Chapter 3

Databases and Data Warehouses: Building Business Intelligence

How Can a Business Increase its Intelligence?

Summary

- Overview of Main Concepts
- Details/Design of a Relational Database
- Creating Business Intelligence using Databases
- Special Issues in Data Management: Ownership and Security

Information processing in a business

- Businesses use IT tools to manage and organize the information they keep.
  - ***Online transaction processing (OLTP)*** – gathering, processing information and updating information as soon as it is entered
  - ***Online analytical processing (OLAP)*** – manipulation of information to support decision making

Information processing in a business

- ***On-line Transaction Processing (OLTP)*** is supported by operational databases & database management systems (DBMS). It
  - collects input information e.g. from sales and orders
  - updates existing information
- ***On-line Analytical Processing (OLAP)*** is supported by data warehouses and data-mining tools. It creates business intelligence for decision making.

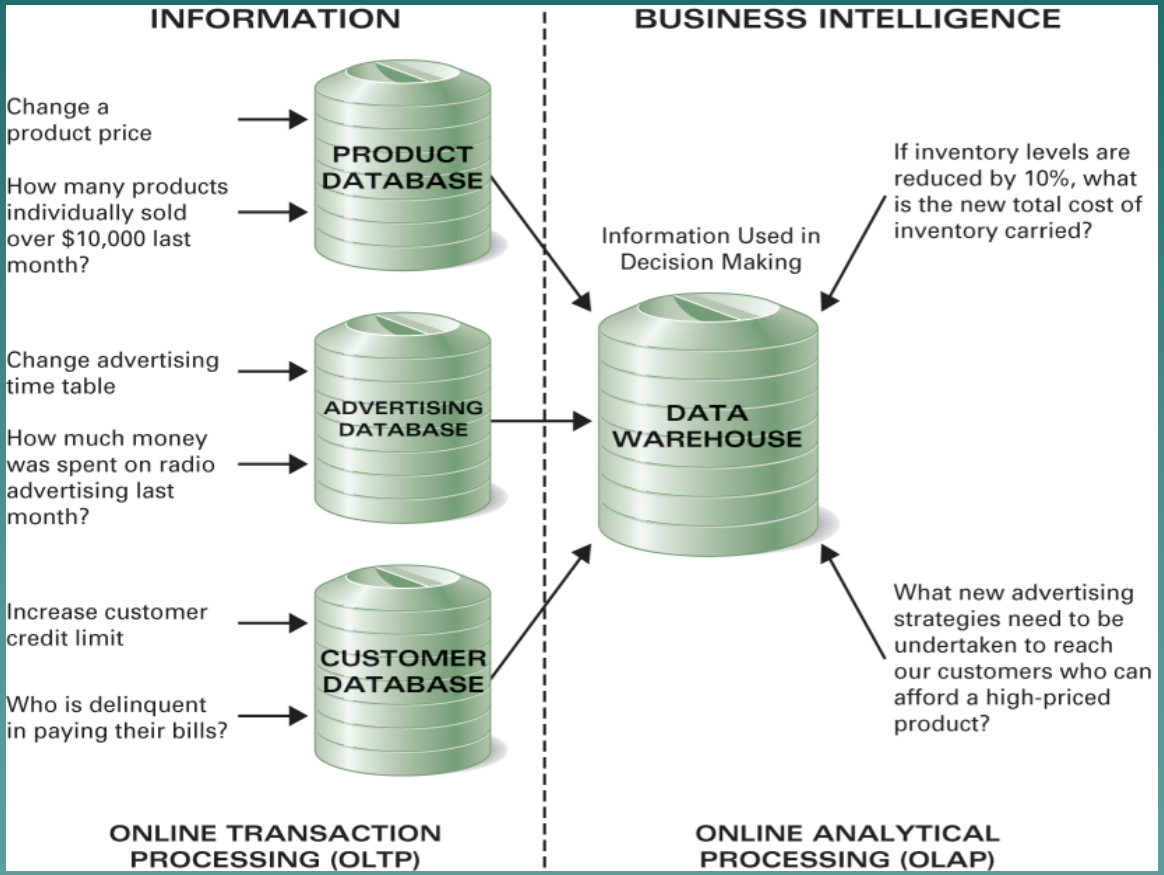
Business Intelligence (BI)

Business intelligence is an organization's collection of information about customers, competitors, business partners, the business environment and internal operations. ***BI*** is necessary for making appropriate business decisions.

To create ***business intelligence***, knowledge workers need:

