

Sistemas de Información Gerencial

Administración de la Empresa Digital

Kenneth C. Laudon y Jane P. Laudon – X Edición

6.1 ORGANIZACIÓN DE DATOS EN UN ENTORNO TRADICIONAL DE ARCHIVO

Un sistema efectivo de información da a los usuarios información exacta, oportuna y relevante. La información exacta está libre de errores. La información es oportuna cuando está disponible en el momento que la requieren los encargados de la toma de decisiones. La información es relevante cuando es útil y adecuada para los tipos de trabajo y decisiones que la necesitan.

Se sorprendería de saber que muchas empresas no cuentan con información oportuna, exacta o relevante porque los datos de sus sistemas de información están mal organizados y con un mantenimiento deficiente. A esto se debe que la administración de datos sea tan importante. Para entender el problema veamos la manera en que los sistemas de información organizan los datos en archivos de computadora y los métodos tradicionales de administración de archivos.

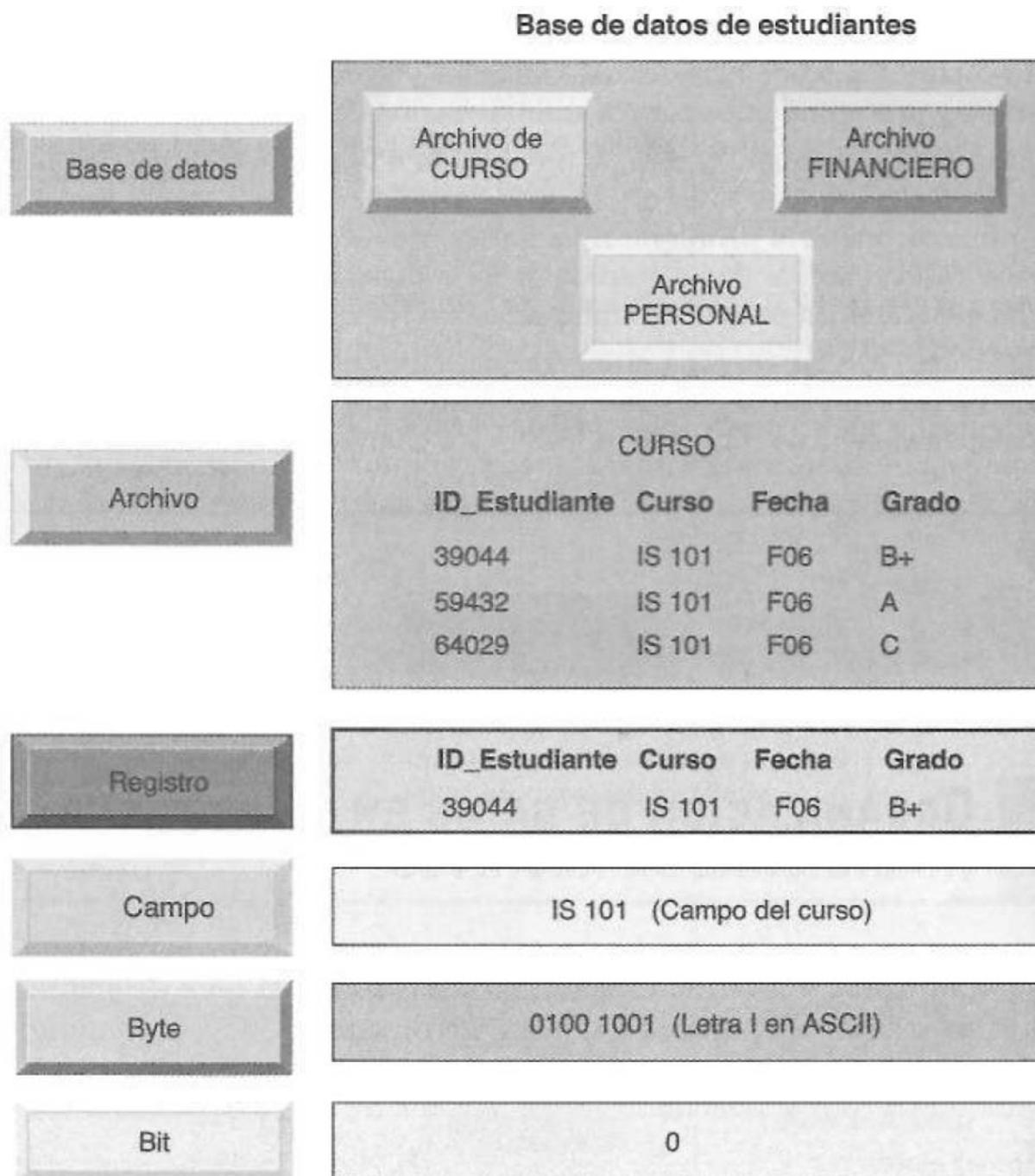
CONCEPTOS DE ORGANIZACIÓN DE ARCHIVOS

Un sistema de cómputo organiza los datos en una jerarquía que empieza con bits y bytes y avanza a campos, registros, archivos y bases de datos (vea la figura 6-1). Un bit representa la unidad más pequeña de datos que puede manejar una computadora. Un grupo de bits, llamado un byte, representa un carácter, el cual puede ser una letra, un número u otro símbolo. Una agrupación de caracteres en una palabra, un grupo de palabras o un número completo (como el nombre o la edad de una persona) se llama **campo**. Un grupo de campos relacionados, como el nombre de un estudiante, el curso que toma, la fecha y el grado, conforma un **registro**; un grupo de registros del mismo tipo se llama **archivo**.

Por ejemplo, los registros de la figura 6-1 podrían constituir el archivo de un curso de un estudiante. Un grupo de archivos relacionados forma una **base de datos**. El archivo de curso del estudiante que se ilustra en la figura 6-1 se podría agrupar con archivos sobre los antecedentes personales y financieros de los estudiantes para crear una base de datos de estudiantes.

Un registro describe una entidad. Una entidad es una persona, lugar, cosa o evento sobre el cual se almacena y conserva información. Cada característica o cualidad que describe a una entidad en particular se llama **atributo**. Por ejemplo, ID_Estudiante, Curso, Fecha y Grado son atributos de la entidad CURSO. Los valores específicos que pueden tener estos atributos se encuentran en los campos del registro que describen la entidad CURSO.

FIGURA 6-1 LA JERARQUÍA DE DATOS



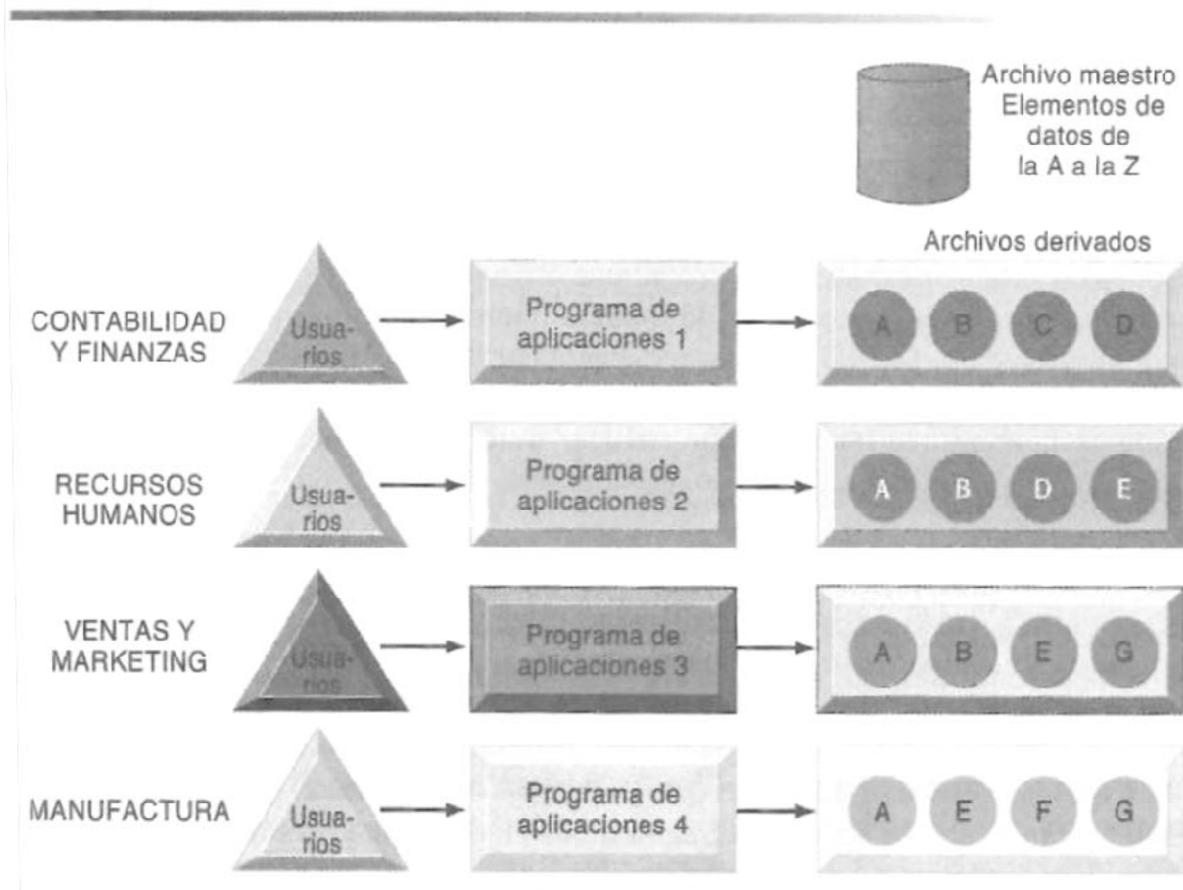
Un sistema de cómputo organiza los datos en una jerarquía que empieza con el bit, el cual representa ya sea un 0 o un 1. Los bits se pueden agrupar para formar un byte para representar un carácter, número o símbolo. Los bytes se pueden agrupar para formar un campo y los campos relacionados se pueden agrupar para formar un registro. Los registros relacionados se pueden conjuntar para formar un archivo y los archivos relacionados se pueden organizar en una base de datos.

PROBLEMAS CON EL ENTORNO TRADICIONAL DE ARCHIVOS

En la mayoría de las organizaciones, los archivos de datos y los sistemas marcaban una tendencia a crecer de manera independiente sin ajustarse a un plan a nivel de toda la empresa. Contabilidad, finanzas, manufactura, recursos humanos y ventas y marketing, desarrollaban sus propios sistemas y archivos de datos. La figura 6-2 ilustra el enfoque tradicional del procesamiento de la información.

Desde luego, para operar, cada aplicación requería sus propios archivos y su propio programa de cómputo. Por ejemplo, el departamento de recursos humanos podía contar con un archivo maestro de personal, un archivo de nómina, un archivo de seguros médicos, un archivo de pensiones, un archivo de listas de correo y así hasta que se creaban decenas, tal vez cientos, de archivos y programas. El departamento de finanzas también podía contar con un archivo de nómina, un archivo de pensiones y una lista maestra de empleados para operar la nómina. Al considerar a la compañía en su conjunto, este proceso conducía a múltiples archivos maestros creados, conservados y operados por divisiones o departamentos separados. Al cabo de 5 o 10 años, la organización se encuentra con una carga de cientos de programas y aplicaciones muy difíciles de mantener y manejar. Los problemas resultantes son redundancia e inconsistencia de datos, dependencia entre los programas y los datos, inflexibilidad, una escasa seguridad de datos y la incapacidad de compartir datos entre aplicaciones.

