

**MODELO PARA LA DEFINICION DE LA ARQUITECTURA
DE DATOS DE LA EMPRESA**



Tesis presentada

por

SILVIA IVETTE AGUILAR MACIAS

Presentada ante la Dirección Académica de la
Universidad Virtual del Instituto Tecnológico
Y de Estudios Superiores de Monterrey

**MAESTRIA EN ADMINISTRACION DE TECNOLOGIAS DE
INFORMACION**

DICIEMBRE DE 1999

MODELO PARA LA DEFINICIÓN DE LA ARQUITECTURA DE DATOS DE LA EMPRESA

Tesis presentada
por

SILVIA IVETTE AGUILAR MACÍAS

Presentada ante la Dirección Académica de la
Universidad Virtual del
Instituto Tecnológico y de Estudios Superiores de
Monterrey
como requisito parcial para optar al título de

**MAESTRO EN ADMINISTRACIÓN DE
TECNOLOGÍAS DE INFORMACIÓN**

Asesor: Dr. Edgar Emmanuel Vallejo Clemente

Sinodales Lic. Horacio Echeverría Cabrera
Dr. Guillermo Rodríguez Abitia

Jurado: Dr. Edgar Emmanuel Vallejo Clemente
Lic. Horacio Echeverría Cabrera
Dr. Guillermo Rodríguez Abitia

Diciembre de 1999
Maestría en Administración de Tecnologías de
Información

DEDICATORIA

Principalmente a mis padres, que me apoyaron a terminar una etapa más en mi vida. Gracias por estar ahí todas esas veces que conté con su incondicional apoyo.

Al MCC. Héctor Aguilar Macías su apoyo incondicional y porque constantemente me enseña las bases del éxito y la excelencia a seguir en la vida.

AGRADECIMIENTOS

Deseo agradecer principalmente a mis padres por su apoyo incondicional y a mi hermano MCC. Héctor Aguilar Macías su apoyo incondicional y porque constantemente me enseña las bases del éxito y la excelencia a seguir en la vida.

Al MTIA. Daniel Peñaloza Moreno por su incondicional y constante apoyo además de sus invaluable enseñanzas de las cuales aprendí tanto a lo largo de la maestría y que a lo largo de mi vida tendré presentes.

Al Ing. Mauricio García Tiscareño, a la Lic. Marcela San Roman Manterola, al Ing. Jorge Eliel Zarate Vázquez y al Lic Marco Tulio Martínez González, por su importante apoyo y empuje profesional además del permitirme ilustrarme con sus enseñanzas durante mis estudios de maestría.

Al Lic. Jose Antonio Delgado Selley por su apoyo profesional a lo largo del desempeño de la maestría.

A mi asesor, Dr. Edgar Emmanuel Vallejo Clemente y a mis sinodales M.A. Juan Norberto Torres Valencia, al Dr. Guillermo Rodríguez Abitia y al Lic. Horacio Echeverría Cabrera por su paciencia, constante apoyo y dirección en la elaboración de esta tesis.

Al Instituto Tecnológico y de Estudios Superiores de Monterrey por sus enseñanzas y su preocupación en la formación de profesionales con alto espíritu de trabajo y de servicio.

RESUMEN

MODELO PARA LA DEFINICIÓN DE LA ARQUITECTURA DE DATOS DE LA EMPRESA

DICIEMBRE DE 1999

SILVIA IVETTE AGUILAR MACÍAS

**LICENCIADA EN SISTEMAS DE COMPUTACIÓN
ADMINISTRATIVA**

**INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE
MONTERREY
CAMPUS ESTADO DE MÉXICO**

**MAESTRO EN ADMINISTRACIÓN DE TECNOLOGÍAS DE
INFORMACIÓN**

**INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE
MONTERREY
CAMPUS ESTADO DE MÉXICO**

Actualmente el contar con datos que conforman información cuando se necesita, donde se necesita y en un formato utilizable, es un factor crítico de éxito para la competitividad de la empresa. El atender estos requerimientos, forma parte de la misión del departamento de Sistemas de Información. La calidad de datos debe planearse y la arquitectura de datos es un producto de la planeación de los datos y como cumplimiento de la misión del área de Sistemas de Información. Es de crucial importancia que los datos posean una organización común que elimine la redundancia y asegure la consistencia de los datos. Entre más completa sea la

directriz al aspecto de los datos de un sistema en la etapa de diseño, se incrementa la producción de sistemas estables.

Debido a esto, resulta relevante realizar una investigación para desarrollar un modelo de apoyo a los analistas y diseñadores de sistemas a definir la arquitectura de datos óptima, que permita dentro de la organización, flexibilidad ante los cambios y necesidades de la misma, asegurando la integridad de los datos a través de la organización a un bajo costo.

Basándonos en la premisa de que los métodos tradicionales de planificación de sistemas para la definición de la arquitectura tecnológica en una empresa, no cubren con oportunidad las necesidades de las organizaciones hoy en día, debido a que inician con la determinación de hardware, para continuar con las aplicaciones que pueden ejecutarse por dicha tecnología y finalizar con el aspecto datos; incrementando costos tanto cuantitativos como cualitativos en la obtención de información intrínseca a la competitividad de la empresa.

El desarrollo de esta tesis, busca proporcionar a las empresas un modelo para la determinación de su arquitectura de datos base como cimiento de la arquitectura tecnológica, que permita la optimización de la modelación de sus datos, buscando integridad y calidad de estos, diferenciando la información necesaria para los tres niveles de la organización: estratégico, táctico y operativo, proporcionando a cada uno información que se requiere, en el momento que se requiere como cimientos de su arquitectura tecnológica.

ÍNDICE DE CONTENIDO

DEDICATORIA	XXXII
AGRADECIMIENTOS	XXXII
RESUMEN	XXXII
ÍNDICE DE CONTENIDO	XXXII
ÍNDICE DE ILUSTRACIONES	XXXII
INTRODUCCIÓN	32
<i>Objetivo y Producto Final de la Tesis</i>	32
<i>Organización del Documento</i>	32
Capítulo 1 Evolución del Manejo de Datos	32
Capítulo 2 Necesidad de una arquitectura de datos.	32
Capítulo 3 Fundamentos de Modelación de Datos.	32
Capítulo 4 Modelo Propuesto para la Definición de la Arquitectura de Datos.....	32
CAPÍTULO 1	32
EVOLUCIÓN DEL MANEJO DE DATOS.....	32
<i>Datos dentro de los programas.</i>	32
<i>Datos fuera de los programas.</i>	32
<i>Sistemas Manejadores de Bases de Datos.</i>	32
<i>El surgimiento del Modelo Relacional</i>	32
<i>Integración de Información a través de Almacén de Datos (Data Warehousing)</i>	32
<i>Minería de Datos (Data Mining)</i>	32
CAPÍTULO 2	32
NECESIDAD DE UNA ARQUITECTURA DE DATOS.....	32
<i>¿Por qué contar con una Arquitectura de Datos?</i>	32
<i>El Posicionamiento y la Información</i>	32
<i>La Misión de los Sistemas de Información en la Organización</i>	32
<i>Calidad de Datos</i>	32
Reingeniería	32
Downsizing	32
Outsourcing	32
Administración de la Calidad Total (TQM Total Quality Management)	32
Integración de Bases de Datos.....	32
Planeación estratégica de Datos e Ingeniería de Información.....	32
CAPÍTULO 3	32
FUDAMENTOS DE MODELACIÓN DE DATOS.....	32
<i>Planeación Estratégica de Bases de Datos.</i>	32
<i>La línea de trabajo de Zachman</i>	32
<i>El Metalenguaje como apoyo a la Arquitectura de Datos</i>	32
<i>Metalenguaje KIF</i>	32
Nivel Carácter	32
Nivel Lexema	32
Nivel Expresiones	32
<i>Metalenguaje CES</i>	32
<i>Reingeniería Automatizada de Programas y Datos a Nivel Empresa Xinotech</i>	32
Acerca de Xinotech.....	32
Alcance de la Tecnología.....	32

Extracción de Modelos a Nivel Empresa	32
Administración Automatizada del Proceso de Reingeniería	32
Transformación Automatizada de Programas y Datos.....	32
Resolución de Problemas	32
Typel y el Año 2000	32
Interfaz del Usuario.....	32
CAPÍTULO 4	32
MODELO PROPUESTO PARA LA DEFINICIÓN DE LA ARQUITECTURA DE DATOS	32
<i>Modelo Propuesto</i>	32
Fase Planeación	32
Fase Modelación del Negocio.....	32
Fase Análisis Sistemas y Tecnologías Actuales.....	32
Fase Definición de Arquitectura de datos.	32
Fase Implantación.....	32
<i>Investigación de Campo</i>	32
Personal Entrevistado.....	32
Visión.....	32
Organización.....	32
<i>Conclusiones</i>	32
REFERENCIAS BIBLIOGRÁFICAS.....	32
CURRICULUM VITAE.....	32

ÍNDICE DE ILUSTRACIONES

FIGURA 1.1 DATOS COMPARTIDOS CON MÚLTIPLES SUBROUTINAS	32
FIGURA 1.2 DATOS LOCALES DENTRO DE SUBROUTINAS.....	32
FIGURA 1.3 UN PROGRAMA ACCEDIENDO UN ARCHIVO.....	32
FIGURA 1.4 COMPARTIR DATOS EN UNA BASE DE DATOS.	32
FIGURA 1.5 MODELO JERÁRQUICO DE BASES DE DATOS.	32
FIGURA 1.6 MODELO RELACIONAL DE BASES DE DATOS.....	32
FIGURA 1.7 ALMACÉN DE DATOS (DATA WAREHOUSE).....	32
FIGURA 1.8 ALMACÉN DE DATOS ENTRE DATOS PRE MINADOS Y MINADOS.....	32
FIGURA 1.9 LOS CUATRO ESPACIOS QUE FORMAN LA BASE DEL SOPORTE A DECISIONES.....	32
FIGURA 2.1 SISTEMAS INDEPENDIENTE CON DATOS DUPLICADOS E INCOMPATIBLES.	32
FIGURA 2.2. EL ÉNFASIS ORGANIZACIONAL ES CAMBIANTE	32
FIGURA 2.3 DATOS DEL ANÁLISIS DE LA SITUACIÓN	32
FIGURA 2.4 FACTORES DE ÉXITO QUE CREAN LA MISIÓN DE SISTEMAS DE INFORMACIÓN	32
FIGURA 2.5 CALIDAD DE DATOS	32
FIGURA 2.6 PLANEACIÓN ESTRATÉGICA DE DATOS E INGENIERÍA DE INFORMACIÓN.....	32
FIGURA 3.1 ANÁLOGÍA DEL OBJETIVO DE LA PLANEACIÓN ESTRATÉGICA DE BASES DE DATOS.	32
FIGURA 3.2 EJEMPLO DE MODELO DE DATOS GENÉRICO ORGANIZACIONAL	32
FIGURA 3.3 FASES DE LA PLANEACIÓN ESTRATÉGICA DE BASES DE DATOS.	32
FIGURA 3.4 PUNTOS DE OPORTUNIDAD TECNOLÓGICA	32
FIGURA 3.5 LÍNEA DE TRABAJO DE ZACHMAN	32
FIGURA 3.6 EJEMPLO DE METALENGUAJE CES EN BASES DE DATOS.....	32
FIGURA 4.1 LA DEFINICIÓN DE UNA ARQUITECTURA DE DATOS.	32
FIGURA 4.2 NIVELES EN LA ORGANIZACIÓN.....	32
FIGURA 4.3 PROCESOS EJEMPLO EN CADA UNO DE LOS NIVELES DE LA ORGANIZACIÓN.....	32
FIGURA 4.4. EJEMPLO DE NIVELES EN UNA ORGANIZACIÓN.....	32
FIGURA 4.5 EJEMPLO DE INFORMACIÓN EN CADA NIVEL DE LA ORGANIZACIÓN	32

INTRODUCCIÓN

Hoy en día el trabajo de todos, requiere de datos y el acceder a estos cada vez es más frecuente. Es común encontrar individuos que utilizan gran parte de su tiempo manipulando información, datos. Debido a esto, el tener la posibilidad de obtener datos cuando se necesita, donde se necesita y en un formato utilizable, es crítico. Entendemos por formato utilizable, cuando los datos pueden ser interpretados en información y no son datos irrelevantes. En pocas palabras, el acceder a datos es el requerimiento número uno para el logro de las metas del negocio.

Conforme una organización crece y se torna más compleja, la administración demanda más de la función de Sistemas de Información. Requieren acceder rápidamente a los datos cuando y en donde sea necesario, en un formato accesible, que pueda ser fácilmente interpretado, con datos correctos y consistentes, a través de cada departamento, maleables al cambio rápido y constante de las condiciones del negocio y compartidos en toda la organización. Estos requerimientos forman parte de la misión del departamento de Sistemas de Información. Proveer datos de calidad a aquellos que lo necesitan. Sin embargo, la calidad de los datos no emerge de la nada, tampoco resulta de enfocarse en la productividad en el desarrollo de aplicaciones. La calidad debe planearse. La arquitectura de datos es un producto de la planeación de la calidad de los datos y el cumplimiento de la misión de Sistemas de Información.

Uno de los puntos primordiales, es la participación de la administración, ya que provee una perspectiva del negocio y credibilidad en el proceso de planeación. Esto es de suma importancia, ya que esta intervención es nivelada conforme se maneja el negocio o los datos, debido a que:

- Un modelo estable (independiente de las fronteras organizacionales, sistemas y procedimientos) es el cimiento de la arquitectura.
- Los datos son definidos antes que las aplicaciones.
- La dependencia de los datos determina la secuencia de implantación de los sistemas aplicativos.

Para prosperar en una economía mundial, los negocios deben de tener la posibilidad de ajustarse y adaptarse a los cambios rápidos. Los ejecutivos requieren de Sistemas de Información que soporten al negocio conforme este cambia. En un ambiente dinámicamente cambiante, los usuarios no pueden permitirse el esperar seis meses, un año o más a que Sistemas de Información responda formalmente a sus peticiones. Las bases de datos deben de ser flexibles y de fácil mantenimiento, para que rápidamente junto con las aplicaciones, permitan ajustarse a los cambios que afecten de manera directa o indirecta a la empresa.

Actualmente los ejecutivos desean y esperan que los datos que reciben sean correctos y consistentes. Los datos deben mantener una integridad completa. No sólo deben de ser correctos dentro de una precisión aceptable, sino consistentes a través de la organización. Para que los datos sean propiamente interpretados y combinados de distintas partes de la organización, es necesario un vocabulario

común y una estandarización de datos. Las diferencias e inconsistencias de significados de datos en la empresa deben eliminarse. Los datos provistos, deben de ser correctos y actualizados. Los ejecutivos reconocen que los datos deben ser compartidos a través de la organización en orden de alcanzar exitosamente las metas de la misma. Los datos también deben compartirse entre los diferentes departamentos y las unidades de negocio, para que esto suceda, los datos deben de existir, mantenerse, administrarse y coordinarse, en forma centralizada. Esto no implica una base de datos, aislada y no distribuida. En su lugar, es de crucial importancia que los datos posean una organización común que elimine la redundancia y asegure la consistencia de los datos. Los ejecutivos buscan: el acceder a datos en formatos utilizables cuando y donde se requiera, la habilidad de adaptarse a los cambios y necesidades del negocio, datos correctos y consistentes y compartir estos datos dentro de la organización, sin elevar los costos de los Sistemas de Información. Los presupuestos de varios dígitos, no son tolerables por mucho tiempo. Los datos deben de ser provistos a un costo accesible y razonable.

Objetivo y Producto Final de la Tesis

Por lo antes expuesto, resulta relevante realizar una investigación para desarrollar un modelo de apoyo a los analistas y diseñadores de sistemas a definir la arquitectura de datos óptima, que permita dentro de la organización, flexibilidad ante los cambios y necesidades de la misma, asegurando la integridad de los datos a través de la organización a un bajo costo.

Organización del Documento

Capítulo 1 Evolución del Manejo de Datos

En el capítulo 1 encontraremos la evolución del manejo de datos iniciando con el manejo de datos globales utilizados dentro de los programas en la programación estructurada, para continuar con datos almacenados en archivos para ser utilizados fuera de los programas y tener un entendimiento del surgimiento de los Sistemas Manejadores de Bases de Datos, contemplando el modelo jerárquico y el surgimiento del modelo relacional, con la integración de información a través de almacenes de datos., finalizando con la minería de datos

Capítulo 2 Necesidad de una arquitectura de datos.

En el capítulo 2 analizamos las razones de contar con una arquitectura de datos en la organización, qué es lo que espera la organización de una arquitectura de datos y como cumple esta con los puntos de la misión del área de Sistemas de Información. Las estrategias que las organizaciones han considerado en los últimos años con el objetivo de encontrar calidad de datos y que significa el obtener calidad de datos.

Capítulo 3 Fundamentos de Modelación de Datos.

En este capítulo analizamos la arquitectura de datos de la organización como resultado objetivo de la planeación estratégica de datos y presentamos las fases que comprende esta planeación, según William H. Inmon [*Data Architecture: The*

Information Paradigm, 1992]. Con el objetivo de entender los diferentes puntos de vista de la organización en el desarrollo de arquitecturas de información, presentamos la línea de trabajo de Zachman [1987]. Por último, el metalenguaje como herramienta de estandarización para la definición de la arquitectura de datos de la organización, analizando KIF, CES y la tecnología Xinotech.

Capítulo 4 Modelo Propuesto para la Definición de la Arquitectura de Datos.

En este último capítulo presentamos las fases que integran el modelo propuesto resultado de la investigación de esta tesis, incluyendo una explicación de cada una de las fases y que aspectos se deben cubrir para la obtención de los productos finales de cada fase, haciendo referencia a los capítulos anteriores. Por último se incluyen los resultados de la investigación de campo realizada con el objetivo de validación del modelo propuesto y finalizando este capítulo con las conclusiones de esta tesis.

CAPÍTULO 1

EVOLUCIÓN DEL MANEJO DE DATOS

La mayoría de los esfuerzos para mejorar el desarrollo de software, se han enfocado en la modularización de procedimientos. Pero existe otro aspecto en el software que mientras es menos obvio, no es menos importante. Este punto lo conforman los datos, la colección de información operada por los procedimientos. Conforme las técnicas de programación modular han evolucionado a través de los años, también se ha incrementado aparentemente que los datos también sean modulares.

Datos dentro de los programas.

Sí un programa requiere solo algunos datos para desempeñar sus tareas, estos datos pueden permanecer de manera segura, disponibles para todas las diferentes subrutinas que constituyen el programa. Este orden es muy conveniente para los programadores, debido a que una colección de datos compartidos, típicamente llamado datos globales, provee una tabla boletín comunal, de la cual las diferentes subrutinas pueden intercambiar información cuando necesiten comunicarla. [Taylor]

Cuándo el número de datos se encuentra en el rango de cientos o miles, esta solución simple, usualmente lleva a errores misteriosos y a un comportamiento impredecible. El problema es qué, compartir datos es una violación a la

programación modular, que requiere que los módulos sean tan independientes como sea posible. Permitir que los módulos interactúen libremente a través de datos compartidos, provoca que las acciones de cualquier otro módulo dependan del comportamiento de todos los demás. En efecto, los datos globales se convierten en grietas de la armonía que la programación estructurada a creado entre las subrutinas.

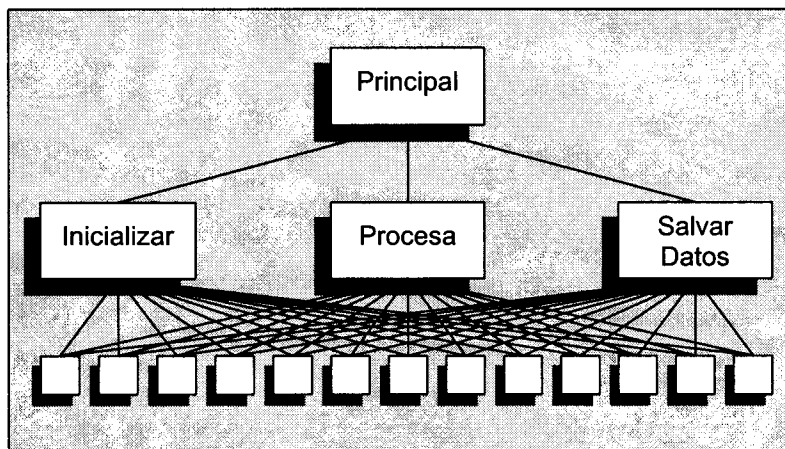


Figura 1.1 Datos compartidos con múltiples subrutinas

La solución a este problema es modularizar los datos a lo largo con los procedimientos. Esto se realiza típicamente, dando a cada subrutina su propio almacenamiento de datos local, de manera que por sí solo logre leer o escribir. Esta estrategia de ocultar información, minimiza interacciones no deseadas entre subrutinas y permite que su diseño y mantenimiento sea más independiente.

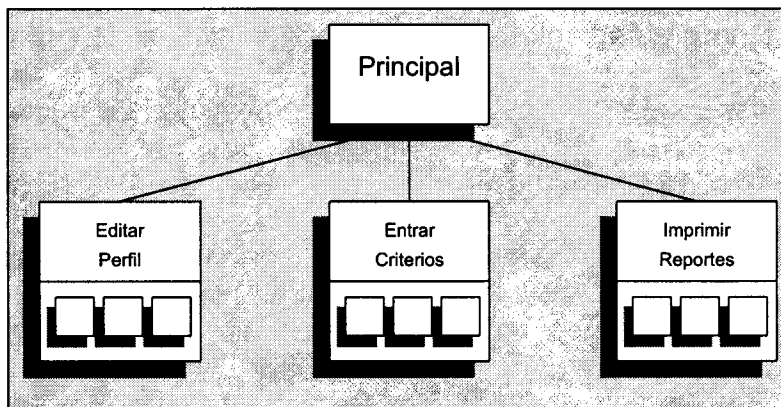


Figura 1.2 Datos locales dentro de subrutinas

Datos fuera de los programas.

Los programas pequeños a menudo requieren solo de algunas entradas y generan salidas que están hechas para ser consumidas inmediatamente. Por ejemplo, un programa para calcular las tablas de amortización, posiblemente acepte un valor base y un periodo de amortización del teclado, para después imprimir una página de cálculos. Los programas de este tipo, no necesitan almacenar datos debido a que trabajan con información fresca cada vez que se ejecutan.

Sin embargo, los programas más grandes, usualmente trabajan con la misma información una y otra vez. Los programas para control de inventarios, sistemas de contabilidad y herramientas de diseño de ingeniería, no pueden funcionar si no cuentan con una manera de preservar la información de una sesión a la siguiente.

La solución más simple a este problema de almacenar información, es contar con un programa que almacene su información en un archivo externo. Cuando el programa finaliza su ejecución, envía los datos a un archivo externo. Cuando se configura el programa nuevamente, recupera la información de un archivo. El uso de un archivo, también permite al programa trabajar con más información que internamente podría mantener, leyendo y escribiendo una pequeña porción del archivo en cualquier momento.

Los archivos externos de datos proveen una solución adecuada para almacenamiento de información, tan larga como los datos, si sólo es accedida por un usuario a la vez. Cuando los datos necesitan ser compartidos entre personas, programas o ambos, surgen nuevos problemas.

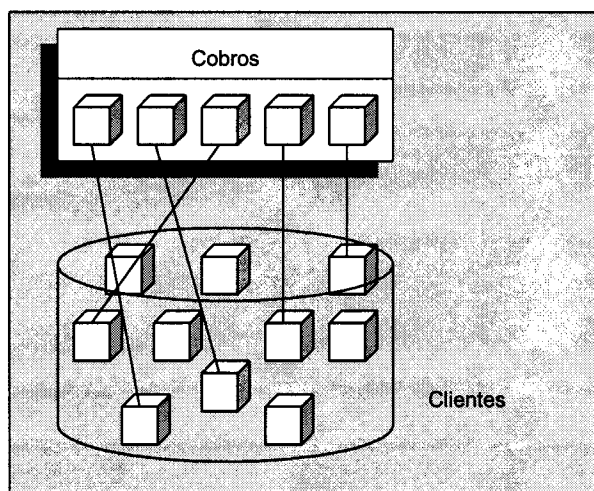


Figura 1.3 Un Programa accediendo un archivo

Sistemas Manejadores de Bases de Datos.

La mayoría de los sistemas basados en archivos no permiten que más de una persona acceda el archivo al mismo tiempo. Permitir el acceder a diferentes personas simultáneamente, presenta la posibilidad de que una persona modifique la información que otros están actualmente utilizando. Prevenir estas confusiones resulta ser un problema técnico bastante difícil, que no fácilmente se soluciona dentro de un sistema basado en archivos. A pesar de que algunos programas viejos continúen utilizando archivos para almacenar información compartida, llamados sistemas manejadores de bases de datos (database management systems (DBMS's)), que son diseñados para manejar simultáneamente el acceder a datos compartidos. [*Object Technology: A Manager's Guide*]

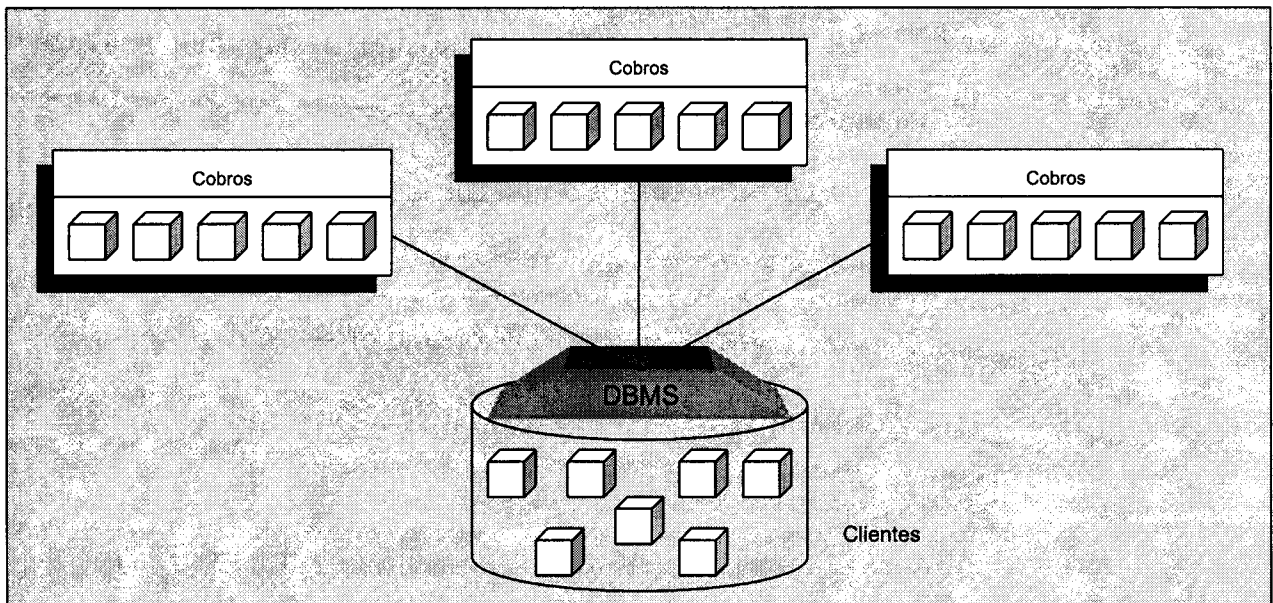


Figura 1.4 Compartir datos en una base de datos.