- data
- information
- appropriate IT tools

Creating Business Intelligence

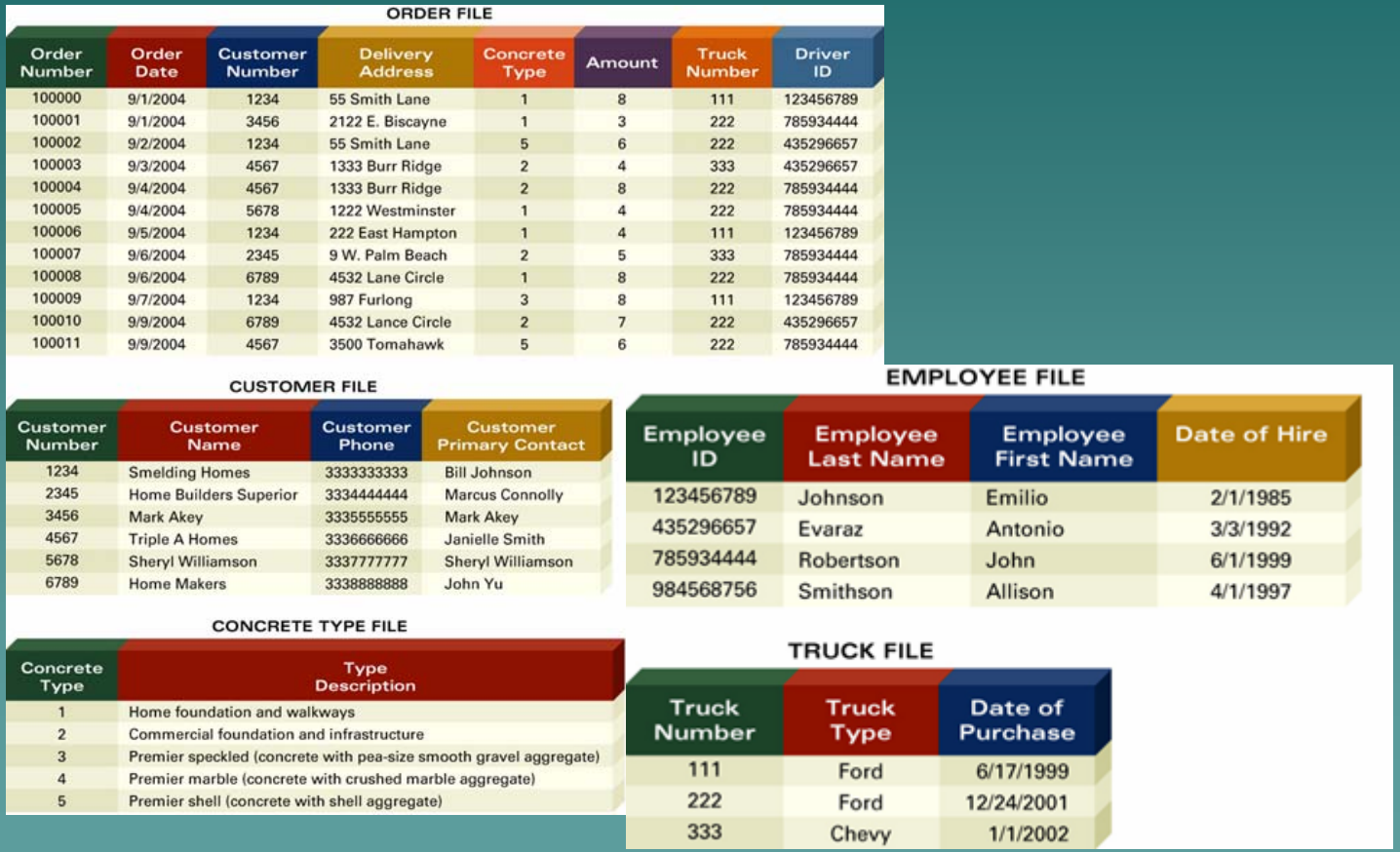


The Relational Database Model

- ***Database*** – collection of information that is organized and accessed according to the logical structure of that information.

- ***Relational database*** – series of logically related 2-D tables or files for storing information in a database. The term ***relation*** refers to the 2-D table or file. A relational database is made up both the information and the logical structure of that information.

Fig 3.2: Solomon's database is made up of five relations



ORDER FILE

Order Number	Order Date	Customer Number	Delivery Address	Concrete Type	Amount	Truck Number	Driver ID
100000	9/1/2004	1234	55 Smith Lane	1	8	111	123456789
100001	9/1/2004	3456	2122 E. Biscayne	1	3	222	785934444
100002	9/2/2004	1234	55 Smith Lane	5	6	222	435296657
100003	9/3/2004	4567	1333 Burr Ridge	2	4	333	435296657
100004	9/4/2004	4567	1333 Burr Ridge	2	8	222	785934444
100005	9/4/2004	5678	1222 Westminster	1	4	222	785934444
100006	9/5/2004	1234	222 East Hampton	1	4	111	123456789
100007	9/6/2004	2345	9 W. Palm Beach	2	5	333	785934444
100008	9/6/2004	6789	4532 Lane Circle	1	8	222	785934444
100009	9/7/2004	1234	987 Furlong	3	8	111	123456789
100010	9/9/2004	6789	4532 Lance Circle	2	7	222	435296657
100011	9/9/2004	4567	3500 Tomahawk	5	6	222	785934444

CUSTOMER FILE

Customer Number	Customer Name	Customer Phone	Customer Primary Contact
1234	Smelding Homes	3333333333	Bill Johnson
2345	Home Builders Superior	3334444444	Marcus Connolly
3456	Mark Akey	3335555555	Mark Akey
4567	Triple A Homes	3336666666	Janielle Smith
5678	Sheryl Williamson	3337777777	Sheryl Williamson
6789	Home Makers	3338888888	John Yu

EMPLOYEE FILE

Employee ID	Employee Last Name	Employee First Name	Date of Hire
123456789	Johnson	Emilio	2/1/1985
435296657	Evaraz	Antonio	3/3/1992
785934444	Robertson	John	6/1/1999
984568756	Smithson	Allison	4/1/1997

CONCRETE TYPE FILE

Concrete Type	Type Description
1	Home foundation and walkways
2	Commercial foundation and infrastructure
3	Premier speckled (concrete with pea-size smooth gravel aggregate)
4	Premier marble (concrete with crushed marble aggregate)
5	Premier shell (concrete with shell aggregate)