FIGURA 6-2 PROCESAMIENTO TRADICIONAL DE ARCHIVOS



El uso de un método tradicional para procesar archivos alienta a cada área funcional de una corporación a desarrollar aplicaciones especializadas y archivos. Cada aplicación requiere un archivo de datos único que probablemente será un subconjunto del archivo maestro. Estos subconjuntos del archivo maestro conducen a redundancia e inconsistencia de datos, inflexibilidad en el procesamiento y recursos de almacenamiento desperdiciados.

Redundancia e inconsistencia de datos

La redundancia de datos es la presencia de datos duplicados en múltiples archivos de datos, de tal manera que los mismos datos están almacenados en más de un lugar. La redundancia de datos ocurre cuando diferentes grupos de una organización capturan de manera independiente la misma pieza de información y la almacenan también de manera independiente de los demás grupos. La redundancia de datos desperdicia recursos de almacenamiento y también conduce a la inconsistencia de datos, en la cual el mismo atributo podría tener valores diferentes. Por ejemplo, en las instancias de la entidad CURSO que se ilustra en la figura 6-1, la Fecha podría actualizarse en algunos sistemas pero no en los demás. El atributo ID_Estudiante podría tener diferentes nombres en los diferentes sistemas de la organización. Por ejemplo, algunos sistemas podrían utilizar ID_Estudiante y otros, sólo ID.

Al utilizar diferentes sistemas de codificación para representar los valores de un atributo se podría generar confusión adicional. Por ejemplo, los sistemas de ventas, inventario y manufactura de un detallista de ropa podrían usar códigos diferentes para representar el tamaño de una prenda. Un sistema podría representar el tamaño de la prenda como "extra grande", en tanto que otro utilizaría el código "XL" con el mismo propósito. La confusión resultante dificultaría que las compañías crearan sistemas de administración de las relaciones con el cliente, de administración de la cadena de suministro o empresariales, que integraran los datos desde fuentes diferentes.

Dependencia entre los programas y los datos

La dependencia entre los programas y los datos se refiere a la estrecha relación entre los datos almacenados en archivos y los programas específicos que se requieren para actualizar y mantener esos archivos, de tal manera que los cambios en los programas requieren cambios a los datos. Todo programa tradicional de cómputo tiene que describir la ubicación y naturaleza de los datos con que trabaja. En un entorno tradicional de archivos, cualquier cambio en un programa de software podría requerir un cambio en los datos a que accede ese programa. Por ejemplo, un programa se podría modificar para aceptar un código postal de nueve dígitos en lugar de cinco dígitos. Si los códigos postales de cinco dígitos se cambiaron a nueve dígitos en el archivo original de datos, entonces otros programas que requieren el código postal de cinco dígitos no funcionarían adecuadamente. Implementar tales cambios de manera apropiada podría costar millones de dólares.

Carencia de flexibilidad

Un sistema tradicional de archivos puede enviar informes programados de rutina después de extensos esfuerzos de programación, pero no puede transmitir informes con fines específicos o responder de manera oportuna a requerimientos imprevistos de información. La información requerida por las solicitudes con fines específicos está en algún lugar del sistema pero podría ser demasiado costoso recuperarla. Varios programadores tendrían que trabajar durante semanas para reunir en un nuevo archivo los elementos de datos requeridos.

Seguridad escasa

Dado que hay poco control o administración de datos, el acceso y la difusión de la información podrían salirse de control. Es posible que la administración no tenga forma de saber quién esté teniendo acceso a los datos de la organización, o incluso modificándolos.

Carencia de compartición y disponibilidad de los datos

Debido a que la información está fragmentada en diferentes archivos y en distintas partes de la organización no se pueden relacionar entre sí, es prácticamente imposible que la información se comparta o se acceda de manera

oportuna. La información no puede fluir libremente a través de las diferentes áreas funcionales o distintas partes de la organización. Si los usuarios encuentran diferentes valores de la misma pieza de información en dos sistemas distintos, tal vez no deseen utilizar estos sistemas porque no pueden confiar en la exactitud de sus datos.

6.2 ENFOQUE DE LAS BASES DE DATOS PARA LA ADMINISTRACIÓN DE DATOS

La tecnología de bases de datos puede reducir muchos de los problemas que origina la organización tradicional de archivos. Una definición más rigurosa de una base de datos es un conjunto de datos organizados para servir eficientemente a muchas aplicaciones al centralizar los datos y controlar su redundancia. En vez de que los datos se almacenen en archivos separados para cada aplicación, se guardan físicamente para que se presenten a los usuarios como si estuvieran almacenados en un solo lugar. Una sola base de datos da servicio a múltiples aplicaciones. Por ejemplo, en vez de que una corporación almacene los datos de sus empleados en sistemas de información separados y en diferentes archivos para personal, nómina y prestaciones, podría crear una sola base de datos común de recursos humanos.

SISTEMAS DE ADMINISTRACIÓN DE BASES DE DATOS

Un **sistema de administración de bases de datos (DBMS)** es el software que permite a una organización centralizar los datos, administrarlos eficientemente y proporcionar, mediante los programas de aplicación, el acceso a los datos almacenados. El DBMS actúa como una interfaz entre los programas de aplicación y los archivos de datos físicos. Cuando el programa de aplicación solicita un elemento de datos, como el sueldo bruto, el DBMS encuentra este elemento en la base de datos y lo presenta al programa de aplicación. Si utilizara los archivos de datos tradicionales, el programador tendría que especificar el tamaño y el formato de cada elemento de datos utilizado en el programa e indicar la computadora en que se localizaran.

Al separar las vistas lógica y física de los datos, el DBMS libera al programador o al usuario final de la tarea de comprender dónde y cómo se almacenan realmente los datos. La vista lógica presenta los datos como los deberían percibir los usuarios finales o los especialistas de la empresa, en tanto que la vista física muestra cómo están organizados y estructurados realmente los datos en un medio de almacenamiento físico.

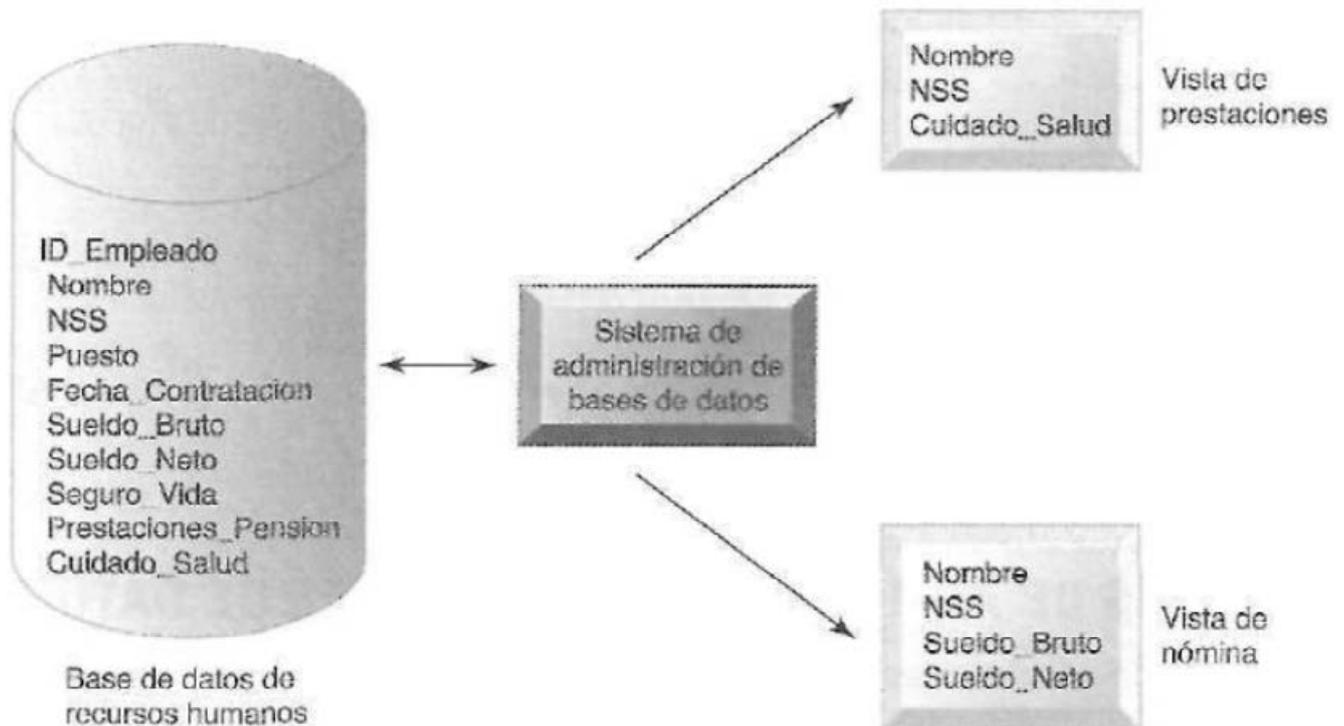
El software de administración de bases de datos hace disponibles las diferentes vistas lógicas de la base de datos física requeridas por los usuarios. Por ejemplo, para la base de datos de recursos humanos ilustrada en la figura 6-3, un especialista en prestaciones podría requerir una vista consistente del nombre, número del seguro social y cobertura del seguro médico del empleado. Un miembro del departamento de nómina podría necesitar datos como el nombre, número del seguro social, sueldo bruto y sueldo neto del empleado. Los datos para todas estas vistas se almacenan en una sola base de datos, donde la organización los puede manejar de una manera más sencilla.

Cómo resuelve un DBMS los problemas del entorno de archivos tradicional

Un DBMS reduce la redundancia y la inconsistencia de datos al minimizar la cantidad de archivos aislados en los cuales se repiten los mismos datos. Quizá el DBMS no permita a la organización eliminar por completo la redundancia de datos, pero puede ayudarle a controlarla. Aun cuando la organización conserve algunos datos redundantes, un DBMS elimina la inconsistencia de los datos porque puede ayudar a la organización a asegurarse de que todas las ocurrencias de los datos redundantes tengan los mismos valores. El DBMS elimina la dependencia entre los programas y los datos, permitiendo a estos últimos ser autosuficientes. El acceso y la disponibilidad de la información se incrementarán y los costos del desarrollo y el mantenimiento de los programas se reducirán porque los usuarios y los programadores pueden

realizar consultas con fines específicos de la información que contiene la base de datos. El DMBS da a la organización la posibilidad de centralizar la administración de datos, su uso y seguridad.

FIGURA 6-3 BASE DE DATOS DE RECURSOS HUMANOS CON MÚLTIPLES VISTAS



Una sola base de datos de recursos humanos proporciona muchas vistas de datos diferentes, según los requerimientos de información del usuario. Aquí se ilustran dos posibles vistas, una para un especialista en prestaciones y una para un miembro del departamento de nómina de la empresa.

