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

INFORMATION

- Change a product price
- How many products individually sold over $10,000 last month?
- Change advertising time table
- How much money was spent on radio advertising last month?
- Increase customer credit limit
- Who is delinquent in paying their bills?

PRODUCT DATABASE

ADVERTISING DATABASE

CUSTOMER DATABASE

ONLINE TRANSACTION PROCESSING (OLTP)

BUSINESS INTELLIGENCE

DATA WAREHOUSE

ONLINE ANALYTICAL PROCESSING (OLAP)

Information Used in Decision Making

If inventory levels are reduced by 10%, what is the new total cost of inventory carried?

What new advertising strategies need to be undertaken to reach our customers who can afford a high-priced product?
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**

<table>
<thead>
<tr>
<th>Order Number</th>
<th>Order Date</th>
<th>Customer Number</th>
<th>Delivery Address</th>
<th>Concrete Type</th>
<th>Amount</th>
<th>Truck Number</th>
<th>Driver ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>100000</td>
<td>9/1/2004</td>
<td>1234</td>
<td>55 Smith Lane</td>
<td>1</td>
<td>8</td>
<td>111</td>
<td>1234966789</td>
</tr>
<tr>
<td>100001</td>
<td>9/1/2004</td>
<td>3456</td>
<td>2122 E. Biscayne</td>
<td>1</td>
<td>3</td>
<td>222</td>
<td>789344444</td>
</tr>
<tr>
<td>100002</td>
<td>9/2/2004</td>
<td>1234</td>
<td>55 Smith Lane</td>
<td>5</td>
<td>6</td>
<td>222</td>
<td>435296657</td>
</tr>
<tr>
<td>100004</td>
<td>9/4/2004</td>
<td>4567</td>
<td>1333 Burr Ridge</td>
<td>2</td>
<td>8</td>
<td>222</td>
<td>789344444</td>
</tr>
<tr>
<td>100005</td>
<td>9/4/2004</td>
<td>5678</td>
<td>1222 Westminster</td>
<td>1</td>
<td>4</td>
<td>222</td>
<td>789344444</td>
</tr>
<tr>
<td>100006</td>
<td>9/5/2004</td>
<td>1234</td>
<td>222 East Hampton</td>
<td>1</td>
<td>4</td>
<td>111</td>
<td>1234966789</td>
</tr>
<tr>
<td>100007</td>
<td>9/6/2004</td>
<td>2345</td>
<td>9 W. Palm Beach</td>
<td>2</td>
<td>5</td>
<td>333</td>
<td>789344444</td>
</tr>
<tr>
<td>100008</td>
<td>9/8/2004</td>
<td>6789</td>
<td>4522 Lane Circle</td>
<td>1</td>
<td>8</td>
<td>222</td>
<td>789344444</td>
</tr>
<tr>
<td>100009</td>
<td>9/7/2004</td>
<td>1234</td>
<td>987 Furlong</td>
<td>3</td>
<td>8</td>
<td>111</td>
<td>1234966789</td>
</tr>
<tr>
<td>100010</td>
<td>9/9/2004</td>
<td>6789</td>
<td>4522 Lance Circle</td>
<td>2</td>
<td>7</td>
<td>222</td>
<td>435296657</td>
</tr>
<tr>
<td>100011</td>
<td>9/9/2004</td>
<td>4567</td>
<td>3500 Tomahawk</td>
<td>5</td>
<td>6</td>
<td>222</td>
<td>789344444</td>
</tr>
</tbody>
</table>

**CUSTOMER FILE**

<table>
<thead>
<tr>
<th>Customer Number</th>
<th>Customer Name</th>
<th>Customer Phone</th>
<th>Customer Primary Contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>1234</td>
<td>Smallding Homes</td>
<td>3333333333</td>
<td>Bill Johnson</td>
</tr>
<tr>
<td>2036</td>
<td>Home Builders Superior</td>
<td>3334444444</td>
<td>Marcus Connolly</td>
</tr>
<tr>
<td>3456</td>
<td>Mark Avey</td>
<td>2395655555</td>
<td>Mark Avey</td>
</tr>
<tr>
<td>4567</td>
<td>Triple A Homes</td>
<td>3390666966</td>
<td>Janelle Smith</td>
</tr>
<tr>
<td>5678</td>
<td>Sheryl Williamson</td>
<td>3377777777</td>
<td>Sheryl Williamson</td>
</tr>
<tr>
<td>6789</td>
<td>Home Makers</td>
<td>3338888888</td>
<td>John Yu</td>
</tr>
</tbody>
</table>

**EMPLOYEE FILE**

<table>
<thead>
<tr>
<th>Employee ID</th>
<th>Employee Last Name</th>
<th>Employee First Name</th>
<th>Date of Hire</th>
</tr>
</thead>
<tbody>
<tr>
<td>123456789</td>
<td>Johnson</td>
<td>Emilio</td>
<td>2/1/1985</td>
</tr>
<tr>
<td>435296657</td>
<td>Evaraz</td>
<td>Antonio</td>
<td>3/3/1992</td>
</tr>
<tr>
<td>785934444</td>
<td>Robertson</td>
<td>John</td>
<td>6/1/1999</td>
</tr>
<tr>
<td>984568756</td>
<td>Smithson</td>
<td>Allison</td>
<td>4/1/1997</td>
</tr>
</tbody>
</table>

**CONCRETE TYPE FILE**

- **Concrete Type**
  - 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**

<table>
<thead>
<tr>
<th>Truck Number</th>
<th>Truck Type</th>
<th>Date of Purchase</th>
</tr>
</thead>
<tbody>
<tr>
<td>111</td>
<td>Ford</td>
<td>6/17/1999</td>
</tr>
<tr>
<td>222</td>
<td>Ford</td>
<td>12/24/2001</td>
</tr>
<tr>
<td>333</td>
<td>Chevy</td>
<td>1/1/2002</td>
</tr>
</tbody>
</table>
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.
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.
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

- Sort using these two buttons
- Find information with the binoculars
- Click here to enter a new record
Report Generator

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

*Example:*

For customer number 4567, display the customer number, employee last name and employee first name.
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.
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.
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](image)

*As in a DBMS, a data warehouse system has an engine responsible for converting your logical requests into their physical equivalent.*

The Data Miner’s Tool Set

LO4 3 -29
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.
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.
Business Intelligence includes both internal and external information.
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.