TRUCK FILE

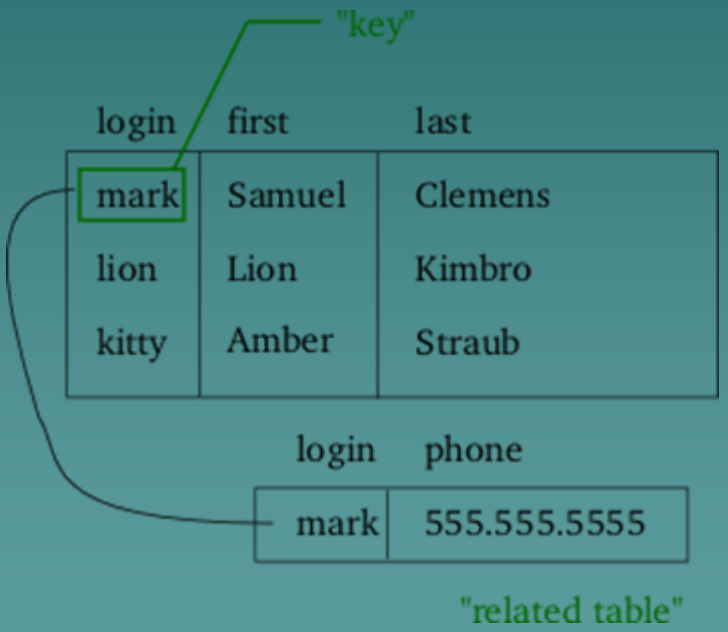
Truck Number	Truck Type	Date of Purchase
111	Ford	6/17/1999
222	Ford	12/24/2001
333	Chevy	1/1/2002

In a relational database, information is organized and accessed according to its logical structure.

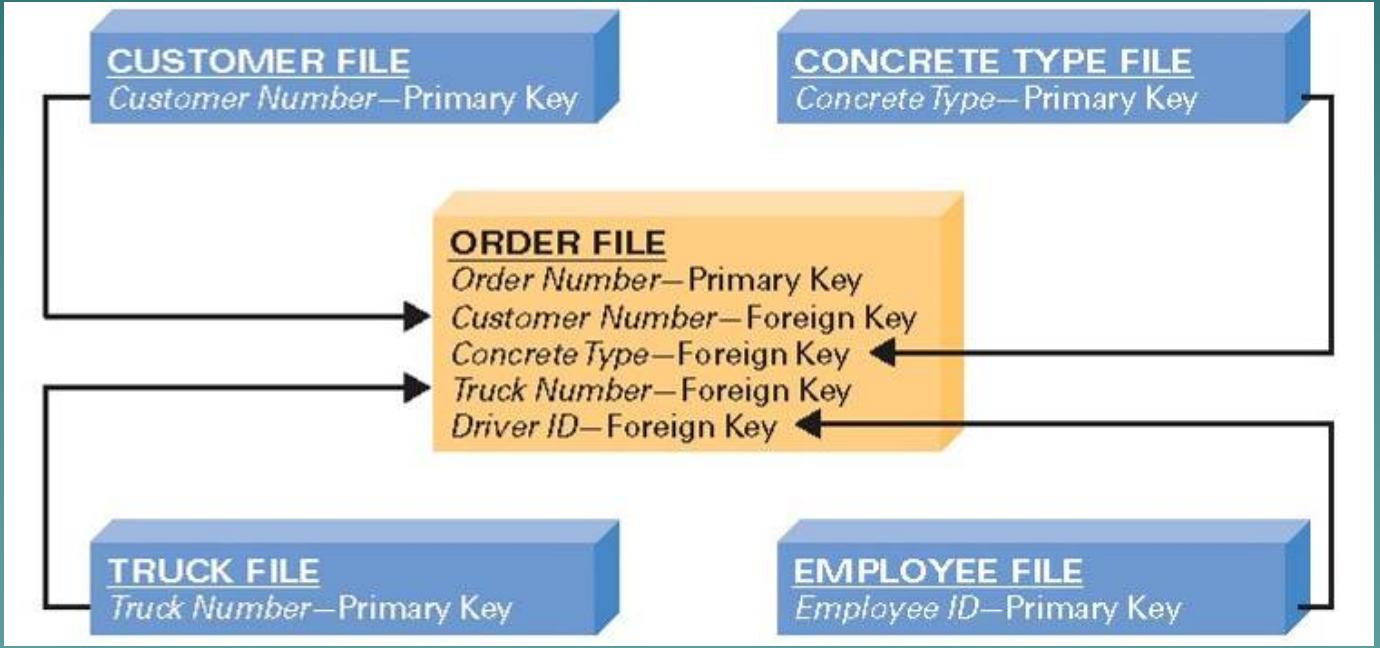
The ***data dictionary*** contains the logical structure for the database information. For example, the data dictionary would specify that a 10-digit *Customer Phone* field is in the *Customer* file.

Before creating relationships between files, you must choose a primary key for each file

- ***Primary key*** – field (or group of fields) that uniquely describes each record in the file e.g. the customer number uniquely identifies a customer in the customer file
- ***Foreign key*** – primary key of one file that appears in another file



Database files are connected using common fields e.g. The Order file and Truck file are connected using Truck Number



Defining the logical structure of a relational database

- ***Integrity constraints*** are rules that help ensure the quality of information e.g. when an order is input into the Order file, the customer must already exist in the Customer file.
- The ***data dictionary*** defines each type of information in a table e.g. format, length.
- ***Foreign keys*** must be found as ***primary keys*** in another file e.g. a *Customer Number* in the *Order* table must also be present in the *Customer* table.