DBMS relacional

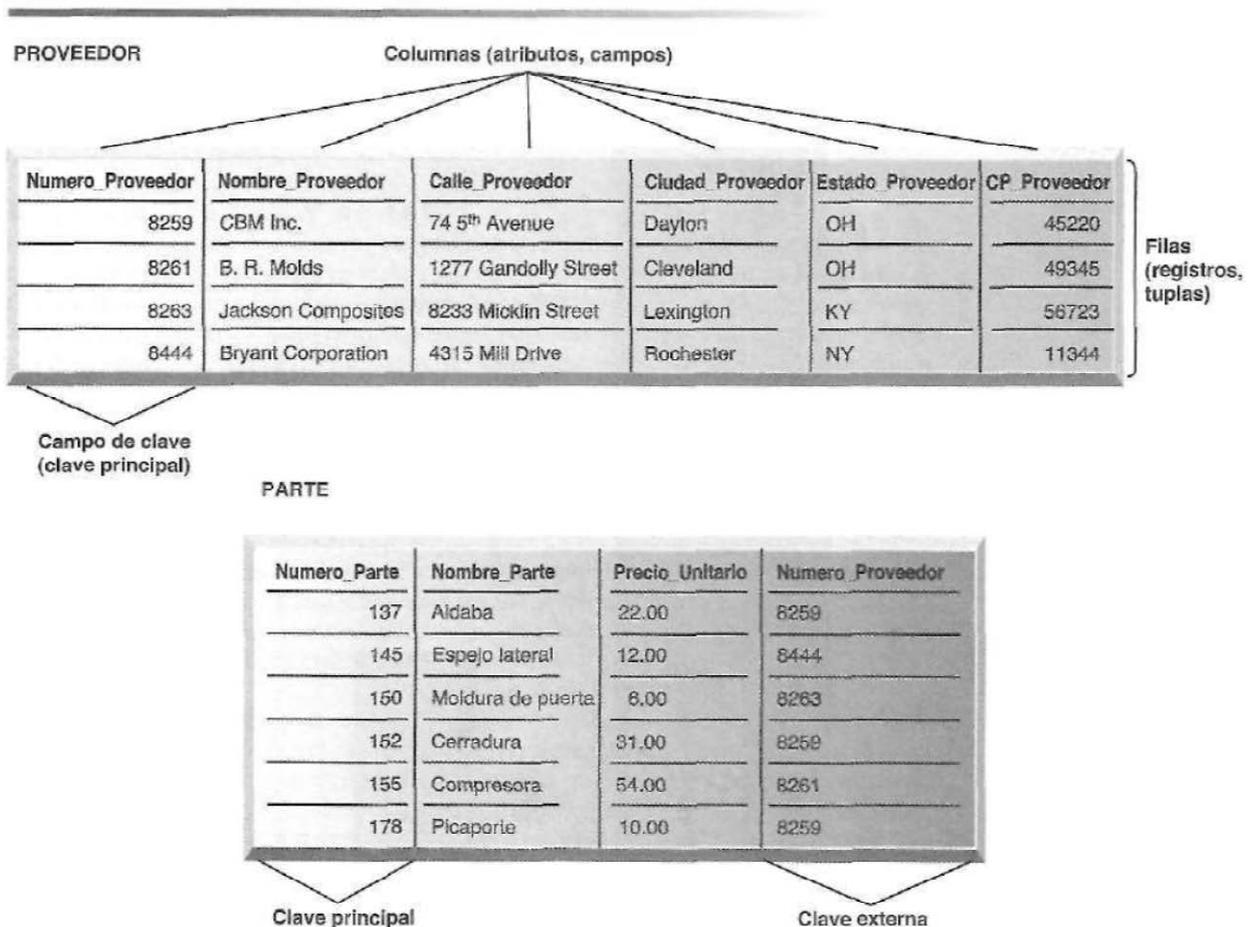
Los DBMS contemporáneos utilizan diferentes modelos de bases de datos para dar seguimiento a las entidades, atributos y relaciones. Actualmente, el tipo más popular de DBMS para PCs, así como para computadoras más grandes y mainframes, es el **DBMS relacional**. Las bases de datos relacionales representan los datos como tablas bidimensionales (llamadas relaciones). Las tablas podrían considerarse como archivos. Cada tabla contiene datos acerca de una entidad y sus atributos. Microsoft Access es un DBMS relacional para sistemas de escritorio, en tanto que DB2, Oracle Database y Microsoft SQL Server son DBMS relacionales para mainframes y computadoras de rango medio. MySQL es un popular DBMS de código abierto, y Oracle Database Lite es un DBMS para pequeños dispositivos de cómputo portátiles.

Veamos cómo una base de datos relacional organiza los datos de proveedores y partes (vea la figura 6-4). La base de datos tiene una tabla independiente para la entidad PROVEEDOR y otra para la entidad PARTE. Ambas tablas constan de una cuadrícula de columnas y filas de datos. Los elementos de datos individuales para cada entidad se almacenan como campos independientes, y cada campo representa un atributo para esa entidad. Los campos de una base de datos relacional también se denominan columnas. Para la entidad PROVEEDOR, el número, nombre, calle, ciudad, estado y código postal del proveedor se almacenan como campos separados dentro de la tabla PROVEEDOR y cada campo representa un atributo para la entidad PROVEEDOR.

La información real sobre un solo proveedor que se encuentra en una tabla se denomina fila. Por lo general, a las filas se les llama **registros**, o en términos muy técnicos, **tuplas**. Los datos de la entidad PARTE tienen su propia tabla separada.

El campo para el Numero_Proveedor de la tabla PROVEEDOR identifica de manera única cada registro con la finalidad de que se pueda recuperar, actualizar o clasificar, y se denomina **campo clave**. Cada tabla de una base de datos relacional tiene un campo designado como **clave principal**. Este campo clave es el identificador único para toda la información en cualquier fila de la tabla y su clave principal no se puede duplicar. Numero_Proveedor es la clave principal para la tabla PROVEEDOR y Numero_Parte es la clave principal para la tabla PARTE. Observe que Numero_Proveedor aparece tanto en la tabla PROVEEDOR como en la tabla PARTE. En la tabla PROVEEDOR, Numero_Proveedor es la clave principal. Cuando el campo Numero_Proveedor aparece en la tabla PARTE se denomina **clave externa** y en esencia es un campo para buscar datos sobre el proveedor de una parte específica.

FIGURA 6-4 TABLAS DE UNA BASE DE DATOS RELACIONAL



Una base de datos relacional organiza los datos en forma de tablas bidimensionales. Aquí se ilustran las tablas para las entidades PROVEEDOR y PARTE que muestran cómo representan a cada entidad y sus atributos. Numero_Proveedor es una clave principal para la tabla PROVEEDOR y una clave externa para la tabla PARTE.

Operaciones de un DBMS relacional

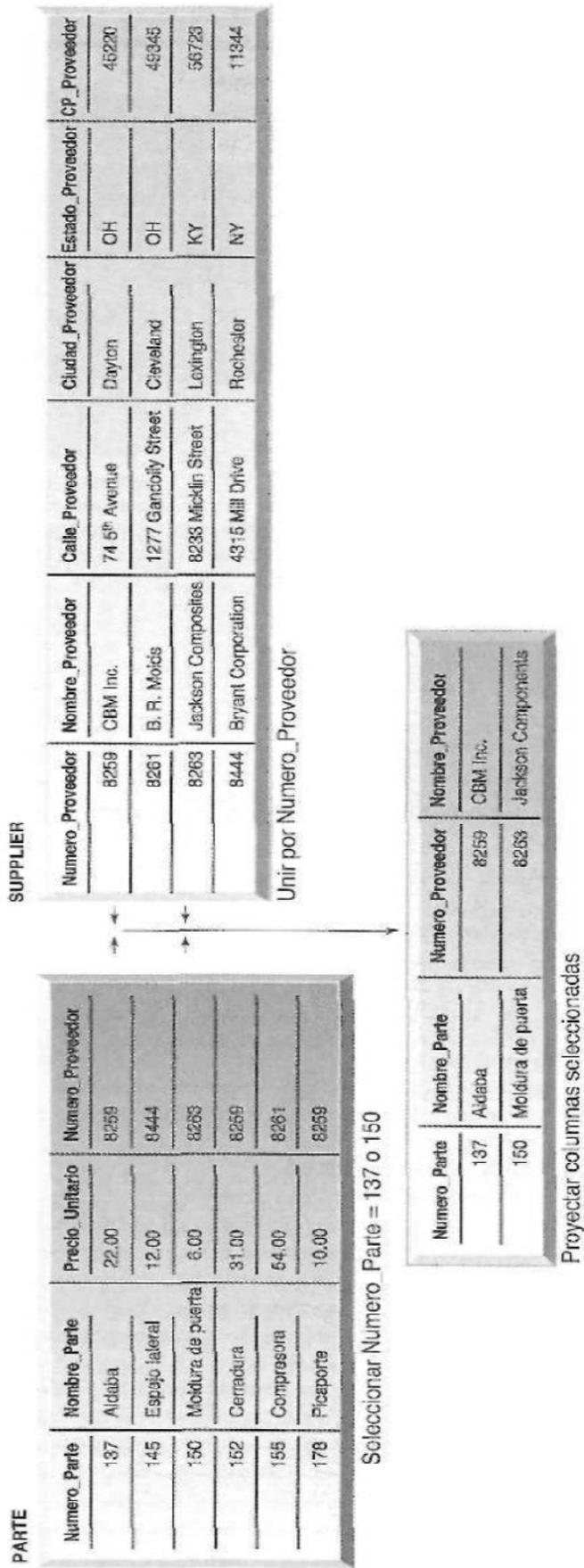
Las tablas de una base de datos relacional se pueden combinar fácilmente para entregar los datos requeridos por los usuarios, siempre y cuando dos tablas compartan un elemento de datos común. Suponga que necesitamos encontrar en esta base de datos los nombres y direcciones de proveedores que puedan suministrarnos la parte número

137 o la número 150. Necesitaríamos información de dos tablas: PROVEEDOR y PARTE. Observe que estos dos archivos tienen un elemento de datos compartido: Numero_Proveedor.

En una base de datos relacional se utilizan tres operaciones básicas, como se muestra en la figura 6-5, para generar conjuntos de datos útiles: seleccionar, proyectar y unir. La operación seleccionar crea un subconjunto que consta de todos los registros (filas) de la tabla que cumplen los criterios establecidos. En nuestro ejemplo queremos seleccionar registros (filas) de la tabla PARTE donde el número de la parte sea a 137 o 150. La operación unir combina tablas relacionales para dar al usuario más información de la que está disponible en las tablas individuales. En nuestro ejemplo deseamos unir la tabla PARTE ahora abreviada (sólo se presentarán las partes 137 o 150) y la tabla PROVEEDOR para formar una sola tabla nueva.