Los programas de los manejadores de bases de datos hacen más que sólo controlar el acceder a datos almacenados en archivos, también almacenan las relaciones entre los diferentes elementos de datos. El entendimiento de cómo la generación de bases de datos va y viene en un ambiente corporativo, es crítico debido a la manera en que se almacenan estas relaciones. Los primeros DBMS's adoptaron el acercamiento de almacenar referencias directas entre los elementos de datos, permitiendo a los datos ser recuperados a través de un proceso llamado acceso de navegación. Este acercamiento soporta extremadamente rápida recuperación de información relacionada, debido a que cada pieza de información almacenada incluye la localización de toda la información relacionada.

Las primeras formas de manejadores de bases de datos conocidas como modelos jerárquicos, representan elementos de datos, llamados registros, en estructuras de árbol. Por ejemplo, un departamento puede incluir registros de los puestos que contiene y del equipo. Cada posición en turno, puede ser asociada con una lista de responsabilidades y una lista de empleados en el departamento que mantienen un cierto puesto.

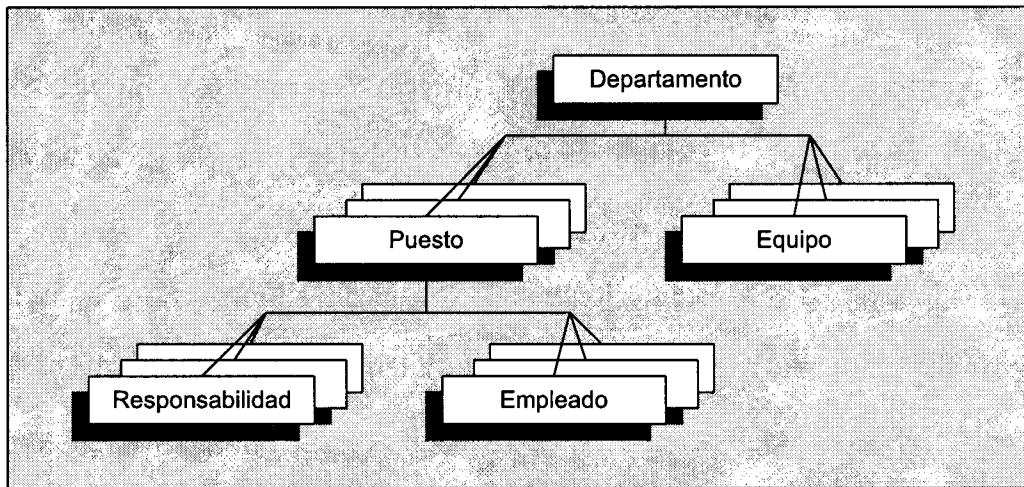


Figura 1.5 Modelo Jerárquico de Bases de Datos.

Una extensión subsecuente de un modelo jerárquico, el modelo de red, permitiendo que los datos sean interconectados libremente, sin requerimiento de que encajen en una simple estructura de árbol. En el ejemplo anterior, cada pieza de equipo puede ser asociada con ambos, un departamento y una lista de empleados que fueron autorizados a utilizarlo. Este tipo de asociaciones no sería permitido en un modelo jerárquico.

El surgimiento del Modelo Relacional

Los modelos de bases de datos jerárquicos y de red permitieron facilidad para representar relaciones complejas entre elementos de datos y proveyeron el acceder rápido a los datos. Sin embargo existía un costo: Acceder a los datos de diferente manera a la ya soportada por las relaciones predefinidas, era muy lento e

ineficiente. Peor aún, es muy difícil modificar las estructuras de datos. Cambiar estas estructuras, requerían de que los administradores del sistema dieran de baja la base de datos y la reconstruyeran, modificando todos los programas de aplicación que utilizaran las nuevas estructuras y entonces, traer todos los componentes del sistema en línea, simultáneamente.

El tiempo fuera de servicio creado por estos cambios era intolerable para los usuarios que tenían que continuar trabajando sin acceder a sus datos o la habilidad de almacenar sus transacciones de negocio. Más aún, estos cambios eran mucho más riesgosos para la empresa. Un error simple en cualquiera de los pasos podría fácilmente corromper los datos corporativos de manera que su costo fuera de millones de dólares. El efecto combinado de estos problemas era que las bases de datos jerárquicas y de red eran raramente modificadas una vez que eran desplegadas.

Una nueva forma de manejador de bases de datos, el modelo relacional, direccionaba estos problemas removiendo la información acerca de las relaciones complejas en la base de datos. Todos los datos son almacenados en tablas simples, mediante relaciones sencillas entre artículos de los datos que se expresan como referencias a valores en otras tablas. Por ejemplo: cada entrada en la tabla de equipo podía contener un valor indicando a que departamento pertenecía. Este acercamiento para ligar los datos en un DBMS se llama acceder asociativo, debido a que considera el contenido, en lugar de la localización, para ligar los elementos. Con el nuevo modelo para acceder datos, el DBMS relacional

también hizo posible una nueva sintaxis, el lenguaje estructurado de consulta (Structured Query Language SQL), que permitía acceder a la información en un RDBMS mediante combinaciones posibles.

A pesar de que el modelo relacional es mucho más flexible que sus predecesores, extrae un precio inclinado debido a esta flexibilidad. La información acerca de las relaciones complejas que fue removida de la base de datos, debe expresarse como procedimientos en cada programa que acceda la base de datos, una violación clara de la independencia requerida por la modularidad del software. El cambio de acceder por navegación a asociativo, también extrae una penalidad seria en cuanto a desempeño, debido a que toma más tiempo localizar información relacionada buscando por contenido en lugar de brincar a una dirección calculada.

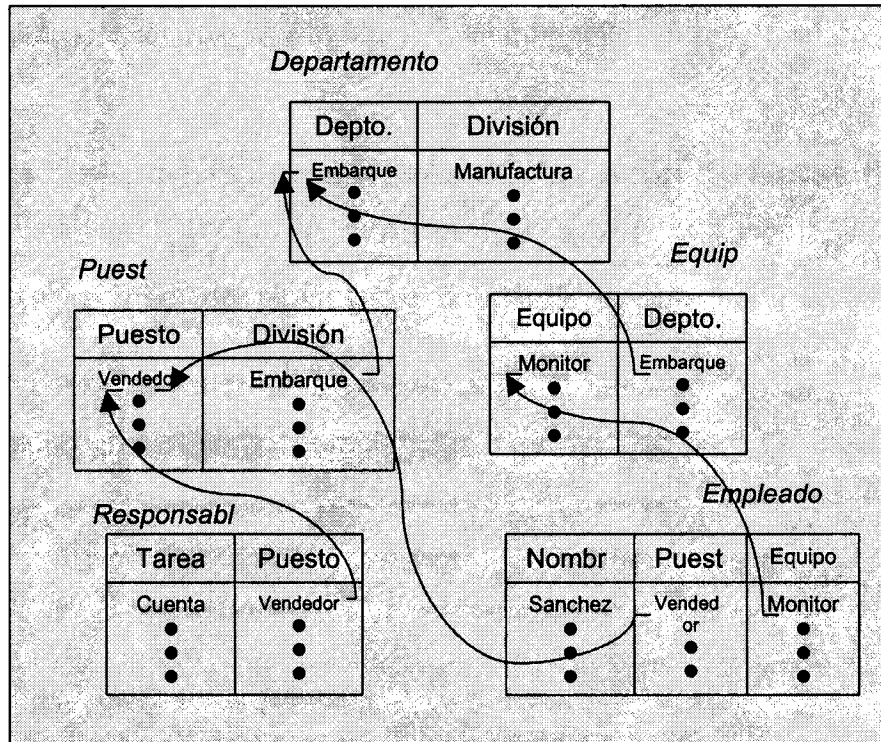


Figura 1.6 Modelo Relacional de Bases de Datos

Las comparaciones han demostrado que las estructuras complejas de datos toman de 100 a 1,000 veces más tiempo para recuperar estructuras de datos de bases de datos relacionales que lo que toman las bases de datos de red. [Taylor]

No obstante estos inconvenientes, la flexibilidad de las bases de datos relacionales a ganado la industria. A pesar de que una vasta cantidad de datos continua siendo administrada por bases de datos jerárquicas, de red y archivos estructurados, estas son herencias de generaciones previas en lugar de reflexiones sobre prácticas actuales. Las bases de datos relacionales representan

la tecnología DBMS dominante hoy en día y virtualmente todas las nuevas aplicaciones son construidas utilizando sistemas relacionales.

Integración de Información a través de Almacén de Datos (Data Warehousing).

El concepto principal de almacenaje de datos es que sea más efectivo el acceder a los datos almacenados para análisis de negocio, separándolos de los datos contenidos en sistemas operacionales. Las razones para separar los datos operacionales de los datos de análisis no han cambiado significativamente con la evolución de los sistemas de almacenaje de datos, excepto que ahora son considerados más formalmente durante el proceso de construcción de sistemas de almacenaje de datos. Los avances en la tecnología y los cambios en la naturaleza del negocio, han permitido que el análisis de negocios sea más sofisticado y más complejo, adicionando la producción de reportes estándares. Hoy en día los sistemas de almacenaje de datos soportan análisis sofisticados en línea, incluyendo análisis de datos multidimensionales. [*Object Technology: A Manager's Guide*]

El mantener datos dispersados a través de varias generaciones de tecnologías DBMS, presenta su propio conjunto de problemas. Es muy difícil para aplicaciones individuales, integrar datos a través de tecnologías tan radicalmente diferentes, lo que significa que obtener cualquier punto de vista administrativo es más difícil que nunca. El acercamiento actual para la resolución de este problema es el almacenaje de datos, una base de datos relacional que se utiliza para tomar datos

seleccionados juntos de diferentes localidades de almacenaje en un repositorio con un mecanismo de acceder común. El almacenaje de datos, por sí solo, no soluciona el problema de integrar datos de fuentes divergentes, pero permite que el problema se solucione una sola vez, utilizando procedimientos habituales para obtener los datos juntos. Una vez que esto se ha realizado, todos los datos combinados pueden ser accedidos de manera uniforme, utilizando SQL y otras herramientas estándar.

Los sistemas de almacenaje de datos son más exitosos cuando se combinan datos de más de un sistema operacional. Cuando los datos necesitan tomarse de más de una fuente aplicativa, es natural que esta integración se dé en una aplicación separada, independiente de las aplicaciones fuente. Los sistemas de almacenaje de datos permiten de manera muy efectiva, la combinación de datos de múltiples fuentes de aplicación, tales como ventas, mercadotecnia, finanzas y producción. Muchas arquitecturas grandes de almacenaje de datos, permiten que las aplicaciones fuente sean integradas en almacenes de datos incrementales.

La razón primaria de combinar datos de múltiples aplicaciones fuente, es la habilidad de referencias cruzadas de datos de estas aplicaciones. Todos los datos de un Almacén de datos típico, se crean en una dimensión de tiempo. El tiempo es el primer filtro criterio para un gran porcentaje de toda la actividad realizada en un almacén de datos. Un analista puede generar consultas de una semana, mes, trimestre, cuarto o año dado. Otra consulta popular en muchas aplicaciones de almacenaje de datos, es la revisión de la actividad año por año. Por ejemplo, es

posible comparar ventas del primer cuarto del año, con ventas del primer cuarto de años anteriores. La dimensión de tiempo en un almacén de datos también sirve como atributo fundamental de referencia cruzada en un almacén de datos. Por ejemplo un analista puede analizar el impacto de una nueva campaña de mercado accediendo las ventas de ciertos meses durante un mismo periodo. La habilidad de establecer y entender la relación entre actividades de diferentes grupos organizacionales dentro de una empresa, es continuamente citado como la mejor característica de los sistemas de almacenaje de datos.

Los sistemas de almacén de datos son útiles, no sólo como plataformas de fusión de datos de múltiples aplicaciones, también es posible integrar múltiples versiones de la misma aplicación. Por ejemplo una organización puede haber migrado a una nueva aplicación estándar de negocio, reemplazando un sistema basado en mainframe. Los sistemas de almacén de datos pueden servir como una plataforma muy poderosa, para combinar datos de viejas y nuevas aplicaciones. Diseñada de manera propia, el almacén de datos puede permitir el análisis de año por año a pesar de cambiar la aplicación operacional base. [*The data warehouse lifecycle toolkit: expert methods for designing, developing, and deploying datawarehouses*]

La razón más importante de separar los datos operacionales de los de negocio, es la degradación del potencial de desempeño en los sistemas operacionales, que puede resultar de los procesos de análisis. El alto desempeño y una rápida respuesta son casi universalmente críticos para los sistemas operacionales. La

pérdida de eficiencia y el costo incurrido con respuestas más lentas sobre las transacciones predefinidas son usualmente fácil de calcular y medir. Por ejemplo, una pérdida de cinco segundos de tiempo de procesamiento es quizá insignificante por sí sola, pero mezcla considerablemente más tiempo y altos costos, una vez que se observa el panorama completo de todas las operaciones juntas. Por otro lado los procesos de análisis de negocio en un almacén de datos son difíciles de predefinir y rara vez necesitan una respuesta rígida en cuanto a requerimientos de tiempo. [*The data warehouse lifecycle toolkit: expert methods for designing, developing and deploying data warehouses*]

El almacenaje de datos es un paso razonable hacia delante, pero es una solución que introduce problemas de sí mismo que aparentemente se han tornado dolorosos. Un problema es que la rigidez de los datos de sistema se ha incrementado, debido a que cualquier cambio en las bases de datos locales requiere modificar el sistema de almacenaje de datos. Otra preocupación es el aspecto tiempo. La información de un almacén de datos siempre se retrasa detrás del tiempo real de operación de datos de la empresa y este retraso difiere cuando las diferentes bases de datos son actualizadas en momentos diferentes. El resultado es que el punto de vista administrativo de la empresa está siempre desactualizado en un cierto grado y puede ser inconsistente temporalmente. Finalmente, existe un problema de escala, los sistemas de almacenaje de datos, tienden a exhibir un crecimiento descontrolado y atenta a administrar este crecimiento utilizando capas de almacenes (warehouses), lo que adiciona complejidad y rigidez a los sistemas resultantes.

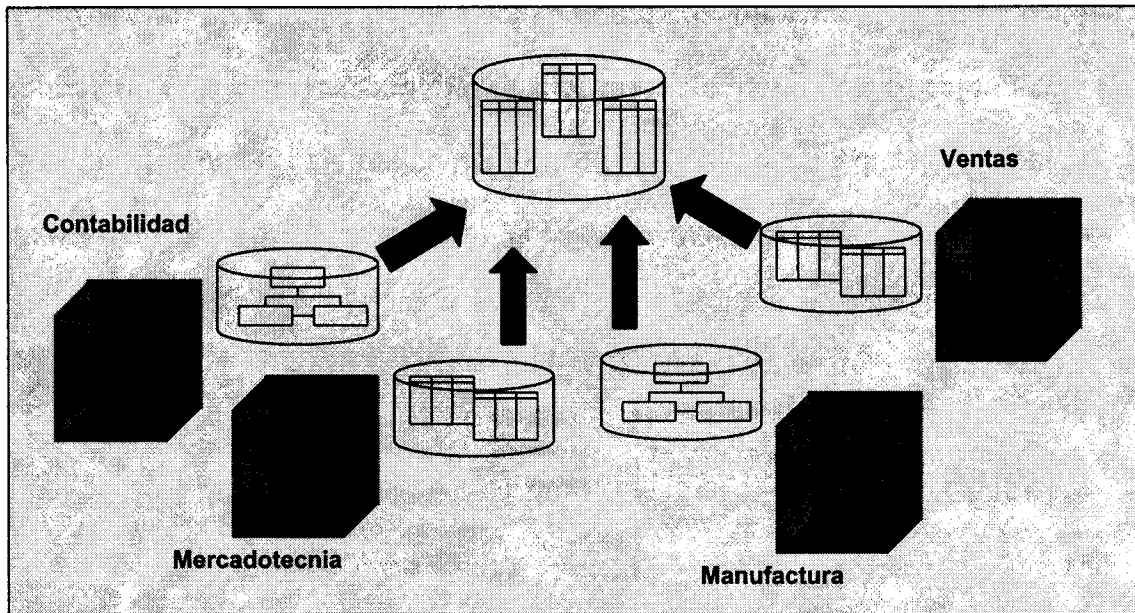


Figura 1.7 Almacén de datos (Data Warehouse)

Minería de Datos (Data Mining)

En los primeros días del data warehousing, la minería de datos era visto como un subconjunto de las actividades asociadas con el almacén de datos. Actualmente los caminos del data warehouse y el data mining se están separando. Mientras el almacén puede ser una buena fuente para los datos a ser minados, la minería ha sido reconocido como una tarea independiente y no más como parte de la colonia del almacén de datos.

De hecho, la minería de datos no sólo ha ganado independencia, sino que está influenciando directa y significativamente el diseño e implantación de grandes almacenes de datos. En los primeros el paradigma de “construya un almacén primero, después mine”, parecía simple e intuitivo y en muchos casos se siguió

por default. Este acercamiento de "construya primero, piense después", fue realmente un paradigma de recolectores de datos, debido a que en muchos casos conducía a la construcción de recolectores de datos cuyo contenido no era fácilmente salvable. La mejor manera de resolver esto, es dejar el almacén de datos entre dos capas de minería, tal como se muestra en la figura 1.8

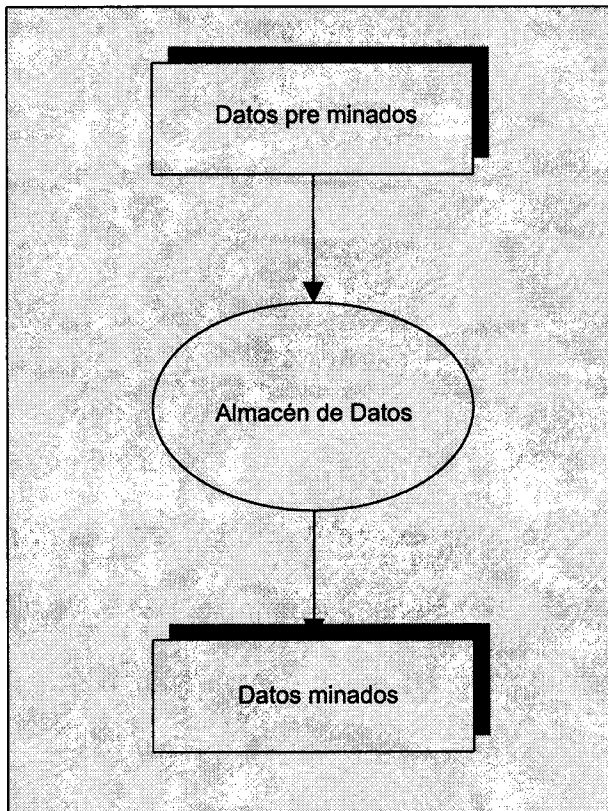


Figura 1.8 Almacén de datos entre datos preminados y minados.

A pesar de que el almacenaje y la minería son sin duda actividades relacionadas, que pueden reforzarse una a otra. La minería de datos requiere de diferentes estructuras de datos, procesos computacionales y abastecedores a diferentes grupos de usuarios que los del almacenaje. Es necesario separar

cuidadosamente estos procesos y entender como difieren en orden de utilizarlos efectivamente.

El propósito de la mayoría de los almacenes de datos es unir grandes montos de datos históricos de diferentes fuentes y utilizarlas para soporte a decisiones. Las actividades desempeñadas para grandes repositorios corporativos, son usualmente diversas, pero continuamente incluyen distintas tareas, tales como consultas, reportes, análisis multidimensional y minería de datos. Estas tareas rompen naturalmente en grupos separados de usuarios, así como procesos computacionales distintos.

El acceder a datos y su análisis, trabajan en diferentes espacios computacionales, los cuales son "espacios derivados". El acceder a operaciones de datos, tales como consultas y reportes, tratan con el espacio de datos, OLAP utiliza el espacio multidimensional y la minería de datos utiliza el espacio de influencia. Los cuatro espacios que forman la base del soporte a decisiones se muestran en la figura 1.9. Estos son espacios de agregación, influencia y variación. Puede ser también utilizado un espacio basado en las relaciones geográficas para cierto tipo de análisis.

Un almacén de datos es entonces, el espacio natural para almacenar el espacio de datos, que es donde se almacena el nivel base de los elementos de datos que más tarde serán analizados para entregar información y justo cuando OLAP no es

visto como un esfuerzo únicamente de almacenaje, la minería de datos es donde se desempeña el análisis que trata con el espacio de influencia.

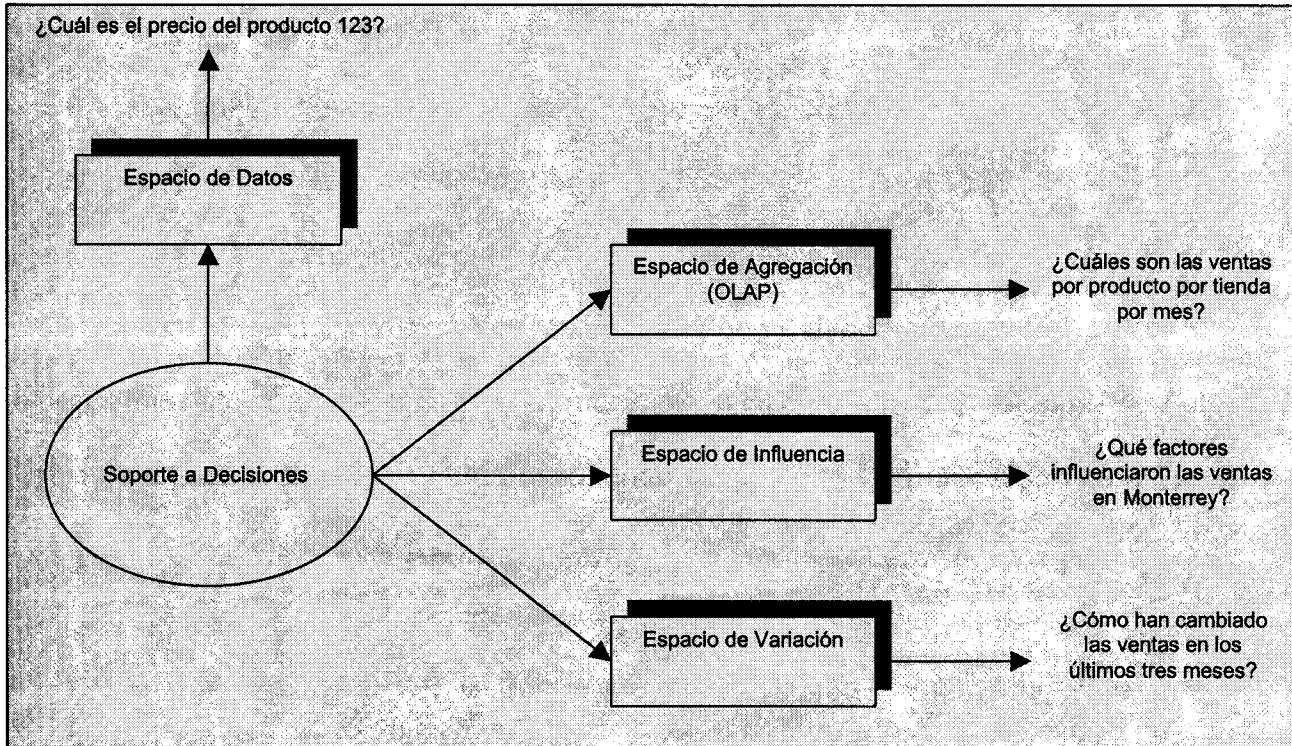


Figura 1.9 Los cuatro espacios que forman la base del soporte a decisiones

Las preguntas planteadas a estos espacios, son inherentemente diferentes. Algunas preguntas como ¿Qué influenció las ventas?, son casi imposibles de contestar directamente en el espacio de datos. Más aún, los espacios derivados son continuamente tan grandes que los factores de influencia no pueden ser fácilmente procesados de antemano, dentro de bases de datos también grandes. Continuamente tenemos que confiar en preprocesamientos parciales de estos espacios, con computaciones dinámicas, con computaciones más allá de las dinámicas para proveer este tipo de respuestas.

Los espacios de datos incluyen toda la información contenida en otros espacios, pero en forma con menor refinamiento. Esto forma las bases para la derivación de los otros espacios, pero de manera menos refinada.

En el proceso de análisis, el espacio de datos continuamente necesita ser enriquecido con semántica adicional, adicionando información acerca de jerarquías y comportamiento periódico. Esta semántica adicional va un poco más allá del modelo relacional simple, pero continuamente ayuda al usuario. De hecho, mes, trimestre, año y ciudad, estado, municipio, forman jerarquías naturales que el modelo relacional simple ignora. Adicionando dicha semántica al espacio OLAP se provee de beneficios adicionales al usuario.

Mientras el espacio OLAP, trata con valores numéricos computados, el espacio de influencia tiene una naturaleza lógica. Este trata con la influencia de grupos específicos de elementos en los demás. Lo que hace más interesante a este espacio es que la información que provee es mucho más utilizable que los otros espacios, debido a que dicha información es típicamente muy general y puede ser referida como conocimiento. Desde que la información uno de los artículos más valorados en el mundo y con la complejidad incremental de la sociedad de hoy en día, la información obtenida por la minería de datos puede ser exponencialmente más valorada que cualquier otro activo.

