing Systems. Retomado el 10 de Mayo, 2001 de ProQuest Direct en el World Wide Web <http://www-cib.mty.itesm.mx>

Moore, J.H.,and M.G.Chang.(1980,Fall)."Design of Decision Support Systems."

Data Base, Vol.12, Nos.1 and 2.

Poe, V, Klaude P., y Brobst S. "Building a data warehouse for decisión support". New Jersey. Prentice Hall. (1998).

Power, D. J. (1997). What is a DSS? The On-Line Executive Journal for Data-Intensive Decision Support 1(3).

Russell, Joy.; " Secrets from the successful". Information Week. En prensa. (1998).

Sprague, R. H. and E. D. Carlson (1982). Building effective decision support systems. Englewood Cliffs, N.J., Prentice-Hall. ISBN 0-130-86215-0.

Sullivan, T. (2001). Business Intelligence keeps tabs on the Net. InfoWorld. . Retomado el 10 de Mayo, 2001, World Wide Web <http://www Infoworld.com>

Treviño G., E. (2002). Sistema de soporte a la decisión y sistemas inteligentes. Disponible: <http://.netmedia.info/netmedia/articulos.php?idsec=32&idart=3680>

Turban, E. (1995). Decision support and expert systems: management support systems. Englewood Cliffs, N.J., Prentice Hall. ISBN 0-024-21702-6.

Turban, E. (2006); Aronson, Leidner, Dorothy; Mclean, Ephraim; Wetherbe, James. "Information Technology for Management". John Wiley & Sons, Inc. 5<sup>th</sup> Edition.

Wolf, C. (1999). Modelamiento Multidimensional. [Libro en línea]. Disponible: <http://www.inf.udec.cl/revista/edicion4/cwolff.htm>

## **Anexos**

### ***Anexo A. Misión, visión, valores y proceso.***

#### **Misión**

Brindar servicios de capacitación que satisfagan las necesidades reales del personal del INEGI y su desarrollo personal, así como promover acciones que impulsen el intercambio de conocimientos y experiencias en el ámbito nacional e internacional, que contribuyan al logro de los propósitos interinstitucionales.

#### **Visión**

Ser altamente competitivos; orientados a satisfacer las necesidades de capacitación de nuestros usuarios, reconocidos como líderes innovadores en este campo, a nivel nacional e internacional, mediante una administración efectiva en un ambiente de trabajo en equipo.

#### **Valores**

Aprendizaje, Compromiso, Comunicación, Honestidad, Justicia, Profesionalismo, Respeto, Responsabilidad, Servicio al cliente, Trabajo en equipo.

#### **Objetivo**

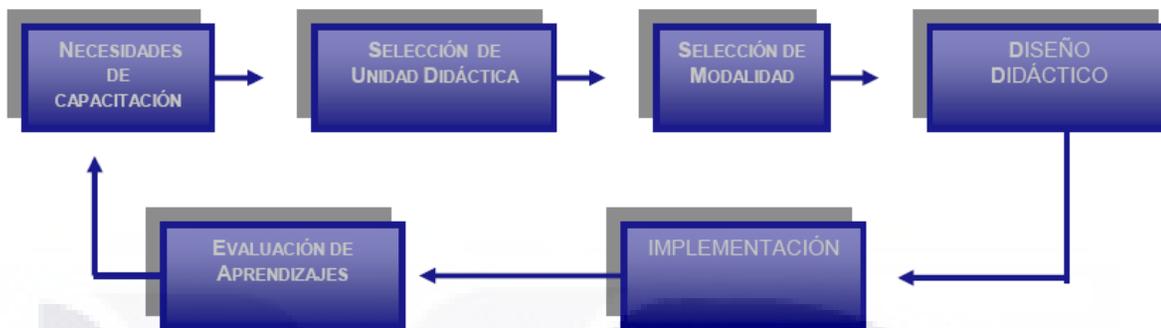
Formar, actualizar y especializar al personal del Instituto para contribuir a la profesionalización del capital humano del INEGI y al desarrollo de programas de investigación y nuevas metodologías en materia estadística y geográfica

#### **Estrategia**

Brindar servicios de capacitación de calidad que atiendan las diversas necesidades de conocimiento del INEGI bajo un enfoque integral y flexible.