La operación proyectar crea un subconjunto que consiste en las columnas de una tabla, lo cual permite al usuario crear nuevas tablas que contengan solamente la información que se requiere. En nuestro ejemplo, necesitamos extraer de la nueva tabla solamente las siguientes columnas: Numero_Parte, Nombre_Parte, Numero_Proveedor y Nombre_Proveedor.

FIGURA 6-5 LAS TRES OPERACIONES BÁSICAS DE UN DBMS RELACIONAL



DBMS jerárquico y de red

Usted podría encontrarse con viejos sistemas que utilicen un DBMS jerárquico o uno de red. Un DBMS jerárquico modela relaciones uno a muchos, en tanto que uno de red modela relaciones muchos a muchos. Los DBMS jerárquicos y los de red ya no se utilizan para construir nuevas aplicaciones para bases de datos. Son mucho menos flexibles que los DBMS relacionales y no soportan consultas con fines específicos, como las consultas de información en lenguaje natural. En el módulo de seguimiento del aprendizaje de este capítulo, localizado en el sitio Web del libro, encontrará más detalles sobre la manera en que funcionan estos DBMS.

Los DBMS relacionales tienen mucho más flexibilidad para proporcionar datos en consultas con fines específicos, combinar información de diferentes fuentes, y agregar nuevos datos y registros sin alterar los programas y las aplicaciones existentes. Sin embargo, estos sistemas se vuelven lentos si requieren acceder una gran cantidad de veces a los datos almacenados en el disco para ejecutar los comandos seleccionar, unir y proyectar. Desde luego, es posible afinar la base de datos para acelerar consultas predeterminadas.

DBMS orientados a objetos

La gran mayoría de aplicaciones actuales requieren bases de datos que tengan la capacidad para almacenar y recuperar registros que no sólo contengan números y caracteres, sino también dibujos, imágenes, fotografías, voz y video. Los DBMS diseñados para organizar datos estructurados en filas y columnas no son muy adecuados para manejar aplicaciones de gráficos o multimedia. Las bases de datos orientadas a objetos son más adecuadas para este propósito. Por ejemplo, una empresa arquitectónica maneja principalmente objetos y dibujos más que registros de texto estructurados.

Un DBMS orientado a objetos almacena los datos y los procedimientos que realizan operaciones sobre estos datos como objetos que se pueden recuperar y compartir de manera automática. Los sistemas de administración de bases de datos orientadas a objetos (OODBMS) se están popularizando porque se pueden emplear para manejar los diversos componentes multimedia o subprogramas de Java que se utilizan en las aplicaciones para la Web, las cuales comúnmente integran la información fragmentada de diversas fuentes.

Aunque las bases de datos orientadas a objetos pueden almacenar tipos de información más complejos que los DBMS relacionales, son relativamente lentas comparadas con los DBMS relacionales para procesar grandes cantidades de transacciones. Ya están disponibles los sistemas híbridos DBMS relacional —orientado a objetos, para proporcionar capacidades de DBMS orientado a objetos y relacional.

CAPACIDADES DE LOS SISTEMAS DE ADMINISTRACIÓN DE BASES DE DATOS

Un DBMS incluye capacidades y herramientas para organizar, manejar y acceder a la información de una base de datos. Las más importantes son su capacidad de definición de datos, su diccionario de datos y su lenguaje de manipulación de datos.

Un DBMS tiene una capacidad de definición de datos para especificar la estructura del contenido de la base de datos. Esta capacidad se podría utilizar para crear tablas de bases de datos y para definir las características de los campos de cada tabla. Esta información acerca de la base de datos podría documentarse en un diccionario de datos. Un diccionario de datos es un archivo automatizado o manual que almacena definiciones de elementos de datos y sus características.

Microsoft Access integra una característica rudimentaria de diccionario de datos que despliega información sobre nombre, descripción, tamaño, tipo, formato y otras propiedades de cada campo de una tabla cuando se trabaja en

la vista Diseño o cuando se emplea el Documentador. Los diccionarios de datos para grandes bases de datos corporativas podrían captar información adicional, como la propiedad (quién es responsable de dar mantenimiento a los datos en la organización), autorización, seguridad, y los individuos, funciones de negocios, programas e informes que utiliza cada elemento de datos (vea la figura 6-6).

Elaboración de consultas e informes

Un DBMS incluye herramientas para acceder y manipular la información de una base de datos. La mayoría de los DBMS tiene un lenguaje especializado llamado lenguaje de manipulación de datos, el cual se utiliza para agregar, cambiar, eliminar y recuperar la información de la base de datos. Este lenguaje contiene comandos que permiten que los usuarios finales y los especialistas en programación extraigan datos de la base de datos para satisfacer requerimientos de información y desarrollar aplicaciones. El lenguaje actual de manipulación de datos más sobresaliente es el Lenguaje de Consultas Estructurado, o SQL. La figura 6-7 ilustra la consulta SQL que produciría la nueva tabla resultante de la figura 6-5. En los módulos de seguimiento de aprendizaje de este capítulo puede averiguar más sobre cómo realizar consultas SQL.

FIGURA 6-6 MUESTRA DE UN INFORME DE UN DICCIONARIO DE DATOS

NOMBRE: SALARIO

DESCRIPCIÓN: *salario anual del empleado*

TAMAÑO: *9 bytes*

TIPO: *N (numérico)*

FECHA DE MODIFICACIÓN: *1/1/05*

PROPIETARIO: *Compensaciones*

ACTUALIZAR SEGURIDAD: *Personal del sitio*

SEGURIDAD DE ACCESO: *Gerente, Planeación de Compensaciones*

Gerente, Planeación de Recursos Humanos

Gerente, Asuntos de Igualdad de Oportunidades

Gerente, Departamento de Prestaciones

FUNCIONES DE NEGOCIOS UTILIZADAS POR: *Compensaciones*

Planeación de Recursos Humanos

Empleo

Seguros

Pensiones y declaraciones 401K

PROGRAMAS QUE LO UTILIZAN: *PI01000*

PI02000

P103000

INFORMES QUE LO UTILIZAN: *Informe de seguimiento de incrementos salariales*

Listado de revisiones salariales

Listado de referencia de pensiones

La muestra del informe del diccionario de datos para una base de datos de recursos humanos proporciona información útil, como el tamaño del elemento de datos, qué programas e informes lo utilizan y cuál grupo de la organización es el propietario responsable de su mantenimiento.

FIGURA 6-7 EJEMPLO DE UNA CONSULTA SQL

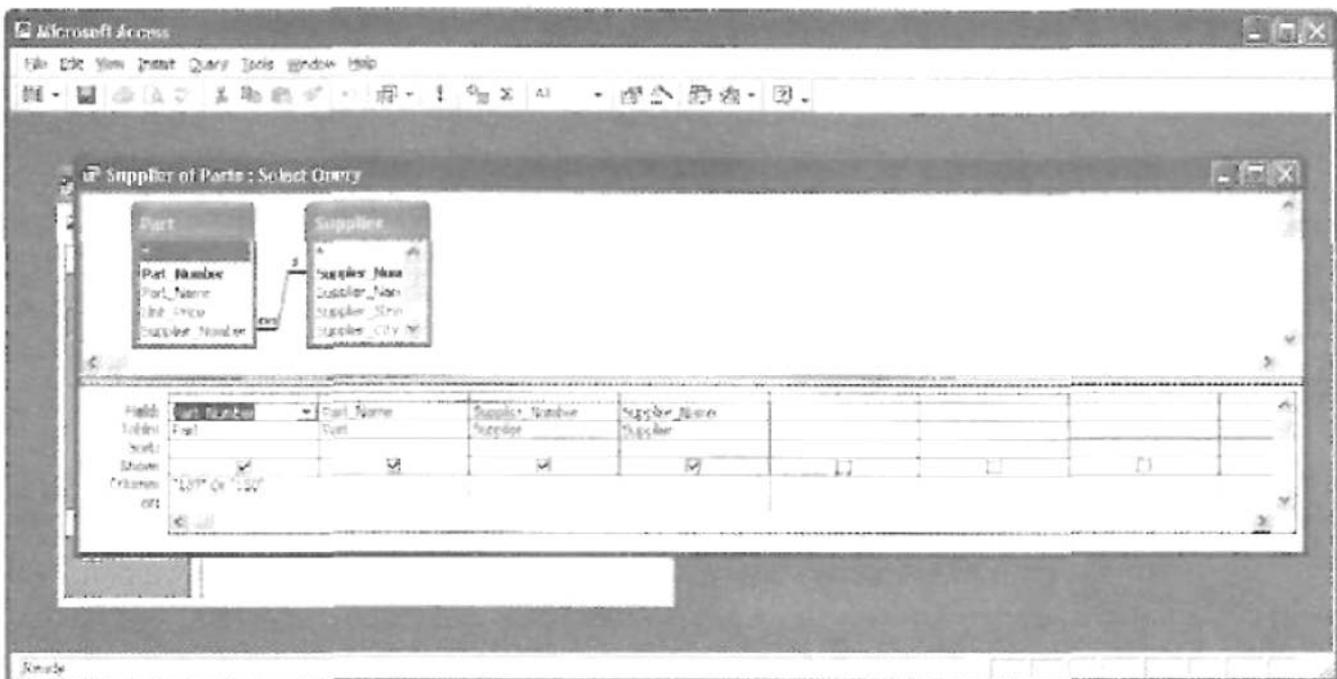
```
SELECT PARTE.Numero_Parte, PARTE.Nombre_Parte, PROVEEDOR.  
Numero_Proveedor, PROVEEDOR.Nombre_Proveedor  
FROM PARTE, PROVEEDOR  
WHERE PARTE.Numero_Proveedor = PROVEEDOR.Numero_Proveedor  
AND Numero_Parte = 137 OR Numero_Parte = 150;
```

Aquí se ilustran las instrucciones de SQL para una consulta que selecciona los proveedores de las partes 137 o 150. Estas instrucciones producen una lista con los mismos resultados que se muestran en la figura 6-5.

Los usuarios de un DBMS para computadoras grandes o de rango medio, como DB2, Oracle o SQL Server, podrían utilizar SQL para recuperar la información que necesitaran de la base de datos. Microsoft Access también utiliza SQL, pero ofrece su propio conjunto de herramientas amigables para el usuario para realizar consultas a la base de datos y para organizar la información de una base de datos en informes más elaborados.

Microsoft Access contiene las características Asistente para consultas y Vista Diseño de consultas que permiten a los usuarios crear consultas identificando las tablas y los campos que desean, así como los resultados, y luego seleccionando las filas de la base de datos que cumplen criterios específicos. A su vez, estas acciones se traducen a instrucciones de SQL. La figura 6-8 ilustra la manera en que se puede construir la misma consulta de SQL para seleccionar partes y proveedores, pero utilizando la Vista Diseño de consultas.

FIGURA 6-8 UNA CONSULTA EN ACCESS



Aquí se ilustra la manera en que se podría construir la consulta de la figura 6-7 por medio de las herramientas de elaboración de consultas de la Vista Diseño de consultas de Access. Se muestran las tablas, campos y los criterios de selección utilizados en la consulta.