Los tres componentes separados del sistema de soporte a decisiones de una empresa son:

- El almacén de datos, donde la montaña de datos corporativos son almacenados en un solo lugar. Aquí los volúmenes de datos son muy altos. Estos diseños son usualmente esquemas de estrella o estructuras altamente normalizadas.
- Los mercados de datos son donde los datos departamentales son almacenados y continuamente se añaden componentes de datos. Los volúmenes de tamaño de los almacenes van de un 15% a un 30%. Estas bases de datos se basan usualmente en diagramas de estrella o están normalizados. Deben encontrarse en el espacio de datos, pero algunas veces se realiza análisis multidimensional.
- La minería de datos es donde se organizan los datos para su análisis y la información es extraída de los datos. Los volúmenes de datos son los mismos que los mercados de datos, pero los datos son ricamente más estructurados y ya no son solo departamentales. Los datos aquí se refieren a objetivos específicos del negocio y son analizados con el propósito de extracción de información.

Mientras las estructuras de datos utilizadas dentro de los almacenes y mercados de datos, son similares, las de la minería de datos son significativamente diferentes. La minería de datos difiere del almacén de datos no solo en el tamaño que maneja, sino también en la estructura de los datos. El contenido de los datos en la minería es a menudo diferente de los datos en el almacén, debido a que es enriquecida frecuentemente con datos externos adicionales que no se encuentran en el almacén. La llave de la arquitectura de la minería de datos es que las teorías existentes de la estructuración de datos no aplican a esta.

La minería de datos puede existir en tres formas básicas:

- Arriba del almacén de datos, como un conjunto de vistas conceptuales
- Junto al almacén de datos como un repositorio separado.
- Dentro del almacén de datos, como un conjunto distinto de recursos.

En algunos casos existe la minería de datos autosuficiente, esto sucede cuando una unidad de negocio dentro de una organización necesita obtener resultados del próximo trimestre, no del próximo año. Cuando toma demasiado construir un almacén corporativo, algunas organizaciones no tienen opción más que tomar el asunto en sus manos, a menudo con resultados excelentes. [*DataMines for DataWarehouses*]

CAPÍTULO 2

NECESIDAD DE UNA ARQUITECTURA DE DATOS

El mundo del desarrollo de sistemas ha cambiado tan rápido, como la tecnología misma. Originalmente existía un análisis basado en requerimientos, donde dominaban los pensamientos de modelación de procesos. Fue entonces cuando iniciaron los análisis y diseños basados en datos.

Las organizaciones continuamente adquirían cierto número de computadoras y era frecuentemente difícil transferir datos entre ellas. El tamaño y la naturaleza de los dispositivos de almacenaje disponibles, eran tal como hacer imposible el manejo de grandes bases de datos integradas. No era para sorprenderse, que el software requerido para administrar dichas bases de datos no existía. Consecuentemente, la situación norma se presenta en la figura 2.1.

Con el desarrollo de nuevos sistemas de alto desempeño de procesamiento de transacciones en línea y el mantenimiento de sistemas viejos orientados al procesamiento por lotes (batch), se hizo realidad el hecho de que el hacer énfasis en la automatización de un solo proceso, era incorrecta. Había mucho más en el desarrollo del sistema, que la especificación de código y de requerimientos. Un poco más allá, mientras los procesos tendían a cambiar con el tiempo, los datos requeridos para desempeñar esos procesos, tendían a permanecer relativamente estables. Entre más completa sea la directriz al aspecto de los datos de un

sistema en la etapa de diseño, se incrementa la producción de sistemas estables. A través de los días de análisis de procesos y análisis y diseño estructurado, el aspecto completo de los datos ha sido tratado como un pensamiento final. De hecho, en los primeros tiempos del desarrollo de sistemas, se contempla a los datos como producto del procesamiento.

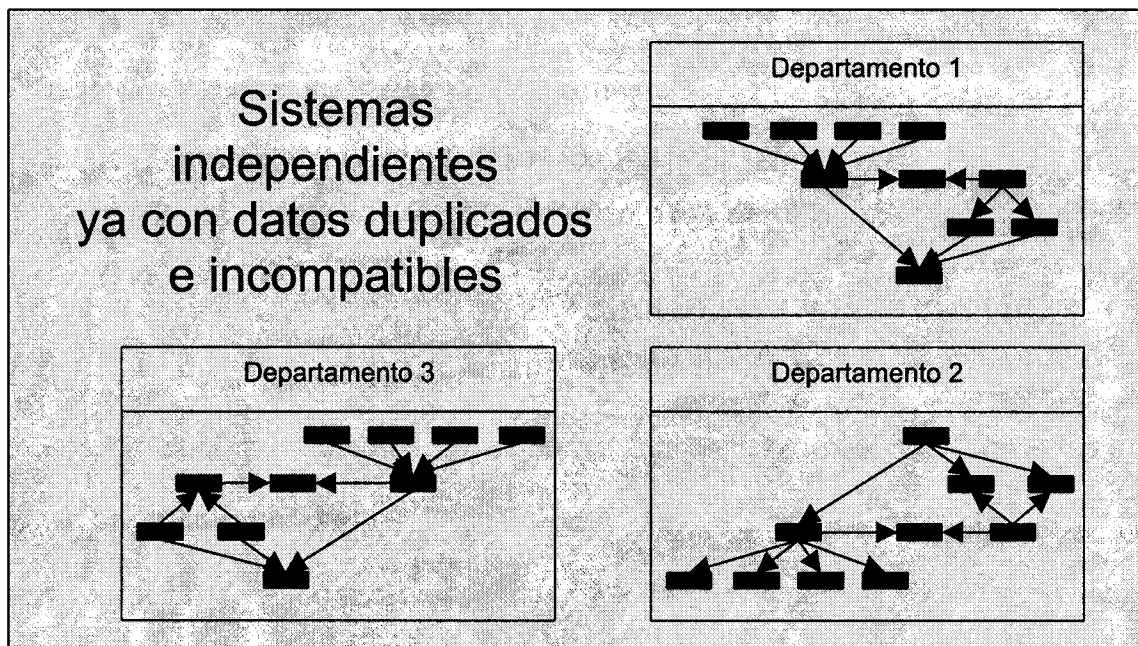


Figura 2.1 Sistemas Independiente con datos duplicados e incompatibles.

Las islas de datos existentes en muchas organizaciones, son parte de los sistemas legados que operan hoy día en la mayoría de las organizaciones que implementaron sistemas hasta los años ochenta. De hecho a la fecha se continúa creando islas de datos. Existen numerosas consecuencias importantes de esta herencia:

Datos acerca de productos, clientes y otras entidades pueden mantenerse en más de un archivo.

Puede resultar extremadamente difícil, reconciliar la información producida por varios sistemas por razones tales como:

- El uso inconsciente de códigos.
- Diferenciación de cortes y procesamientos.
- La eficiencia y efectividad de las operaciones intra e inter organizacionales son forzadas debido a las barreras resultantes del flujo ligero de datos.

¿Por qué contar con una Arquitectura de Datos?

Los crecientes problemas ocasionados por los sistemas autosuficientes y las islas de datos asociadas, llegan a ser rigurosamente aparente, desde una consideración tipo informe de la Figura 2.2, la cual muestra el cambiante énfasis organizacional. Históricamente el comando jerárquico y control de la organización fue un lugar común, tal como se indica en la organización vertical de la Figura 2.2. Sin embargo, dichas organizaciones son relativamente sin responsiva e inflexibles. Ahora en muchos sectores se han establecido nuevos arreglos que son mucho más flexibles y con más responsiva a las demandas de un mercado cambiante. La figura 2.2, utiliza el término horizontal para describir una forma de dichos arreglos organizacionales. El término red debió ser de igual manera utilizado. [*Data Structure and Managemet*]

Un requerimiento llave de las organizaciones red, es que debe existir un flujo libre de datos a través de la organización y que a todo el personal de la organización se le debe presentar una vista consistente de la información que requieren para realizar su trabajo efectivamente. La calidad de la arquitectura de datos de la empresa y la efectividad de su implantación, determina que tan bien pueden fluir los datos a través de la organización. [Architecture]

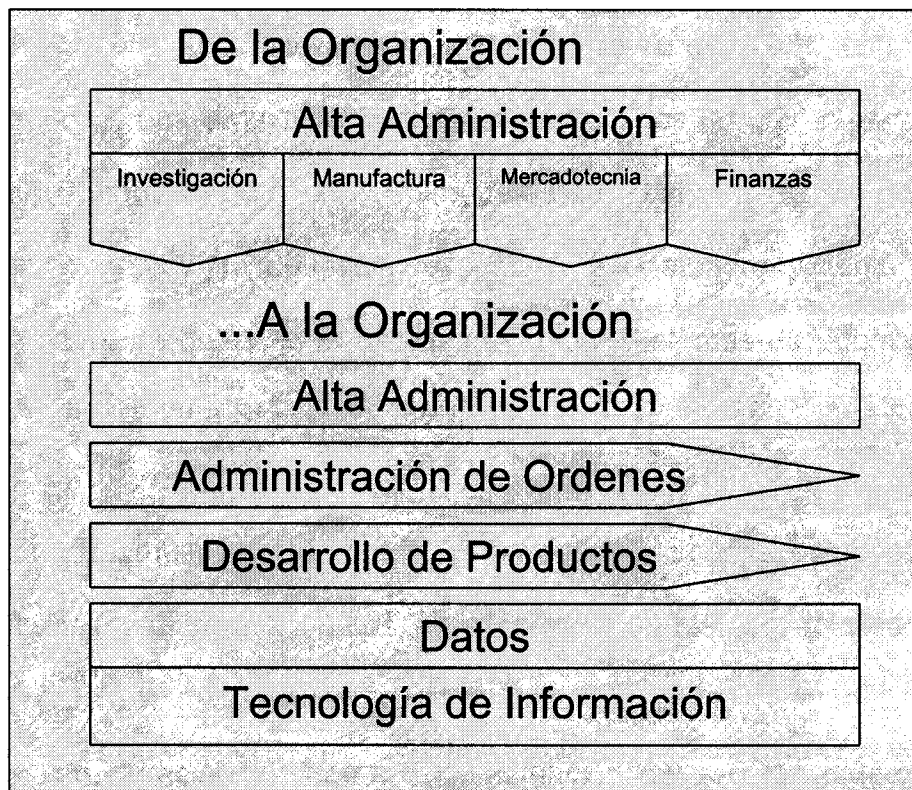


Figura 2.2. El énfasis organizacional es cambiante

De tal manera la estructura de datos puede tener una influencia significativa en la habilidad de la organización de ser flexible y responsiva ante las demandas de un mercado cambiante. Una estructura de datos fuerte no es suficiente por sí misma para asegurar que una organización es flexible y responsiva, pero una

estructura de datos pobre, seguramente dará lugar constantemente a represiones serias en la habilidad de la organización de archivar estándares de desempeño mundial, es parecido a hacer imposible el archivo de dichos estándares. Es por esto, que las arquitecturas de datos son importantes desde cierta perspectiva.

El paradigma de la información es la arquitectura de datos, que es el resultado de una evolución y de alguna manera, se explica así mismo, debido a que muchos de sus componentes sobrevivientes o moldeados por la evolución, son familiares. Este resultado es una arquitectura, descrita en términos de datos, pero aplicable más allá de la narrativa de que solo los datos se ajustan a las necesidades de información genérica y presupuestos de la empresa de manera eficiente y efectiva. Esta arquitectura de datos es el paradigma de los Sistemas de Información y la tecnología para sistemas estables y efectivos.

El término de arquitectura de datos, se ha definido como el componente de la línea de trabajo del recurso datos, que contiene todas las actividades y los productos de esas actividades relacionados con la identificación, nombre, definición, estructuración, calidad y documentación del recurso datos para una empresa.

Una arquitectura de datos común, es una arquitectura de datos que provee un contexto dentro del cual todos los datos son definidos de manera que se determine su contenido y significado real, esto permite que puedan ser integrados en un recurso de datos formal y de lectura compartida que soporte las

necesidades de información. Los Sistemas de Información que prosperan, son aquellos que se encuentran en armonía con la arquitectura de datos.

El departamento de defensa de los Estados Unidos de Norte América, dice qué: "Una arquitectura es definida en IEEE 610.2 como la estructura de componentes, sus relaciones, sus principios, las guías gobernantes de su diseño y evolución sobre el tiempo". La arquitectura de datos de una empresa contiene el procesamiento de transacciones en línea y (OLTP On-Line Transaction Processing) el procesamiento analítico en línea (OLAP On-Line Analytical Processing). Los datos de OLTP soportan las operaciones y los datos de OLAP producen inteligencia.

Los datos y la información son activos de extremo valor a la empresa. La arquitectura de datos, define una infraestructura para proveer datos de alta calidad y consistentes, donde se necesita y cuando se necesita. Esta infraestructura asiste de manera complementaria a los requerimientos de datos, para hacerlos fácilmente accesibles y entendibles por usuarios finales autorizados y por sistemas aplicativos.

Los resultados de un estudio realizado por la Sociedad de Administración de Información (The Society of Information Management SIM) y el Centro de Investigación de Sistemas de Información gerencial de la Universidad de Minnesota (MIS Research Center at the University of Minnesota), publicados en

Diciembre de 1991, muestran que los aspectos más importantes que enfrenta la administración de Sistemas de Información actualmente, son:

- El desarrollo de una arquitectura de información, hace efectivo el uso de datos como recurso.
- El mejoramiento de la planeación estratégica de los Sistemas de Información.

El proceso de arquitectura, inicia con un entendimiento de la empresa y de los datos que constituyen su infraestructura de información. A manera de ser más útil, los Sistemas de Información deben de derivarse de ésta base de conocimiento acerca de la empresa.

Generalmente hablando, una arquitectura define qué es lo que se necesita, y un plan soporta cuando esta arquitectura será implementada. Las arquitecturas por si solas pueden proveer definiciones útiles y estándares, además de proveer ideas.

En cualquier compañía, negocio o industria, los ejecutivos enfrentan grandes retos para alcanzar sus metas. Sin discusión alguna, la información provee las respuestas a las necesidades de éxito. La gente a todos niveles de la organización, necesita de información para llevar a cabo su trabajo de manera correcta. Cuando se pregunta: Con respecto a la información, *¿Qué es necesario para completar sus objetivos o realizar su trabajo?* Presidentes, Vicepresidentes y los niveles que reportan a estos, responden las siguientes cinco mismas respuestas: [Inmon]

1. Acceder a datos en un formato utilizable cuando y donde sea necesario.

Virtualmente, el trabajo de cualquier persona requiere datos y el acceder a estos es la requisición más frecuente. Es común encontrar individuos que gastan más del 50% de su tiempo manejando información –buscando, encontrando, ordenando, examinando, copiando, guardando, cambiando, completando, y enviando datos. Consecuentemente, el que sea posible acceder datos cuando sea necesario, donde sea necesario y en un formato utilizable es crítico. El que se encuentre en un formato utilizable, significa que es posible leer los datos e interpretarlos en información y no convertirlos en datos irrelevantes. Pura y simplemente, el acceder a datos es el requerimiento número uno para alcanzar las metas del negocio.

2. Habilidad para adaptarse a las necesidades cambiantes del negocio.

Para prosperar en una economía mundial, los negocios modernos deben ser capaces de ajustarse y adaptarse rápidamente a cambios. Los ejecutivos requieren de Sistemas de Información que soporten al negocio conforme este cambia. En un ambiente dinámico de negocios, los usuarios no pueden darse el lujo de esperar seis meses, un año, o más a que el área de Sistemas de Información atienda de manera formal sus requisiciones. Las bases de datos y las aplicaciones que se utilicen dentro de la organización, deben ser flexibles y de fácil mantenimiento, de manera que rápidamente sea posible acoplarse a los cambios de nuevos productos, mercados, fusiones y adquisiciones, tecnología, regulaciones gubernamentales y cambios competitivos.

3. Datos Exactos y Consistentes

Los ejecutivos desean y esperan que los datos que reciben sean exactos y consistentes. Los datos deben tener una alta integridad, no solo deben ser correctos dentro de una precisión aceptable también deben ser consistentes a través de la organización. Para datos que deben ser propiamente interpretados y combinados en toda la organización, es necesario un vocabulario común o una estandarización de datos. Por ejemplo, suponga que el gerente divisional de una gran compañía necesita tomar una decisión para asignar recursos de venta sobre la base de niveles de venta. El gerente podría preguntar a las cabezas de las áreas de ventas, mercadotecnia, manufactura, distribución y finanzas, *¿Cuáles son nuestras ventas?* y recibiría cuatro diferentes respuestas significativas. aunque ninguna de estas figuras de *ventas* sería incorrecta desde la perspectiva de la persona que provee la información, cada uno determinó la información de manera diferente. Esto es, la palabra *ventas* puede tener cuatro diferentes significados en la empresa. Estas diferencias e inconsistencias deben de ser eliminadas. Los datos provistos deben de ser exactos y sobretodo actualizados al día.

4. Compartir datos dentro de la organización.

Los ejecutivos reconocen que los datos deben de ser compartidos dentro de la organización, para alcanzar exitosamente las metas del negocio. Los datos deben de ser compartidos entre los departamentos y las unidades de la organización y para que esto suceda, los datos deben de ser centralizados

y coordinados en su administración. Sin embargo, este enunciado no implica una base de datos centralizada. En su lugar es crucial que los datos posean una organización común que elimine la redundancia y asegure la consistencia de los datos, donde ésta deba estar.

5. Contener Costos

Los ejecutivos desean los cuatro puntos anteriores sin escalar los costos de los Sistemas de Información. Los presupuestos de varios dígitos reservados para Sistemas de Información, no son ampliamente aceptados. Los datos deben proveerse a un costo razonable y accesible.

A los individuos con perfil en cómputo o con experiencia práctica extensa en la utilización de sistemas procesadores de transacciones en organizaciones medianas y grandes, les es más difícil adquirir estas ideas. Sin embargo, muchos gerentes de nivel medio alto, no tienen experiencia. Actualmente no es posible considerar, que entiendan inmediatamente el significado de la calidad de la estructura de datos en una organización para esa habilidad de competir en los mercados globales actuales, lo que demanda incrementalmente la demanda de clase mundial de estándares de desempeño.

Las razones por las que los gerentes de nivel medio alto podrían interesarse en este tópico incluyen:

- Mejorar la eficiencia y efectividad del negocio.

- Habilitar una rápida respuesta a las necesidades cambiantes del negocio.
- Preocuparse acerca del impacto que las estructuras de datos pobres están teniendo en una organización y un interés en ¿cómo las oportunidades pueden maximizarse y minimizando los riesgos?

Desde la perspectiva de la administración de Sistemas de Información, las metas y los aspectos incluyen:

- Entender que información es estratégica y que información tiene un alto potencial, parecido a tornarse estratégica.
- Crear una estructura de datos estable e integra.
- Mejorar la calidad de la información y los sistemas.

Estos dos conjuntos de razones para que los gerentes tomen interés en este tópico son complementarios. Sin embargo, es muy importante para el personal de Sistemas de Información que considere el tópico desde la perspectiva generalmente de los gerentes de nivel medio alto y no lo considere desde la perspectiva técnica.

Esto refuerza la noción implícita en otras ideas varias. La importancia crítica del trabajo en conjunto entre el personal de Sistemas de Información y los gerentes de nivel medio alto, para identificar las oportunidades y acordar en cual es la mejor manera de asignar recursos de Sistemas de Información; En conjunto, la consideración de las estructuras de datos de una organización, siempre resultará en la dedicación de algunos recursos a desarrollar e implementar una

arquitectura. Sin embargo, los pasos de la implantación serán determinados por las prioridades del negocio.

Muchas de las organizaciones tienen un gran monto de datos variables dentro de sus sistemas, que frecuentemente no están disponibles para obtener los valores completos de los datos, debido a que los datos se encuentran pobremente estructurados. Sin embargo, es posible obtener adicionalmente, valor significativo adicional, de los datos contenidos en sistemas existentes sin remplazar estos sistemas.

El Posicionamiento y la Información

El establecimiento de la estrategia de mercado, parte integral del posicionamiento de una empresa en el mercado, se basa en la revisión corporativa de sus mercados (por línea de producto) y de sus capacidades, fortalezas y debilidades para competir. Las oportunidades y las acciones necesarias para eliminar las debilidades internas y capitalizar las de la competencia se analizan frente a la información básica de los modelos de posicionamiento, para determinar fuerza e impacto. Establecer una estrategia de mercado implica tomar decisiones sustentadas en una base amplia con relación a las posiciones que la compañía y sus productos aspiran a alcanzar en el mercado. Cuando sea necesario, esta estrategia incluirá también las decisiones relacionadas con el desarrollo de nuevos productos.

En la actualidad, las decisiones corporativas para el posicionamiento en el mercado pueden tener un alcance novísimo más amplio debido a la globalización, tomando en cuenta factores como la competencia global. Las sociedades internacionales y el desarrollo de activos internacionales. Para garantizar el enfoque apropiado, todos los proyectos corporativos deben acceder de inmediato a la estrategia de mercado más actualizada de la compañía y a los planes individuales de mercado para cada producto y/o servicio, esto es a información. En los niveles más bajos de los proyectos de cambio, este tipo de información debe incluirse con los detalles acerca de lo que la compañía expresa en su publicidad, lo que está haciendo la competencia y de los cuales son las actitudes de sus clientes. [*Reingeniería: Cómo Aplicarla con Éxito en los Negocios*]

La etapa más sistemática del posicionamiento es la recopilación de datos. Las clases de información reunidas por el posicionamiento y utilizadas en los proyectos de cambio, aparecen en la figura 2.1. Este es necesariamente un proceso detallado que requiere tiempo, en especial cuando las operaciones son muy grandes y antiguas. Sin embargo, el tiempo y el esfuerzo pueden reducirse y el efecto de los datos puede aumentar mediante el empleo de tecnología. Cualquier sistema computarizado de textos, como el procesador de palabras y los diagramas son mejores que utilizar sólo papel. Una vez que la recopilación de datos ha concluido, éstos deben mantenerse actualizados. Por lo demás, esta actividad no es sólo importante como entrada al esfuerzo de reingeniería, sino que a través de su aplicación creativa tiene muchos usos en el respaldo a las actividades diarias y a la toma de decisiones. Esta utilización, no obstante requiere

flexibilidad y apoyo con base en computadoras, Sistemas de Información, que permitan la explotación de datos.

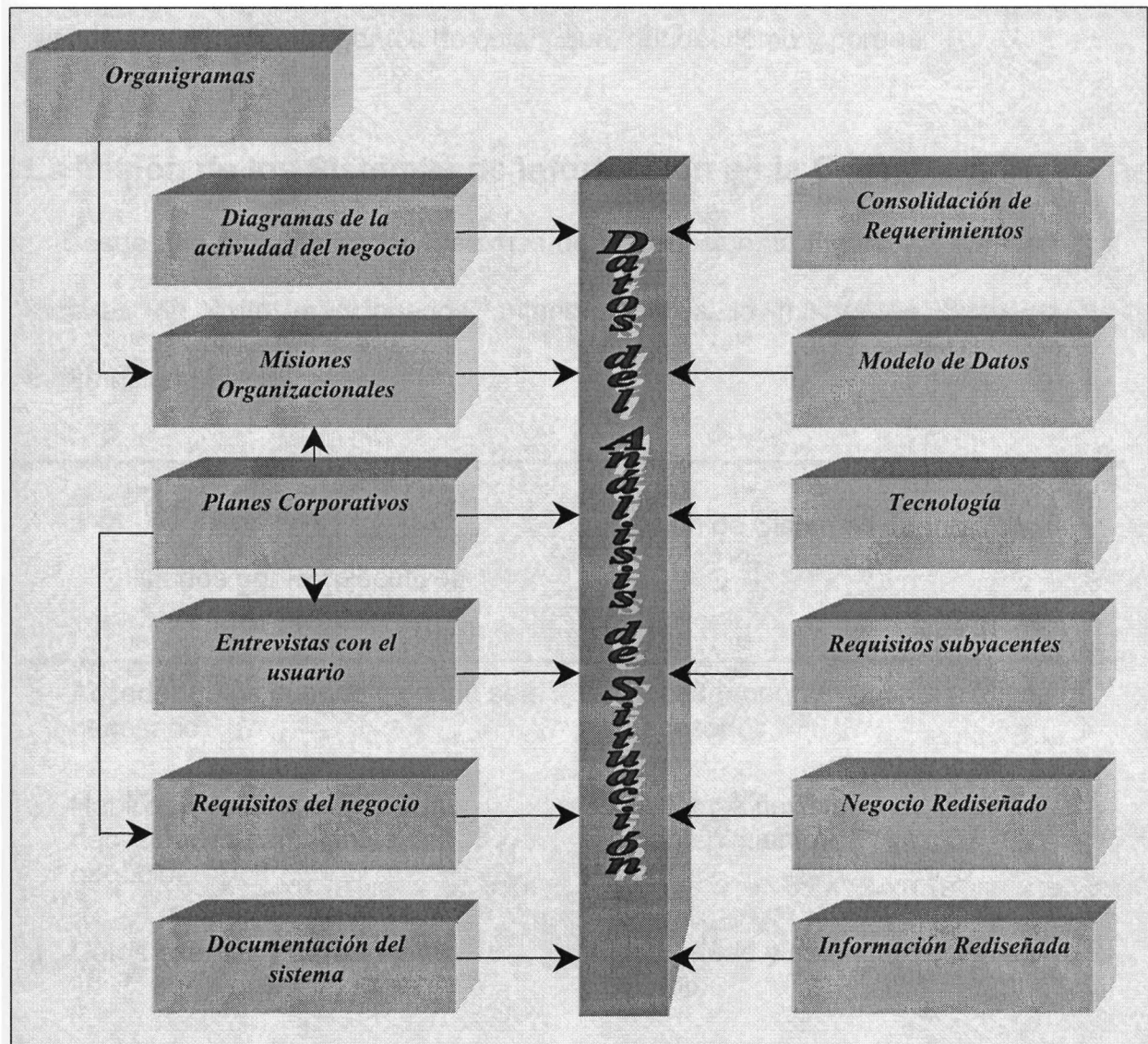


Figura 2.3 Datos del Análisis de la Situación

La información incluida en la conformación de una arquitectura de datos, debe dirigirse a cada faceta de la operación. Deberá incluir modelos de flujo de trabajo,

organigramas, enunciados de la misión corporativa, modelos de proceso de negocios, política de negocios y reglas establecidas, modelos de funciones (diagramas de relación) y planes corporativos. Esta información contiene las respuestas a los interrogantes de quien, qué, dónde, cómo y porqué.