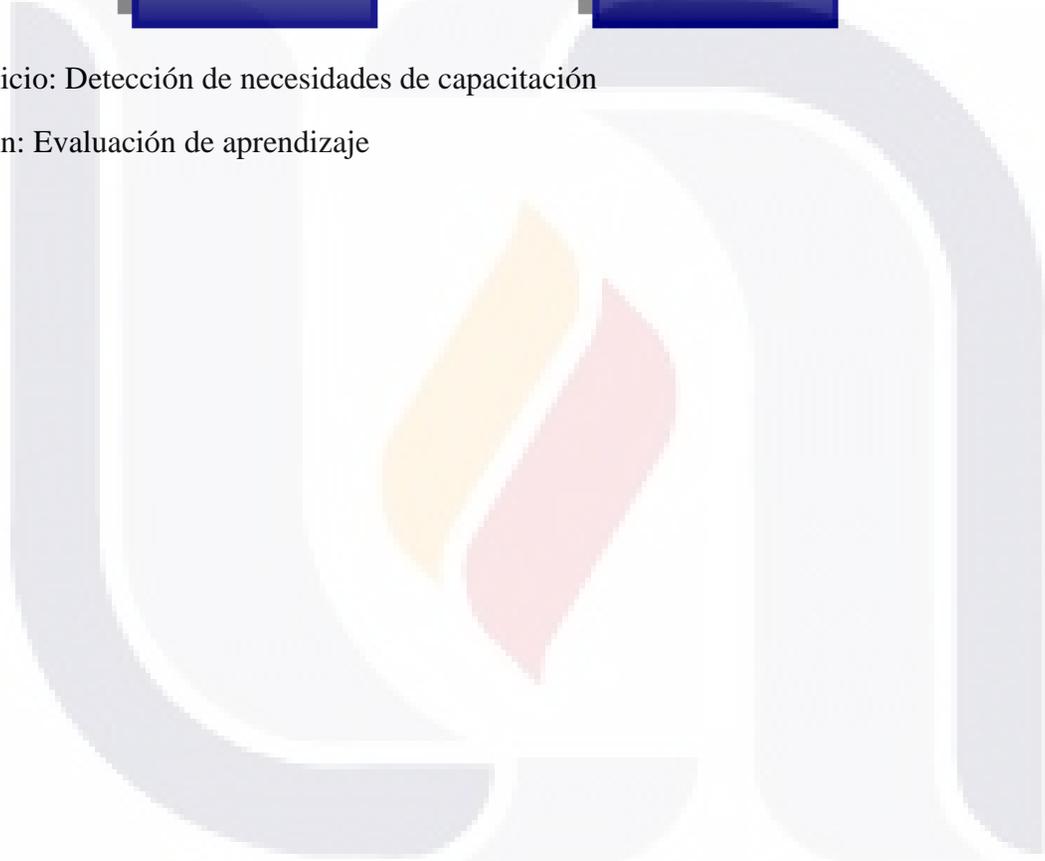
### Proceso

Nombre: Satisfacer necesidades de capacitación del INEGI.



Inicio: Detección de necesidades de capacitación

Fin: Evaluación de aprendizaje



## ***Anexo B. Proceso de registro en el sistema transaccional***

La Dirección de Capacitación realiza anualmente un levantamiento de necesidades de capacitación de las diferentes áreas del INEGI con la finalidad de formar su programa de trabajo.

El programa de trabajo contiene los cursos que se tiene que realizar durante el siguiente año al levantamiento, así como el número de personas que serán atendidos.

Una vez autorizado el programa de trabajo por el comité correspondiente se registran las necesidades de capacitación en un sistema de información llamado Sistema Automatizado de Capacitación (SAC).

El SAC es un sistema de información transaccional mediante el cual la Dirección de Capacitación administra la capacitación del INEGI a nivel nacional. En el SAC se registran:

- Catálogos de cursos
- Se definen las características del curso tales como el objetivo, temario, duración, etc.
- Eventos
- Son una instancia de los cursos, en los cuales se registra el lugar de donde se impartirá el curso, el instructor, periodo, horario, etc.
- Instructores
- Se registran las personas que han participado como instructores en algún curso de capacitación, así como el curso que pueden impartir.
- Participantes

- TESIS TESIS TESIS TESIS TESIS
- Contiene a las personas del Instituto que han recibido un curso de capacitación o que están programados a uno de ellos. Se conserva el historial de los cursos en los que ha participado desde 1997.