Microsoft Access y otros DBMS incluyen características para generar informes que permitan desplegar los datos de interés en un formato más estructurado y refinado del que sería posible con la sola consulta. Crystal Reports es un popular generador de informes para grandes DBMS corporativos, aunque también se puede utilizar con Access. Este

último también cuenta con características para desarrollar aplicaciones para sistemas de escritorio. Incluye herramientas para crear pantallas de captura de datos, informes y para desarrollar la lógica para el procesamiento de transacciones.

DISEÑO DE BASES DE DATOS

Para crear una base de datos usted debe entender las relaciones entre los datos, el tipo de datos que se conservarán en la base de datos, cómo se utilizarán y cómo tendrá que cambiar la organización para manejar los datos desde una perspectiva de la empresa como un todo. La base de datos requiere tanto un diseño conceptual como un diseño físico. El diseño conceptual, o lógico, de una base de datos es un modelo abstracto de la base de datos desde una perspectiva de negocios, en tanto que el diseño físico muestra la manera real en que la base de datos está organizada en dispositivos de almacenamiento de acceso directo.

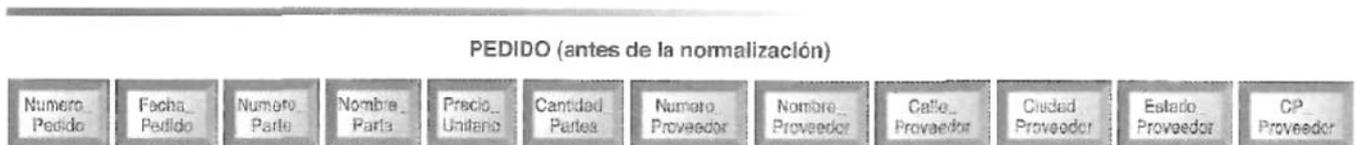
Normalización y diagramas entidad-relación

El diseño conceptual de la base de datos describe la manera en que se agruparán los elementos en la base de datos. El proceso de diseño identifica las relaciones entre los elementos de datos y la manera más eficiente de agruparlos para que cumplan los requerimientos de información de la empresa. El proceso también identifica los elementos de datos redundantes y las agrupaciones de elementos de datos requeridos por programas de aplicación específicos. Los grupos de datos se organizan, refinan y simplifican hasta que surge una vista lógica general de las relaciones entre todos los elementos de datos en la base de datos.

Para utilizar de manera efectiva un modelo de base de datos relacional, es necesario simplificar las agrupaciones complejas de datos para minimizar elementos de datos redundantes y complicadas relaciones muchos a muchos. El proceso para crear estructuras de datos pequeñas, estables, aunque flexibles y adaptables a partir de grupos complejos de datos se denomina normalización. Las figuras 6-9 y 6-10 ilustran este proceso.

En el negocio específico que se modela aquí, un pedido puede contener más de una parte, pero cada parte la surte solamente un proveedor. Si construyéramos una relación llamada PEDIDO con todos los campos incluidos aquí, tendríamos que repetir el nombre y la dirección del proveedor de cada parte del pedido, aun cuando el pedido contenga partes de un proveedor individual. Esta relación contiene lo que se llama grupos de datos repetitivos porque un solo pedido puede contener una gran cantidad de partes de un proveedor en particular. Una manera más eficiente de ordenar los datos es dividir PEDIDO en relaciones más pequeñas, cada una de las cuales describe a una sola entidad. Si vamos paso a paso y normalizamos la relación PEDIDO, el resultado son las relaciones que se ilustran en la figura 6-10. En el primer módulo de seguimiento del aprendizaje de este capítulo encontrará más acerca de la normalización, los diagramas entidad-relación y el diseño de bases de datos.

FIGURA 6-9 RELACION SIN NORMALIZAR PARA UN PEDIDO



Una relación sin normalizar contiene grupos repetidos. Por ejemplo, puede haber muchas partes y proveedores para cada pedido. Existe sólo una correspondencia uno a uno entre Numero_Pedido y Fecha_Pedido.

Los diseñadores de bases de datos documentan sus modelos de datos con un diagrama entidad-relación, el cual se ilustra en la figura 6-11. Este diagrama muestra la relación entre las entidades PEDIDO, ARTICULO_LINEA, PARTE y

PROVEEDOR. Los cuadros representan entidades. Las líneas que conectan los cuadros representan relaciones. Una línea que conecta dos entidades y termina con dos marcas cortas designa una relación uno a uno. Una línea que conecta dos entidades y termina con una pata de cuervo y una marca corta indica una relación uno a muchos. La figura 6-11 muestra que un PEDIDO puede contener muchos ARTICULO_LINEA. (Una PARTE se puede ordenar muchas veces y aparecer una gran cantidad, de veces como artículo de línea en un solo pedido.) Cada PARTE puede tener sólo un PRO-VEEDOR, pero el mismo PROVEEDOR puede suministrar una gran variedad de PARTES.

FIGURA 6-10 TABLAS NORMALIZADAS CREADAS A PARTIR DE PEDIDO



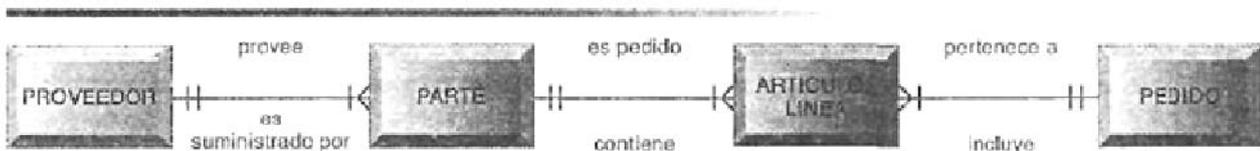
Después de la normalización, la relación original PEDIDO ha sido dividida en cuatro relaciones más pequeñas. La relación PEDIDO queda con sólo dos atributos y la relación ARTICULO_LINEA tiene una clave combinada, o concatenada, conformada por Numero_Pedido y Numero_Parte.

No está de más recalcarlo: si la empresa no implementa su modelo de datos correcto, el sistema será incapaz de servir bien a la empresa. Los sistemas de la compañía no serán tan efectivos como debieran porque tendrán que trabajar con datos que podrían ser inexactos, incompletos o difíciles de recuperar. Tal vez la lección más importante que se puede aprender de este curso es la comprensión de los datos de la organización y de la manera como se deben representar en una base de datos.

Distribución de Bases de datos

El diseño de bases de datos también abarca la manera en que se distribuyen los datos. Se pueden diseñar sistemas de información con una base de datos centralizada que sea utilizada por un procesador central único o por múltiples procesadores en una red cliente/servidor. Alternativamente, la base de datos puede ser distribuida. Una base de datos distribuida es aquella que se almacena en más de un lugar físico.

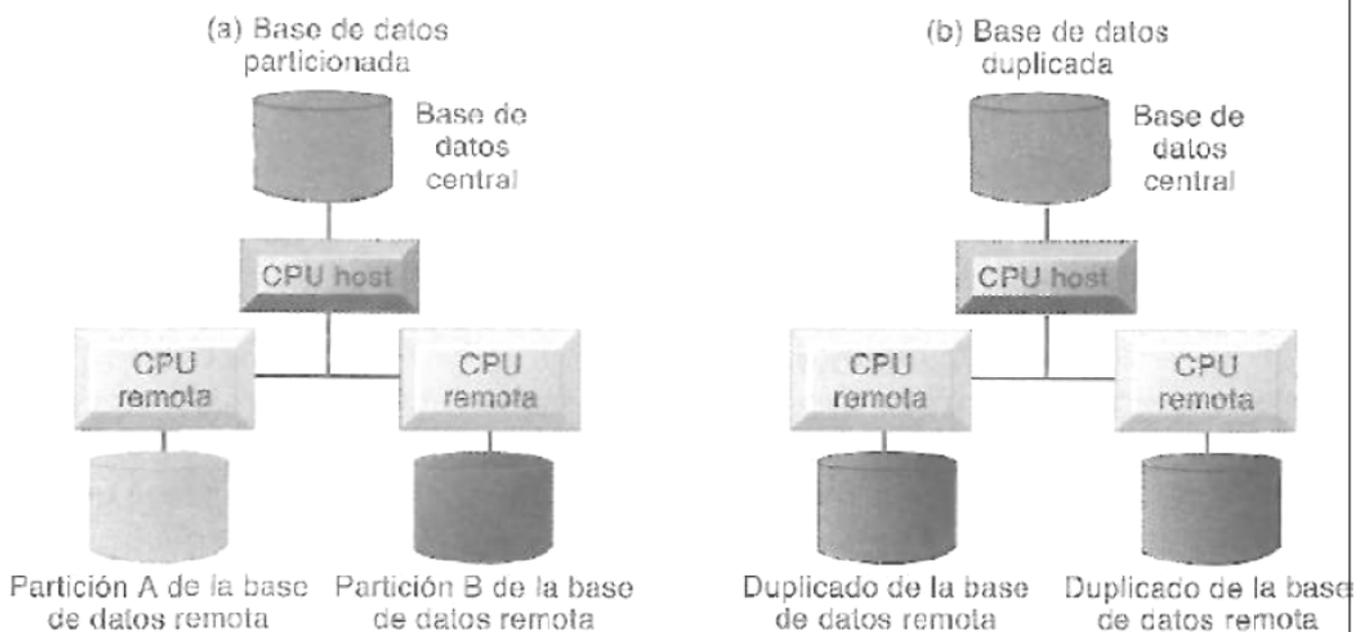
FIGURA 6-11 DIAGRAMA ENTIDAD-RELACIÓN



Este diagrama muestra las relaciones entre las entidades PEDIDO, ARTICULO_LINEA, PARTE y PRO-VEEDOR que podrían utilizarse para modelar la base de datos de la figura 6-10.

Hay dos formas principales de distribuir una base de datos (vea la figura 6-12). En una base de datos particionada, unas partes de la base de datos se almacenan y mantienen físicamente en un lugar y otras partes se almacenan y mantienen en otros lugares (vea la figura 6-12a) de manera que cada procesador remoto tenga los datos necesarios para abastecer su área local. Los cambios en los archivos locales se pueden justificar con la base de datos central a partir de lotes, con frecuencia por la noche. Otra estrategia es replicar (es decir, duplicar por completo) la base de datos central (vea la figura 6-12b) en todas las ubicaciones remotas. Por ejemplo, Lufthansa Airlines reemplazó su base de datos de mainframe centralizada con una base de datos replicada para que la información estuviera disponible de manera más inmediata para los despachadores de vuelos. Cualquier cambio que se haga al DBMS de Lufthansa en Frankfurt se replica automáticamente en Nueva York y en Hong Kong. Esta estrategia también requiere que la actualización de la base de datos central se haga en horas de menos carga.

FIGURA 6-12 BASES DE DATOS DISTRIBUIDAS



Existen formas alternativas para distribuir una base de datos. La base de datos central se puede particionar (a) de tal manera que cada procesador remoto tenga los datos necesarios para satisfacer sus propias necesidades locales. La base de datos central también se puede replicar (b) en todas las ubicaciones remotas.