La Misión de los Sistemas de Información en la Organización

Desde una perspectiva de Sistemas de Información, la Figura 2.4 compara los factores de éxito mencionados anteriormente a la misión de Sistemas de Información:

5 factores de éxito listados por los ejecutivos	Misión de Sistemas de Información
<input type="checkbox"/> Acceder datos cuando y donde sea necesario.	<input type="checkbox"/> Acceder oportunamente a los datos necesarios.
<input type="checkbox"/> Habilidad de adaptación a las necesidades cambiantes del negocio.	<input type="checkbox"/> Sistemas flexibles y de fácil mantenimiento.
<input type="checkbox"/> Datos exactos y consistentes	<input type="checkbox"/> Integridad y estandarización de datos.
<input type="checkbox"/> Compartir datos dentro de la organización.	<input type="checkbox"/> Integración Datos/Sistemas
<input type="checkbox"/> Contener Costos.	<input type="checkbox"/> Efectividad del costo

Figura 2.4 Factores de éxito que crean la misión de Sistemas de Información

Estos cinco factores forman la misión de Sistemas de Información en las organizaciones. La planificación de la arquitectura de datos de la empresa, es el primer paso en la vía de logro de esta misión, y es el único tipo de proceso de planeación que a largo plazo puede alcanzar la misma.

Calidad de Datos

Sí uno pudiera utilizar una palabra que describa estos cinco componentes, sería *calidad*. En otras palabras, proveer calidad a aquellos que la necesitan es la misión de los Sistemas de Información. Este concepto ha sido objeto de mucha atención; W. Edwards Deming y otros, han expuesto la longitud en la calidad y productividad del negocio.

Sí aceptamos que la misión de Sistemas de Información es proveer *calidad en los datos*, entonces los 14 puntos para calidad de Deming, podrían ser utilizados. La Figura 2.5 interpreta los 14 puntos de Deming, para lograr calidad en los datos.

Un principio importante a notar es que los autores como Deming, Juran y Conway quienes escribieron acerca de calidad, han declarado que la productividad resulta del enfoque en la calidad, pero la calidad no ocurre a través de perseguir la productividad. Esto es un punto vital porque diversas metodologías de desarrollo de sistemas y herramientas, tales como CASE y los productos de AD/Cycle han sido explícitamente aclamados hacia mejorar la productividad del desarrollo de los sistemas. De acuerdo a las enseñanzas de Deming, enfocándonos en la

productividad de los sistemas por si sola, no necesariamente resultan en la calidad de datos ni habilita la misión de Sistemas de Información

Los 14 puntos de calidad de Deming	14 puntos de calidad de Datos
1 Crear una constancia de propósito hacia mejora (misión y plan)	Desarrollar una constancia para la administración de recursos de datos
2 Adoptar la nueva filosofía (no aceptar retrasos, defectos y errores)	Administrar los datos como activo, compromiso de compartir datos e integridad de datos.
3 Eliminar con la dependencia de inspecciones masivas (Requiere evidencia estadística)	Desarrollar medidas para calidad de datos.
4 Utilizar una fuente única	Establecer una estrategia de migración basada en la creación de datos (Sistemas de datos fuente)
5 Constantemente mejorar la producción del sistema y servicio; encontrar problemas	Entender el negocio (modelo funcional del negocio, planes del negocio)
6 Instituir métodos modernos basados en el entrenamiento del trabajo.	Instituir programas de entrenamiento sobre Administración de datos como recurso.
7 Instituir métodos modernos de supervisión.	Debe identificarse un líder para Administración de Datos como Recurso; promover el trabajo en equipo; eliminar los cálculos de desempeño a corto plazo.
8 Desaparecer el miedo.	Enfoque en metas a largo plazo. (las arquitecturas, claramente establecen la directriz a futuro, alimentando el sentido de seguridad de trabajo, por lo tanto reduciendo el miedo)

9	Romper las barreras entre departamentos	Desarrollar modelos de negocio, arquitecturas y planes que crucen las fronteras de la organización
10	Eliminar las metas numéricas, publicidad y axiomas para nuevos niveles de productividad sin métodos.	Desarrollar estándares y forzar mecanismos que aseguren la calidad de los datos.
11	Eliminar estándares de trabajo que prescriben citas.	Seguir con estándares a través del liderazgo y la responsabilidad para la calidad de los datos (proveer incentivos de adaptación)
12	Promover el orgullo de los trabajadores.	Utilizar métodos, técnicas y herramientas para actualizar la posición de las responsabilidades.
13	Instituir programas vigorosos de educación y entrenamiento.	Equipo con la función o comité de aseguramiento de la calidad.
14	Crear estructuras que empujen los 13 puntos antes mencionados.	La administración debe comprometerse a estos principios de calidad de datos (establecer administración del recurso de datos, reorganización, autorización)

Figura 2.5 Calidad de Datos

La calidad no sucede simplemente, esta debe de ser planeada y diseñada en los productos de Sistemas de Información, de ahí que debe ser referida propiamente como planificación para la calidad de datos. [*Enterprise Architecture Planning: Developing a Blueprint for data, Applications and Technology*]

Algunas de las ventajas obtenidas de establecer una arquitectura de datos a nivel empresa son:

- El enfoque es en el uso estratégico de la tecnología para la administración de datos como un activo.

- El uso de un vocabulario estándar, facilita la comunicación y reduce la inconsistencia y redundancia de datos.
- La documentación incrementa el entendimiento del negocio.
- Los modelos pueden ser utilizados para explicar el negocio y asegurar el impacto de los cambios del mismo.
- Las políticas de toma de decisiones pueden ser revisadas.
- Considera la integración de sistemas actuales con los nuevos.
- Permite un acercamiento comprensivo, objetivo e imparcial.
- En cuanto a costo – beneficio, una solución a largo plazo considera la tasa de retorno.
- Involucra una estrategia de migración factible con logros a corto plazo.
- Facilita el asegurar los beneficios e impactos de nuevos sistemas y software.
- Permite el fácil acoplamiento de cambios dinámicos del negocio, tales como: fusiones, adquisiciones, nuevos productos, líneas de negocio, etc.
- La participación de la administración gerencial provee la perspectiva del negocio, su credibilidad y confidencialidad, además de eliminar el mito del desarrollo de sistemas.

En lo que se refiere a los beneficios a obtener de una planificación de sistemas, a continuación se presenta una lista de estos:

- Más responsabilidad hacia las necesidades de los clientes.
- Reducción de los costos de entrada de datos.
- Incrementa la productividad del personal, que permite incrementar el nivel de negocio y mantiene los costos.
- Mejora las habilidades, lo que incrementa el entusiasmo y la lealtad.

-
- ❑ Eficientar el mantenimiento de sistemas, lo que significa un mejoramiento en el servicio.
 - ❑ Las arquitecturas eliminan la complejidad y las costosas interfaces entre sistemas incongruentes.
 - ❑ Las decisiones administrativas en todas las áreas funcionales serán basadas en datos seguros y a tiempo, llevando a varias mejoras y medidas de ahorro.
 - ❑ El usuario final accede directo a los datos compartidos.
 - ❑ Los nuevos sistemas se desarrollan más rápido y a un costo menor, debido a los datos comunes, código común y una fase de requerimientos más corta.
 - ❑ Facilidad en la evaluación y selección de paquetes de software.
 - ❑ Utilización efectiva del repositorio y productos CASE.

Muchas empresas y organizaciones hoy en día, utilizan DBMS (Data Base Management Systems) y muchas adquieren más de uno. La redundancia incontrolada, la inconsistencia en definiciones de datos, el rápido incremento en almacenamiento, tiempo de respuesta inaceptable y altos costos de mantenimiento son algunos de los tópicos que enfrentan las organizaciones utilizando DBMS's. El simple hecho de adquirir un DBMS, no alcanza beneficios tales como:

- ❑ Independencia física y lógica.
- ❑ Mejoramiento en el desempeño, tiempo de respuesta.
- ❑ Minimizar los costos de almacenamiento.
- ❑ Reducción de la redundancia de datos.

- Integridad de datos.
- Búsquedas rápidas y disponibilidad.
- Asegurar la privacidad y seguridad.
- Adaptación a futuros requerimientos.
- Estructuras de datos flexibles y simples.
- Estandarización.
- Agilizar el desarrollo de sistemas

Estos beneficios dan a lugar un ahorro en costos, sin embargo un DBMS es únicamente una herramienta para acceder a datos y para su almacenamiento. Estas promesas no pueden alcanzarse simplemente con el uso de un DBMS. El camino a la calidad de los datos involucra planeación y cambio. [*Oracle Design*]

La Calidad debe de planearse. La planificación de la arquitectura de la empresa, es un acercamiento moderno para la planeación de la calidad de los datos, alcanzando la misión de los Sistemas de Información. [*Developing Effective Technology Plans*]

Los planes efectivos de tecnología son a corto plazo, no a largo plazo. Un plan a cinco años es muy largo. La tecnología está cambiando a un ritmo tan rápido, que es casi imposible planear que tipo de tecnología estará disponible dentro de cinco años, a partir de ahora. Inclusive en muchos aspectos tecnológicos los planes a un año son inservibles. [*See*]

Peterson (1989) nos recuerda que una organización, simplemente no puede conocer que es lo que hace y que intenta realizar, a menos que establezca y supervise periódicamente sus metas y el tiempo en que mide ese logro de metas. El logro de estos requerimientos asegura que la planeación estratégica ocurra.

Antes de obtener éxito con un plan de tecnología, es necesario entender los conceptos rudimentarios de lo que debe ser un plan. Las definiciones difieren de la persona con la que estamos hablando. Quizá unificar esta amenaza es “la manera en que lo hacemos” es la mejor.

Cook (1988) Afirma que los mejores ejemplos de planeación, son los basados en más de una intuición colectiva del comité de planeación.

Un modelo funcional del negocio es un conocimiento base de qué es el negocio y qué información es utilizada para conducir el negocio. Los acercamientos tradicionales a la planificación de sistemas no inician con una definición de un todo del negocio, pero es usual que inicien con un grupo de analistas de sistemas que se dirigen a las oficinas de los ejecutivos, tome asiento y pregunte *¿Qué sistemas requiere?*. Algunos nuevos acercamientos vagamente utilizaron variaciones de los acercamientos tradicionales, reemplazando la pregunta tradicional *¿Qué sistemas requiere?*, Por *¿Cuáles son sus factores críticos de éxito?*. Lo que dirige a establecer prioridades por construir Sistemas de Información basados en los factores críticos de éxito de los ejecutivos. Sin embargo, lo que es crítico el día de hoy, quizá no lo sea el próximo año. El efecto neto de ambas preguntas es el

mismo. Preguntas como *¿Qué sistemas requiere?, ¿Cuáles son sus factores críticos de éxito?, ¿Qué sistemas necesita?* Conllevan a los Sistemas de Información que direccionan los requerimientos a corto plazo, no a términos de calidad de datos.

La arquitectura de datos soporta las necesidades del negocio, no los requerimientos de los individuos y sin tomar en cuenta las limitaciones de las fronteras de la organización. De ésta manera la etiqueta de dirección del negocio es apropiada.

Tradicionalmente el primer paso era determinar qué sistemas son requeridos para soportar el negocio y que harán cada una de estas aplicaciones. El segundo paso era determinar qué datos era necesario procesar. Sin embargo, el enfoque que tomaremos tiene como punto primordial la determinación de los datos que soportan al negocio antes que las aplicaciones.

En el acercamiento tradicional de planificación de sistemas, después de la pregunta *¿Qué sistemas son necesarios para alcanzar los objetivos?*. Los analistas de sistemas cuestionan *¿Qué tan importantes son?* o *¿Qué sistemas deben implementarse primeramente?*. Al final, la persona que mantiene el último control sobre el presupuesto es quien da prioridad a sus sistemas, involucrando aspectos de visión y comandos de poder político. Por lo tanto, las requisiciones de esta persona son las que se acomodan en primer lugar seguido de las prioridades de sistemas de aquellos que mantienen una posición descendente de influencia.

Después de analizar esto, el definir una arquitectura de datos permite que la dependencia de datos sea quien establece las prioridades además de la secuencia ideal en que las aplicaciones deben de ser desarrolladas e implementadas. La dependencia de datos está basada en un principio fundamental que dice que deben desarrollarse las aplicaciones que crean datos, antes de las aplicaciones que necesitan datos. Este acercamiento no es muy aceptado por las corrientes tradicionales.

Dentro de las funciones más importantes para cumplir con la misión de sistemas, es identificar y atender las necesidades de la organización en cuanto a información se refiere, por lo tanto para darnos cuenta de que existe una integración pobre en la organización, podemos contemplar los siguientes 10 puntos:

- Tanto empleados como clientes protestan sobre su impotencia para obtener información
- Se le da mantenimiento inapropiado al equipo.
- Tenemos problemas conformando empleados con tareas.
- No estamos seguros del proceso en el que se encuentra un producto.
- Tanto empleados, como clientes y proveedores reciben correspondencia en la dirección equivocada.
- No tenemos idea de los niveles de inventario y perdemos seguimiento de nuestros clientes y proveedores.
- Algunas veces nuestros clientes reciben dos facturas.

- Se realizan sumalizaciones reorganizaciones, y cálculos redundantes antes de que se entreguen reportes, además de captura redundante de datos significativos
- Se retrasan reportes administrativos.
- No es posible recuperar datos de las bases de datos organizacionales.

Sí encontramos estos síntomas dentro de la organización, es necesario un reforzamiento en la en la integración de información. La integración de Información es la reestructuración de la arquitectura de Sistemas de Información dentro de una organización; y esto es más que fusionar hardware, software y sistemas de comunicación en una nueva unidad operacional consolidada, donde debemos incluir siempre un análisis estratégico de requerimientos. Los beneficios de esta integración se ven reflejados al incrementar la habilidad de la organización para competir globalmente, facilitando la toma de decisiones descentralizada, reduciendo el efecto de esparcimiento geográfico de las funciones corporativas y brindando el acceder rápido y fácil; además de incrementar la sensibilidad organizacional ante las necesidades del cliente y eliminando las actividades del negocio, que al paso del tiempo se hacen innecesarias o nocivas a la misión vigente. El problema de los sistemas heredados.

Existen varias estrategias para realizar esta integración de sistemas, entre las cuales tenemos: Reingeniería, Downsizing, Outsourcing, Total Quality Management, e Integración estratégica de bases de datos. A continuación se describen las ideas principales de cada una de estas estrategias:

Reingeniería

Es el proceso de analizar, modelar y modificar radicalmente una organización existente, de manera que sus actividades son capaces de entregar calidad en sus productos y servicios, además de ser una manera de hacer aerodinámica la forma en que se procesa la información y eliminar los sistemas heredados.

Downsizing

Involucra un escalamiento hacia abajo sobre la plataforma de hardware, reemplazando los Sistemas de Información centralizados con componentes micro o mini, ligados unos a otros por una red local. Las tecnologías típicas utilizadas en downsizing son: la arquitectura cliente-servidor y los clientes ligeros.

Outsourcing

Contratar varias funciones, tales como la captura de datos, el desarrollo de sistemas, la programación, medios de dirección, recuperación por desastre y administración de telecomunicaciones con proveedores de servicio externos.

Administración de la Calidad Total (TQM Total Quality Management)

Filosofía donde los procesos para conceptualizar, produciendo y entregando productos que constantemente están siendo mejoradas constantemente para afrontar las demandas de los clientes.

Integración de Bases de Datos.

Proceso para asegurar la integridad de las bases de datos y la estabilidad. La meta es mejorar la eficiencia organizacional y proveer la era competitiva.

Planeación estratégica de Datos e Ingeniería de Información

Se combina la estrategia corporativa con el diseño de bases de datos y el análisis de sistemas, enfocando y explotando las oportunidades del diseño de bases de datos.

La planeación estratégica de sistemas puede ser utilizada como ventaja competitiva. Los acercamientos de la planeación típica de Sistemas de Información y la planeación estratégica corporativa, no proveen un punto de inicio o dirección para aplicar la tecnología estratégicamente y la mayoría de los Sistemas de Información son creados sobre tecnología de bases de datos. En la figura 2.6, podemos observar una analogía del objetivo de la planeación estratégica de bases de datos.



Figura 2.6 Planeación estratégica de datos e Ingeniería de Información.

CAPÍTULO 3

FUDAMENTOS DE MODELACIÓN DE DATOS

Planeación Estratégica de Bases de Datos.

Una de las principales actividades de la planeación estratégica de bases de datos, es la modelación de datos genérica de la empresa, este modelo de datos de la empresa o modelo de datos a macro nivel, describe la arquitectura de datos a un nivel muy alto.

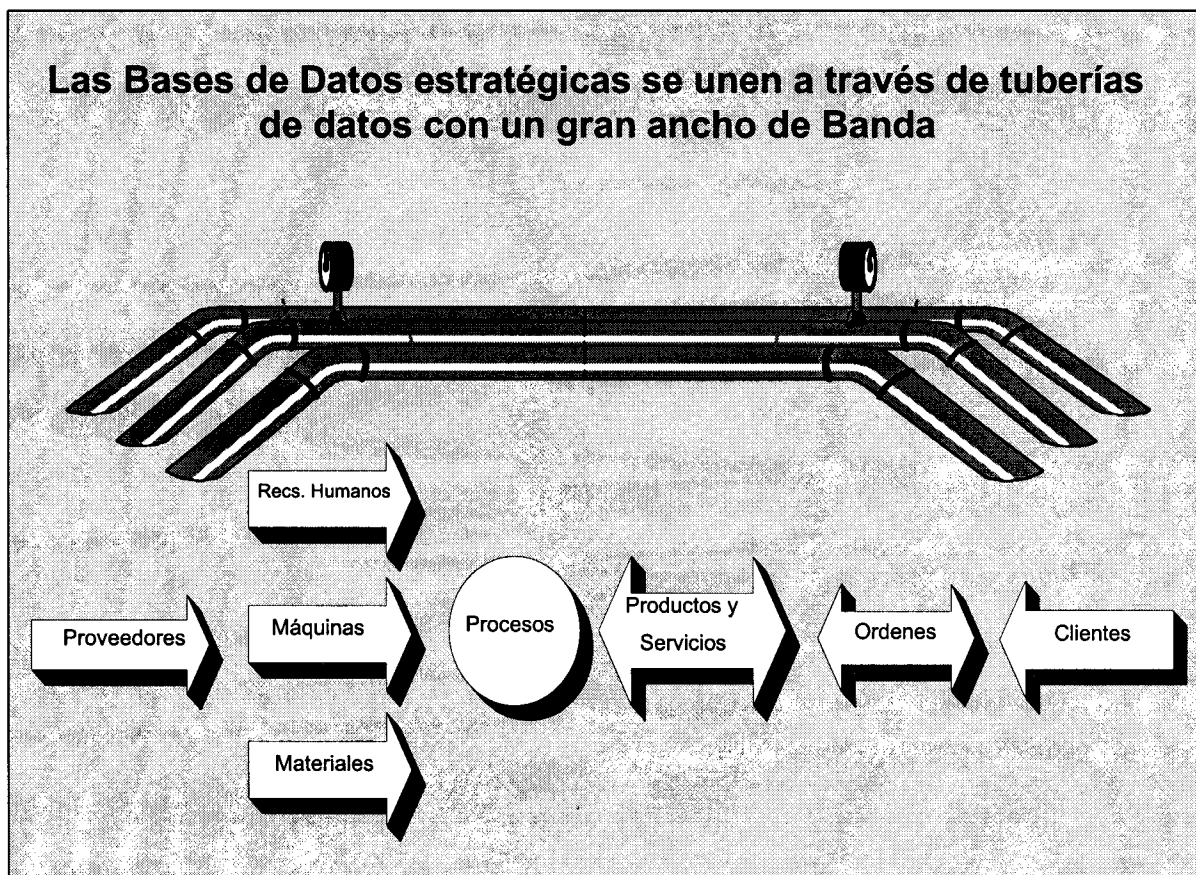


Figura 3.1 Analogía del objetivo de la planeación estratégica de bases de Datos.

La modelación de datos genérica de la empresa, provee un punto focal que asiste para identificar las oportunidades tecnológicas para resolver problemas y desarrollar la competitividad. Este modelo de datos ayuda a determinar en que negocio está la empresa y cómo dirigirá la empresa el negocio.

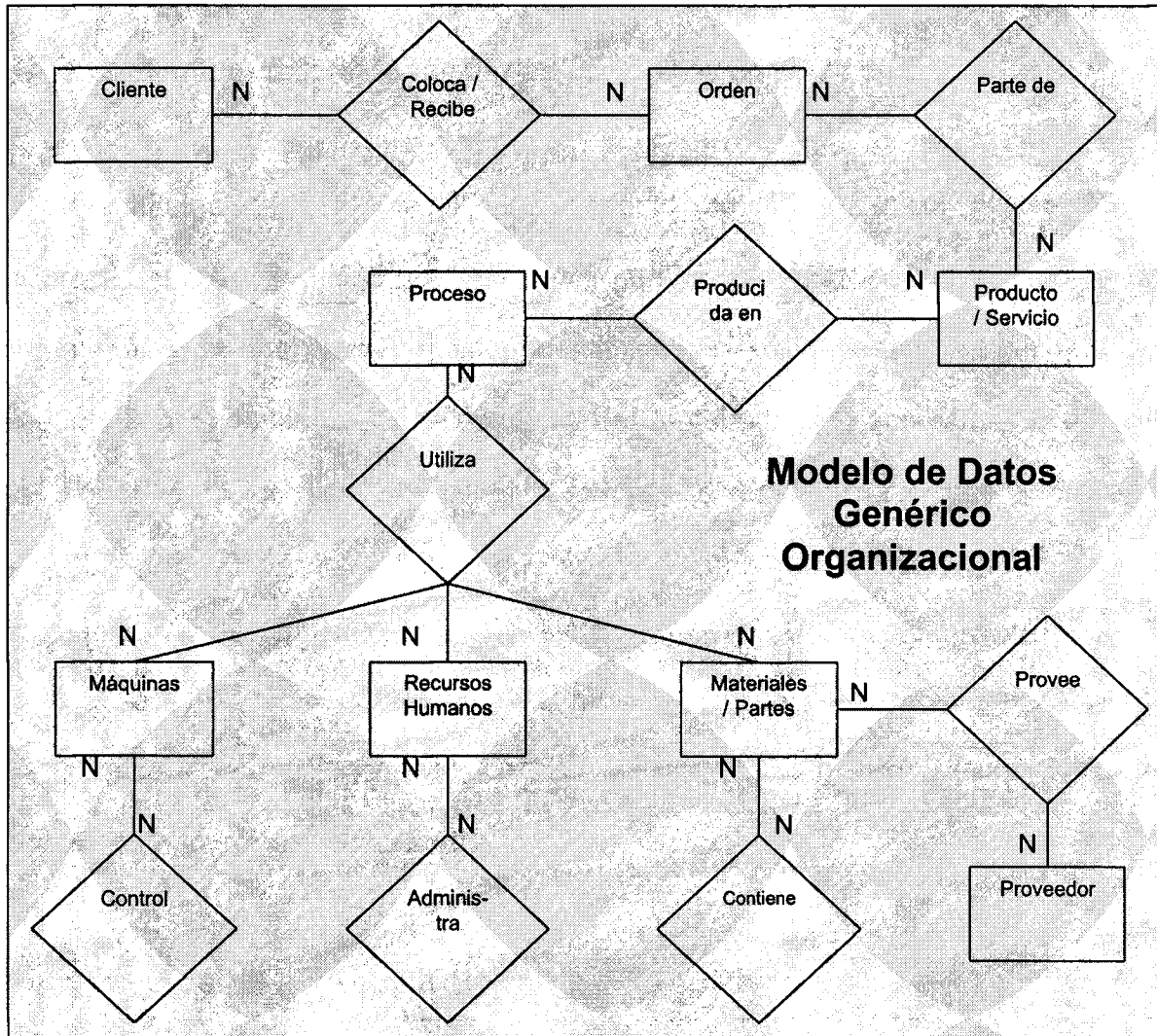


Figura 3.2 Ejemplo de Modelo de datos Genérico Organizacional

Para esta modelación, es posible utilizar tanto el acercamiento Top-Down (de arriba para abajo) como el Bottom-Up (de abajo para arriba). La modelación utilizando el acercamiento Top-Down se da cuando los directivos de nivel medio alto de la empresa, delinear las entidades a nivel macro y las descomponen en entidades y relaciones detalladas. La desventaja de este acercamiento es que la entrada puede tener grandes retrasos por individuos que tratan con las actividades del día a día de la empresa, además de que no hay un punto de inicio.

La modelación utilizando el acercamiento bottom-up, inicia con el nivel operacional, donde los supervisores, administradores de nivel medio y trabajadores con el conocimiento, identifican las entidades y determinan los requerimientos de las bases de datos. La desventaja de este acercamiento es que las entradas de los directivos de nivel medio alto retrasan hasta que el proceso este cerca de terminarse, siendo que estas personas son los accionistas de las bases de datos organizacionales y de los Sistemas de Información estratégicos.