Para darle seguimiento al programa de trabajo, el Depto. de sistemas de información de capacitación realiza un reporte mensual el cual es entregado al director de capacitación y los enlaces que se tiene a nivel nacional (ver anexo D). Es decir, se requiere que pase un mes para conocer los resultados y en su caso poder tomar decisiones al respecto.

Anualmente se realiza un reporte similar al mensual salvo que éste contiene el acumulado de dicho año.

Recientemente se han realizados algunas vistas en la base de datos transaccional con el objetivo de que los enlaces y administradores de la capacitación a nivel nacional puedan realizar consultas a través de Excel para cubrir sus necesidades de información y puedan tomar decisiones en su ámbito correspondiente. Lo anterior ha permitido que se puedan tomar medidas de manera más oportuna y no tener que depender de oficinas centrales (Aguascalientes) para que la generación las consultas a la base de datos.

El departamento de sistema de información de capacitación, realiza consultas no planeadas a petición de los enlaces y administradores de la capacitación, tales como:

- Porcentaje de ocupación con el objetivo de identificar espacios disponibles y poder programar algún curso solicitado.
- Cursos vigentes para actualizarlos.
- Historial académico de alguna persona en particular.
- Identificar posibles instructores, etc.

## **Anexo C. Entrevistas**

**Nombre de la Compañía:** INEGI

**Nombre de la persona entrevistada:** Mario Alberto Rivera Díaz

**Nombre del área:** Dirección de Capacitación.

**Funciones principales del puesto:**

- **Proponer los procedimientos y estrategias generales para el desarrollo del Programa de Capacitación Institucional en Oficinas Centrales, Direcciones Regionales y Coordinaciones Estatales.**
- **Establecer líneas generales de orientación de la capacitación para los SNEIG.**
- **Coordinar la supervisión de la ejecución del programa de capacitación Institucional, así como establecer los mecanismos de evaluación de la capacitación en el Instituto.**
- **Someter a la autorización del Consejo Directivo de Capacitación el Programa Anual de Capacitación institucional.**
- **Establecer contactos con las áreas internas del INEGI para proponer estrategias y líneas de acción que permitan coordinar el programa de capacitación institucional, conforme a las metas y objetivos propuestos.**
- **Instrumentar mecanismos de intercambio en materia de capacitación con organismos gubernamentales e instituciones educativas y con agencias de estadística y geografía de otros países y organismos internacionales.**
- **Coordinar la administración general de las acciones inherentes a la ejecución del Programa de Capacitación Institucional.**

**Día y hora de la entrevista:** 9 de julio de 2007, de las 10:00 a las 11:30 hrs.

### **1. ¿Cuál es la misión de su área?**

Brindar servicios de capacitación que satisfagan las necesidades reales del personal del INEGI y su desarrollo personal, así como promover acciones que impulsen el intercambio de conocimientos y experiencias en el ámbito nacional e internacional, que contribuyan al logro de los propósitos interinstitucionales.

### **2. ¿Cuál es la visión de su área?**

Ser altamente competitivos; orientados a satisfacer las necesidades de capacitación de nuestros usuarios, reconocidos como líderes innovadores en este campo, a nivel nacional e internacional, mediante una administración efectiva en un ambiente de trabajo en equipo.

### 3. ¿A qué se dedica principalmente su área?

- a) La Dirección de Capacitación es responsable de elaborar, en coordinación con las Unidades Administrativas, una cartera de cursos en los temas de Estadística, Geografía, Informática, Difusión y Administración, en base a la cual las Direcciones Generales determinen cuáles de ellos deberán incorporarse al programa que se someterá al Consejo Directivo de Capacitación para su aprobación.
- b) La Dirección de Capacitación elabora esquemas para medir el impacto de la capacitación en el desarrollo de los procesos sustantivos y el resultado será dado a conocer a los mandos directivos y al Consejo Directivo de Capacitación.
- c) La Dirección de Capacitación es la responsable de la organización de cursos, talleres, maestrías, diplomados, etc., atendiendo los requerimientos de las áreas.
- d) La Dirección de Capacitación en conjunto con las áreas sustantivas elabora un programa de capacitación en el extranjero para personal del Instituto, que contemple financiamiento por parte de organismos internacionales.