Los sistemas distribuidos reducen la vulnerabilidad de un solo sitio central masivo. Incrementan el servicio y el nivel de respuesta para los usuarios locales y con frecuencia pueden correr en computadoras más pequeñas y menos caras. Sin embargo, las bases de datos locales en ocasiones pueden desviarse de los estándares y definiciones centrales de datos y plantear problemas de seguridad por la amplia distribución del acceso a datos delicados. Los diseñadores de bases de datos necesitan ponderar estos factores en sus decisiones.

6.3 USO DE BASES DE DATOS PARA MEJORAR EL DESEMPEÑO EMPRESARIAL Y LA TOMA DE DECISIONES

Las empresas utilizan sus bases de datos para dar seguimiento a las transacciones básicas, como el pago a proveedores, el procesamiento de pedidos, el seguimiento de los clientes y el pago a los empleados. Pero también necesitan las bases de datos para suministrar información que ayudará a la empresa a manejar los negocios de manera

más eficiente, y que ayudará a los gerentes y a los empleados a tomar mejores decisiones. Si una empresa desea saber cuál producto es el más popular o quién es su cliente más rentable, puede encontrar la respuesta en los datos.

Por ejemplo, al analizar los datos de compras de sus clientes con tarjeta de crédito, Louise's Trattoria, una cadena de restaurantes de Los Ángeles, se dio cuenta de que la calidad era más importante que el precio para la mayoría de sus clientes, quienes tenían educación universitaria y les gustaba el buen vino. En respuesta a esta información, la cadena introdujo platos vegetarianos, más variedad de mariscos y vinos más caros, elevando sus ventas en más de 10 por ciento.

En una empresa grande, con robustas bases de datos o enormes sistemas para cada función, como manufactura, ventas y contabilidad, se requieren capacidades y herramientas especiales para analizar extensas cantidades de datos y para acceder datos desde múltiples sistemas. Estas capacidades incluyen el almacenamiento de datos, la minería de datos y herramientas para acceder bases de datos internas a través de la Web.

ALMACENES DE DATOS

Suponga que requiere información concisa y confiable acerca de las operaciones actuales, las tendencias y los cambios a través de toda la empresa. Si usted trabajara en una empresa grande, podría dificultársele obtener esta información porque con frecuencia los datos se conservan en sistemas separados, como ventas, manufactura o contabilidad. Algunos de los datos que usted necesitara podrían estar en el sistema de ventas, y otros fragmentos en el sistema de manufactura. Muchos de estos sistemas son antiguos y heredados, que utilizan tecnologías de administración de datos obsoletas o sistemas de archivos en los cuales es difícil para los usuarios acceder a la información.

Usted podría pasar cantidades de tiempo excesivas localizando y recopilando los datos que necesitara, o podría verse obligado a tomar sus decisiones con base en conocimiento incompleto. Si requiriera información sobre tendencias, también podría tener problemas para encontrar los datos acerca de sucesos pasados porque la mayoría de las empresas sólo ponen a disposición inmediata sus datos actuales. El almacenamiento de datos resuelve esos problemas.

¿Qué es un almacén de datos?

Un almacén de datos es una base de datos que almacena datos actuales e históricos de potencial interés para los encargados de la toma de decisiones de toda la empresa. Los datos se originan en muchos sistemas de transacciones operativas esenciales, como los sistemas de ventas, cuentas de clientes y manufactura, y podrían incluir datos derivados de transacciones en sitios Web. El almacén de datos consolida y estandariza información de diferentes bases de datos operativas con el propósito de que la información se pueda utilizar a través de la empresa para el análisis y la toma de decisiones por parte de la administración.

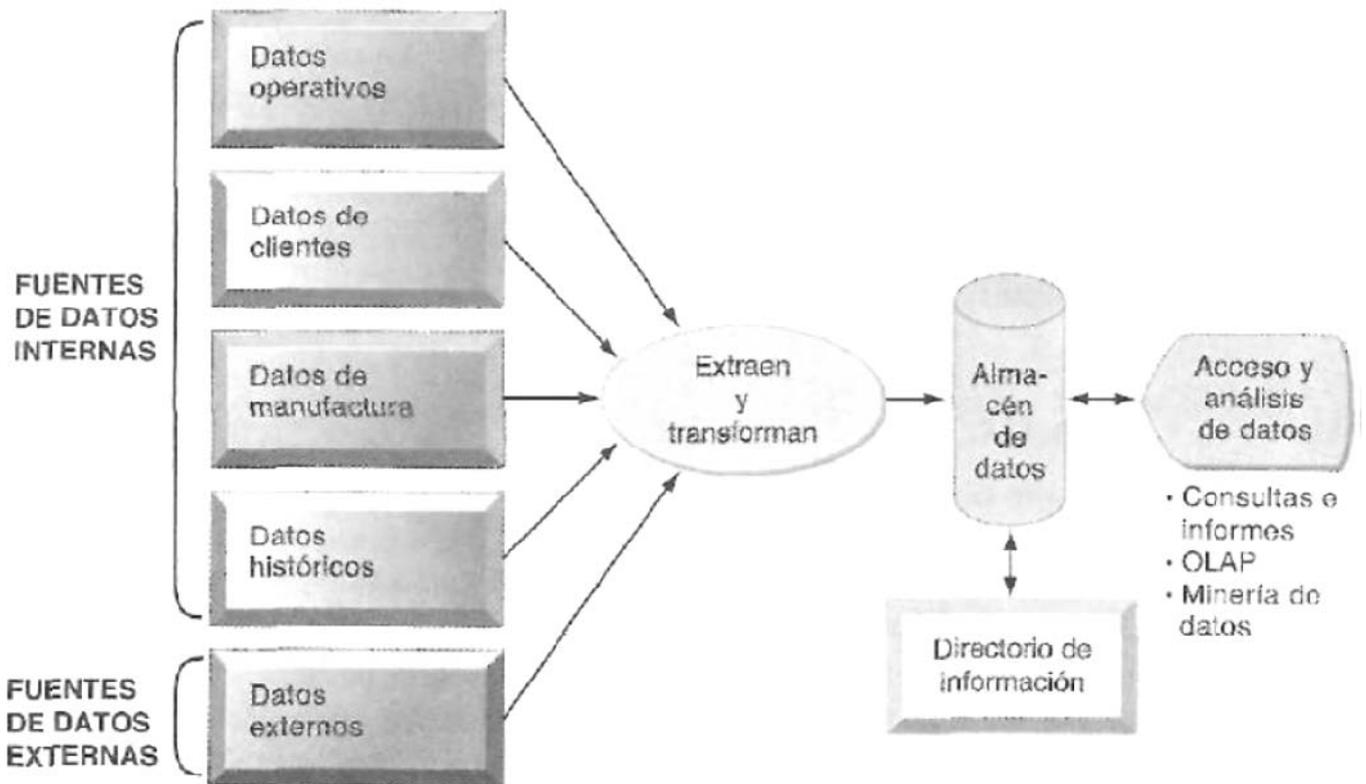
La figura 6-13 ilustra la manera en que funciona un almacén de datos. El almacén de datos pone los datos a disposición de quien los necesite, pero no se pueden modificar. Un sistema de almacén de datos también proporciona una variedad de herramientas de consultas con fines específicos y estandarizadas, herramientas analíticas y características gráficas para la elaboración de informes. Muchas empresas utilizan portales de intranets para poner la información de los almacenes de datos a disposición de toda la empresa.

Mercados de datos

Con frecuencia, las compañías construyen almacenes de datos a nivel general de la empresa, en los que un almacén de datos central abastece a toda la organización, o crean almacenes descentralizados más pequeños llamados mercados de datos. Un mercado de datos es un subconjunto de un almacén de datos en el que una parte resumida o altamente enfocada de los datos de la organización se coloca en una base de datos separada para una población

específica de usuarios. Por ejemplo, una compañía puede desarrollar mercados de datos de marketing y ventas para manejar la información de los clientes. Por lo común, un mercado de datos se enfoca en una sola área objetivo o línea de negocios, por lo que usualmente se puede construir más rápido y a un costo más bajo que un almacén de datos a nivel de toda la empresa.

FIGURA 6-13 COMPONENTES DE UN ALMACÉN DE DATOS



El almacén de datos extrae datos actuales e históricos de múltiples sistemas operativos dentro de la organización. Estos datos se combinan con datos de fuentes externas y se reorganizan en una base de datos central diseñada para elaborar informes y análisis para la administración. El directorio de información suministra información a los usuarios acerca de los datos disponibles en el almacén de datos.

INTELIGENCIA DE NEGOCIOS, ANÁLISIS DE DATOS MULTIDIMENSIONALES Y MINERÍA DE DATOS

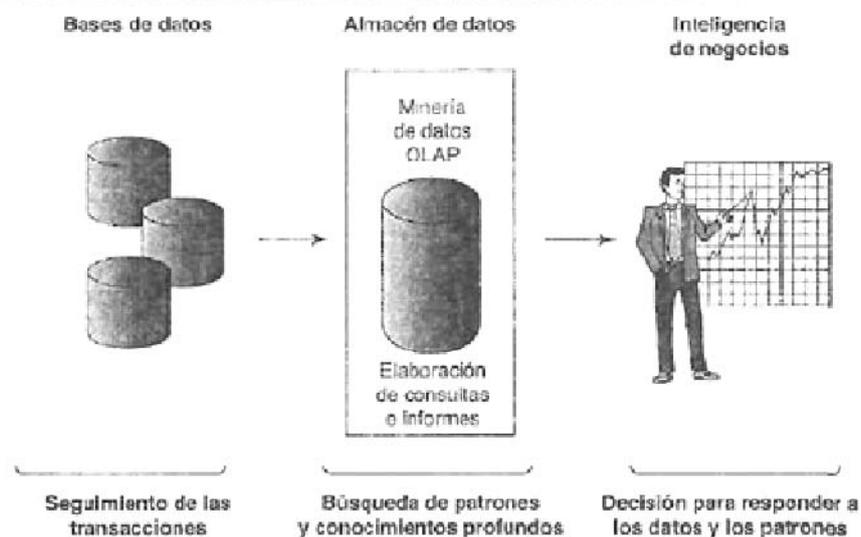
Una vez que se han capturado y organizado los datos en almacenes de datos y en mercados de datos, están a disposición para análisis más profundos. Una serie de herramientas permite a los usuarios analizar estos datos para descubrir nuevos patrones, relaciones y conocimientos profundos para orientar la toma de decisiones. Estas herramientas para consolidar, analizar y dar acceso a extensas cantidades de datos para ayudar a los usuarios a tomar mejores decisiones de negocios con frecuencia se denominan inteligencia de negocios (BI). Las principales herramientas para la inteligencia de negocios incluyen software para consultas e informes de bases de datos, herramientas para análisis de datos multidimensionales (procesamiento analítico en línea) y minería de datos.

Cuando pensamos en la inteligencia como se aplica a los humanos, por lo general, pensamos en la capacidad de la gente para combinar el conocimiento aprendido con nueva información y cambiar comportamientos con la idea de

alcanzar el éxito en su actividad o adaptarse a una nueva situación. De manera similar, la inteligencia de negocios da a las empresas la capacidad de acumular información; desarrollar conocimiento sobre clientes, competidores y operaciones internas, y cambiar la manera de tomar decisiones para lograr una rentabilidad más alta y otras metas de negocios.