El modelo de datos genérico, consiste de ocho meta-niveles, macro-niveles o entidades superclase, estas son:

- Cliente
- Ordenes
- Productos / Servicios
- Procesos
- Recursos Humanos

- Máquinas
- Materiales Crudos
- Proveedores

En la figura 3.3, podemos observar las fases de la planeación estratégica de Bases de Datos, con las actividades específicas y consideraciones especiales en cada una de las fases. [Inmon]

Fase 1	<p>Entendiendo el sistema Organizacional</p> <ul style="list-style-type: none"> □ Recolectar información acerca del historial de la empresa y su cultura. □ Identificar productos y servicios. □ Identificar el perfil del cliente □ Identificar factores de producción (gente, materiales crudos, máquinas) □ Desarrollar una visión de las entradas y salidas de los procesos y crear un diagrama de flujo de datos de contexto a alto nivel <p>Consideraciones básicas estructurales para la formulación de la estrategia.</p> <ul style="list-style-type: none"> □ Perfil de la administración actual <ul style="list-style-type: none"> □ Metas y objetivos Organizacionales. □ Valores, Metas personales y compromisos □ Habilidades y estilo administrativo.
---------------	--

	<ul style="list-style-type: none"> □ Productos y Servicios <ul style="list-style-type: none"> □ Descripción de productos y servicios producidos. □ Descripción de los procesos de producción. <ul style="list-style-type: none"> □ Batch o procesos continuos, intensidad laboral, materiales en crudo y maquinaria. □ Perfil de recursos. <ul style="list-style-type: none"> □ Posición actual capital, disponibilidad de capital, características de la fábrica (edad, dirección, restauración, sofisticación). □ Características de los recursos humanos: edad, habilidades, educación y disponibilidad. □ Materiales crudos, merma y almacenamiento. □ Sofisticación de la maquinaria. □ Perfil del Mercado. <ul style="list-style-type: none"> □ Características de los clientes, dinámicas de mercado, abastecimiento y demanda, imagen de la compañía y producto, segmento del mercado, penetración en el mercado y ciclo de vida del producto. □ Características de los proveedores <ul style="list-style-type: none"> □ Número de proveedores, disponibilidad de materiales, relación con la competencia. □ Perfil Ambiental <ul style="list-style-type: none"> □ Regulaciones gubernamentales, impuestos.
Fase 2	Determinar la Dirección Organizacional

	<ul style="list-style-type: none"> □ Políticas para los clientes <ul style="list-style-type: none"> □ ¿Qué segmentos de clientela se buscan? □ ¿Qué nivel de servicio al cliente será provisto? □ Políticas de proceso y productos <ul style="list-style-type: none"> □ ¿Qué productos o líneas de producto se ofrecerán? □ ¿Cómo y cuando se producirán los productos? □ ¿Se maquilarán los productos? □ Políticas de distribución y promoción <ul style="list-style-type: none"> □ ¿Mediante que mecanismo, qué tan rápido y cuando se distribuirán los productos y servicios? □ ¿Qué estrategias y salidas se utilizarán? □ Políticas a proveedores. <ul style="list-style-type: none"> □ ¿Cómo, mediante quien y cuándo se ordenaran los materiales? □ ¿Qué tipo de relaciones contractuales existirán con los proveedores? □ ¿Cómo se medirá la calidad de los materiales? □ Políticas de énfasis competitivo. <ul style="list-style-type: none"> □ ¿Cómo se explotarán las ventajas únicas de la compañía? □ Políticas de precios y créditos. <ul style="list-style-type: none"> □ ¿Qué precios y estructura de crédito debe ser seleccionada? □ Políticas de Financiamiento. <ul style="list-style-type: none"> □ ¿Qué es una proporción aceptable de débito? □ ¿Cómo se adquirirán nuevas fuentes de capital?
--	---

	<ul style="list-style-type: none"> □ ¿Qué tipo de políticas de dividendos perseguirá la compañía? □ Políticas de Inversión. □ ¿Cómo se asignarán los recursos al desarrollo de productos, manufactura, recursos humanos y expansión?
<p>Fase 3</p>	<p>Identificar la cardinalidad entre entidades meta-nivel</p> <ul style="list-style-type: none"> □ Provee una penetración en las operaciones y complejidad organizacional. □ Es una buena manera de medir “la intensidad informativa” <ul style="list-style-type: none"> □ Las compañías con intensidad informativa tienen grandes números de proveedores y clientes, numerosos productos y partes y un gran número de pasos en el proceso de manufactura. □ Las compañías con intensidad informativa, pueden utilizar las tecnologías de información como un arma competitiva
<p>Fase 4</p>	<p>Utilice el Modelo de Datos Genérico para Identificar los Puntos de Oportunidad Tecnológica (POT)</p> <ul style="list-style-type: none"> □ Desarrolla soluciones basadas en Bases de Datos y tecnología para incrementar la eficiencia de entradas y salidas. □ Desarrolla soluciones basadas en Bases de Datos y tecnología para diferenciar productos de la competencia. □ Desarrolla soluciones basadas en Bases de Datos y tecnología para mejorar la posición competitiva. <p>Realice preguntas relacionadas con ¿Cómo la tecnología puede</p>

utilizarse para mejorar la eficiencia y la competitividad?. Refiérase a la figura 3.4, Puntos de Oportunidad Tecnológica. La siguiente lista de preguntas puede utilizarse como una lista de chequeo

- ¿Qué información del cliente se debe obtener?
- ¿Cómo pueden utilizarse las tecnologías de información para soportar la recolección y mantenimiento de información de clientes?
- ¿Cómo pueden utilizarse las tecnologías de información para desarrollar productos y servicios para diferentes segmentos de mercado?
- ¿Qué información se necesita para soportar la entrada de las ordenes?
- ¿Cómo pueden las tecnologías de información utilizarse para soportar el proceso de entrada de ordenes?
- ¿Qué información es necesaria para soportar el servicio al cliente?
- ¿Cómo pueden utilizarse las tecnologías de información para soportar el servicio al cliente?
- ¿Qué tipo de información debe obtenerse para mejorar el proceso de producción?
- ¿Qué indicadores de calidad debe monitorear la compañía?
- ¿Cómo pueden utilizarse las tecnologías de información para soportar el control de calidad?
- ¿Qué información es necesaria para soportar el despliegue y mantenimiento de recursos humanos?
- ¿Cómo pueden utilizarse las tecnologías de información para

	<p>soportar los recursos humanos?</p> <ul style="list-style-type: none"> □ ¿Qué información es necesaria para soportar la calendarización de producción? □ ¿Cómo pueden emplearse las tecnologías de información para soportar la calendarización de producción? □ ¿Cómo debe controlarse y monitorearse el mantenimiento a las máquinas? □ ¿Cómo pueden utilizarse las tecnologías de información para soportar el mantenimiento a las máquinas? □ ¿Cómo debe la compañía ordenar a los proveedores? □ ¿Cómo debe monitorearse la calidad de los proveedores? □ ¿Cómo pueden utilizarse las tecnologías de información para soportar los Sistemas de Información de proveedores?
<p>Fase 5</p>	<p>Desarrollar el Modelo de Datos de la Empresa</p> <ul style="list-style-type: none"> □ Necesidad basada en gran involucramiento □ El modelo de datos puede ser un subconjunto del modelo de datos empresarial. □ Minimizar la larga guía de tiempo entre la construcción del modelo de datos y las aplicaciones consiguientes. <ul style="list-style-type: none"> □ Regla del 80/20 – 80% del beneficio de la modelación de datos, será realizada con el 20% de esfuerzo. □ Regla de los 6 meses – limitar el proceso de modelación de datos a no más de 6 meses.

Figura 3.3 Fases de la planeación estratégica de bases de datos.

Dentro de las ventajas de una planeación estratégica de Bases de Datos tenemos que provee una ilustración gráfica para identificar aquellos puntos de oportunidad tecnológica dentro de la empresa, además de que asiste para el entendimiento de cómo las nuevas tecnologías encajan con la infraestructura existente y añade una estructura a la actividad conceptual del proceso de planeación estratégica de la organización.

Esta planeación estratégica provee una plataforma de comunicación entre directivos, staff y accionistas relevantes, forzando a los gerentes de alto nivel a buscar dentro de las operaciones por oportunidades estratégicas y alienta al desarrollo de una perspectiva estratégica hacia los requerimientos de información de la organización.

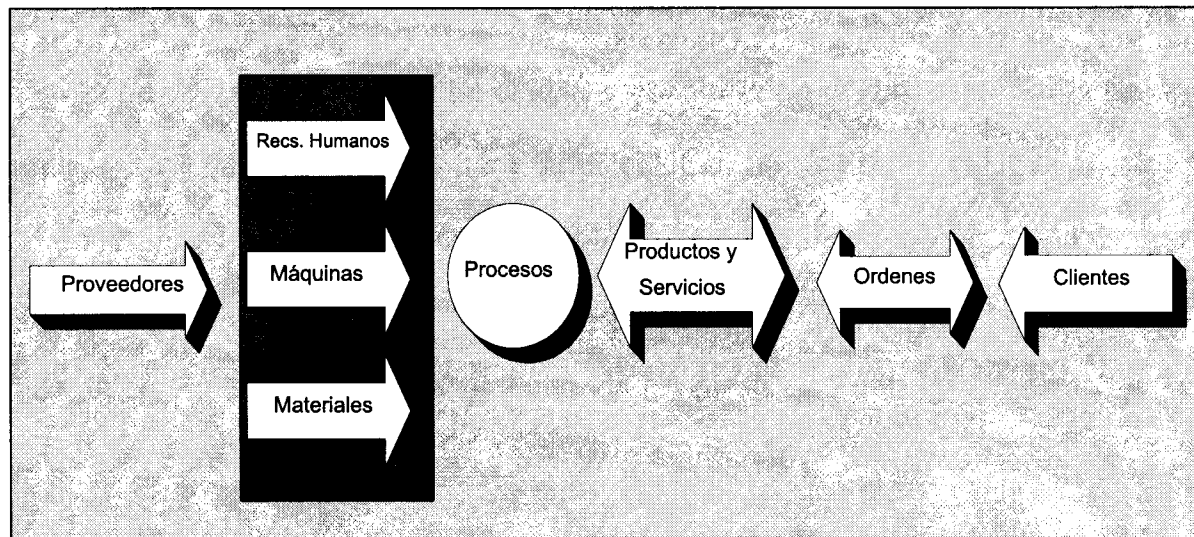


Figura 3.4 Puntos de Oportunidad Tecnológica

Los factores críticos de éxito de la planeación estratégica de bases de datos, son algunas áreas críticas que la organización tiene que atender en orden de ser exitosa y estos factores deben implementarse en términos de metas operacionales que son monitoreadas utilizando las tecnologías de información. Por ejemplo: si es esencial que nos concentremos en el diseño y características del producto, se debe incrementar la porción del mercado y la satisfacción del cliente.

La línea de trabajo de Zachman

En 1987, John Zachman publicó un documento en *IBM Systems Journal*, el cual se considera como un trabajo clásico sobre conceptos de arquitectura de Sistemas de Información. El documento de Zachman fue significativo por diferentes razones. Primero, identifica una línea de trabajo de seis niveles de arquitecturas, iniciando con los niveles conceptuales desde el punto de vista estratégico y desde el punto de vista del dueño, a través del diseño y construcción de un sistema. Zachman explica su línea de trabajo para arquitecturas de Sistemas de Información, utilizando una analogía al proceso de planeación, bosquejo y construcción de una casa. Por esta razón, él utiliza los términos: punto de vista del dueño, del diseñador y del constructor, incrementando los niveles de detalle apropiados para cubrir el propósito.





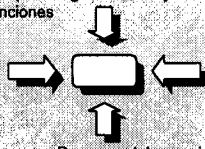
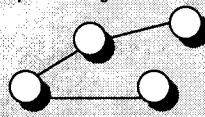
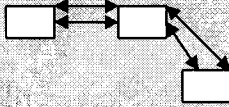
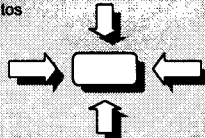

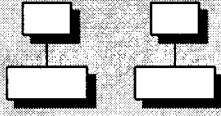
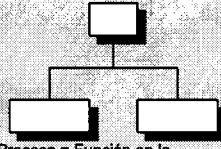
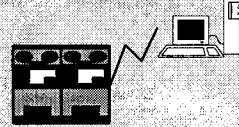



Tal como Zachman apunta, es importante considerar la analogía de que el proceso de desarrollo de un sistema de información, es muy parecido a la construcción de una casa, que requiere de varios niveles o iteraciones de arquitecturas. Como primer punto uno debe decidir en que tipo de casa se desea

habitar, como será, cuantos cuartos deberá tener, etc. Durante esta fase de planeación, uno puede encontrarse con una variedad de arquitecturas y bosquejos antes de decidir que opción tomar. Pero el hecho es que, las arquitecturas tienen un pequeño significado real o valor hasta que los planes se solidifican para contratar al que se encargará de la construcción de la casa. Para construir la casa ambos, la arquitectura y los planes para su implantación.

El segundo aspecto significativo de la línea de trabajo de Zachman fue, una clara definición y distinción entre los tres tipos de arquitecturas: datos, procesos (aplicaciones) y redes (tecnología). La línea de trabajo de Zachman fue un descubrimiento al identificar claramente los niveles difiriendo el detalle del propósito de arquitecturas y las tres dimensiones: datos, aplicaciones y tecnología. La figura 3.5 es una representación de la línea de trabajo de Zachman.

El Metalenguaje como apoyo a la Arquitectura de Datos

El metalenguaje, siendo una herramienta para la estandarización de la modelación, basado en el fondo matemático de este y de acuerdo a lo antes expuesto por Zachman, quien nos dice que para la construcción de las arquitecturas de la organización tenemos que considerar los diferentes puntos de vista de las áreas involucradas en el proyecto. La base matemática de esta herramienta, apoya la modelación de la arquitectura de datos de la organización de manera objetiva, eliminando de manera sustancial la subjetividad resultante de cualquier proyecto de análisis.

	<i>Datos</i>	<i>Funciones</i>	<i>Red</i>
<p>Objetivos / Alcance</p> <p><i>Punto de Vista Estratégico</i></p>	<p>Lista de aspectos importantes para el negocio</p>  <p>Entidad = Clase de aspectos del negocio</p>	<p>Lista de procesos que desempeña el negocio</p>  <p>Proceso = Clase de procesos del negocio</p>	<p>Lista de locaciones donde opera el negocio</p>  <p>Nodo = Locación del negocio</p>
<p>Modelo del Negocio</p> <p><i>Punto de Vista del dueño</i></p>	<p>Ejemplo: Diagrama Entidad - Relación</p>  <p>Entidad = Entidad del Negocio Relación = Regla del Negocio</p>	<p>Ejemplo: Diagrama de Flujo de Funciones</p>  <p>Proceso = Procesos del negocio E/S = Recursos del Negocio (Incluyendo Información)</p>	<p>Ejemplo: Red Logística</p>  <p>Nodo = Unidad del Negocio Liga = Relación del Negocio (Organización, Producto e Información)</p>
<p>Modelo del Sistema de Información</p> <p><i>Punto de Vista del diseñador</i></p>	<p>Ejemplo: Modelo de Datos.</p>  <p>Entidad = Entidad de datos Relación = Relación de datos</p>	<p>Ejemplo: Diagrama de Flujo de Datos</p>  <p>Proceso = Función de aplicación E/S = Vistas del usuario (Conjunto de elementos de datos)</p>	<p>Ejemplo: Arquitectura de sistemas distribuidos</p>  <p>Nodo = almacenamiento de info. (Procesador, almacenamiento, etc.) Liga = línea de caracteres.</p>
<p>Modelo Tecnológico</p> <p><i>Punto de Vista del constructor</i></p>	<p>Ejemplo: Diseño de Datos.</p>  <p>Entidad = Segmento / Ranglón Relación = Apuntador / Llave</p>	<p>Ejemplo: Organigrama.</p>  <p>Proceso = Función en la computadora E/S = Pantallas/Formatos</p>	<p>Ejemplo: Arquitectura del Sistema</p>  <p>Nodo = Hardware /Software Liga = Especificaciones de línea.</p>
<p>Representaciones Detalladas</p> <p><i>Punto de Vista fuera de contexto</i></p>	<p>Ejemplo: Descripción del diseño de datos</p>  <p>Entidad = Campos. Relación = Direcciones</p>	<p>Ejemplo: Programa</p>  <p>Proceso = Lenguaje. E/S = Bloques de Control</p>	<p>Ejemplo: Arquitectura de la red.</p>  <p>Nodo = Dirección. Liga = Protocolos.</p>
<p>Sistema Funcionando</p>	<p>Ejemplo: Datos</p>	<p>Ejemplo: Función</p>	<p>Ejemplo: Comunicaciones</p>

E/S = Entradas / Salidas

Figura 3.5 Línea de trabajo de Zachman

Metalinguaje KIF

El formato de intercambio de conocimiento (Knowledge Interchange Format (KIF)) es un lenguaje diseñado para utilizarse en el intercambio de conocimiento entre diversos sistemas de computo (creados por diferentes programadores, en momentos diferentes, en lenguajes diferentes, etc.).

KIF no fue creado con la intención de ser una representación interna de conocimiento dentro de un sistema o dentro de un conjunto de sistemas (sin embargo el lenguaje puede utilizarse con este propósito también).

Típicamente, cuando un sistema lee un conocimiento basado en KIF, convierte los datos en su forma interna propia (estructuras de apuntadores, arreglos, etc.). Toda la computación se realiza utilizando estas formas internas. Cuando un sistema necesita comunicarse con otro sistema, mapea sus estructuras internas de datos en KIF.

El propósito de KIF es bruscamente análogo al del postscript. Postscript es utilizado por sistemas que dan formato a gráficas y texto que comunican información acerca de documentos a impresoras. Aunque no es tan eficiente como una representación especializada para documentos, ni tan entendibles como un despliegue de ayuda especializado, Postscript es una representación cuya lectura es accesible para los programadores, que facilita el desarrollo independiente de programas de estructuración e impresoras. Mientras KIF no es tan eficiente como una representación especializada para el conocimiento, ni tan entendible como un despliegue especializado (cuando se imprime en forma de lista), también es un

lenguaje de fácil lectura para los programadores y por lo tanto facilita el desarrollo independiente de programas de manipulación del conocimiento. [*Knowledge Interchange Format*]

Las siguientes características son esenciales para el diseño de KIF:

1. El lenguaje tiene semántica declarativa. Es posible entender el significado de expresiones en el lenguaje sin apelación a un interprete para manipular esas expresiones. De esta manera, KIF difiere de otros lenguajes que son basados en interpretes específicos, como Emcyn y Prolog.
2. El lenguaje es lógicamente comprensivo, provee de sentencias lógicas arbitrarias. De esta manera, difiere de lenguajes de bases de datos relacionales (como SQL) y lenguajes de programación lógica (como Prolog).
3. El lenguaje provee la representación de conocimiento sobre conocimiento. Esto permite al usuario hacer representación del conocimiento de decisiones explícitas y permite al usuario introducir nuevas representaciones de conocimiento sin modificar el lenguaje.

En adición a estas características esenciales, KIF está diseñado para maximizar las siguientes características (a la magnitud posible mientras conserve los rasgos precedidos)

- Desempeño. KIF no tiene la intención de utilizarse dentro de programas como una representación o lenguaje de comunicación, debe ser utilizado para ese propósito si así se desea.

-
- **Comprensible.** KIF no tiene como intención primaria un lenguaje para interacción con humanos, la comprensión humana facilita su uso describiendo semántica de representación de lenguaje, se utiliza como un lenguaje de publicación, por ejemplo bases de conocimiento, su utilización en la asistencia a humanos con problemas de traducción de basados en conocimiento.

Así como en varios lenguajes orientados a las computadoras, la sintaxis de KIF es fácilmente descrita en tres capas. Primero tenemos los caracteres básicos del lenguaje. Estos caracteres pueden ser combinados para formar lexemas y finalmente, los lexemas del lenguaje pueden ser combinados para formar expresiones legales gramáticamente. Sin embargo ésta división en capas no es estrictamente esencial para la especificación de KIF, simplifica la descripción de la sintaxis lidiando con los espacios en blanco en el nivel lexema y eliminando el detalle del nivel de expresión.

La base de la semántica para KIF, es la conceptualización del mundo en términos de objetos y relaciones entre estos objetos.

Un universo de discusión, es un conjunto de todos los objetos existentes en el mundo, presuntos o supuestos. La notación de objetos utilizados aquí es bastante extensa. Los objetos pueden ser concretos (ejemplo: un átomo de carbón específico, Confusio, el sol) o abstracto (ejemplo: el número 2, el conjunto de enteros, el concepto de justicia).

Los objetos pueden ser primitivos o compuestos (ejemplo: un circuito que consiste de varios subcircuitos). Los objetos pueden ser ficticios (ejemplo: un unicornio, Sherlock Holmes).

Diferentes usuarios de un lenguaje de representación declarativa como KIF, es probable que tengan universos de discurso diferente. Por otro lado, en KIF está conceptualmente establecido, que en cada universo de disertación sea requerido incluir ciertos objetos básicos.

Los siguientes objetos básicos deben suceder en cualquier universo de disertación.

- Todos los números, reales y complejos.
- Todos los caracteres ASCII
- Todas las cadenas finitas de caracteres ASCII.
- Palabras. Las palabras son objetos por sí solos en el universo de disertación, sobre las cosas que representan.
- Todas las listas finitas de objetos en el universo de disertación.
- *raíz*. Un objeto distinguido que ocurre como el valor de un parcial cuando una función es aplicada a argumentos cuya función no hace sentido.

Recuerde que para estos elementos básicos, el usuario puede adicionar cualquier objeto no-básico que parezca útil.

En KIF, el parentesco entre objetos toma la forma de relación. Formalmente, una relación es definida como un conjunto arbitrario de listas finitas de objetos (de

longitudes posiblemente variables). Cada lista es una selección de objetos que juntamente satisfacen una relación. Por ejemplo, la relación $<$, contiene la lista $\langle 2,3 \rangle$, indicando que 2 es menor que 3.

Una función es un tipo especial de relación. Para cada secuencia finita de objetos (llamada argumentos), una función asocia un objeto único (llamado valor). Más formalmente, una función es definida como el conjunto finito de listas de objetos, una por cada combinación de argumentos posibles. En cada lista, los elementos iniciales son los argumentos y el elemento final es el valor. Por ejemplo, la función $1+$, contiene la lista $\langle 2,3 \rangle$, indicando que el sucesor de 2 es 3.

Note que ambas funciones y relaciones son definidas como conjuntos de listas. De hecho, cada función es una relación. Sin embargo, no cada relación es una función. En una función, no pueden existir dos listas que desacuerden en sólo el último elemento. Esto sería equivalente a la función, teniendo dos valores para una combinación de argumentos. Por contraste, en una relación, pueden existir cualquier número de listas que concuerden en todo, menos el último elemento. Por ejemplo, la lista $\langle 2,3 \rangle$ es un miembro de la función $1+$ y no hay otra lista de longitud 2, con 2 como su primer argumento, existe solo un sucesor de 2. En contraposición, la relación $<$, contiene las listas $\langle 2,3 \rangle$, $\langle 2,4 \rangle$, y en adelante, indicando que 2 es menor que 3, 4, 5 y en adelante.

Muchos matemáticos requieren que las funciones y las relaciones tengan paridad fija, requieren que todas las listas comparando una relación, tengan la

misma longitud. Las definiciones aquí permitidas con paridad variable, son perfectamente aceptadas para que en una función o relación contenga listas de diferentes longitudes. Por ejemplo, la relación $<$, contiene las listas $\langle 2,3 \rangle$ y $\langle 2,3,4 \rangle$, reflejando el hecho de que 2 es menor que 3 y el hecho de que 2 es menor que 3 y 3 es menor que 4. Esta flexibilidad no es esencial, pero es en extremo conveniente y posee problemas teóricos insignificantes.

En KIF, todas las funciones son totales, existe un valor para cada combinación de argumentos. A manera de permitir a un usuario expresar la idea de que una función no tiene significado para ciertos argumentos, KIF asume que existe un objeto especial indefinido en el universo y provee la constante raíz para referirse a este objeto.

El valor de un término funcional sin una variable de secuencia terminante, se obtiene de aplicar la función denotada por la función constante en el término a los objetos, denotados por los argumentos.

Por ejemplo, el valor del término $(+ 2 3)$ se obtiene aplicando la función de adición (la función denotada como $+$) a los números 2 y 3 (los objetos denotados por las constantes 2 y 3) para obtener el valor 5, que es el valor del objeto constante 5.

Sí un término funcional tiene una variable de secuencia terminante, el valor se obtiene de aplicar la función a la secuencia de argumentos formada por los valores

de los términos que preceden a la secuencia variable y los valores en la secuencia, denotados por la variable de secuencia.

Asuma por ejemplo que la variable secuencia @1, tiene un valor de secuencia 2, 3, 4. Entonces, el valor del término (+ 1 @1) se obtiene de aplicar la función de adición a los números 1, 2, 3 y 4, para obtener el valor 10, el cual es el valor de la constante de objeto 10.

Una sentencia simple relacional sin una variable de secuencia terminante, es verdadera si y solo si la relación es denotada por la relación constante en la sentencia, es verdadera en los objetos denotados por los argumentos.

Igualmente, al ver una relación como un conjunto de tuplas, decimos que la sentencia es verdadera, si y solo si, la tupla de objetos, formada por los valores de los argumentos es miembro de un conjunto de tuplas, denotado por la relación constante.

Sí una sentencia relacional termina en una variable de secuencia, la sentencia es verdadera, si y solo si, la relación contiene la tupla consistente de los valores de los términos que preceden a la variable secuencia junto con los objetos en la secuencia denotada por la variable.

Una ecuación es verdadera si y solo si, los términos en la ecuación refieren al mismo objeto en el universo de disertación.

Una desigualdad es verdadera si y solo si, los términos en la ecuación refieren a distintos objetos en el universo de disertación.

Nivel Carácter

El alfabeto de KIF, consiste de siete bloques de datos. En este documento, nos referimos a los bloques de datos KIF vía su código ASCII usual como caracteres (dado en ISO 646:1983)

Los caracteres KIF son clasificados como letras mayúsculas, minúsculas, dígitos, caracteres alfanuméricos (los caracteres no alfabéticos son utilizados de la misma manera que las letras), caracteres especiales, espacio y otros caracteres (cualquier caracter ascii que no esté en ninguna de las categorías anteriores).

mayúsculas	:	:	=		A		B		C		D		E		F		G		H		I		J		K		L		M	
					N		O		P		Q		R		S		T		U		V		W		X		Y		Z	
minúsculas	:	:	=		a		b		c		d		e		f		g		h		i		j		k		l		m	
					n		o		p		q		r		s		t		u		v		w		x		y		z	
dígitos	:	:	=		0		1		2		3		4		5		6		7		8		9							
alfa	:	:	=		!		\$		%		&		*		+		-		.		/		<		=		>		?	
					@		_		~																					
especiales	:	:	=		"		#		'		()		,		\		^		'									
blancos	:	:	=		space		tab		return		linefeed		page																	

Un caracter normal puede ser un caracter de mayúscula, minúscula, un dígito o un caracter alfa.

normal	:	:	=	mayúscula	minúscula	dígito	alfa
--------	---	---	---	-----------	-----------	--------	------

Nivel Lexema

El proceso de convertir caracteres en lexemas se llama análisis léxico. La entrada a este proceso es un flujo de caracteres y su salida es un flujo de lexemas.