Database Management System (DBMS)

A ***DBMS*** is used to specify the logical requirements for a database as well as provide the means by which information is accessed.

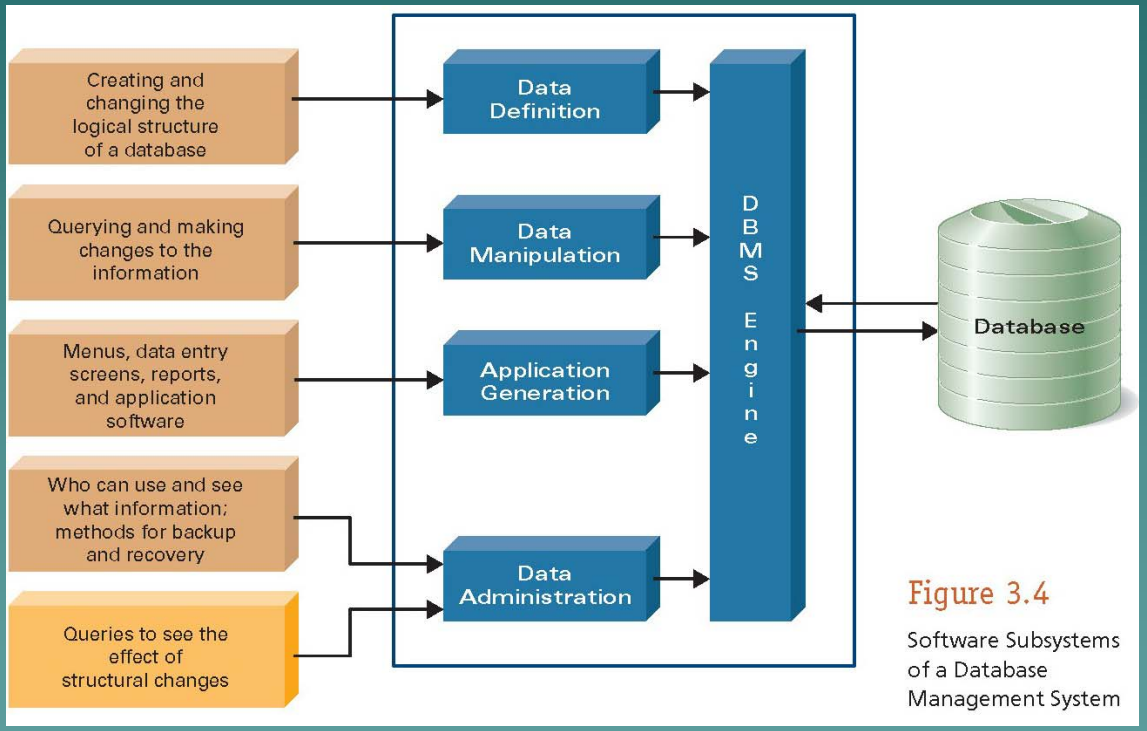


Figure 3.4
Software Subsystems of a Database Management System

A database management system has five software components.

1. DBMS engine
2. Data definition subsystem
3. Data manipulation subsystem
4. Application generation subsystem
5. Data administration subsystem

First component of a DBMS: the DBMS Engine

The ***DBMS engine*** accepts ***logical*** requests from other DBMS subsystems, converts them into the ***physical*** equivalents, and accesses the database and data dictionary on a storage device.

- ***Physical view*** – how information is physically arranged, stored, and accessed on a storage device e.g. customer table is stored on a particular track and sector on the computer's hard drive
- ***Logical view*** – how the knowledge worker arranges and accesses information
  e.g. may sort the customer file by customer ID

Second component of a DBMS: the data definition subsystem

*The **data definition subsystem*** helps the user create and maintain the data dictionary. It defines the structure of the files in the database

When creating the data dictionary, the logical structure of the database is defined. The following are examples of logical properties that can be specified for a field in the data dictionary:

- Field name e.g. SIN
- Data type e.g. alphabetic
- Default value e.g. 000-000-000

Third component of a DBMS: the data manipulation subsystem

- The ***data manipulation subsystem*** helps you add, change, and delete information in a database and query it to find valuable information. It is the main means by which a user works with the information in a database.
- The ***data manipulation subsystem*** includes views, report generators, query-by-example tools, and structured query language

View

The data manipulation subsystem's ***view*** allows you to see the contents of a database file, make changes, and query the database to find information.