Por ejemplo Harrah's Entertainment, la segunda empresa de apuestas más grande en su industria, analiza continuamente los datos sobre sus clientes recopilados cuando la gente juega en sus máquinas tragamonedas o cuando acude a sus casinos y hoteles. El departamento de marketing de Harrah's utiliza esta información para construir un perfil de apuestas detallado, con base en el valor continuo de un cliente en particular para la empresa. Esta información orienta las decisiones de la administración sobre la manera de cultivar a los clientes más rentables, animar a esos clientes a gastar más y atraer más clientes con un potencial alto de generación de ingresos. La inteligencia de negocios ha mejorado tanto las utilidades de Harrah's que se ha convertido en la pieza central de la estrategia de negocios de la empresa.

FIGURA 6-14 INTELIGENCIA DE NEGOCIOS



Una serie de herramientas analíticas trabajan con la información almacenada en las bases de datos para encontrar patrones y conocimientos profundos para ayudar a los gerentes y empleados a tomar mejores decisiones para una mejoría en el desempeño de la organización.

La figura 6-14 ilustra la manera en que funciona la inteligencia de negocios. Las bases de datos operativas de la empresa dan seguimiento a las transacciones que se generan por el manejo del negocio. Estas bases de datos alimentan información al almacén de datos. Los gerentes utilizan las herramientas de inteligencia de negocios para encontrar patrones y significados en los datos. A continuación, los gerentes aprovechan lo que han aprendido del análisis de los datos tomando decisiones de inteligencia de negocios más informadas.

Esta sección le presentará las tecnologías y herramientas de inteligencia de negocios más importantes. En el capítulo 12 le proporcionaremos más detalles acerca de las aplicaciones de inteligencia de negocios.

Procesamiento analítico en línea (OLAP)

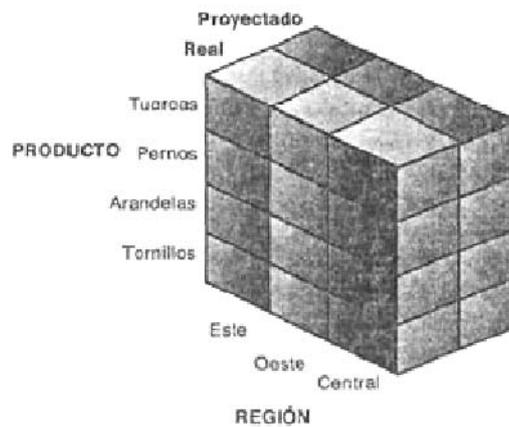
Suponga que su empresa vende cuatro diferentes productos —tuercas, pernos, arandelas y tornillos— en las regiones Este, Oeste y Central. Si usted quisiera hacer una pregunta bastante directa, como la cantidad de arandelas vendidas durante el trimestre pasado, podría encontrar fácilmente la respuesta consultando su base de datos de ventas.

¿Pero qué pasaría si quisiera saber la cantidad de arandelas vendidas en cada una de sus regiones de ventas y comparar los resultados reales con las ventas proyectadas?

Para obtener la respuesta, usted necesitaría el procesamiento analítico en línea (OLAP). OLAP soporta el análisis de datos multidimensionales, el cual permite a los usuarios ver los mismos datos en diferentes formas utilizando múltiples dimensiones. Cada aspecto de la información —producto, precio, costo, región o periodo— representa una dimensión diferente. Así, un gerente de producto podría utilizar una herramienta de análisis de datos multidimensionales para saber cuántas arandelas se vendieron durante junio en el Este, cómo se compara esto con el mes anterior y con el mismo mes del año pasado, y cómo se compara con el pronóstico de ventas. OLAP permite a los usuarios obtener respuestas en línea a preguntas específicas como las anteriores en un lapso de tiempo sumamente rápido, aun cuando los datos estén almacenados en bases de datos bastante grandes, como las cifras de ventas de varios años.

La figura 6-15 muestra un modelo multidimensional que podría crearse para representar productos, regiones, ventas reales y ventas proyectadas. Una matriz de ventas reales se podría apilar sobre una matriz de ventas proyectadas para formar un cubo con seis caras. Si gira 90 grados el cubo en una sola dirección, la cara que se muestra será el producto contra las ventas reales y las proyectadas. Si gira 90 grados el cubo de nueva cuenta, verá la región contra las ventas reales y las proyectadas. Si lo gira 180 grados a partir de la vista original, verá las ventas proyectadas y el producto contra la región. Los cubos se pueden anidar dentro de cubos para construir vistas de datos complejas. Una empresa podría utilizar una base de datos multidimensional especializada o una herramienta que genere vistas de datos multidimensionales en bases de datos relacionales.

FIGURA 6-15 MODELO DE DATOS MULTIDIMENSIONALES



La vista que se muestra es producto contra región. Si usted gira 90 grados el cubo, la cara que se mostrará es producto contra ventas reales y proyectadas. Si lo vuelve a girar, tendrá una vista de la región contra ventas reales y proyectadas. Se pueden tener otras vistas.

Minería de datos

Las consultas de bases de datos tradicionales responden preguntas como "¿Cuántas unidades del producto número 403 se embarcaron en febrero de 2007?" OLAP, o análisis multidimensional, soporta solicitudes de información mucho más complejas, como "Comparar las ventas del producto 403 respecto de lo planeado, por trimestre y región de

ventas, de los dos años anteriores". Con OLAP y el análisis de datos orientado a consultas, los usuarios necesitan tener una buena idea sobre la información que están buscando.

La minería de datos está más orientada al descubrimiento de información. La minería de datos proporciona conocimientos acerca de los datos corporativos que no se podrían obtener con OLAP, al encontrar patrones y relaciones ocultas en robustas bases de datos e infiriendo reglas a partir de ellos para predecir el comportamiento futuro. Los patrones y las reglas se utilizan para orientar la toma de decisiones y predecir el efecto de tales decisiones. Los tipos de información que se pueden obtener a partir de la minería de datos incluyen asociaciones, secuencias, clasificaciones, agrupaciones y pronósticos.

- Las **asociaciones** son ocurrencias enlazadas a un solo evento. Por ejemplo, un estudio de los patrones de compras de un supermercado podría revelar que cuando se compran hojuelas de maíz también se compra, en un 65 por ciento de las veces, una bebida de cola, pero cuando hay alguna promoción se compra una bebida de cola en 85 por ciento de las veces. Esta información ayuda a los gerentes a tomar mejores decisiones porque se han dado cuenta de la rentabilidad de las promociones.
- En las **secuencias**, los eventos se enlazan con el paso del tiempo. Por ejemplo, podríamos encontrar que si se compra una casa, 65 por ciento de las veces se comprará un nuevo refrigerador dentro de las siguientes dos semanas, y que al mes de la compra de la casa habrá un 45 por ciento de veces en que se comprará un horno.
- La **clasificación** reconoce patrones que describen el grupo al cual pertenece un elemento por medio del análisis de los elementos existentes que se han clasificado y por medio de la inferencia de un conjunto de reglas. Por ejemplo, las empresas de tarjetas de crédito o las telefónicas se preocupan por la pérdida de clientes permanentes. La clasificación ayuda a descubrirlas características de los clientes que se podrían perder y puede aportar un modelo para ayudar a los gerentes a predecir quiénes son estos clientes e idear campañas especiales para retenerlos
- El **agrupamiento** funciona de una manera semejante a la clasificación cuando aún no se han definido grupos. Una herramienta de minería de datos puede descubrir diferentes agrupamientos dentro de los datos, como encontrar grupos de afinidad para tarjetas bancarias o particionar una base de datos en grupos de clientes por sus características demográficas y el tipo de sus inversiones personales.
- Aunque estas aplicaciones incluyen predicciones, la **generación de pronósticos** utiliza las predicciones de una manera distinta. Emplea una serie de valores existentes para pronosticar qué otros valores habrá. Por ejemplo, la elaboración de pronósticos podría encontrar patrones en los datos para ayudar a los gerentes a estimar el valor futuro de variables continuas, como cifras de ventas.

Estos sistemas realizan análisis de alto nivel sobre patrones o tendencias, pero también pueden profundizar para ofrecer más detalles cuando sea necesario. Existen aplicaciones de minería de datos para todas las áreas funcionales de una empresa, al igual que para el gobierno y el trabajo científico. Uno de los usos populares de la minería de datos es ofrecer análisis detallados de patrones existentes en los datos de clientes para realizar campañas de marketing o para identificar clientes rentables.

Por ejemplo, Virgin Mobile Australia utiliza un almacén de datos y minería de datos para incrementar la lealtad del cliente y presentar nuevos servicios. El almacén de datos consolida en una enorme base de datos la información de su sistema empresarial, su sistema de administración de las relaciones con el cliente y los sistemas de facturación al cliente. La minería de datos ha permitido a la administración determinar el perfil demográfico de nuevos clientes y relacionarlo con los equipos que hayan comprado. También ha ayudado a la administración a evaluar el desempeño de cada tienda y de las campañas en el punto de venta, las reacciones del consumidor a nuevos productos y servicios, los índices de deserción de clientes y los ingresos generados por cada cliente.

El **análisis predictivo** utiliza técnicas de minería de datos, datos históricos y suposiciones sobre condiciones futuras para predecir resultados de eventos, como la probabilidad de que un cliente responderá a una oferta o comprará un producto específico. Por ejemplo, la división estadounidense de The Body Shop International pie utilizó análisis predictivo con su base de datos de clientes de catálogo, la Web y de las tiendas para identificar clientes que tuvieran más probabilidad de realizar compras por catálogo. Esa información ayudó a la empresa a construir una lista de correos más precisa y dirigida para sus catálogos, con lo cual mejoró la tasa de respuesta a los correos con ofertas de catálogos y a los ingresos por ventas de catálogo.

La minería de datos es una herramienta poderosa y rentable, pero plantea riesgos para la protección de la privacidad individual. La tecnología de minería de datos puede combinar información de muchas fuentes diversas para crear una "imagen de datos" detallada sobre cada uno de nosotros —nuestros ingresos, nuestros hábitos de conducir, nuestras aficiones, nuestras familias y nuestros intereses políticos. En el capítulo 4 se analiza la cuestión sobre si se debe permitir a las empresas recopilar esta información detallada acerca de los individuos. La Sesión Interactiva sobre Organizaciones explora el debate acerca de si las grandes bases de datos que almacenan perfiles de ADN utilizados en la lucha contra el crimen representan una amenaza a la privacidad y al bienestar social.