La función de un analista léxico es cíclica. Lee caracteres de una cadena hasta encontrar un caracter que no puede ser combinado con caracteres previos para formar un lexema legal. Cuando esto sucede, da como resultado el lexema correspondiente a la lectura previa de caracteres y comienza nuevamente el proceso con el nuevo caracter. Los espacios en blanco causan una ruptura en el proceso de análisis léxico, pero de otra manera se descarta.

Existen cinco tipos de lexema en KIF – lexemas especiales, palabras, referencias a caracteres, cordones de caracteres y bloques de caracteres.

Cada caracter especial forma su propio lexema. No puede combinarse con otro caracter para formar lexemas más complejos, excepto a través de la sintaxis de escape, descrita a continuación.

Una palabra es una secuencia contigua de caracteres normales u otros caracteres precedidos por un caracter de escape \.

palabra : : = | normal | palabra normal | palabra\caracter |

Es posible incluir el caracter \ en una palabra precediéndola por otra ocurrencia de \. Dos ocurrencias contiguas de \, es interpretada como una sola ocurrencia. Por ejemplo, la cadena A \\ \ 'B corresponde a la palabra consistente de cuatro caracteres: A, \, ', y B.

Excepto por caracteres que siguen a \, el análisis léxico de palabras es insensible. El lexema de salida para cualquier palabra, corresponde al lexema obtenido de convertir todas las letras a las que no le siguiera \ a su mayúscula equivalente. Por ejemplo, la palabra abc y la palabra ABC, apuntan al mismo lexema. La palabra a\bc apunta al mismo lexema, tal como la palabra A\bC, la cual no es el mismo lexema para la palabra ABC, ya que el segundo caracter es minúscula.

Un caracter referencia consiste de caracteres #, \ y cualquier otro caracter. Las referencias a caracteres, nos permiten referirnos a caracteres como caracteres y diferenciarlos de un caracter símbolo, el cual puede referirse a otros objetos.

caracter referencia : : = | #\caracter |

Un caracter cadena es una serie de caracteres adjuntos entre comillas. El caracter de escape \, es utilizado para permitir la inclusión de comillas y el mismo caracter \ dentro de dichos cordones.

cadena : : = | "entre comillas" |

entre comillas : : = | vacío | caracter cadena | entre comillas |

entre comillas | \ caracter |

caracter cadena : : = | caracter – {“,\} |

Algunas veces es deseable agrupar una secuencia de bits o caracteres arbitrariamente sin imponer caracteres de escape, ya sea para codificar imágenes, audio o video en formatos especiales. Los bloques de caracteres permiten esta clase de agrupamiento a través del uso de prefijos que especifican cuantos de los siguientes caracteres son elegibles para agruparlos de esta manera. Un bloque de caracteres consiste del caracter #, seguido del código decimal de un entero positivo N, el caracter q o Q y después n caracteres arbitrariamente.

bloque : : = | # int (n) q caracter ^ n | # int (n) Q caracter ^ n |

Para propósitos de análisis gramatical, es útil subdividir la clase de palabras un poco más así como variables, operadores y constantes.

Una variable es una palabra donde el primer caracter es ? o @. Una variable que comienza con ?, es llamada una variable individual. A una variable que comienza con @ se le conoce como variable de secuencia.

variable : : = | variable individual | Variable de secuencia |

variable individual : : = | ? palabra |

variable secuencia : : = | @ palabra |

Los operadores son utilizados para formar expresiones complejas de varias clases. Existen tres tipos de operadores en KIF, de término, de sentencia y de definición. Los operadores de término son utilizados para formar condiciones complejas. Los operadores de sentencia y operadores de usuario, son utilizados

para formar sentencias complejas. Los operadores de definición son utilizados para formar definiciones.

operador	:	:	=		op. término		op. sentencia		op. definición	
op. término	:	:	=		valor		lista		cita	
op. sentencia	:	:	=		dominio		=		/=	
					para todos		existe		no	
									y	
									o	
									=>	
									<=	
									<=>	
op. definición	:	:	=		definición		definición		definición	
					objeto		función		relación	
					:=		:->		:<=	
									definición	
									lógica	
									:=>	

Todas las demás palabras, son llamadas constantes

Constante : : = Palabra – Variable - operador

Semánticamente existen cuatro categorías de constantes en KIF: constantes objeto, constantes función, constantes relación y constantes lógicas. Las constantes objeto son utilizadas para denotar objetos individuales. Las constantes función denotan funciones sobre esos objetos. Las constantes relación denotan relaciones. Las constantes lógicas expresan condiciones acerca del mundo y son ya sea, verdaderas o falsas. KIF es raro entre los lenguajes lógicos en los que no hay distinción sintáctica entre estos cuatro tipos de constantes; cualquier constante puede usarse donde cualquier otra constante sea usada. Las diferencias entre estas categorías de constantes es enteramente sintáctica.

Nivel Expresiones

Las expresiones legales de KIF son formadas de lexemas de acuerdo a las reglas presentadas en esta sección. Existen tres tipos de expresiones de

separación en el lenguaje – términos, sentencias y definiciones. Los términos son utilizados para denotar objetos en el mundo que se está describiendo; las sentencias son utilizadas para expresar hechos acerca del mundo y las definiciones son utilizadas para definir constantes. Las definiciones y las sentencias son llamadas formas. Un conocimiento base, es un conjunto finito de formas.

Existen nueve tipos de términos en KIF – variables individuales, constantes, referencia a caracteres, cadenas, bloques de caracteres, términos funcionales, términos lista, citas y términos lógicos. Las variables individuales, las constantes, las referencias a caracter, cadenas y bloques fueron discutidas anteriormente.

término	:	:	=	variable individual	constante	referencia a caracter	cadena	bloque
				término funcional	término lista	término cita	término lógico	

Un término funcional implícito, consiste de una constante y un número arbitrario de términos argumento, concluyendo con una variable de secuencia opcional y rodeado de paréntesis. Note que no existe restricción de sintaxis en el número de argumentos término; las restricciones de paridad en KIF, son tratadas semánticamente.

término funcional : : = (término constante * [variable secuencial])

Un término explícito funcional, consiste del valor operador y uno o más términos argumento, concluyendo con una variable de secuencia opcional y rodeado de paréntesis.

término funcional : : = (valor término término * [variable secuencial])

Un término lista, consiste de un operador lista de y una lista finita de términos, concluyendo con una variable de secuencia opcional y rodeado de paréntesis.

término lista : : = (lista de término * [variable secuencial])

Una equidad consiste del operador = y dos términos y una inequidad consiste de los operadores /= y dos términos. Una sentencia implícita relacional consiste de una constante y un número arbitrario de términos argumento, terminado de una variable de secuencia opcional. Tal como en términos funcionales, no existe restricción sintáctica en el número de términos argumento en la sentencia relacional.

Sentencia relacional : : = (constante término* [variable secuencial])

Una sentencia relacional explícita, consiste del operador mantiene y uno o más argumentos, terminado por una variable opcional de secuencia y rodeado de paréntesis.

Sentencia relacional : : = (mantiene término término* [variable secuencial])

Un conocimiento base en un conjunto de formas finitas. Es importante mantener en cuenta que el conocimiento base, es un conjunto de sentencias, no una sentencia y por lo tanto el orden de formas dentro de un conocimiento base no es menos importante. El orden puede tener valores heurísticos para programas deductivos sugiriendo un orden en donde utilizar aquellas sentencias, sin embargo, este acercamiento implícito al intercambio de conocimiento radica fuera de la definición de KIF.

Metalenguaje CES

El interés sobre utilización de métodos empíricos de la comunidad de lenguaje de ingeniería dio inicio a la demanda de un cuerpo a gran escala. Existen numerosos esfuerzos sobre obtención de datos en ambos lados del Atlántico para proveer el acceso a una cobertura ancha a ambos recursos monolingües y bilingües, de tamaño suficiente para trabajos orientados a datos, incluyendo en Estados Unidos de América al Consorcio Lingüístico de Datos, el cuerpo de Iniciativa de Europeo (European Corpus Initiative (ECI)), ICAME, el Cuerpo Nacional Británico (British National Corpus (BNC)), y la Asociación de Recursos Europeos de Lenguaje (European Language Resources Association (ELRA)). La rápida multiplicación de dichos esfuerzos ha hecho crítico para la comunidad de lenguaje de ingeniería crear un conjunto de estándares para codificar.

CES, (Corpus Encoding Estándar), está diseñado satisfacer óptimamente el uso del lenguaje de investigaciones y aplicaciones de ingeniería, en orden de utilizarlo como un conjunto de estándares, trabajados en lenguaje natural de aplicaciones de procesamiento. . El CES es una aplicación de SGML (ISO 8879:1986, Procesamiento de Información--Texto y Sistemas de Oficina--lenguaje de encarecimiento general) conforme a las especificaciones de TEI, guías para Codificación Electrónica de Texto e Iniciativa de Intercambio sobre Codificación de Texto.

El CES especifica el nivel mínimo de codificación que se debe cumplir para ser considerado estándar, en términos de una representación descriptiva (señalando

información estructural y tipográfica), así como arquitecturas generales (tal como maximizar el uso de texto en bases de datos). También provee especificaciones de codificación para notación lingüística, junto con la arquitectura de datos del cuerpo lingüístico.

CES se desarrollo de acuerdo al método bottom up, comenzando con un mínimo de especificaciones y expansión basado en retroalimentación resultado de su uso y la entrada de la comunidad de investigación en general.

CES aplica al cuerpo monolingüe incluyendo textos de una variedad de lenguajes Europeos de este y oeste, así como el cuerpo multi-lingüe y paralelo comprendiendo textos en cualquiera de estos lenguajes.

CES esta proyectado para utilizarse en cuerpos de codificación de lenguaje de ingeniería, incluyendo todas las áreas de procesamiento natural de lenguaje, traducción máquina, lexicografía, etc.

El cuerpo se utiliza en lenguaje de ingeniería para obtener evidencia real del lenguaje, tanto cuantitativa como cualitativa. La evidencia cualitativa consiste de ejemplos que pueden ser utilizados para la construcción de léxico y gramática computacional, léxicos multi-lenguajes y bancos de términos para lexicografía, etc. La información cualitativa consiste de estadísticas que indican frecuencias o utilización de características de lenguaje. Estas estadísticas pueden ser solo utilizadas. Estas estadísticas pueden utilizarse para asistir a la lexicografía,

determinar estadísticas de traducción, etc. Las operaciones del cuerpo para propósitos de lenguaje de ingeniería incluyen la extracción de sub-cuerpos, búsquedas sofisticadas, incluyendo extracciones, generación de concordancia generación de listas de elementos lingüísticos, etc. Y la determinación de estadísticas tales como la frecuencia de información, promedios, etc. [*Corpus Encoding Standard*]

En la figura 3.6 encontramos un ejemplo de cómo CES se utiliza como metalenguaje en base de datos.

```
<!ENTITY % ISOLat1 PUBLIC "ISO 8879-1986//ENTITIES Added Latin 1//EN">
%ISOLat1;

<!ENTITY % ISOLat2 PUBLIC "ISO 8879-1986//ENTITIES Added Latin 2//EN">
%ISOLat1;

<!ENTITY % ISOGrk1 PUBLIC "ISO 8879-1986//ENTITIES Greek Letters//EN">
%ISOGrk1;

<!ENTITY % ISOGrk2 PUBLIC "ISO 8879-1986//ENTITIES Monotoniko
Greek//EN">
%ISOGrk2;

<!ENTITY % ISOCyr1 PUBLIC "ISO 8879-1986//ENTITIES Russian
Cyrillic//EN">
%ISOCyr1;

<!ENTITY % ISOCyr2 PUBLIC "ISO 8879-1986//ENTITIES Non-Russian
Cyrillic//EN">
%ISOCyr2;

<!ENTITY % ISOPUB PUBLIC "ISO 8879-1986//ENTITIES Publishing//EN">
%ISOPUB;

<!ENTITY % ISONUM PUBLIC "ISO 8879-1986//ENTITIES Publishing//EN">
%ISONUM;
```

Figura 3.6 Ejemplo de Metalenguaje CES en bases de datos

Reingeniería Automatizada de Programas y Datos a Nivel Empresa Xinotech

La tecnología de Xinotech se ha construido para dar soporte, a nivel empresarial, a la reingeniería de sistemas COBOL y Año 2000. Altamente adaptable a través de los metalenguajes de Xinotech, esta tecnología provee soluciones de reingeniería de un solo ciclo, para problemas múltiples, incluyendo el Año 2000, integración a la moneda común Europea Euro, y estandarización bajo el sistema internacional de identificación en la bolsa de valores (ISIN), que es administrado por organizaciones como la CUSIP de Standard & Poor en USA, SICOVAM en Francia, Telekurs, y Wertpapier Mitteilungen en Alemania. 2001 es adaptable a soluciones de problemas específicos del cliente. Esta tecnología provee soporte a soluciones múltiples para el Año 2000, tales como expansión a cuatro dígitos, compresión, técnicas de ventana, y puentes de datos. Esta tecnología provee soporte a fases de reingeniería como análisis de impacto, minería de datos, y conversión automatizada de componentes tales como programas COBOL, archivos VSAM y secuenciales, bases IMS y DB2, CICS, BMS, JCL, SORT, y DEC DCL. 2001 es adaptable a otros lenguajes y componentes. [*Reingeniería Automatizada de Programas y Datos a Nivel Empresa*]

Acerca de Xinotech

Xinotech Research fue fundada en 1983 en Minneapolis, Minnesota, USA, con un enfoque exclusivo a la investigación, desarrollo, y comercialización de tecnología para el análisis y la transformación automatizada de programas, con el objetivo de proveer soporte para la reingeniería y modernización de aplicaciones pre-existentes. Durante los últimos cinco años, Xinotech ha sido seleccionada nacionalmente para recibir contratos de investigación de parte de las agencias de investigación de mas empuje en los Estados Unidos de América (16 contratos de 7 agencias diferentes). Estas agencias federales se rigen por un proceso selectivo muy riguroso con el objetivo de identificar las tecnologías mas avanzadas e innovadoras en los Estados Unidos, en este caso, en el área de reingeniería de software.

Por ejemplo, la Defense Advanced Research Projects Agency (DARPA), el organismo central para la investigación y desarrollo del Departamento de Defensa de los Estados Unidos (DoD), seleccionó a Xinotech en 1991 y nuevamente en 1993, como la tecnología de mas avance en el área de transformación automática de programas. Y en 1996, después que esta tecnología ha sido evaluada en proyectos de reingeniería, DARPA seleccionó a Xinotech una vez más para otro contrato que proveerá avances continuos de investigación y desarrollo por los siguientes tres años.

Otras agencias de los Estados Unidos que han seleccionado nacionalmente a Xinotech y le han otorgado contratos de investigación en esta área incluyen, la National Science Foundation (NSF), Ada Joint Program Office (AJPO), Edwards Air Force Base, y Navy Research and Development Center (NRaD). Asimismo, Xinotech ha desarrollado acuerdos de cooperación de investigación con el ARMY Missile Command Software Engineering Directorate.

La tecnología de Xinotech es única por el hecho de que está soportada por cuatro metalenguajes. XML, el Metalenguaje de Xinotech, es usado para describir las reglas sintácticas de lenguajes tales como COBOL, y SQL, y también las reglas de formato y de estandarización estética textual de programas fuentes. XSSL, el Metalenguaje de Especificación Semántica de Xinotech, es usado por Xinotech para describir las reglas de semántica de estos lenguajes de programación.

XPAL, el Metalenguaje de Abstracción de Planes de Xinotech, es un lenguaje declarativo, de restricciones lógicas, altamente especializado, y provee soporte a la especificación de planes (patrones, clichés), que se usan para describir las características y anomalías en aplicaciones pre-existentes, así como las reglas de transformación de tales planes.

Estos tres metalenguajes hacen a esta tecnología altamente adaptable. La creación por Xinotech de nuevas reglas, permiten que esta tecnología pueda:

- Proveer soporte a múltiples lenguajes y bases de datos y componentes, tales como COBOL, IMS, MFS, DB2, SQL, CICS, JCL, EasyTrieve, etc.
- Resolver múltiples problemas de reingeniería, tales como el Año 2000, conversión a la nueva moneda europea Euro, transformación a nuevos modelos de bases, migración y conversión.

Adaptación es accesible al usuario a través de Typel, el Metalenguaje de Xinotech para el Reconocimiento y Evolución de Tipos de Datos. A pesar de su nombre un poco complicado, Typel es muy sencillo e intuitivo, y no requiere más conocimiento de parte del usuario que el que ya posee acerca de COBOL y las aplicaciones a convertir. Typel permite que el usuario describa los tipos de estructuras de datos en la DATA DIVISION de COBOL que necesitan ser reconocidos en las aplicaciones pre-existentes, y como estas estructuras de datos necesitan ser transformadas.

La tecnología de Xinotech es adaptada completamente para resolver problemas específicos, antes de ser entregada al cliente. Sin embargo, los usuarios mismos pueden, por ejemplo, hacer uso de Typel para corregir, estandarizar, o consolidar formatos heterogéneos de fechas, e incorporar soluciones tales como expansión de años a cuatro dígitos, compresión, técnicas de ventanas y puentes. Usuarios pueden hacer uso de Typel para corregir problemas adicionales que involucran la modernización de estructuras de datos.

Typel es muy intuitivo y no requiere conocimiento adicional fuera del área de COBOL, el problema y la solución deseada, y las aplicaciones a convertir. Typel

es muy poderoso. Con Typel, por ejemplo, usuarios pueden tomar la solución de Xinotech para el Año 2000, y complementarla con soluciones adicionales para conversión de moneda, extensión de zonas postales, números de cuenta, y otros problemas específicos del usuario.

Alcance de la Tecnología

La tecnología de Xinotech provee soporte para la reingeniería de software a nivel empresarial, en las siguientes áreas:

- Manejo de Modelos a Nivel Empresa
 - Extracción de modelos empresariales que ostentan componentes heterogéneas inter-relacionadas, tales como programas de COBOL, IMS, DB2, archivos o ficheros secuenciales, MFS, CICS, JCL.
 - Modelado gráfico para asistir en la toma de decisiones a todo nivel: gerencial, administrativo, y técnico.
 - Análisis de impacto y propagación a nivel empresa. Las características del problema identificado son propagadas automáticamente a través de componentes heterogéneos, para que estructuras de datos relacionadas en otros componentes puedan ser reconocidas y transformadas. Por ejemplo, los efectos de un campo problema en un programa de COBOL, pueden entonces ser identificados y relacionados en pantallas BMS, SORTs, JCL's, bases de datos y archivos, o bien en otros programas COBOL.
 - Desglose del modelo en aplicaciones separadas o proyectos de reingeniería separados.
- Administración Automatizada del Proceso de Reingeniería

- Soporte automatizado para la reingeniería incremental y creación de puentes
- Soporte para minería de datos, pruebas, y generación de datos de prueba.
- Transformación Automática de Programas y Datos
 - Conversión automatizada de programas COBOL, bases IMS y DB2, archivos, JCL y procedimientos SORT, pantallas y generadores de reportes
 - Soporte para la reintegración y coexistencia
 - Detección automatizada y respeto por campos previamente corregidos
- Resolución de Problemas
 - Solución integrada para problemas múltiples (Año 2000, integración a la moneda común Europea, y otros)
 - Soluciones múltiples para cada problema específico (para Año 2000, estas soluciones incluyen expansión a cuatro dígitos, compresión, ventanas y puentes)
 - Adaptaciones a nivel usuario para problemas específicos del cliente

Extracción de Modelos a Nivel Empresa

El análisis de impacto en un proyecto de reingeniería no puede ser llevado a cabo efectivamente sin antes hacer un control de inventario de los componentes en la aplicación o empresa. A su vez, dicho control de inventario no se puede efectuar sin antes extraer el modelo original que describe las relaciones entre los componentes que constituyen el modelo de informática de la empresa. Son estas relaciones las que dictarán las necesidades de inventario.

El Modelo Informático de Empresa que es extraído por las herramientas de Xinotech revela las múltiples relaciones de relevancia entre los componentes heterogéneos que constituyen el modelo completo de informática de la empresa.

Este modelo provee las siguientes características distintivas:

- El modelo es extraído automáticamente de los componentes y programas fuentes, así nunca se vuelve obsoleto: siempre se puede recrear a partir de los programas fuentes.
- El modelo puede ser usado para representar componentes arbitrarios de cualquier clase. Este modelo no está restringido a representar sólo tipos predeterminados de componentes, y puede ser fácilmente extendido para dar soporte a otros tipos adicionales de componentes. Puede ser usado para representar múltiples tipos de bases de datos, programas en lenguajes diferentes, tales como COBOL y FORTRAN, múltiples sistemas de diálogos y pantallas, múltiples lenguajes de control de tareas, etc. En la actualidad, el modelo provee soporte para COBOL (IMS, VMS, II, 370, 74, 85, Micro Focus, Object COBOL), VSAM files, IMS, DB2, IMS y pantallas MFS, CICS y pantallas BMS, JCL, SORTs, y DEC VAX DCL.
- El modelo puede ser usado para representar relaciones arbitrarias entre componentes. Por ejemplo: ¿Cuáles son los segmentos de IMS que este programa lee o modifica? ¿Cuáles son todos los programas que acceden un segmento IMS o una tabla DB2?. Este modelo no está restringido a representar solo cierto tipo de relaciones y puede ser extendido para representar otras relaciones de relevancia.
- Construcción progresiva de modelos. No todos los componentes en el sistema de informática de una empresa necesitan ser tratados de una vez. El modelo es extraído de los componentes que han sido importados por las herramientas de Xinotech. Asimismo, el modelo es actualizado cuando componentes adicionales son importados.

-
- El modelo provee soporte para control de inventario.
 - El modelo provee soporte gráfico para la manipulación jerárquica de componentes. El desplegar componentes a diferentes niveles de abstracción reduce la complejidad y facilita la comprensión. Por ejemplo, para ciertos fines, el usuario puede estar interesado en las relaciones que gobiernan la conducta de una base de datos entera; para otros fines, en las relaciones detalladas que gobiernan cada segmento dentro de la base.
 - Algunas de estas jerarquías son explícitas en el sistema informático de la empresa, tales como el hecho que una base IMS esta compuesta de segmentos, o que un programa de COBOL esta compuesto de párrafos.
 - Otras jerarquías pueden ser comprendidas implícitamente sólo por el usuario, tales como las agrupaciones de programas que forman una aplicación. Con el modelo Xinotech, estas agrupaciones se pueden hacer explícitas por el usuario, con el fin de entender mejor el análisis de impacto, y para distribuir mejor la carga de trabajo durante el diseño de los proyectos para el proceso incremental de reingeniería.
 - Análisis de impacto a nivel empresa. El análisis de impacto provee estadísticas de alcance global a nivel de toda la empresa.

El Modelo Informático de Empresa de Xinotech provee los siguientes beneficios:

- Facilita la comprensión completa del sistema informático de la empresa a través de representaciones gráficas concisas.
- Provee los mecanismos para monitoreo y control de etapas en el proceso de reingeniería de software.
- Provee asistencia a todos los niveles en la toma de decisiones:

- Nivel de Gerencia. Esto incluye soporte para evaluar impacto y para establecer prioridades a las varias aplicaciones que necesitan ser modernizadas. Las decisiones de como desglosar el proceso global de reingeniería en "enjambres" o subsistemas menores, van a ser influenciadas por decisiones gerenciales.
- Nivel administrativo. Esto incluye soporte para la partición del sistema informático en proyectos de reingeniería manejables y hasta donde sea posible, independientes.
- Nivel de Ingeniería. Esto incluye soporte para facilitar la transición de fase de cada proyecto de reingeniería; desde control de inventario hasta reintegración y coexistencia.

Administración Automatizada del Proceso de Reingeniería

El Modelo Informático de Empresa de Xinotech provee soporte en el área de administración del proceso de reingeniería:

- Provee soporte para la reingeniería incremental. A través de las herramientas de Xinotech, todo el sistema empresarial de informática puede ser modelado gráficamente. Las herramientas para elaborar modelos de Xinotech pueden entonces ser usadas para asistir en el proceso interactivo y semi-automatizado de desglosar las relaciones complejas de todo el sistema para producir los núcleos o "enjambres" de componentes que formarán los proyectos independientes de reingeniería.
- El progreso de este proceso incremental de reingeniería es mantenido automáticamente en el Modelo Informático de Empresa. Esta característica releva al usuario de tareas complejas de coordinación, tales como el saber si, cuando y cómo bases de datos específicas tendrán que ser convertidas automáticamente, o bien que archivos

secuenciales deberán ser mantenidos en formatos duales y por cuanto tiempo, etc.

- Provee soporte automatizado para la transición de etapas de proyectos de reingeniería, de acuerdo a las siguientes fases: control de inventario, análisis de impacto y verificación, minería de datos, conversión, prueba, y reintegración y coexistencia. Por ejemplo:
 - Transición a la fase de coexistencia de subsistemas convertidos y no convertidos. Cuando el subsistema está listo para reingreso a la producción, se automatizan los cambios necesarios para introducirlo al sistema empresarial híbrido (que contiene subsistemas convertidos y no convertidos). Estos cambios incluyen el uso de procedimientos JCL convertidos que dan soporte a dicha coexistencia, mantenimiento de archivos en formatos duales, etc.

Transformación Automatizada de Programas y Datos

La meta del análisis de impacto es la de obtener información exacta para producir el nivel más alto posible de automatización durante la fase de transformación de programas y datos. Las características principales de esta fase de transformación incluyen:

- Conversión automatizada de programas COBOL
- Conversión automática de bases de datos IMS. Esto incluye:
 - La generación automática de nuevas definiciones (schemas) de bases (en DBD para IMS, y SQL para DB2). Estas nuevas definiciones incluirán los nuevos formatos corregidos de todos los campos problema en los segmentos o tablas.

-
- La transformación de la base de datos en sí, del formato antiguo al nuevo, incluyendo en este proceso la corrección de los valores en campos problema.
 - Programas para dar soporte al mantenimiento simultáneo de bases de datos duales durante la coexistencia.
 - Conversión automatizada de archivos o ficheros secuenciales.
 - Conversión automatizada de procedimientos JCL y SORT.
 - Conversión automatizada de pantallas y formas.
 - Soporte automatizado para reintegración y coexistencia.
 - Detección automatizada y respeto por campos previamente corregidos.
 - Correcciones automatizadas exactas para los problemas analizados, respetando casos especiales de acuerdo a la taxonomía del problema.
 - Soporte para la transformación masiva. A través del Modelo Informático de Empresa, el usuario puede gráficamente seleccionar todos los programas o aplicaciones que necesitan ser convertidos automáticamente.