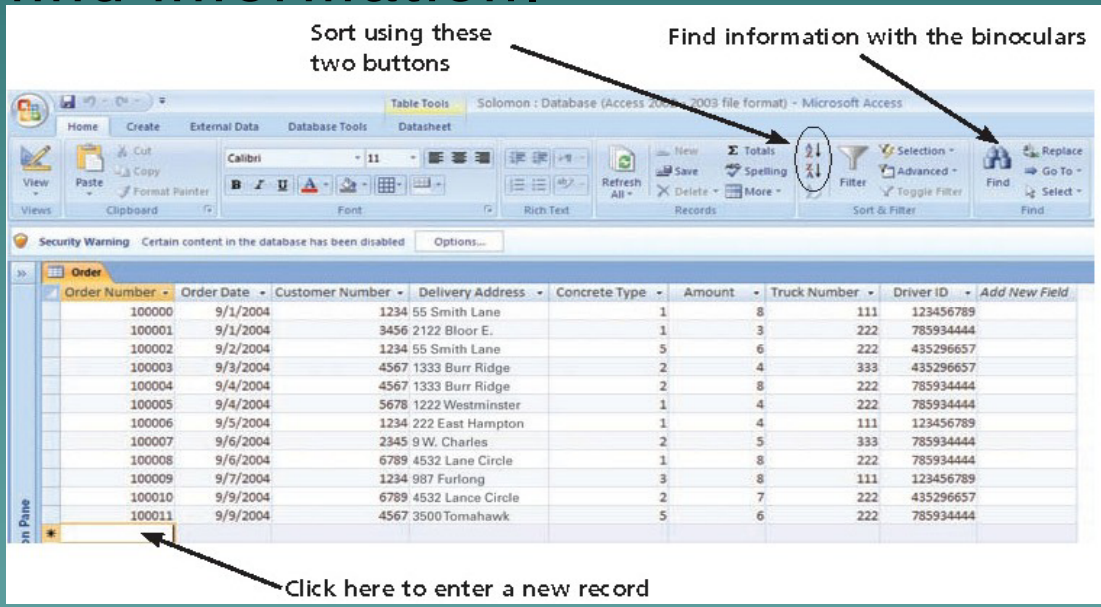
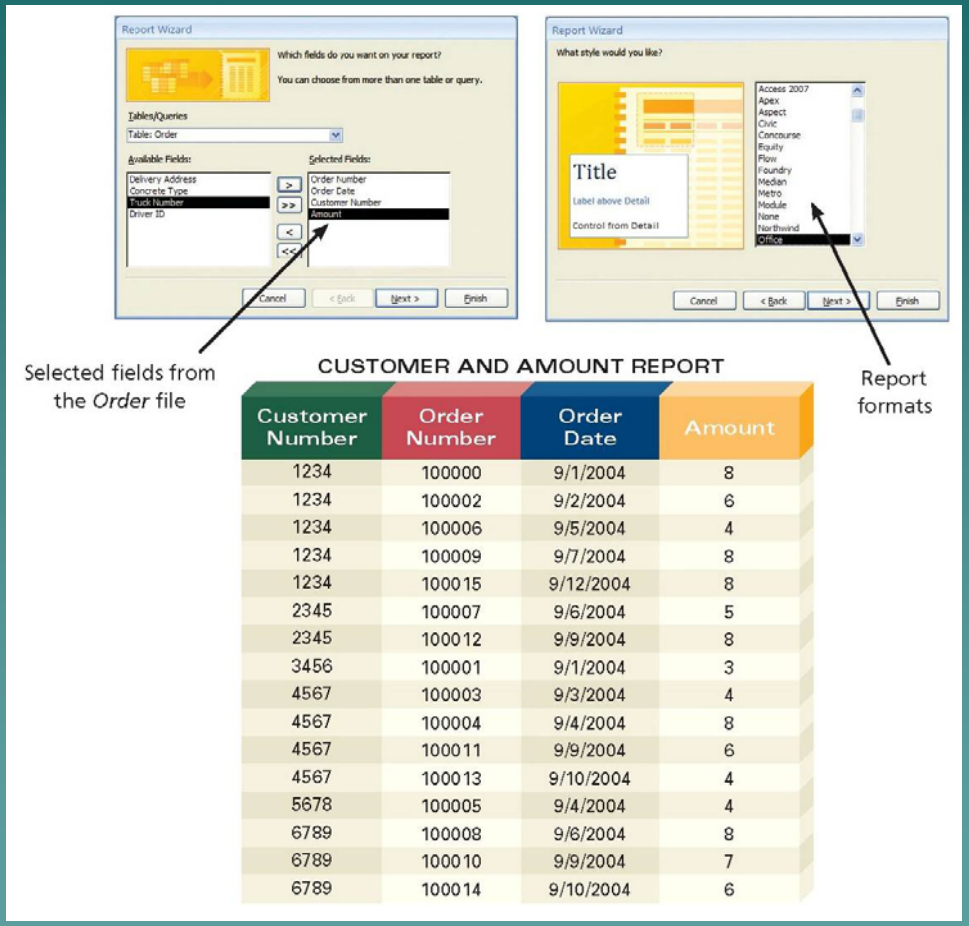


Figure 3.5

A View in Microsoft Access

Report Generator

The data manipulation subsystem's ***report generator*** helps you quickly define a report's format specifying what information you want displayed on a report.



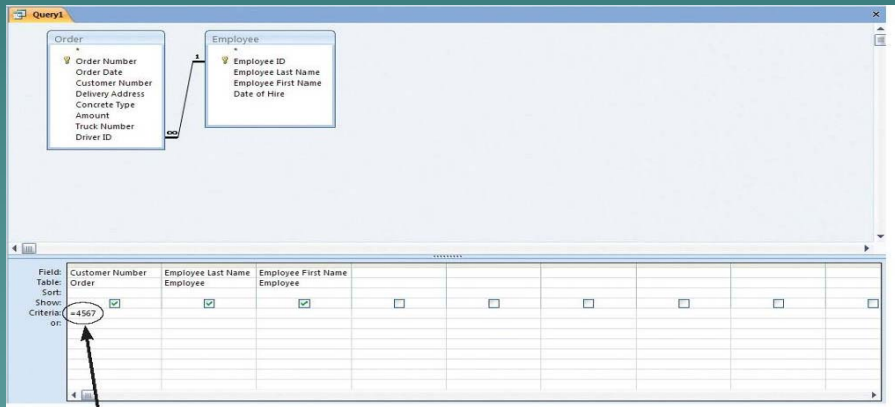
CUSTOMER AND AMOUNT REPORT

Customer Number	Order Number	Order Date	Amount
1234	100000	9/1/2004	8
1234	100002	9/2/2004	6
1234	100006	9/5/2004	4
1234	100009	9/7/2004	8
1234	100015	9/12/2004	8
2345	100007	9/6/2004	5
2345	100012	9/9/2004	8
3456	100001	9/1/2004	3
4567	100003	9/3/2004	4
4567	100004	9/4/2004	8
4567	100011	9/9/2004	6
4567	100013	9/10/2004	4
5678	100005	9/4/2004	4
6789	100008	9/6/2004	8
6789	100010	9/9/2004	7
6789	100014	9/10/2004	6