BASES DE DATOS Y LA WEB

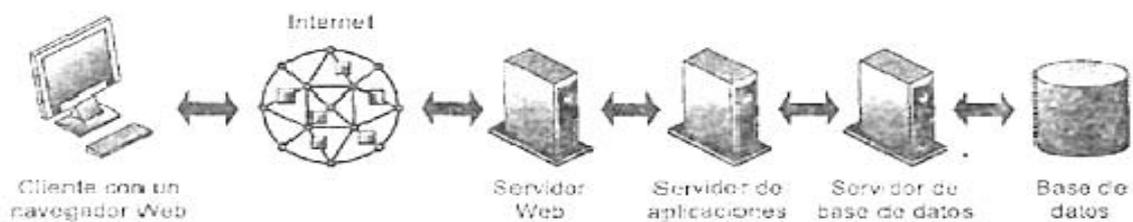
¿Alguna vez ha intentado utilizar la Web para colocar un pedido o ver un catálogo de productos? Si es; así, probablemente estuvo empleando un sitio Web enlazado a una base de datos interna de alguna corporación. Muchas empresas utilizan ahora la Web para poner parte de la información de sus bases de datos internas a disposición de sus clientes y socios de negocios.

Por ejemplo, suponga que un cliente con un navegador Web desea buscar información de precios en la base de datos de un detallista en línea. La figura 6-16 ilustra la manera en que ese cliente podría acceder a la base de datos interna del detallista a través de la Web. El usuario accede al sitio Web del detallista por medio de Internet, utilizando un navegador Web instalado en su PC del cliente. El navegador Web del usuario solicita datos a la base de datos de la organización, utilizando comandos de HTML para comunicarse con el servidor Web.

Puesto que muchas bases de datos "de respaldo" no pueden interpretar comandos escritos en HTML, el servidor Web pasa las solicitudes de datos a software que traduce los comandos HTML en SQL para que los pueda procesar el DBMS que trabaja con la base de datos. En un entorno cliente/servidor, el DBMS reside en una computadora dedicada denominada servidor de base de datos. El DBMS recibe las solicitudes de SQL y suministra los datos requeridos. El middleware transfiere información desde la base de datos interna de la organización de regreso al servidor Web para que se entregue en forma de página Web al usuario.

La figura 6-16 muestra que el middleware que funciona entre el servidor Web y el DBMS podría ser un servidor de aplicaciones ejecutándose en su propia computadora dedicada (vea el capítulo 5). El software servidor de aplicaciones maneja todas las operaciones de las aplicaciones, incluyendo el procesamiento de transacciones y el acceso a datos, entre las computadoras con navegadores y las aplicaciones de negocios o las bases de datos de respaldo de una empresa. El servidor de aplicaciones toma solicitudes del servidor Web, ejecuta la lógica de negocios para procesar transacciones con base en dichas solicitudes, y proporciona la conectividad a los sistemas o bases de datos de respaldo de la organización. De manera alternativa, el software para manejar estas operaciones podría ser un programa personalizado o un script de CGI. Un script de CGI es un programa compacto que utiliza la especificación Interfaz de Compuerta de Enlace (CGI) para procesar los datos en un servidor Web.

FIGURA 6.16 ENLACE DE LAS BASES DE DATOS INTERNAS CON LA WEB



Los usuarios acceden a la base de datos interna de una organización a través de la Web, utilizando sus PCs de escritorio y un navegador.

Existen varias ventajas al utilizar la Web para acceder a las bases de datos internas de una organización. Primero, un navegador Web es más sencillo de utilizar que las herramientas de consulta propietarias. Segundo, la interfaz Web requiere pocos cambios o ninguno a la base de datos interna. Cuesta mucho menos agregar una interfaz Web a un sistema heredado que rediseñar y volver a construir el sistema para mejorar el acceso de los usuarios.

El acceso a bases de datos corporativas a través de la Web está creando nuevas eficiencias, oportunidades y modelos de negocios. ThomasNet.com proporciona un directorio en línea actualizado de más de 650,000 proveedores de productos industriales, como químicos, metales, plásticos, hule y equipo automotriz. Con el nombre anterior de Thomas Register, la empresa acostumbraba enviar enormes catálogos impresos con esta información. iGo.com es una empresa en Internet que vende baterías y accesorios para teléfonos celulares y dispositivos de cómputo. Su sitio Web enlaza a una base de datos relacional de gran tamaño que contiene información de productos como baterías y periféricos para casi todas las marcas y modelos de teléfonos celulares y dispositivos electrónicos portátiles.



El sitio Web de Internet Movie Database se enlaza a una enorme base de datos que incluye resúmenes, información de repartos y biografías de actores de casi todas las películas que se han filmado.

6.4 ADMINISTRACIÓN DE LOS RECURSOS DE DATOS

La implementación de una base de datos es sólo el comienzo. Para asegurarse de que los datos de su negocio permanecerán exactos, confiables y disponibles con facilidad para quienes los requieran, su empresa necesitará políticas y procedimientos especiales para la administración de datos.

ESTABLECIMIENTO DE UNA POLÍTICA DE INFORMACIÓN

Todas las empresas, grandes y pequeñas, necesitan una política de información. Los datos de su empresa son un recurso importante, y a usted no le agrada que los demás hagan con ellos lo que se les antoje. Usted necesitará contar con reglas sobre la manera en que se organizarán y mantendrán los datos, y quién tendrá autorización para verlos o modificarlos.

Una política de información específica las reglas de la organización para compartir, distribuir, adquirir, estandarizar, clasificar e inventariar la información. Una política de información establece procedimientos y responsabilidades específicos, que identifican cuáles usuarios y unidades de la organización pueden compartir información, dónde se puede distribuir la información y quién es responsable de actualizar y mantener la información. Por ejemplo, una política de información típica podría especificar que sólo miembros seleccionados de los departamentos de nómina y de recursos humanos podrían tener autorización para modificar y ver datos delicados sobre empleados, como sus salarios o sus números de seguro social, y que estos departamentos son responsables de garantizar que tales datos sean exactos.

Si usted está en una empresa pequeña, tal vez los propietarios o los gerentes establezcan la política de información. En una organización grande, el manejo y la planeación de la información como un recurso corporativo con frecuencia requiere una función formal de administración de datos. La administración de datos es responsable de las políticas y procedimientos específicos por medio de los cuales se pueden manejar los datos como un recurso organizacional. Estas responsabilidades incluyen desarrollar la política de información, planificar los datos, supervisar el diseño lógico de la base de datos y el desarrollo del diccionario de datos, y vigilar la manera en que los especialistas en sistemas de información y los grupos de usuarios finales utilizan los datos.

Tal vez haya escuchado el término gobierno de la información utilizado para describir muchas de estas actividades. Promovido por IBM, el gobierno de la información se encarga de las políticas y procesos para manejar la disponibilidad, usabilidad, integridad y seguridad de los datos empleados en una empresa, con un énfasis especial en promover la privacidad, la seguridad, la calidad de los datos y el cumplimiento de las regulaciones gubernamentales.

Una organización grande también tendrá un grupo de diseño y administración de bases de datos dentro de la división de sistemas de información corporativa, responsable de definir y organizar la estructura y contenido de la base de datos, así como de su mantenimiento. En estrecha cooperación con los usuarios, el grupo de diseño establece la base de datos física, las relaciones lógicas entre los elementos, y las reglas de acceso y los procedimientos de seguridad. Las funciones que desempeña se conocen como administración de la base de datos.

ASEGURAMIENTO DE LA CALIDAD DE DATOS

Una base de datos y una política de información bien diseñadas contribuirán a garantizar que la empresa tenga la información que necesita. No obstante, se deben emprender acciones adicionales para asegurar que la información de las bases de datos de la organización sea exacta y confiable.

¿Qué pasaría si el número telefónico o el saldo de la cuenta de un cliente fueran incorrectos? ¿Cuál sería el impacto si la base de datos tuviera un precio erróneo para el producto que usted vendió? Los datos imprecisos, a

destiempo o inconsistentes en relación con otras fuentes de información conducen a tomar decisiones incorrectas, devolución de productos e incluso pérdidas financieras.

De acuerdo con Forrester Research, 20 por ciento del correo y de las entregas de paquetes comerciales en Estados Unidos fueron devueltos debido a nombres o direcciones incorrectos. Los consultores de Gartner Group reportaron que más de 25 por ciento de la información crítica de las bases de datos de grandes empresas de Fortune 1000 es imprecisa o está incompleta, incluyendo códigos y descripciones de productos equivocados, descripciones de inventario deficientes, datos financieros erróneos, información incorrecta de proveedores y datos incorrectos de empleados. Gartner cree que los datos de clientes pierden calidad a un ritmo de 2 por ciento mensual, lo cual hace que la calidad de datos deficiente sea uno de los principales obstáculos para administrar con éxito las relaciones con el cliente (Gage y McCormick, 2005; Klau, 2003).

Piense en todas las ocasiones en que ha recibido varias piezas de la misma publicidad por correo directo durante el mismo día. Muy probablemente éste sea el resultado de que su nombre esté varias veces en una base de datos. Su nombre podría haber sido mal escrito o quizá utilizó su segundo nombre en una ocasión y no en otra, o tal vez la información se introdujo inicialmente en un formulario en papel y no se digitalizó de manera adecuada en el sistema. Debido a estas inconsistencias, ¡la base de datos lo trataría a usted como si fueran diferentes personas! Con frecuencia recibimos correo repetitivo dirigido a Laudon, Lavdon, Lauden o Landon.

Si una base de datos está debidamente diseñada y se han establecido estándares de datos para toda la empresa, los elementos de datos duplicados o inconsistentes deberán ser mínimos. Sin embargo, la mayoría de los problemas de calidad de datos, como nombres mal escritos, números traspuestos o códigos incorrectos o faltantes, se derivan de errores durante la captura de datos. La incidencia de tales errores se está incrementando a medida que las empresas desplazan sus negocios a la Web y permiten a los clientes y proveedores introducir datos en sus sitios Web que actualizan directamente sus sistemas internos.

Antes de que una nueva base de datos entre en funcionamiento, las organizaciones tienen que identificar y corregir los datos erróneos y establecer mejores rutinas para editar datos una vez que sus bases de datos estén operando. El análisis de la calidad de datos suele empezar con una auditoría de la calidad de los datos, que es un sondeo estructurado de la exactitud y nivel de completitud de los datos que hay en un sistema de información. Las auditorías de la calidad de los datos se pueden realizar sondeando archivos de datos completos, sondeando muestras de los archivos de datos o sondeando a los usuarios finales acerca de su percepción de la calidad de datos.

La depuración de datos, también conocida como limpieza de datos, consiste en actividades para detectar y corregir datos de una base de datos que sean incorrectos o incompletos, tengan un formato inadecuado o sean redundantes. La depuración de datos no sólo corrige errores sino que también impone la consistencia entre diferentes conjuntos de datos originados en sistemas de información separados. NASCAR, descrita en el caso con que inicia el capítulo, necesitaba utilizar la depuración de datos para eliminar inconsistencias y errores en los datos provenientes de múltiples fuentes cuando generó su amplia base de datos de aficionados. Existe software de depuración de datos especializado para sondear automáticamente archivos de datos, corregir errores en los datos e integrar los datos en un formato consistente para toda la empresa.

La Sesión Interactiva sobre Administración describe algunas empresas que padecieron problemas de calidad de datos y la manera en que trataron de solucionarlos. Estos problemas eran multifacéticos. A medida que lea el caso, busque los factores administrativos, organizacionales y tecnológicos que produjeron estos problemas de calidad de datos y la manera en que los enfrentaron.