Chapter 13 Business Intelligence and Data Warehouses

Objectives
• In this chapter, you will learn:
  – How business intelligence is a comprehensive framework to support business decision making
  – How operational data and decision support data differ
  – What a data warehouse is, how to prepare data for one, and how to implement one
  – What star schemas are and how they are constructed
  – What data mining is and what role it plays in decision support
  – About online analytical processing (OLAP)
  – How SQL extensions are used to support OLAP type SQLs

13.1 The Need for Data Analysis
• Managers must be able to track daily transactions to evaluate how the business is performing
• Strategies should be developed to meet organizational goals using operational databases
• Data analysis provides information about short-term tactical evaluations and strategies

13.2 Business Intelligence
• Definition: Comprehensive, cohesive, integrated tools and processes to
  – Capture, collect, integrate, store, and analyze data
  – Generate information to support business decision making
• Framework that allows a business to transform:
  – Data into information
  – Information into knowledge
  – Knowledge into wisdom

| TABLE 13.1 Solving Business Problems and Adding Value with BI Tools and Solutions |
| COMPANY | PROBLEM | BENEFIT |
| MOIN (Manufacturer of bathroom and kitchen fixtures and supplies) | Information generation very limited and time-consuming. How to extract data using a 3GL

  known to only five people.

  Response time unacceptable for managers’ decision-making purposes. | Provided quick answers to ad hoc questions for decision-making. Provided access to data for decision-making purposes. Received in-depth view of product performance and customer margins. |

  NASDAQ (Largest U.S. electronic stock market trading organization) | Inability to provide real-time ad hoc query and standard reporting to executives, business analysts, and other users. Excessive storage costs for many terabytes of data. | Reduced storage cost by moving to multilayer storage solution. Implemented new data warehouse center with support for ad hoc query and reporting and near-real-time data access for end users. |

  Sega of America, Inc. (Interactive entertainment systems and video games) | Needed a way to rapidly analyze great amount of data.

  Needed to track advertising, coupons, and rebates associated with effects of pricing changes.

  Used to do it with Excel spreadsheets, leading to human errors. | Eliminated data-entry errors. Identified successful marketing strategies to dominate interactive entertainment niches.

  Used product analysis to identify better markets/product offerings. |
Business Intelligence

- Implementing BI captures not only business data (internal and external), but also metadata
- BI involves the following general steps:
  1. Collecting and storing operational data
  2. Aggregating the operational data into decision support data
  3. Analyzing decision support data to generate information
  4. Presenting such information to the end user to support business decisions
  5. Making business decisions, which in turn generate more data that is collected, stored, etc. (restarting the process)
  6. Monitoring result to evaluate outcomes of the business decisions (providing data to be collected, stored, etc.)

13.3 Business Intelligence Architecture

- Composed of data, people, processes, technology, and management of components
- Focuses on strategic and tactical use of information
- Multiple tools from different vendors can be integrated into a single BI framework
  - Check Figure 13.1 in p. 517 for BI framework
  - Governance is a method or process of government
- Key performance indicators (KPI)
  - Measurements that assess company’s effectiveness or success in reaching goals, check p. 517 for examples of KPI
- Master Data Management
  - a collection of concepts, techniques, and processes for the proper identification, definition, and management of data elements within an organization

Check Table 13.2 in p. 518 for description of these components
Check Table 13.3 in p. 519-520 for samples of BI tools
13.4 Decision Support Data

- **Operational data**
  - Mostly stored in relational database
  - Optimized to support transactions representing daily operations

- **Decision support data** differs from operational data in three main areas:
  - Time span
  - Granularity
    - drill-down and roll-up to different levels of aggregation
  - Dimensionality

---

**TABLE 13.4** Contrast Operating and DSS Data Characteristics

<table>
<thead>
<tr>
<th>CHARACTERISTIC</th>
<th>OPERATIONAL DATA</th>
<th>DSS DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data currency</td>
<td>Current operations&lt;br&gt;Real-time data</td>
<td>Historic data&lt;br&gt;Snapshot of company data&lt;br&gt;Time component (week/month/year)</td>
</tr>
<tr>
<td>Granularity</td>
<td>Atomic-detailed data</td>
<td>Summarized data</td>
</tr>
<tr>
<td>Summarization level</td>
<td>Low; some aggregate yields</td>
<td>High; many aggregation levels</td>
</tr>
<tr>
<td>Data model</td>
<td>Highly normalized&lt;br&gt;Mostly relational DBMS</td>
<td>Non-normalized&lt;br&gt;Complex structures&lt;br&gt;Some relational, but mostly multidimensional DBMS</td>
</tr>
<tr>
<td>Transaction type</td>
<td>Mostly updates</td>
<td>Mostly query</td>
</tr>
<tr>
<td>Transaction volumes</td>
<td>High update volumes</td>
<td>Periodic loads and summary calculations</td>
</tr>
<tr>
<td>Transaction speed</td>
<td>Updates are critical</td>
<td>Retrievals are critical</td>
</tr>
<tr>
<td>Query activity</td>
<td>Low to medium</td>
<td>High</td>
</tr>
<tr>
<td>Query scope</td>
<td>Narrow range</td>
<td>Broad range</td>
</tr>
<tr>
<td>Query complexity</td>
<td>Simple to medium</td>
<td>Very complex</td>
</tr>
<tr>
<td>Data volumes</td>
<td>Hundreds of megabytes, up to gigabytes</td>
<td>Hundreds of gigabytes, up to terabytes</td>
</tr>
</tbody>
</table>