Query-by-Example Tool

The data manipulation subsystem's ***QBE tool*** helps you identify the files in which to look, fields to query or display and the selection criteria to be used.

e.g. for customer number 4567, display the customer number, employee last name and employee first name



Customer Number	Employee Last Name	Employee First Name
4567	Evaraz	Antonio
4567	Robertson	John
4567	Robertson	John
4567	Robertson	John

Figure 3.7
Using a Query-by-Example to Find Information

Structured Query Language

- The data manipulation subsystem's ***SQL*** is a standardized fourth-generation query language. It is used by most database management systems and accomplishes the same thing as QBE.
- SQL performs a query by creating a statement with the 3 parts SELECT...FROM...WHERE.

- IT professionals often use SQL to specify the criteria for searching through a database.

Fourth component of a DBMS: the application generation subsystem

The ***application generation subsystem*** is used by IT professionals to enhance transaction processing. It includes tools for creating data entry screen layouts as well as programming languages and interfaces for the DBMS.

Fifth component of a DBMS: the Data Administration Subsystem

The ***data administration subsystem*** is used by database administrators to manage the overall database environment. This subsystem provides facilities for:

- Backup and recovery
- Security management
- Query optimization
- Concurrency control
- Change management

Components of the data administration subsystem

- ***Backup and recovery*** – creation of a second copy of the information so that it can be recovered should there be a problem with the database
- ***Security management*** - control of who has access to what information
- ***Query optimization*** - means by queries are enhanced and response time is minimized

Components of the data administration subsystem

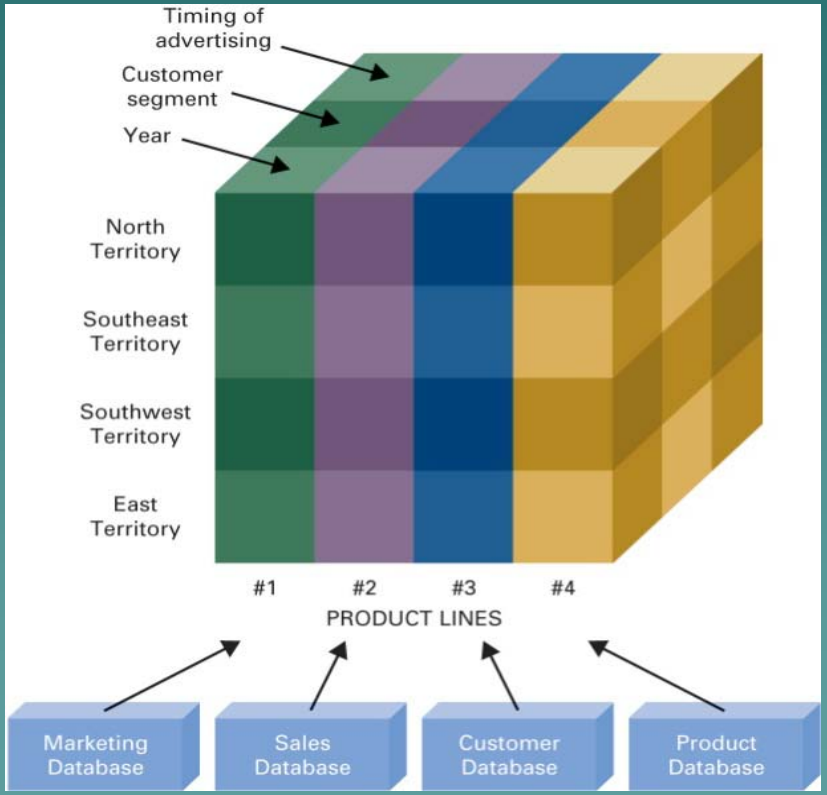
- ***Reorganization*** - physically rearranging the storage of the information so it matches how it is usually accessed
- **Concurrency control** - *addressing how the database is updated should many users want to change the same information at the same time*
- ***Change management*** - assessing the impact of proposed changes to the database's structure

Data Warehouse

- A data warehouse is a large collection of information gathered from many operational databases. It creates business intelligence for business analysis and decision making.
- A data warehouse contains summarized information. It does not contain every detail of every transaction e.g. every sale.
- Data warehouses support on-line analytical processing (OLAP) but not on-line transaction processing (OLTP).

A data warehouse is multidimensional.

A data warehouse has more than 2 dimensions. In this hypercube you could retrieve product information by product line and region (columns & rows), by year (1st layer), by customer segment (2nd layer) and by timing of advertising (3rd layer).



Data-Mining Tools

Data-mining tools are used to query information in a data warehouse.