### 4. ¿Tienen sistemas transaccionales que soporten las actividades que se realizan en el área?

La Dirección de Capacitación cuenta con un sistema transaccional llamado Sistema de Administración de la Capacitación (SAC). Este sistema se utiliza a nivel nacional por los administradores de la capacitación, mediante el cual se definen y actualiza lo siguiente:

- Catálogo de cursos
- Instructores
- Eventos
- Participantes
- Exámenes diagnósticos y de conocimiento
- Historial académico
- Capacidades del personal, etc.

En la intranet del Instituto se puede consultar los servicios de Historial académico y Programación de cursos del personal del INEGI, los cuales acceden a la base de datos transaccional de Capacitación.

**5. Estos sistemas transaccionales. ¿Dónde guardan la información que recaban, que formato tienen?**

La base de datos que se utiliza en el SAC es SQL server 2000 y se tiene planeado para finales de este año migrarla a SQL server 2005.

**6. ¿Cuándo usted requiere información referente a las actividades que realizan, cuanto se tarda en obtenerla?**

Existe un departamento de sistemas de información de capacitación, el cual es responsable del procesamiento de datos del área. Dicho departamento atiende a nivel nacional las peticiones tanto de las consultas planeadas y las que no lo son.

Este departamento genera mensualmente un reporte con la información solicitado por el director de capacitación y los administradores de la misma, bajo un formato preestablecido. Este mismo formato es aplicado al final del año para el cierre del programa de trabajo.

**7. ¿La información que maneja en su área es administrada?, en qué forma?**

La información es administrada por el departamento de sistemas de información de capacitación.

**8. ¿Considera que la información puede ser consultada en cualquier momento por cualquiera que así lo requiera?**

En este momento, solo puede ser consultada por las personas que conocen el modelo de la base de datos y tiene acceso al mismo.

**9. ¿Tiene algún plan para mejorar la consulta de información de su área?**

A principio de este año, se realizaron unas vistas en la base de datos transaccional con el objetivo de que los administradores de la capacitación a nivel nacional pudieran realizar sus consultas a

través de Excel y atendieran directamente sus necesidades de información para la toma de decisiones. Lo anterior se apoyo con una guía y cursos para la consulta la información.

**10. ¿Actualmente la actual administración de información le permiten tomar decisiones en el momento que así lo requiera?**

No directamente, sino mediante la solicitud de reportes al departamento correspondiente.

**11. ¿Qué diferentes perfiles tiene el personal de su área?**

Los perfiles del personal de la Dirección de Capacitación y los administradores de la misma a nivel nacional cuentan con un perfil de administración, pedagogía, estadística o geografía y pocos de ellos cuentan con un perfil informático.



## Anexo D. Reportes varios de capacitación

Eventos y Egresados programados y realizados por Unidad Ejecutora

Octubre

Programa de Capacitación alineado a

Capacidades

Clave	Unidad Ejecutora	Eventos				Participantes			
		Programado	Realizado	Porcentaje	Diferencia	Programado	Realizado	Porcentaje	Diferencia
A0	OFICINAS CENTRALES	69	8	11.59%	61	1035	92	8.89%	943
B0	DIR. REGIONAL NOROESTE	5	0	0.00%	5	75	0	0.00%	75
B1	Coord. Estatal Sonora	1	7	700.00%	-6	15	48	320.00%	-33
B2	Coord. Estatal Baja California Norte	1	0	0.00%	1	15	0	0.00%	15
B3	Coord. Estatal Sinaloa	0	0	N.A.	0	0	0	N.A.	0
B4	Coord. Estatal Baja California Sur	1	2	200.00%	-1	20	15	75.00%	5

Eventos cancelados ámbito y área de conocimiento

Octubre de 2006

Oficina	Eventos	Estado	Área	Causa
OFICINAS CENTRALES	4	Cancelado	Administración	Reestructuración del curso.
	4	Cancelado	Administración	Otra causa
	1	Cancelado	Estadística	Aula no disponible.
	5	Cancelado	Estadística	Instructor no disponible.
	3	Cancelado	Geografía	Otra causa
	4	Cancelado	Informática	Aula no disponible.
	3	Cancelado	Informática	Aula cedida a otro evento de mayor prioridad.
	2	Cancelado	Informática	Instructor no disponible.