Resolución de Problemas

La tecnología de reingeniería de Xinotech provee soporte para la resolución de problemas específicos de reingeniería en forma flexible y comprensiva.

- Soluciones a múltiples problemas. Estos problemas incluyen:
 - Año 2000
 - Conversión a la moneda común europea Euro
 - Conversión a estándares financieros internacionales, tales como el sistema numérico de identificación corporativa en las bolsas de valores (ISO 6166 ISIN) que es administrado por agencias como la CUSIP de

Standard & Poor en USA, SICOVAM en Francia, Telekurs in Suiza, y Wertpapier Mitteilungen en Alemania.

- Problemas específicos del usuario, tales como la adición en las direcciones de clientes de subcampos para correo electrónico o número de fax, o la adición de nuevos subcampos para extender o generalizar números de cuentas.
- Soluciones múltiples para cada problema específico. Para la mayoría de los problemas de reingeniería, tales como Año 2000, es menester tener acceso a soluciones múltiples, para entre las cuales seleccionar la que mejor se ajuste a cada entorno en particular. Algunas situaciones pueden requerir el uso de soluciones alternativas que puedan ser aplicadas a diferentes porciones de un sistema, y tal estrategia requiere que tales soluciones coexistan bajo un mismo entorno. Para Año 2000, por ejemplo, puede ser deseable que ciertos campos fecha dentro de determinados archivos sean expandidos, mientras otros campos fechas sean dejados intactos y manipulados a través de técnicas de ventanas. Esto requiere que la tecnología inserte operaciones puente para transferir valores entre estos campos. Para Año 2000, las siguientes soluciones están disponibles:
 - Expansión de dos a cuatro dígitos.
 - Compresión de dos a cuatro dígitos.
 - Técnica de cálculo de ventana corrediza.
 - Técnica de cálculo de ventana fija.
 - Técnicas de puente.
- Solución integrada para múltiples problemas. Cada fase del proceso de reingeniería se efectuara abordando todos los problemas a resolverse, resultando así en un solo ciclo de reingeniería, simplificando así el proceso y reduciendo el alto costo de múltiples iteraciones. Por ejemplo, cada programa COBOL se va a transformar para corregir de una vez

todos los problemas considerados en el proceso. El resultado es que cada programa COBOL es analizado, transformado, y probado solo una vez; además, cada base de datos será transformada sólo una vez.

- Subclasificación de tipos de datos. La tecnología de Xinotech para el reconocimiento de problemas permite la subclasificación de criterios adicionales de reconocimiento para detectar casos especiales, permite la adaptación de estos criterios por el usuario, y permite usar estos criterios en el proceso de propagar decisiones a otros campos para eliminar el número de decisiones interactivas. Por ejemplo, no es suficiente saber que un campo es una fecha; es también importante para una conversión correcta, el clasificar fechas de acuerdo a sus ventanas de valores (e.g. de 1920 a 2019, o de 1980 a 1079).
- La verificación de campos afectados es eficiente y, efectuada a través de un solo punto de decisión, es propagada al resto de los componentes. Esta verificación puede ser interactiva, o serial, así como individual o colectiva. El nivel de automatización en la verificación es ajustable.
- Detección automatizada y respeto por campos problema previamente corregidos. Las herramientas de Xinotech están diseñadas para separar "campos problema que no son problema", quizás porque fueron corregidos previamente. Alternativamente, el usuario puede desear transformar estos campos, con el único fin de estandarizar su formato, convenciones de nombres, o uso.
- Integración a bibliotecas estándares. Opcionalmente, Xinotech puede proveer soporte para la integración automatizada a nuevas bibliotecas estándares, tales como bibliotecas de manejo de fechas y de moneda.
- Adaptación a problemas específicos del cliente. La tecnología de Xinotech se presta para resolver otros problemas adicionales típicos, como la migración de VSAM o IMS a DB2, o bien problemas específicos o particulares del usuario.

Typel y el Año 2000

Típicamente, Xinotech, sus subsidiarias, o sus proveedores de servicio, adaptan 2001 a los requerimientos específicos del cliente. Opcionalmente, 2001 puede ser adaptado por el usuario mismo, a través de Typel, el Metalenguaje de Xinotech para el Reconocimiento y Evolución de Tipos de Datos. Typel puede ser usado para crear o adaptar una solución, especificando las reglas que describen la evolución de las estructuras de datos pertinentes a esta solución. Typel es general, simple, e intuitivo. Para usarlo en aplicaciones COBOL, por ejemplo, el analista o programador COBOL no requiere experiencia adicional fuera de su conocimiento de COBOL, de la aplicación, del problema, y de la solución deseada. Asimismo, Typel puede ser usado para resolver diferentes problemas. Sin embargo, para simplificar la explicación en esta sección, Typel es descrito exclusivamente en términos de COBOL y el Año 2000.

El concepto de resolución de problemas con Typel, está basado en la clasificación de las estructuras de datos en aplicaciones pre-existentes de acuerdo a un Catálogo de Tipos de Datos definido por Xinotech o por el usuario. Una vez clasificadas, estas estructuras pueden ser transformadas, corregidas, estandarizadas, y modernizadas.

Las reglas que pueden ser descritas en Typel se clasifican así:

-
- Reglas de reconocimiento. Estas son las reglas que describen como el Catálogo de Tipos de Datos habrá de ser usado para clasificar las estructuras en las aplicaciones pre-existentes.
 - Reglas de transformación de tipos de datos. Estas reglas describen como las estructuras de datos, una vez clasificadas, habrán de ser transformadas. Estas reglas de transformación permiten la implantación de soluciones escogidas por el usuario, incluyendo la estandarización y unificación de formatos heterogéneos en aplicaciones diversas.
 - Reglas de transformación de referencias. Estas reglas describen como referencias (en el PROCEDURE DIVISION) a las estructuras originales de datos necesitan ser transformadas para acomodar a las nuevas estructuras modernizadas.
 - Reglas de transformación de literales. Estas reglas complementan las reglas de transformación de referencias. Typel también provee soporte para el reconocimiento y transformación de centinelas, o literales con valores especiales usados para detectar condiciones excepcionales.
 - Reglas de conversión de datos. Estas reglas son usadas para describir como los datos en sí, almacenados en archivos y bases de datos, deberán de ser cambiados a su nuevo formato. Estas reglas de conversión son usadas para la migración de datos a sus nuevos formatos, y para crear puentes entre nuevos y viejos formatos.
 - Puentes a procedimientos cerrados. Typel también provee soporte para la descripción de procedimientos que no pueden o deben ser convertidos, causando así la creación automática de puentes.

Las estrategias de Xinotech para reconocer y clasificar estructuras de datos durante la reingeniería (de Año 2000), consisten en: convenciones de nombres, estructura y composición, formato, contenido, contexto, y uso. Para que las estructuras de datos puedan ser propiamente corregidas, es necesario que

subclasificaciones adicionales sean usadas, para reconocer por ejemplo, diversas ventanas de valores, o intervalos de tiempo.

2001 provee soporte para la detección de limitaciones de tamaño, tales como tamaño de bloques en disco, para que soluciones alternativas puedan ser incorporadas. Xinotech provee soporte para la selección de formatos de fecha alternos, que pueden resultar en la corrección del problema dentro de las limitaciones específicas de tamaño.

Además de las reglas generales de transformación basadas en clasificación por tipos, 2001 provee soporte para la selección de transformaciones para que campos individuales puedan ser corregidos y atender así a requerimientos específicos de campos especiales.

En ocasiones, puede ser necesario que ciertos campos intercambien valores con otros campos de formato incompatible. Por ejemplo, éste puede ser el caso:

- Porque el usuario puede haber seleccionado la transformación de algunos de estos campos a formatos diferentes.
- Por causa de restricciones de uso, tales como llaves de bases de datos.
- Por causa de funciones predefinidas que requieren formatos específicos como parámetros.
- Por causa del uso de campos con propósitos conflictivos.

La tecnología de Xinotech toma en cuenta la existencia de tales conflictos de tipos y provee soluciones que permiten la transferencia de valores o puentes a través de estos tipos heterogéneos.

La solución de Xinotech para la conversión de aplicaciones a la nueva moneda europea Euro, provee soporte para la detección y corrección automatizada de campos moneda. La solución de Xinotech también incorpora la conversión al standard ISO para el soporte de múltiples monedas, y soporte para actualizar cálculos al manejo aritmético de monedas mixtas.

Toda la funcionalidad presente en 2001 para Año 2000, también se encuentra presente en Euro\$olve, la solución Xinotech para integración al Euro. Esto incluye reconocimiento de tipos y subtipos, manejo de literales y centinelas, restricciones de tamaño, transformaciones de concurrencia múltiple, puentes de valores, etc.

Otro ejemplo de la generalidad de la solución Xinotech es la conversión al standard ISO 6166, o el International Securities Identification Numbering System (ISIN), que es el formato internacional para identificación corporativa en las bolsas de valores. Entre las agencias que administran la asignación de estos códigos, se incluyen: la CUSIP de la Standard & Poor en USA, SICOVAM en Francia, Telekurs en Suiza, y Wertpapier Mitteilungen en Alemania.

Interfaz del Usuario

Xinotech provee una interfaz moderna creada usando UNIX Motif. El usuario controla o gradúa el nivel de interactividad en el proceso de la toma de decisiones. Los reportes de análisis de impacto son generados en español.

Así mismo, las relaciones entre los componentes de las aplicaciones pueden ser desplegadas gráficamente. Por ejemplo, el usuario puede pedir gráficas que desplieguen el uso de bases de datos o segmentos a través de los programas en las aplicaciones, las relaciones entre programas producidas por llamadas a subrutinas compartidas, las relaciones entre archivos y procedimientos SORT, etc.

Con un "click" del ratón, las herramientas Xinotech permiten la navegación directa de nodos en las gráficas, a las posiciones correspondientes en los programas fuentes.

Estos modelos gráficos pueden ser manipulados para desplegar sólo la información relevante a los objetivos del momento, y se pueden imprimir a color para documentación. Los modelos gráficos son un vehículo efectivo para tomar decisiones en una forma concisa.

CAPÍTULO 4

MODELO PROPUESTO PARA LA DEFINICIÓN DE LA ARQUITECTURA DE DATOS

El proceso de definición de la arquitectura en una empresa, está formado de cuatro capas, tal como nos muestra la figura 4.1 [*Data Architecture: The Information Paradigm, 1992*]

♦ **Primer Nivel – ¿Dónde iniciamos?**

Planeación. Dentro de este punto, es importante que desde un principio se tome el camino correcto, incluyendo la metodología a utilizar, quien debe de estar involucrado y qué conjunto de herramientas será utilizado. Esto conlleva a producir un plan de trabajo para la definición de la arquitectura y asegura el compromiso administrativo a seguir cada una de las fases.

♦ **Segundo Nivel - ¿Dónde estamos hoy en día?**

Modelación del negocio. Implica un conocimiento base acerca del sistema y de la información utilizada en la conducción del negocio.

Sistemas y Tecnología Actuales Define la existencia en cuanto a sistemas aplicativos y plataformas tecnológicas que los soportan hoy en día. Este es un inventario a un nivel sumario de los sistemas aplicativos, los datos y las plataformas tecnológicas, que provee una línea base para planes de migración a largo plazo.

♦ **Tercer Nivel - ¿Dónde deseamos estar en un futuro?**

Arquitectura de Datos. Define los tipos clave de datos necesarios para soportar el negocio.

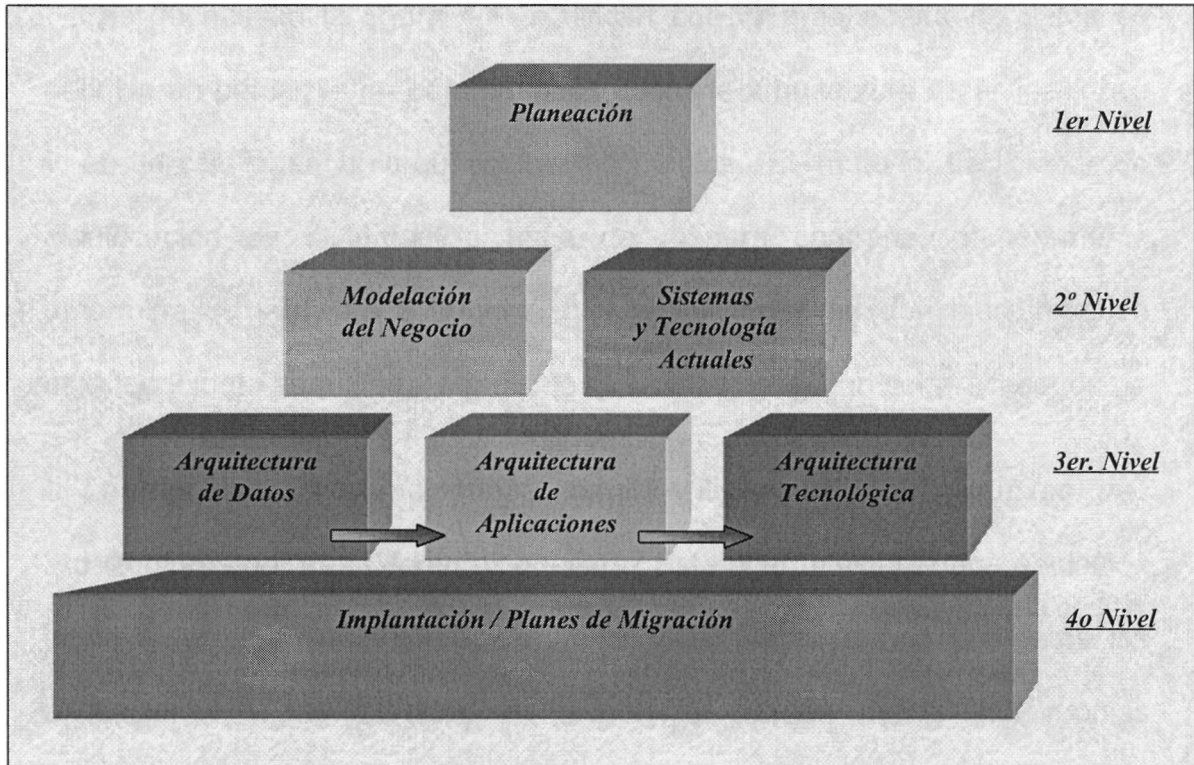


Figura 4.1 La definición de una arquitectura de datos.

Arquitectura de Aplicaciones. Define los tipos clave de aplicaciones necesarias para administrar los datos que soportan las funciones del negocio.

Arquitectura Tecnológica. Define las plataformas tecnológicas necesarias para proveer un ambiente para las aplicaciones que administran los datos y soportan las funciones del negocio.

Las flechas en la figura 4.1 significan que la arquitectura de datos es definida en primer lugar, para después definir la arquitectura de aplicaciones y en tercer lugar la arquitectura tecnológica. El método tradicional de planificación de sistemas lo hace de manera contraria, primeramente determina el hardware a adquirir, qué aplicaciones pueden ejecutarse en esta tecnología adquirida y por último qué datos necesitan ser procesados.

La siguiente pregunta puede surgir: *¿Es posible definir estas tres arquitecturas por tres diferentes equipos?* Este acercamiento tiene apelación intuitiva debido a que el traslape entre las fases ahorra tiempo. Los datos, las aplicaciones y los expertos en tecnología pueden definir sus propias arquitecturas, maximizando la productividad del equipo y minimizando las disputas interpersonales.

Entonces, las tres arquitecturas pueden ser combinadas, reconciliadas e integradas en un conjunto cohesivo de cianotipos. Sin embargo, en la práctica, este acercamiento no funciona. las arquitecturas individuales se tornan en incompatibles o demasiado complejas para combinarse, debido a las siguientes razones:

1. Los equipos tienen diferentes perspectivas y niveles de entendimiento del negocio.
2. Las arquitecturas se traslapan.
3. Existe un incentivo compatible mínimo o no existe.
4. Sin la existencia de una arquitectura de datos, no existe una fundación firme para la definición de aplicaciones y tecnología.
5. Las metodologías, herramientas y consultores son diferentes, cada equipo desea hacerlo a su propia manera.
6. Separar equipos apunta una inhabilidad de los tres grupos a trabajar cooperativamente, lo que indicaría una cultura no favorable para la implantación de ambientes de datos compartidos.

◆ **Cuarto Nivel - ¿Cómo llegamos ahí?**

Implantación / Planes de Migración. Define la secuencia de implantación de las aplicaciones, la calendarización para implantación, un análisis costo – beneficio y propone una ruta clara para la migración de donde estamos ahora y a donde queremos llegar.

La arquitectura de datos es de suma importancia ya que la calidad de los datos es uno de los factores críticos de éxito del departamento o área de sistemas de cualquier empresa. Esta consiste de entidades de datos, las cuales tienen atributos y relaciones con otras entidades de datos.

Las definiciones de la arquitectura de datos se convierten en estándares a utilizar subsecuentemente en las fases de diseño de los sistemas, comúnmente referidas como diseño lógico de base de datos, diseño físico de base de datos y creación de base de datos.

Modelo Propuesto

Fase Planeación

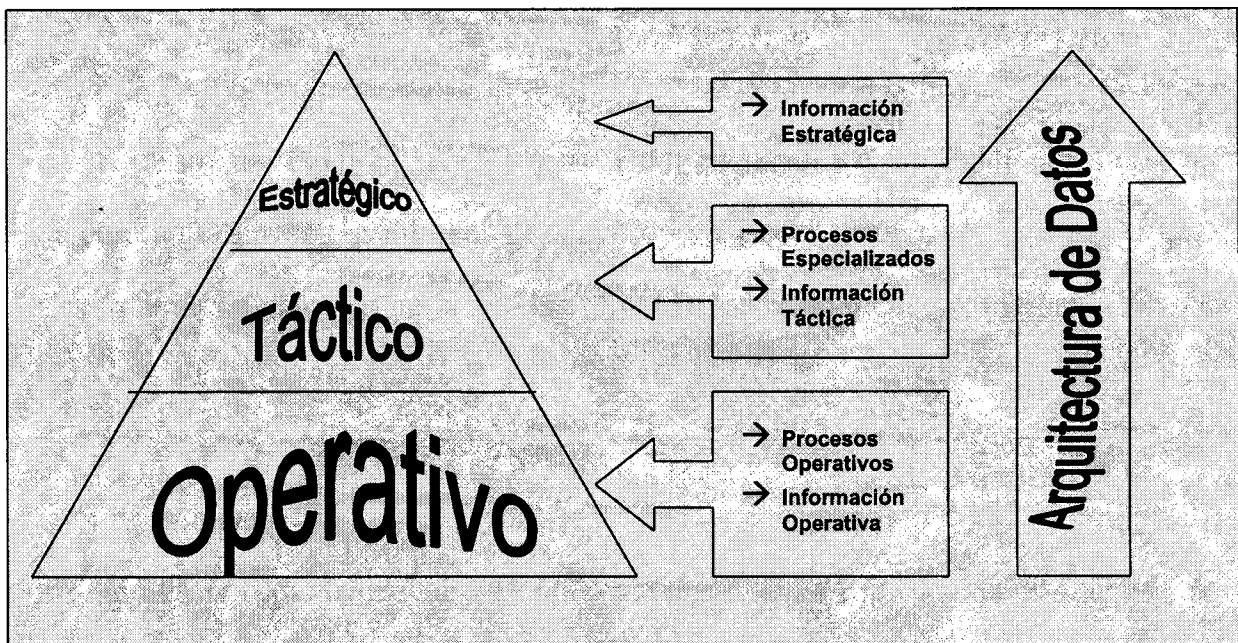


Figura 4.2 Niveles en la organización

Es muy importante involucrar a cada una de las partes que integran el negocio, debido a que como podemos observar en la figura 4.2, el conjunto de procesos operativos que integran la parte operativa del negocio, producen información operativa, y esos dos puntos brindan información que sirve de entrada la mayoría de las veces para procesos más especializados que conforman la parte táctica del negocio y como último nivel, tenemos a la parte estratégica quien utiliza como entrada a sus procesos, en su mayoría toma de decisiones, información resultado de la parte operativa y la parte táctica del negocio. La selección del personal involucrado en esa parte es decisión del negocio, una vez que se tengan claramente definidos cada uno de los niveles y la importancia del proyecto.

En esta etapa de planeación, es necesario iniciar con la definición del negocio, cual es la misión del negocio, cual es su visión, sus planes a corto y largo plazo, su estructura organizacional, sus procesos esenciales o core, etc. El proceso de definición de una arquitectura de datos es un proyecto cuyo plazo depende de varios factores, el tamaño de la empresa, la cantidad de sistemas legados, y básicamente de la definición de sus procesos. El personal que integre el equipo para la definición de la arquitectura de datos, debe estar familiarizado totalmente con los puntos antes mencionados sobre el negocio.

Por lo tanto hasta el momento tenemos los puntos importantes dentro de la planeación:

- La formación de equipos, involucrando la parte operativa, táctica, estratégica del negocio y la parte de tecnología.
- El conocimiento de la empresa, su misión, su visión, sus planes a corto y largo plazo.
- Cada una de las actividades que conforman la arquitectura:
 - Modelación del negocio
 - Sistemas Actuales
 - Modelación de la arquitectura de datos
 - Implantación.

El producto final de esta etapa, son planes de trabajo que identifiquen tiempos y responsables de cada una de actividades que conformaran el proyecto de definición de arquitectura de datos. Una gran ayuda de los puntos o actividades a cubrir dentro de este proyecto se encuentran en el capítulo 3, apoyados con la figura 3.3.

Fase Modelación del Negocio

En esta parte es esencial una completa definición de los procesos que involucran cada uno de los niveles de la empresa. Iniciando con la identificación de los procesos operativos, tal como se muestra en la figura 4.3

Dentro de la modelación del negocio tenemos que identificar claramente cuales son los procesos operativos el negocio, sus entradas y salidas de información, es muy importante clarificar las excepciones de cada uno de estos procesos, para después identificar que salidas de información de cada uno de estos procesos

operativos, conforman la entrada de información a procesos tácticos y de igual manera, es necesario identificar que salidas de información de procesos tácticos, conforman la entrada de información al proceso de toma de decisiones en la parte estratégica del negocio.



Figura 4.3 Procesos ejemplo en cada uno de los niveles de la organización

Esta modelación se logra a través de diferentes herramientas, sin embargo siendo que el equipo del proyecto se conforma de personal con diferentes grados de entendimiento, una herramienta como lo son los diagramas de flujo serán entendibles para todo el equipo.

La parte operativa del negocio, debe identificar la cadena de procesos que conforma la parte operativa, además de los flujos de información que se ven involucrados en los procesos. Es de suma importancia identificar las excepciones que se encuentran en cada uno de los procesos y la manera en que pueden afectar estos flujos de información.

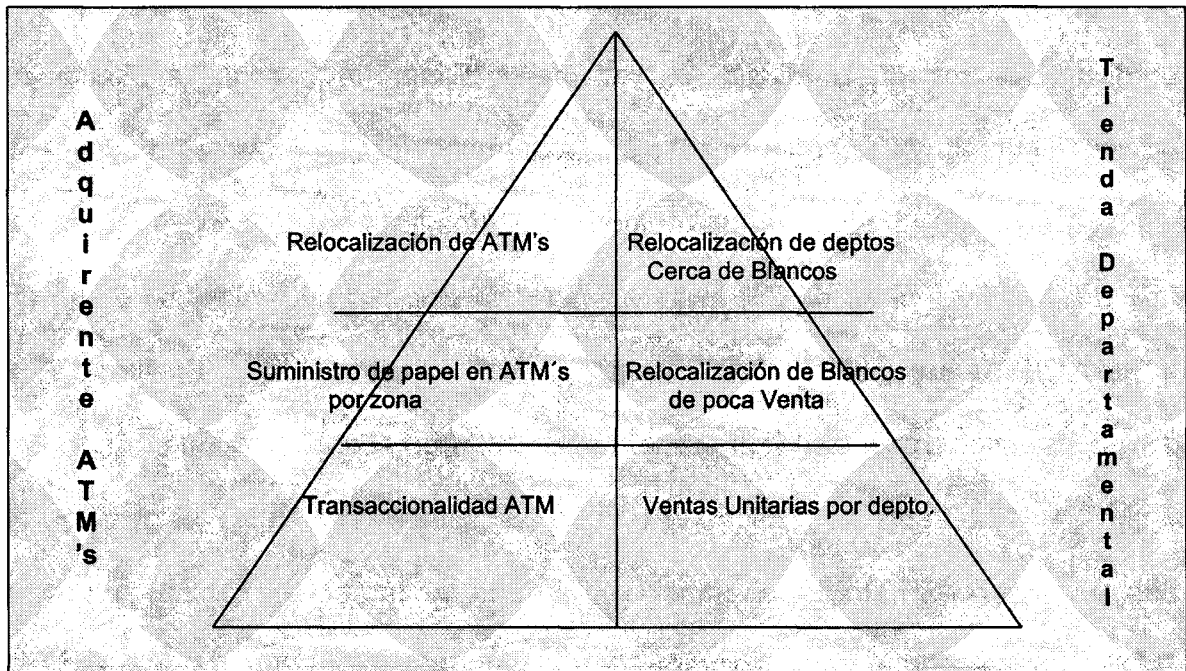


Figura 4.4. Ejemplo de Niveles en una Organización

En caso de ser necesario, se deben utilizar árboles de decisión para la construcción de los diferentes caminos que estas excepciones provocan en la operatividad y básicamente en la información. En la figura 4.4 podemos observar dos ejemplos de procesos de acuerdo a sus niveles en una tienda departamental y en un adquirente de cajeros automáticos. Supongamos en el adquirente atm's que

su parte operativa como principal proceso operativo tiene la transaccionalidad en cajeros automáticos, es decir, cada una de las transacciones que se ejecutan en un cajero automático, la entrada al proceso táctico de suministros, no es cada una de esas transacciones, sino en que cajeros se debe de mantener un continuo surtido de papel, debido al conjunto de transacciones que se tiene en determinado cajero y la parte estratégica analizará si esta zona donde se tiene un mayor numero de transacciones atm, es necesario añadir cajeros automáticos o en su caso relocalizarlos. De igual manera en una tienda departamental, la parte operativa mantendrá información respecto a cada una de las ventas que se realiza en cada uno de los departamentos que conforman la tienda, el conjunto de estas ventas es la entrada de información para el nivel táctico de la organización y de esta manera decidir aquellos productos por ejemplo de poca venta en el departamento de blancos, para finalmente en la parte estratégica decidir la relocalización de departamentos. Estos dos ejemplos son simples. Nuevamente hacemos hincapié en la definición de excepciones que por la misma operatividad, indican diferentes caminos al manejo de información. Por ejemplo ¿qué pasa? Sí el cajero no pertenece al adquirente, no forma parte de su red de cajeros, solo adquiere sus transacciones, ¿la información debe tomar otro camino?, En el caso de la tienda departamental ¿qué pasa? Sí los productos están a concesión. En fin, existe un gran número de excepciones que den diferentes caminos a la información y todas estas excepciones junto con el detalle de información necesaria para cada uno de los procesos operativos, tácticos y estratégicos, deben de contenerse en esta etapa.