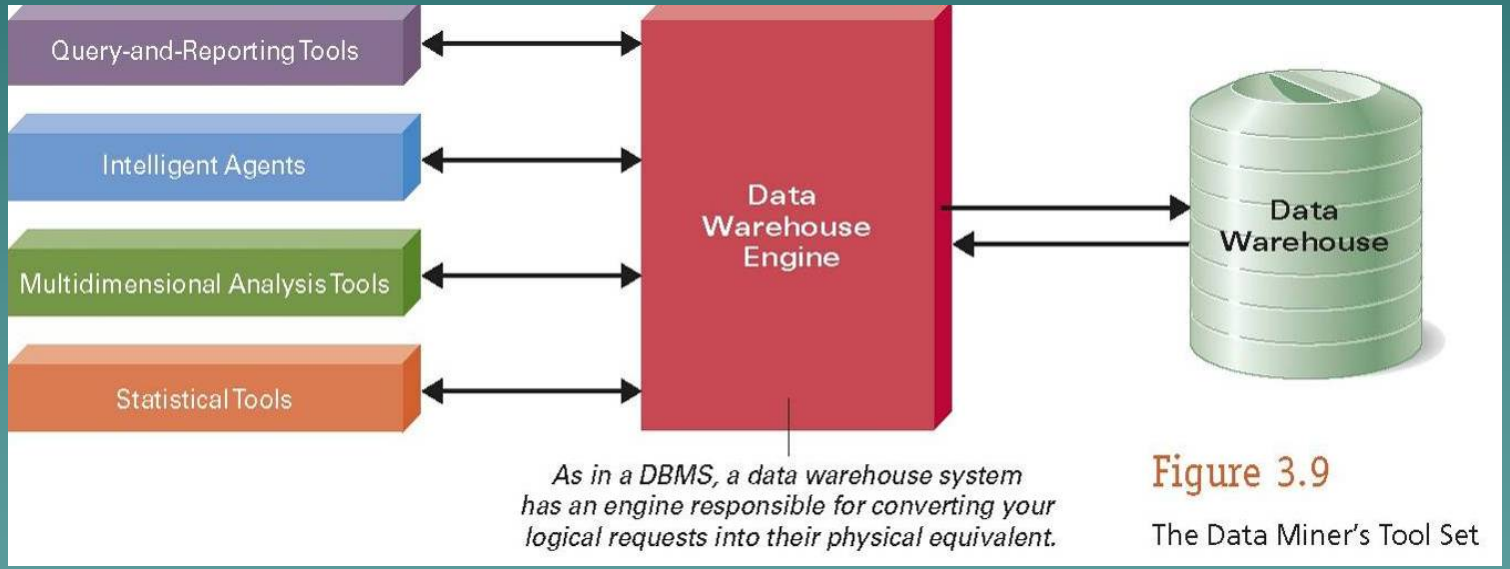


Figure 3.9

The Data Miner's Tool Set

Data-Mining Tools

- ***Query-and-reporting tools*** (e.g. QBE tools, SQL, and report generators) - making simple queries and reports
- ***Intelligent agents*** (e.g. neural networks, fuzzy logic) - using artificial intelligence when analyzing processes and looking for trends
- ***Multidimensional analysis tools*** - viewing multidimensional information from a different perspective (like turning the cube to slice it differently)
- ***Statistical tools*** (e.g. time-series analysis, regression analysis) - applying mathematical models to data warehouse information

Data Marts

*A **data mart*** is a portion of the data warehouse pertaining to a specific focus. For example, marketing information could be put into a data mart for the marketing department.

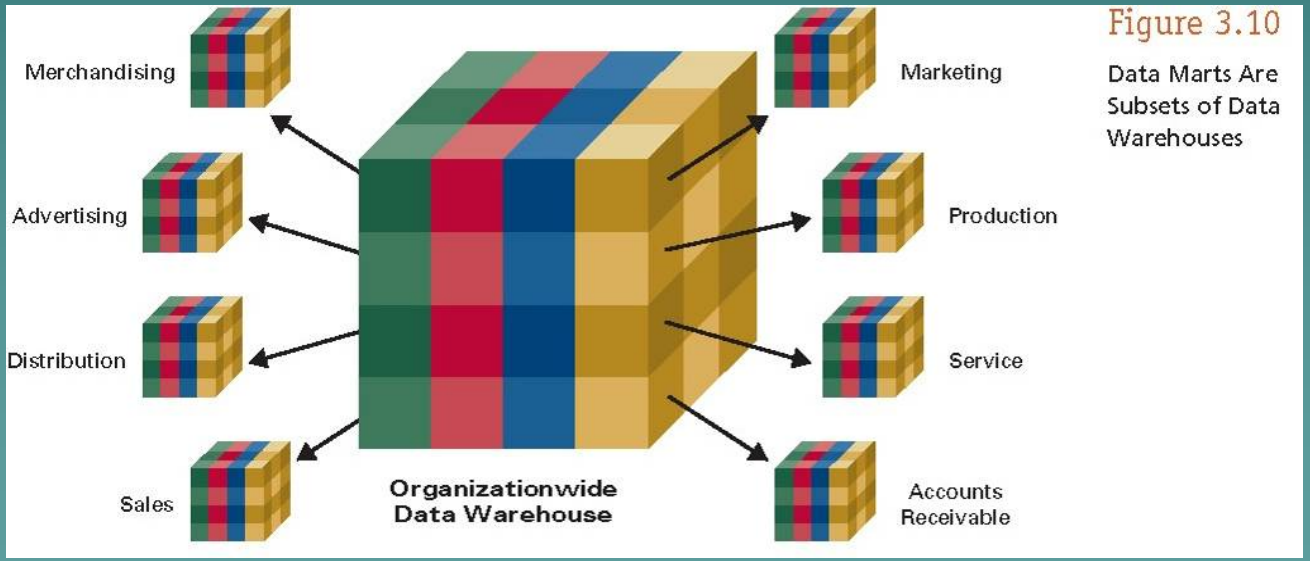


Figure 3.10
Data Marts Are Subsets of Data Warehouses

Back to Business Intelligence

- ***Business intelligence (BI)*** is the collection of information from internal operations as well as information about customers, competitors, business partners, and the competitive environment.
- BI focused solely on the external competitive environment is called ***competitive intelligence***.
- Knowledge workers may use BI to understand the organization's capabilities, market trends, implications of certain actions by competitors, etc.