Eventos y Egresados programados y realizados por Unidad

Ejecutora

Octubre de 2006

Clave	Unidad Ejecutora	Eventos								
		Programado	Realizado					Total	Porcentaje	Diferencia
			Del Programa	Adicional al Programa	Asociado a capacidades	No asociado a capacidades				
A0	OFICINAS CENTRALES	69	2	28	8	22	30	43.48%	39	
B0	DIR. REGIONAL NOROESTE	5	0	0	0	0	0	0.00%	5	
B1	Coord. Estatal Sonora	1	0	9	7	2	9	900.00%	-8	

Eventos y Egresados por

Ámbito

Octubre de 2006

Ámbito	Área de conocimiento	Número de cursos	Número de participantes
<b>OFICINAS CENTRALES</b>	Administración	6	78
	Estadística	4	84
	Geografía	5	73
	Informática	13	138
	Inglés	2	9
<b>OFICINAS ESTATALES</b>	Administración	14	161
	Estadística	7	66
	Geografía	23	174
	Informática	46	415
<b>OFICINAS REGIONALES</b>	Administración	15	196
	Estadística	1	23
	Geografía	13	117
	Informática	18	171
	Inglés	1	12
<b>TOTAL</b>		<b>168</b>	<b>1717</b>

Eventos y Egresados por Área de conocimiento y  
 Unidad Ejecutora  
 Octubre de 2006

Oficina	Área	Número de cursos	Número de participantes
<b>OFICINAS CENTRALES</b>	Administración	6	78
	Estadística	4	84
	Geografía	5	73
	Informática	13	138
	Inglés	2	9
<b>Coord. Estatal Sonora</b>	Estadística	1	7
	Informática	8	56
<b>Coord. Estatal Baja California Sur</b>	Administración	1	12
	Geografía	2	15
<b>DIR. REGIONAL NORTE</b>	Geografía	6	54
	Informática	2	17
<b>Coord. Estatal Durango</b>	Geografía	3	15
	Informática	1	12
<b>Coord. Estatal Chihuahua</b>	Administración	2	20
	Estadística	1	9
	Geografía	1	7
<b>Coord. Estatal Zacatecas</b>	Geografía	2	10
	Informática	1	4

### **Anexo E. Portafolio de análisis requeridos**

- Porcentaje de ocupación con el objetivo de identificar espacios disponibles y poder programar algún curso solicitado.
- Cursos vigentes para actualizarlos.
- Historial académico de alguna persona en particular.
- Identificar posibles instructores.
- Cursos que corresponden a las diferentes capacidades necesarias en el instituto.
- Identificar los cursos programados y cancelados
- Identificación de cursos por granuralidad de tiempo.
- Cantidad de cursos por área de conocimiento y modalidad
- Identificar cursos a los cuales les faltan manuales, guías didácticas y apoyos.
- Identificar cursos por ámbito o región.
- Cantidad de cursos de acuerdo al nivel de puesto de los trabajadores
- Identificar causas principales por las que se cancelan eventos.
- Identificar motivos principales por los que los participantes no asisten.