Siendo que al ir subiendo de nivel, la parte táctica es podríamos decir el conjunto de información de la parte operativa al igual que la parte estratégica es la suma de la parte táctica del negocio. En esta parte la clave se encuentra en la definición de métricas por la parte estratégica. Ya que esta definición dará la pauta para la información que servirá de soporte a la toma de decisiones da rumbo a la empresa y que afecta directamente su competitividad. Esta definición de métricas podríamos afirmar que es la última etapa de la modelación del negocio. Cada uno de los niveles tiene métricas, sin embargo aquellas que se definan por la parte estratégica del negocio, son las que darán la pauta para decidir si vamos por buen camino, al momento de que se aclare que se cuenta con toda la información necesaria para soportar dichas métricas.

Tal como lo explicamos en los ejemplos, de la unidad se obtendrán métricas y cada una de estas irá subiendo de nivel, por ejemplo en el caso de la tienda departamental, el nivel operativo mantiene registro de cada una de las ventas que se realizan en cada uno de los departamentos, pero ejemplos de métricas a este nivel, serian: ventas por empleado, por producto de cada uno de los departamentos, ya que dentro de la parte operativa podríamos contemplar la parte de gestión.

Para el caso del nivel táctico en este mismo ejemplo, podríamos tener en un periodo dado las ventas por línea de productos, en cada uno de los departamentos, además de contemplar los puntos de reorden del proveedor y tomar decisiones sí es que estos deben de cambiar. Otro ejemplo en esta parte es

el análisis de costo, ya que dependiendo de las ventas podríamos definir oferta demanda de ciertos productos, o como lo comentamos en el ejemplo, relocalización de ciertos productos, o quizá el lanzamiento de una nueva campaña de publicidad.

En la parte estratégica no vamos a tener métricas que contemplen el conjunto de ventas unitarias a ese nivel, sino el conjunto de ventas que nos indiquen la competencia con el mercado en ciertos departamentos, el nivel de ventas que tiene la empresa.

Fase Análisis Sistemas y Tecnologías Actuales

En esta parte necesitamos conocer la tecnología con la que cuenta la empresa y de ser posible seccionarla de acuerdo a los niveles antes vistos. Tal segmentación debe incluir la definición de los sistemas con los que cuenta la organización sistemas de administración de información, Sistemas de Información ejecutiva, sistemas de soporte a decisiones y todos aquellos sistemas operacionales con los que cuenta la empresa. Es importante aclarar que parte de importante de la definición de arquitectura de datos de la empresa, comprende la contenida o que deberán contener los sistemas de soporte a la decisión de la organización en cada uno de sus niveles, ya que como se explicó anteriormente, cada uno de los niveles toma decisiones a diferentes rangos de tiempo o niveles o jerarquías dentro de la empresa.

La definición de los sistemas con los que cuenta la organización, debe incluir al igual que en la modelación del negocio las entradas y salidas de información de cada una de las tecnologías actuales y un punto crucial en este caso son las validaciones que se realizan a la información, en este punto observaremos que todas aquellas excepciones de las que hablábamos en la modelación del negocio se cumplan de acuerdo a las validaciones que contengan los sistemas y tecnologías con los que cuenta la organización. Estas validaciones dan forma a la información y esta forma es necesario contemplarla en la arquitectura de datos de la empresa.

Esta fase tiene como producto final, la producción de un inventario de los sistemas y tecnologías con los que cuenta la organización, incluyendo la plataforma tecnológica en la que la empresa sostiene su operatividad.

Fase Definición de Arquitectura de datos.

Este es el punto crucial, la definición de la arquitectura de datos en base a lo antes expuesto, esta es la realización de lo antes planteado, la modelación del negocio, donde se analizarán las entradas y salidas de cada uno de los procesos tanto operativos, tácticos y estratégicos y la producción de información de cada uno de ellos, sus entradas y sus salidas y lo más importante de este punto, las posibles excepciones en cada uno de los procesos, en cada uno de los niveles.

Una vez modelados los procesos del negocio, es necesario levantar el inventario de los sistemas y tecnologías que sostienen la plataforma operativa de la empresa, incluyendo todas aquellas transformaciones que tiene la información en cada uno de estos procesos.

Una vez finalizados estos dos puntos es necesario conjuntarlos, identificando donde interviene cada uno de los sistemas junto con su tecnología en los procesos de la empresa analizando cada una de las validaciones y excepciones que se determinaron anteriormente.

En esta parte tal como lo hemos visto a través del proceso tenemos en el equipo personal tanto operativo como técnico, aspecto importante que durante el desarrollo de todo el proyecto no se pierda alguno de los puntos clave de información del manejo del negocio.

En la figura 4.5 tenemos un ejemplo continuando con la tienda departamental de esta definición. Podemos observar como en el nivel operativo se cubre el detalle de un tipo de información, las ventas unitarias de cada uno de los departamentos, en el nivel táctico podemos observar que se utiliza la información sumariada de estas ventas, es decir, se utiliza la sumariación de ese detalle, las ventas totales de cada uno de los departamentos ya que en el ejemplo que hemos utilizado a lo largo de esta parte podemos decir que es la información necesaria para el departamento de mercadotecnia por ejemplo para lanzar una campaña de publicidad y en la parte estratégica se analizarán las ventas totales de la tienda.

Uno de los puntos más importantes en esta parte es la definición del factor tiempo ya que cada uno de los niveles tanto operativo, como táctico y estratégico tienen necesidades diferentes en cuanto a rangos de tiempo se refiere dentro del manejo de la información. Podríamos asumir que los rangos de tiempo para cada uno de los niveles van de acuerdo al detalle de la información, sin embargo la arquitectura de datos de la organización, debe ser capaz de cumplir con el nivel más bajo en cada uno de los niveles de la organización, ya que uno de los objetivos de esta definición es cumplir con las necesidades de información donde y cuando se necesite y cómo se necesite.

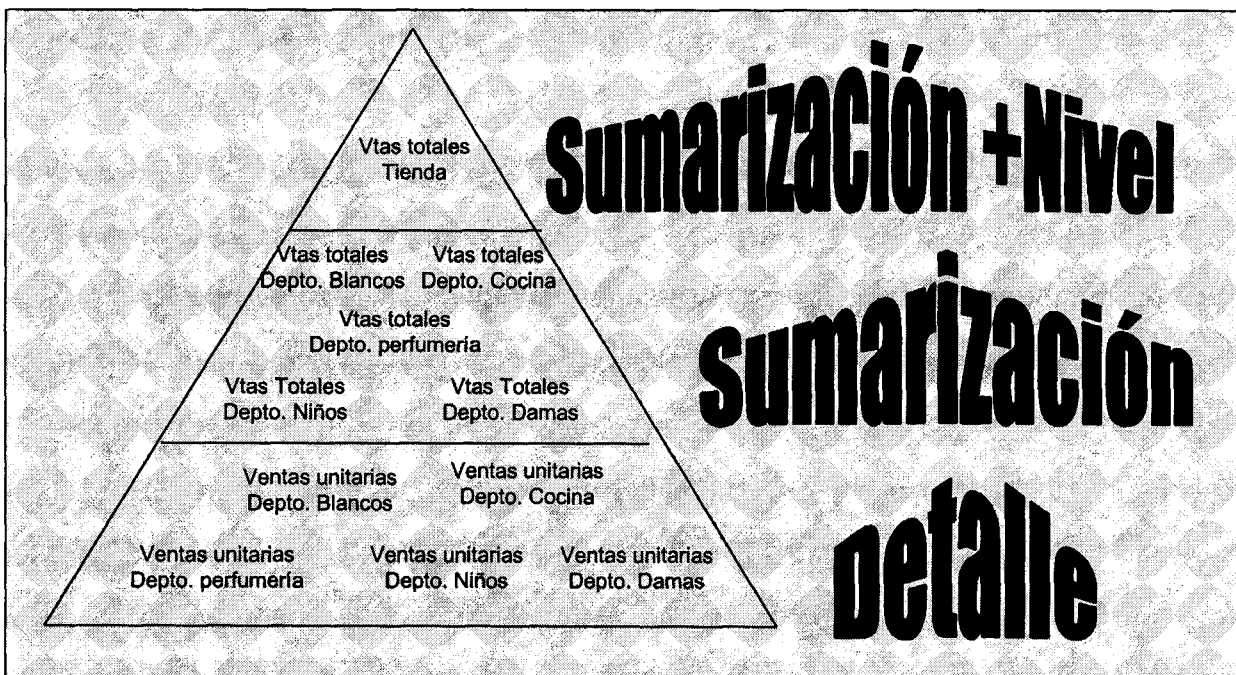


Figura 4.5 Ejemplo de información en cada nivel de la organización

Una vez que se han definido todos los datos que soportarán la parte operativa, táctica y estratégica de la empresa, es necesario realizar el modelo de datos que soporte esta. Varias de las herramientas para la modelación de datos definida de la arquitectura se explican en el capítulo 3 de éste documento. La utilización de diferentes herramientas es decisión del negocio, ya que depende del personal técnico que se encuentre involucrado en el equipo y de las necesidades de la empresa. Nuevamente reiteramos que la parte fundamental de la definición de una arquitectura de datos está en la modelación del negocio, donde encontraremos todos los datos que conforman información en cada uno de los niveles que realizan información y a su vez esta se convierte en datos de entrada para nuevos procesos a otro nivel. Sin olvidar la parte de la definición de los sistemas actuales y plataformas tecnológicas que conforman los procesos de la empresa donde se define la transformación actual de la empresa.

Hasta el momento se ha hablado de una organización existente donde ya existe una definición de procesos y una tecnología que los soporta, es posible que esto no suceda, ya que el hecho de la inexistencia de sistemas hereditarios facilita la definición de una arquitectura de datos, sin embargo actualmente una empresa que inicia sin una automatización de procesos, no le es posible manejar grandes volúmenes de información.

La parte fundamental es la definición de procesos en la parte de modelación del negocio, ya que sin una determinación clara por parte de las personas

involucradas en estos procesos, se corre el riesgo de cometer errores en la definición que afectarán los demás niveles de definición en la organización.

Fase Implantación

La parte final es la implantación de la arquitectura de datos definida, donde se decide paso a paso como vamos a construir la arquitectura de datos definida. Tal como se realizó la modelación de datos en la parte de definición de arquitectura de datos, ahora tenemos que definir por donde empezaremos a construirla en la organización, que no afecte la operación y básicamente esta construcción se debe de realizar mediante interfaces y sistemas paralelos que no afecten la operación diaria de la empresa.

Siguiendo la misma línea de niveles que hemos mencionado, es importante iniciar con la estructura de datos que apoya la parte operativa del negocio, donde tendremos los datos a más bajo nivel, siguiendo con las sumarización correspondiente para el nivel táctico y finalizando con la estructura de datos que apoyará a la parte estratégica del negocio. La creación de dicha estructura basada en la modelación de acuerdo a la arquitectura de datos definida es donde se involucra totalmente la parte técnica involucrada en el proyecto.

De esta manera se cumple el objetivo de definir e implantar una arquitectura de datos que apoye la parte operativa, táctica y estratégica de la empresa.

Investigación de Campo

Con el objetivo de realizar una validación del modelo antes expuesto, se realizó una investigación de campo. Como inicio a esta investigación, se realizó una presentación del modelo propuesto en esta tesis, para continuar con entrevistas a personal involucrado en los tres niveles organizacionales: operativo, táctico y estratégico, además de personal del área de sistemas en diferentes organizaciones de la Ciudad de México, tales como:

- ◆ **Consultoría de Sistemas Monterrey, S.A. de C.V:**
- ◆ **Dow Corning de México, S.A. de C.V.**
- ◆ **Grupo Financiero Bancomer, Enlace Bancomer**
- ◆ **IPN - UPIICSA**
- ◆ **Next, S.A. de C.V.**
- ◆ **QAD Sistemas Integrados, S.A. de C.V:**
- ◆ **Servicios Electronicos Globales, S. A. de C. V.**
- ◆ **Xanum Consultores, S.A. de C.V.**
- ◆ **Xerox, S.A. de C.V.**

Con el objetivo de validar el modelo de acuerdo al esquema de los diferentes puntos de vista involucrados en las organizaciones, a continuación se presenta un cuadro de referencia de acuerdo al personal a quien se presentó el modelo propuesto.

MODELO PROPUESTO PARA LA DEFINICIÓN DE LA ARQUITECTURA DE DATOS

Personal Entrevistado	Visión	Organización
Ing. Alberto Rugeiro	Tecnología	Dow Corning de México, S.A. de C.V.
Ing. Claudia Leticia Lozano Beristain	Sistemas	Servicios Electronicos Globales S. A. de C. V.
MTIA. Daniel Peñaloza Moreno	Sistemas	Xanum Consultores, S.A. de C.V.
Lic. Eduardo Aguilar Rodríguez	Sistemas	Consultoría de Sistemas Monterrey, S.A. de C.V.
Lic. Erika Flores González	Sistemas	Servicios Electronicos Globales S. A. de C. V.
Lic. Jorge Zarate Vázquez	Sistemas	Xanum Consultores, S.A. de C.V.
Lic. José Antonio Delgado Selley	Estrategia Organizacional	Servicios Electronicos Globales S. A. de C. V.
Ing. José Luis García Rodríguez	Sistemas	IPN - UPIICSA
Lic. Marcela San Roman Manterola	Táctica Organizacional	Next, S.A. de C.V.
Lic. Marco Tulio Martínez González	Sistemas	Servicios Electronicos Globales S. A. de C. V.
Lic. María Luisa Osorio	Operativa	Servicios Electronicos

Antonio	Organizacional	Globales S. A. de C. V.
Lic. María Teresa Ramírez Bueno	Operativa Organizacional	Grupo Financiero Bancomer Enlace Bancomer
Ing. Miguel Angel Martínez	Táctica Organizacional	Servicios Electronicos Globales S. A. de C. V.
Ing. Ramón Alejandro García Aguilar	Tecnología	QAD Sistemas Integrados, S.A. de C.V.
Ing. Raúl Hernández Nagore	Tecnología	Xerox, S.A. de C.V.

A continuación se presentan los comentarios obtenidos de dichas entrevistas, en orden alfabético de acuerdo a la empresa:

Lic. Eduardo Aguilar Rodríguez

Vicepresidente, Consultoría de Sistemas Monterrey, S.A. de C.V:

"El modelo presentado permite la reducción de costos sobre tecnologías de información, debido a la planeación de los cimientos de la tecnología de la organización, permitiendo la implantación de la tecnología adecuada y sobre todo del personal adecuado a dicha tecnología"

Ing. Alberto Rugerio

Consultor Técnico, Dow Corning de México, S.A. de C.V.

"El modelo propuesto obliga una visión integral de la organización, ya que contempla la visión de la empresa a nivel operativo, táctico, estratégico, la visión de sistemas y la visión de tecnología, y mediante

el uso de metalenguaje se elimina la subjetividad resultado de la contemplación de las diferentes visiones de la organización"

Lic. María Teresa Ramírez Bueno
Consultor Grupo Financiero Bancomer

"El modelo presentado permite a la organización fundamentar su arquitectura tecnología en la parte fundamental, los datos, apoyando no sólo a cierto nivel de la empresa, sino a todos sus niveles"

Ing. José Luis García Rodríguez
Profesor de tiempo completo de la Licenciatura en Ciencias de la Informática, Maestría de Ciencias de la Informática IPN - UPIICSA y Docente de la Universidad Tecnológica de México en las áreas de Ingeniería e Informática.

"Este modelo puntea la jerarquía de información como base del MIS general de la organización, permite identificar con oportunidad las bases de la infraestructura tecnológica de la empresa"

Lic. Marcela San Roman Manterola
Gerente Administrativo Next, S.A. de C.V.

"La reingeniería de procesos como estrategia corporativa, debe complementarse con una reingeniería tecnológica, que permita a la empresa contar con la información táctica, estratégica y operativa, tal como lo presenta el modelo"

Ing. Ramón Alejandro García Aguilar
Consultor Técnico, QAD Sistemas Integrados, S.A. de C.V:

"El modelo presentado permite una visión completa de la organización, además de contemplar metalenguaje permitiendo la utilización de diferentes tecnologías y apoyar la reingeniería de información de la organización"

Lic. José Antonio Delgado Selley

Director Planeación Estratégica, Servicios Electronicos Globales S. A. de C. V.

"El modelo propuesto permite la consideración de la información a nivel corporativo, para el apoyo y soporte de la planeación que permita a la empresa una continuidad competitiva, además de que busca brindar con oportunidad información para toma de decisiones"

Ing. Miguel Angel Martínez

Subdirección del área de Planeación Estratégica, Servicios Electronicos Globales S. A. de C. V.

"Uno de los factores críticos de éxito actualmente para las organizaciones, es no solo poseer información, sino saber utilizarla inteligentemente, pero no se puede utilizar la información que no se tiene y este modelo permite contemplar el universo de datos a utilizar en la organización"

Lic. María Luisa Osorio Antonio

Subdirección del área de Autorización Voz, Servicios Electronicos Globales S. A. de C. V.

"La toma de decisiones correctas en la empresa, debe estar sustentada por información confiable, resultado de las operaciones de la empresa,

la cual debe subir de nivel hasta llegar al nivel estratégico, tal como lo presenta el modelo"

Ing. Claudia Leticia Lozano Beristain

Gerente Sistemas Transaccionales, Servicios Electronicos Globales S. A. de C. V.

"El contar con el esquema de datos a procesar y manipular, antes de desarrollar aplicaciones, apoya la calidad del área de sistemas, aminorando tiempo de respuesta a las necesidades del usuario"

Lic. Marco Tulio Martínez González

Analista de Negocios, Servicios Electronicos Globales S. A. de C. V.

"Los modelos de datos permiten contemplar redundancias y eliminarlas cuando se cuenta con un modelo ya sea inicial o en el momento del análisis de lo que contempla la infraestructura de sistemas de la organización"

Lic. Erika Flores González

Analista Programador, Servicios Electronicos Globales S. A. de C. V.

"La programación de las aplicaciones es más sencilla cuando se cuenta con los modelos de datos a utilizar en las aplicaciones, tal como lo presenta el modelo"

MTIA. Daniel Peñaloza Moreno

Director General, Xanum Consultores, S.A. de C.V.

"El modelo aquí expuesto tiene como principio la calidad de los datos que conforman la información intrínseca a la competitividad de la

organización, sustentando los tres niveles que la integran y apoyando de manera total a la misión de los sistemas de información y la tecnología de la empresa"

Ing. Jorge Zarate Vázquez

Vicepresidente, Xanum Consultores, S.A. de C.V.

"El modelo propuesto es flexible, ya que resulta de una planeación integral facilitando la adaptación de los cambios que tenga la organización en su ambiente permitiendo así la evolución y crecimiento de la empresa"

Ing. Raul Hernández Nagore

Consultor, Xerox, S.A. de C.V.

"La reducción de tiempos de implantación es notoria debido a la utilización de modelos desarrollados en metalenguaje, ya que facilita la base a desarrollar como modelo de datos de la organización"

Conclusiones

De acuerdo a la premisa planteada en esta tesis, de que los métodos tradicionales de planificación de sistemas para la definición de la arquitectura tecnológica en una empresa, no cubren con oportunidad las necesidades de las organizaciones hoy en día, debido a que inician con la determinación de hardware, para continuar con las aplicaciones que pueden ejecutarse por dicha tecnología y finalizar con el aspecto datos; incrementando costos tanto cuantitativos como cualitativos en la obtención de información intrínseca a la competitividad de la

organización y en base a la investigación realizada para definir un modelo para la definición de la arquitectura de datos como primer acercamiento, antes de definir la arquitectura aplicativa y tecnológica de la empresa, siguiendo los lineamientos de la misión de sistemas de información y en base a la investigación presentada, podemos concluir que el modelo propuesto resultante, permite apoyar a los diseñadores y analistas de sistemas, una arquitectura de datos que permita flexibilidad ante los cambios y necesidades de los tres niveles que integran la empresa, además de apoyar la tecnología aplicativa, cuidando la calidad total de los datos y bajo los estatutos que marca la misión de sistemas, permitiendo el ofrecer información en el momento que se requiera, con la rapidez que se requiera y en el formato que se requiera, sin incurrir en altos costos, debido a que los datos fundamentan la arquitectura que brinda servicio al usuario de la organización.

El modelo propuesto define las fases que debe contemplar en la definición de la arquitectura de datos de la empresa, sin especificar las partes específicas de cada una de estas, obteniendo como producto final de este trabajo, la base para la continuidad de este trabajo en el desarrollo de la metodología para la definición de la arquitectura de datos de la organización.

REFERENCIAS BIBLIOGRÁFICAS

- **AMERICAN NATIONAL STANDARD**, "Knowledge Interchange Format" <http://logic.stanford.edu/kif/dpans.html>, Abril 4 1999
- **ANDERSON**, Larry S., **PERRY**, John F., "Technology Planning: Receipe for Success", <http://www.nctp.com/tp.receipe.htm>, Octubre 10 1998
- **ANONIMO**, "Corpus Encoding Standard", <http://www.cs.vassar.edu/CES>, Noviembre 10 1999
- **ANONIMO**, "Architecture", <http://www.dbstuff.com/architecture.htm>, Noviembre 6 1998
- **ANONIMO**, "Chapter 4: Data Architecture", <http://www.state.nc.us/IRM/techarch/data.htm>, Noviembre 6 1998
- **ANONIMO**, "Reingeniería Automatizada de Programas y Datos a Nivel Empresa" <http://www.xinotech.com/enterp-reeng.es.html>, Noviembre 9 1999
- **BATINI**, Carlo, **CERI**, Stefano, **NAVATHE**, Shamkant, B., "Diseño conceptual de bases de datos: un enfoque de entidades- interrelaciones", 1994
- **BIELAWSKI**, Larry, **LEWAND** Robert, "Intelligent systems design: integrating, expert systems, hypermedia, and database technologies", USA, 1991
- **BLAHA**, Michael, **PREMERLANI**, William, "Objet-Oriented modeling and design for database applications", USA, 1998
- **BRACKETT**, Michael H., "Data Sharing: Using a Common Data Architecture", John Willey & Sons, Inc, Primera Edición, United States of America, 1994
- **COMDEX** "Conference Recordings", Otoño 1989
- **DEVLIN**, Barry, "Data warehouse: from architecture to implementation", United States of America, 1997
- **ENSOR**, Dave, **STEVENSON**, Ian, "Oracle Design", USA, 1997
- **FLORES** Ivan, "Data Structure and Management", segunda edición, USA, Prentice Hall, 1977
- **FLORES**, Ivan, "Data Base architecture", United States of America, 1981
- **GUILLENSON**, Mark L., "Strategic planning, systems analysis, and database design: the coninous flow approach", USA, 1984
- **HALPIN**, T.A., "Conceptual Schema & relational database design", USA, 1995
- **HANSEN**, Gary W., "Database management and design", USA, 1996
- **HAWRYSZKIEWYCZ**, I.T., "Relational database design: an introduction", USA, 1990

-
- **HEINCKIENS**, Peter M., "Building Scaleble database applications: Object-orientes design, architectures and implementation", United States of America, 1998
 - **HERNANDEZ** Michael J., "Database design for mere mortals: a hands-on guide to relational database design", USA, 1997
 - **INFORMATION DISCOVERY, INC**, "DataMines for DataWarehouses", <http://www.datamining.com/dm4dw.htm>, Noviembre 22, 1999
 - **INMON**, William H., "Data Architecture: The Information Paradigm", John Willey & Sons, Segunda edición, UNITED STATES OF AMERICA, 1992
 - **INMON**, William H., **CAPLAN**, Jeffrey H., "Information Systems Architecture, Devlopment in the 90's", John Willey & Sons, Primera edición, USA, 1992
 - **KIMBAL**, Ralph, "The data warehouse lifecycle toolkit: expert methods for designing, developing, and deploying datawarehouses", John Willwy & Sons, primera edición, USA, 1997
 - **KIMBAL**, Ralph, "The data warehouse toolkit: Practical techiques for building dimensional data warehouses", John Willwy & Sons, primera edición, USA, 1995
 - **KORTH**, Henry, "Database System Concepts", Segunda edición, USA, Mc.Graw-Hill, 1991
 - **KROENKE**, David, "Database Processing: Fundamentals, design, implemantation", USA, 1990
 - **NIEDERMAN**, Fred, **BRANCHEAU**, James H., **WETHERBE**, James C, "Information Management Systems Management Issues for the 1990s", MIS Quartely, Volume 14, Number 4, December 1991
 - **ROB**, Peter, **CORONEL**, Carlos, "Database Systems: design, implemtation, and management", USA 1993
 - **ROMAN**, Steven, "Access Database design & programming", USA, 1997
 - **SEE**, John, "Developing Effective Technology Plans", <http://www.nctp.com/john.see.htm>, Octubre 10 1998
 - **SPEWAK**, Steven H., **HILL**, Steven C., "Enterprise Architecture Planning: Developing a Blueprint for Data, Applications and Technology", Primera edición, UNITED STATES OF AMERICA, 1992.
 - **TAYLOR**, David A. "Object Technology: A Manager's Guide", USA, 1997
 - **TEOREY**, Tobey J. "Database modeling and design: The entity-relationship approach", USA, 1990
 - **TEOREY**, Tobey J. "Database modeling and design: The fundamental principles", USA, 1994