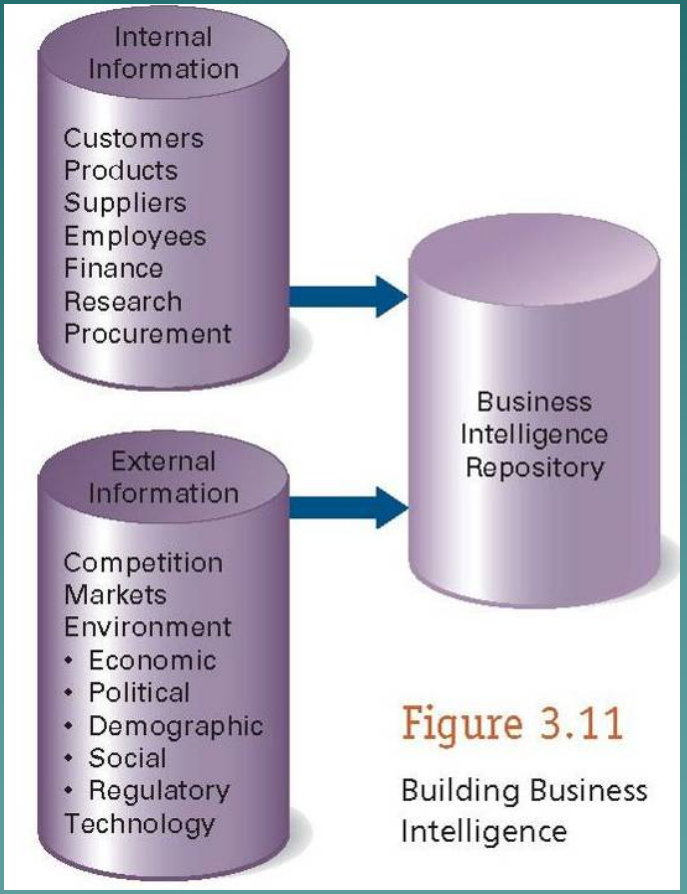


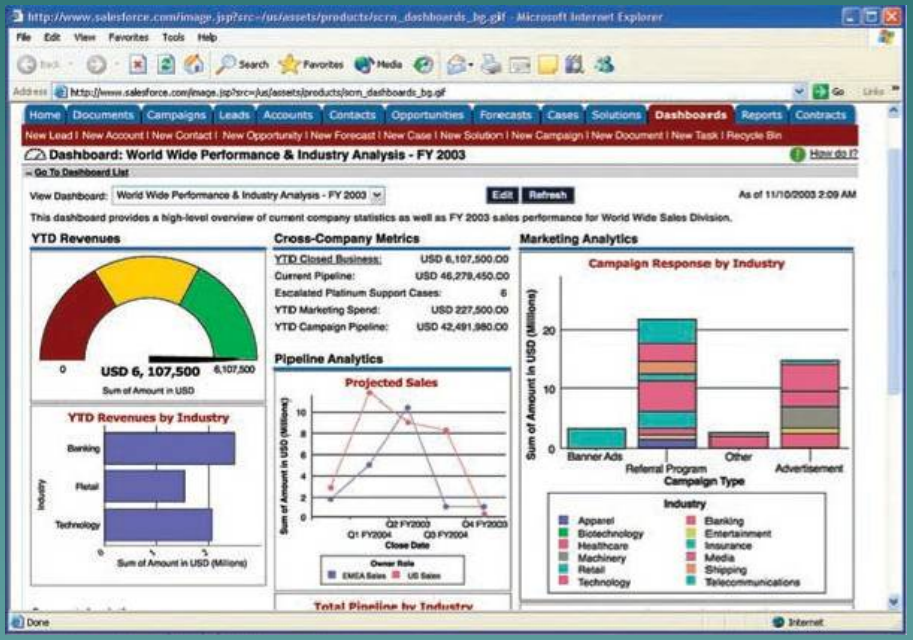
Figure 3.11

Building Business Intelligence

- ***Business Intelligence includes both internal and external information***

Viewing Business Intelligence

A ***digital dashboard*** shows a snapshot of information gathered from many sources. It presents results in a format tailored to the needs of the user.



Special Issues in Data Management

- ***Data administration*** plans for, oversees the development of, and monitors the information resource in an organization. It must coordinate with the organization's strategic direction.
- ***Database administration*** *addresses* the more technical and operational aspects of managing and organization's information kept in databases, data warehouses, and data marts.
- Access to information must be controlled.
- Every effort must be made to ensure that an organization's information is accurate.

Can companies keep your personal information private and secure?

- Databases are large repositories of detailed information.
- Information is a valuable commodity. A large portion of that information is personal.
- Organizations must protect their information from theft and loss.
- There are people who want to steal your personal information from the companies you do business with.