---

**Decision Support Database Requirements**

- Specialized DBMS tailored to provide fast answers to complex queries
- Four main requirements:
  - Database schema
  - Data extraction and loading
  - End-user analytical interface
  - Database size
- **Database schema**
  - Must support complex data representations
    - Bitmap indexes, data partitioning, non-normalized
    - Must contain aggregated and summarized data
  - Queries must be able to extract multidimensional time slices
**Decision Support Database Requirements**

- End-user analytical interface
  - One of most critical DSS DBMS components
  - Permits user to navigate through data to simplify and accelerate decision-making process

- Database size
  - In 2005, Wal-Mart had 260 terabytes of data in its data warehouses
  - DBMS must support **very large databases** (VLDBs)
    - Might be required to use advanced hardware, such as disk arrays, symmetric multiprocessor (SMP), or massively parallel processor (MPP)

**13.5 The Data Warehouse**

- Integrated, subject-oriented, time-variant, and nonvolatile collection of data
  - Provides support for decision making

- Usually a read-only database optimized for data analysis and query processing

- Requires time, money, and considerable managerial effort to create
### The Data Warehouse (continued)

- **Data mart**
  - Small, single-subject data warehouse subset
  - More manageable data set than data warehouse
  - Provides decision support to small group of people
  - Typically lower cost and lower implementation time than data warehouse

### Decision Support Architectural Styles

- Provide advanced decision support features
- Some capable of providing access to multidimensional data analysis
- Complete data warehouse architecture supports:
  - Decision support data store
  - Data extraction and integration filter
  - Specialized presentation interface
### 13.6 Online Analytical Processing

- Advanced data analysis environment that supports:
  - Decision making
  - Business modeling
  - Operations research

- OLAP systems Share four main characteristics:
  - Use multidimensional data analysis techniques
  - Provide advanced database support
  - Provide easy-to-use end-user interfaces
  - Support client/server architecture

### Multidimensional Data Analysis Techniques

- Data are processed and viewed as part of a multidimensional structure
  - Particularly attractive to business decision makers

- Augmented by the following functions:
  - Advanced data presentation functions
  - Advanced data aggregation, consolidation, and classification functions
  - Advanced computational functions
  - Advanced data modeling functions
Advanced Database Support

- Advanced data access features include:
  - Access to many different kinds of DBMSs, flat files, and internal and external data sources
  - Access to aggregated data warehouse data
  - Advanced data navigation
  - Rapid and consistent query response times
  - Maps end-user requests to appropriate data source and to proper data access language (SQL)
  - Support for very large databases

Easy-to-Use End-User Interface

- Advanced OLAP features more useful when access is simple

- Many interface features are “borrowed” from previous generations of data analysis tools
  - Already familiar to end users
  - Makes OLAP easily accepted and readily used
Client/Server Architecture

- Provides framework for design, development, implementation of new systems
  - Enables OLAP system to be divided into several components that define its architecture
  - OLAP is designed to meet ease-of-use as well as system flexibility requirements

OLAP Architecture

- Operational characteristics’ three main modules:
  - Graphical user interface (GUI)
  - Analytical processing logic
  - Data-processing logic
- Designed to use both operational and data warehouse data
- In most implementations, data warehouse and OLAP are interrelated and complementary
- OLAP systems merge data warehouse and data mart approaches

![OLAP client/server architecture](image1)

![OLAP server arrangement](image2)
Interrelated and Complementary data warehouse and OLAP system

Relational OLAP

- Uses relational databases and relational query tools
  - Stores and analyzes multidimensional data

- Adds following extensions to traditional RDBMS:
  - Multidimensional data schema support within RDBMS
  - Data access language and query performance optimized for multidimensional data
  - Support for very large databases

<table>
<thead>
<tr>
<th>NORTH</th>
<th>SOUTH</th>
<th>EAST</th>
<th>WEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

The GUI front end runs on the client computer and passes data-analysis requests to the ROLAP server. The GUI receives data reply from the ROLAP server and formats them according to the end user’s presentation needs.
Multidimensional OLAP

- Extends OLAP functionality to multidimensional database management systems (MDBMSs)
  - MDBMS end users visualize stored data as a 3D data cube
  - Data cubes can grow to $n$ dimensions, becoming hypercubes
  - To speed access, data cubes are held in memory in a cube cache
  - Must handle sparsity effectively to reduce processing overhead and resource requirement

Relational vs. Multidimensional OLAP

- Selection of one or the other depends on evaluator’s vantage point
- Proper evaluation must include supported hardware, compatibility with DBMS, etc.
- ROLAP and MOLAP vendors working toward integration within unified framework
- Relational databases use star schema design to handle multidimensional data