## Anexo F. Código de la aplicación del ETL

```
Protected Sub Page_Load(ByVal sender As
Object, ByVal e As System.EventArgs)
Handles Me.Load
    Borrar_Hechos()
    Llenar_EstadoParticipante()
    Llenar_Oficina()
    Llenar_eventos()
    Llenar_edo_eventos()
    Llenar_tipo_evento()
    Llenar_Reconocimiento()
    Llenar_Area()
    Llenar_Programa()
    Importar_Participantes()
    Importar_Cursos()
End Sub

Sub Llenar_Programa()
    Dim sql As String
    Dim ds As Data.DataSet
    sql = "DELETE FROM d_programa"
    clase.sql2005(sql)
    sql = "select * from tc_programas"
    ds = clase.sqlds(sql)
    Dim registro As DataRow
    For Each registro In
ds.Tables(0).Rows
        sql = "insert into d_programa
(id_programa,programa) " & _
            " values (' &
registro.Item("cve_programa") & "',' &
registro.Item("programa") & "'"
            clase.sql2005(sql)
    Next
End Sub

Sub Llenar_Area()
    Dim sql As String
    Dim ds As Data.DataSet
    sql = "DELETE FROM d_area"
    clase.sql2005(sql)
    sql = "select * from tc_areas "
    ds = clase.sqlds(sql)
    Dim registro As DataRow
    For Each registro In
ds.Tables(0).Rows
        sql = "insert into d_area
(id_area,area,sustantiva) " & _
            " values (' &
registro.Item("cve_area") & "',' &
registro.Item("area") & "',' &
registro.Item("sustantiva") & "'"
            clase.sql2005(sql)
    Next
End Sub

Sub Llenar_Reconocimiento()
    Dim sql As String
    Dim ds As Data.DataSet
    sql = "DELETE FROM
d_reconocimiento"
    clase.sql2005(sql)
    sql = "select * from
tc_reconocimientos"
    ds = clase.sqlds(sql)
    Dim registro As DataRow
    For Each registro In
ds.Tables(0).Rows
        sql = "insert into
d_reconocimiento
(id_reconocimiento,reconocimiento) " & _
            " values (' &
registro.Item("cve_reconocimiento") &
 "',' & registro.Item("reconocimiento") &
 "'"
            clase.sql2005(sql)
    Next
End Sub

Sub Llenar_tipo_evento()
    Dim sql As String
    Dim ds As Data.DataSet
    sql = "DELETE FROM d_tipo_evento"
    clase.sql2005(sql)
    sql = "select * from
tc_tipos_evento"
    ds = clase.sqlds(sql)
    Dim registro As DataRow
    For Each registro In
ds.Tables(0).Rows
```

```

        sql = "insert into
d_tipo_evento (id_tipo_evento,tipo_evento)
" & _
        " values ('" &
registro.Item("cve_tipo_evento") & "','" &
registro.Item("tipo_evento") & "')"
        clase.sql2005(sql)
    Next
End Sub
Sub Llenar_edo_eventos()
    Dim sql As String
    Dim ds As Data.DataSet
    sql = "DELETE FROM d_edo_evento"
    clase.sql2005(sql)
    sql = "select * from
tc_edo_eventos"
    ds = clase.sqlds(sql)
    Dim registro As DataRow
    For Each registro In
ds.Tables(0).Rows
        sql = "insert into
d_edo_evento (id_edo_evento,estado) values
('" & registro.Item("cve_edo_evento") &
',''" & registro.Item("edo_evento") & "')"
        clase.sql2005(sql)
    Next
End Sub
Sub Llenar_eventos()
    Dim sql As String
    Dim ds As Data.DataSet
    sql = "DELETE FROM d_evento"
    clase.sql2005(sql)
    sql = "select * from veventos"
    ds = clase.sqlds(sql)
    Dim registro As DataRow
    Dim id, evento As String
    For Each registro In
ds.Tables(0).Rows
        id =
Trim(registro.Item("cve_evento"))
        evento =
registro.Item("evento")
        sql = "insert into d_evento
(id_evento,evento) values ('" & id & "','"
& evento & "')"
        clase.sql2005(sql)
    Next
End Sub
Sub Llenar_estadoParticipante()
    Dim sql As String
    Dim ds As Data.DataSet
    sql = "DELETE FROM
d_estadoParticipante"
    clase.sql2005(sql)
    sql = "select * from
tc_edo_participantes"
    ds = clase.sqlds(sql)
    Dim registro As DataRow
    Dim id As Integer
    Dim estado As String
    For Each registro In
ds.Tables(0).Rows
        id =
Int(registro.Item("cve_edo_participante"))
        estado =
registro.Item("edo_participante")
        sql = "insert into
d_estadoParticipante (id_estado,estado)
values (" & id & "','" & estado & "')"
        clase.sql2005(sql)
    Next
End Sub
Sub Llenar_Oficina()
    Dim sql As String
    Dim ds As Data.DataSet
    sql = "DELETE FROM d_oficina"
    clase.sql2005(sql)
    sql = "select * from tc_oficinas"
    ds = clase.sqlds(sql)
    Dim registro As DataRow
    Dim id, oficina, ciudad As String
    For Each registro In
ds.Tables(0).Rows
        id =
registro.Item("cve_oficina")
        oficina =
registro.Item("oficina")
        ciudad =
registro.Item("ciudad")
        sql = "insert into d_Oficina
(id_oficina,oficina,ciudad) values ('" &
id & "','" & oficina & "','" & ciudad &
"')"
        clase.sql2005(sql)
    Next
End Sub

```

```

Next
End Sub

Sub Importar_Participantes()
    Dim sql As String
    Dim ds As Data.DataSet
    sql = "select a.*, b.inicio from
vparticipantes a, veventos b where
a.cve_evento=b.cve_evento"
    ds = clase.sqllds(sql)
    Dim registro As DataRow
    Dim anio, estado As Integer
    Dim mes, evento, oficina As String
    For Each registro In
ds.Tables(0).Rows
        anio =
Year(registro.Item("inicio"))
        mes =
Month(registro.Item("inicio"))
        evento =
Trim(registro.Item("cve_evento"))
        oficina =
registro.Item("cve_oficina")
        estado =
Int(registro.Item("cve_edo_participante"))
        sql = "insert into
F_participante " & _
        "
(id_estado,id_evento,id_oficina,id_anio,id
_mes) " & _
        "values (" & estado & ",'" &
evento & "','" & oficina & "'," & anio &
"','" & mes & "')"
        clase.sql2005(sql)
    Next
End Sub

Sub Importar_Cursos()
    Dim sql As String
    Dim ds As Data.DataSet
    sql = "select * from veventos"
    ds = clase.sqllds(sql)
    Dim registro As DataRow
    For Each registro In
ds.Tables(0).Rows
        sql = "insert into F_curso " &
        "
(id_edo_evento,curso,id_tipo_evento,id_rec
onocimiento,id_area,id_oficina,id_anio,id_
_mes,id_programa,duracion) " & _
        "values ('" &
registro.Item("cve_edo_evento") & "','" &
registro.Item("evento") & _
        "','" &
registro.Item("cve_tipo_evento") & "','" &
registro.Item("cve_reconocimiento") & _
        "','" &
registro.Item("cve_area") & "','" &
registro.Item("cve_oficina") & "','" &
Year(registro.Item("inicio")) & "','" &
Month(registro.Item("inicio")) & _
        "','" &
registro.Item("cve_programa") & "','" &
registro.Item("duracion") & "')"
        clase.sql2005(sql)
    Next
End Sub

Sub Borrar_Hechos()
    Dim sql As String
    sql = "DELETE FROM F_curso"
    clase.sql2005(sql)
    sql = "DELETE FROM F_participante"
    clase.sql2005(sql)
End Sub
Sub

```

## Anexo G. Código MDX

### Procesar cubo Cursos

```
<Batch ProcessAffectedObjects="true" xmlns="http://schemas.microsoft.com/analysisservices/2003/engine">
  <Parallel>
    <Process xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:dll2="http://schemas.microsoft.com/analysisservices/2003/engine/2"
xmlns:dll2_2="http://schemas.microsoft.com/analysisservices/2003/engine/2/2">
      <Object>
        <DatabaseID>Analysis Services Project2</DatabaseID>
        <CubeID>Cursos</CubeID>
      </Object>
      <Type>ProcessFull</Type>
      <WriteBackTableCreation>UseExisting</WriteBackTableCreation>
    </Process>
  </Parallel>
</Batch>
```

### Procesar dimensión Evento

```
<Batch ProcessAffectedObjects="true" xmlns="http://schemas.microsoft.com/analysisservices/2003/engine">
  <Parallel>
    <Process xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:dll2="http://schemas.microsoft.com/analysisservices/2003/engine/2"
xmlns:dll2_2="http://schemas.microsoft.com/analysisservices/2003/engine/2/2">
      <Object>
        <DatabaseID>Analysis Services Project2</DatabaseID>
        <DimensionID>D Evento</DimensionID>
      </Object>
      <Type>ProcessUpdate</Type>
      <WriteBackTableCreation>UseExisting</WriteBackTableCreation>
    </Process>
  </Parallel>
</Batch>
```

### Procesar cubo Participante

```
<Batch ProcessAffectedObjects="true" xmlns="http://schemas.microsoft.com/analysisservices/2003/engine">
  <Parallel>
    <Process xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:dll2="http://schemas.microsoft.com/analysisservices/2003/engine/2"
xmlns:dll2_2="http://schemas.microsoft.com/analysisservices/2003/engine/2/2">
      <Object>
        <DatabaseID>Analysis Services Project2</DatabaseID>
```

```
<CubeID>Prototipo</CubeID>  
</Object>  
<Type>ProcessFull</Type>  
<WriteBackTableCreation>UseExisting</WriteBackTableCreation>  
</Process>  
</Parallel>  
</Batch>
```

