

Information Systems

Prepared by Dr. Amjad Qtaish

IT372
Information Systems

Lecture 02



Module 2: Information Technology Infrastructure



Information Systems

Dr. Amjad Qtaish

*Faculty of Informatics
Midocean University*

Chapter 5: Databases and Information Management



Information Systems (IT372)



An Information Technology Infrastructure :

- Data Management
- Data Modeling
- Relational Database Model
- Database Management Systems (DBMS)
- Database Administration
- Database Applications
- Database and Web
- Managing Data Resources
- Case Study
- Summary





An Information Technology Infrastructure :

- Define general data management concepts and terms
- Identify the advantages of database approach and describe relational database model
- Identify the role and functions of DBMS
- Identify current database applications
- Identify the role of Business Intelligence





- Hierarchy of data



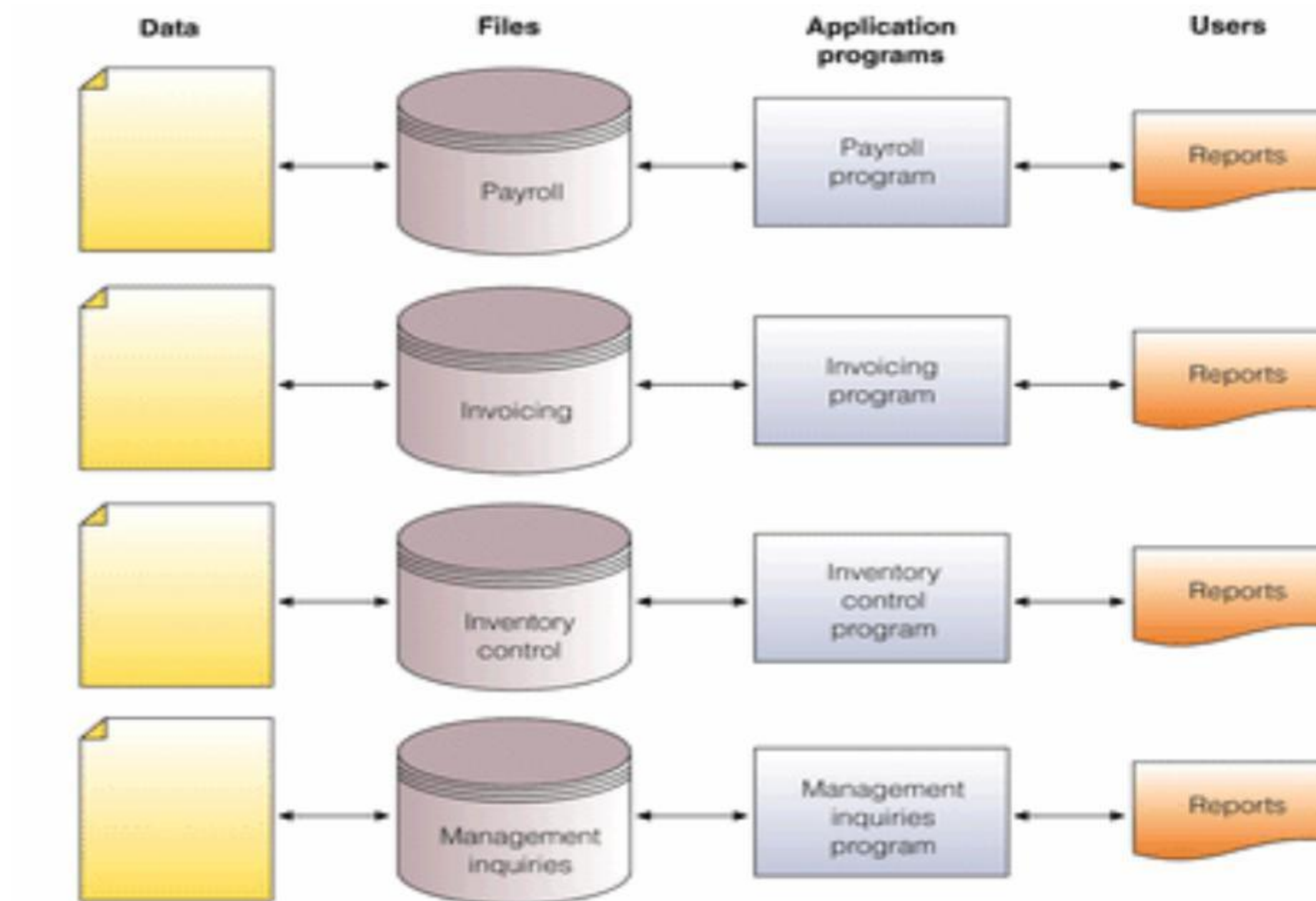
Hierarchy of data	Example
Database	Personnel file Department file Payroll file (Project database)
Files	098 - 40 - 1370 Fiske, Steven 01-05-1985 549 - 77 - 1001 Buckley, Bill 02-17-1979 005 - 10 - 6321 Johns, Francine 10-07-1997 (Personnel file)
Records	098 - 40 - 1370 Fiske, Steven 01-05-1985 (Record containing SSN, last and first name, hire date)
Fields	Fiske (Last name field)
Characters (Bytes)	1000100 (Letter F in ASCII)



- **Building blocks of hierarchy**
- **bit** (the smallest unit of data) has only two values - 1 or 0
- **bytes** - 8 bits make up one byte, which represents one character like the letter A
- **field** (or in a database attribute), represents a combination of bytes that make up one aspect of a business object (i.e. last name, invoice number, age)
- **record** - a collection of related data fields (i.e. name/address/phone information for one student)
- **file** (or in a database an entity) - a collection of related records (all students in MIS213)
- **database** - a group of similar items (all students and faculty in Cameron School of Business)



- **TRADITIONAL (File Based)** approach to data management





- **PROBLEMS**

- **Data Redundancy**

- Duplication of data, same data is stored in multiple locations
 - Data inconsistency, same attributed have different names or values
 - Updating problems

- **Program Data Dependence**

- Changes in program require changes in data

- **Lack of Flexibility**

- difficult and expensive process to retrieve ad-hoc reports

- **Lack of Sharing**

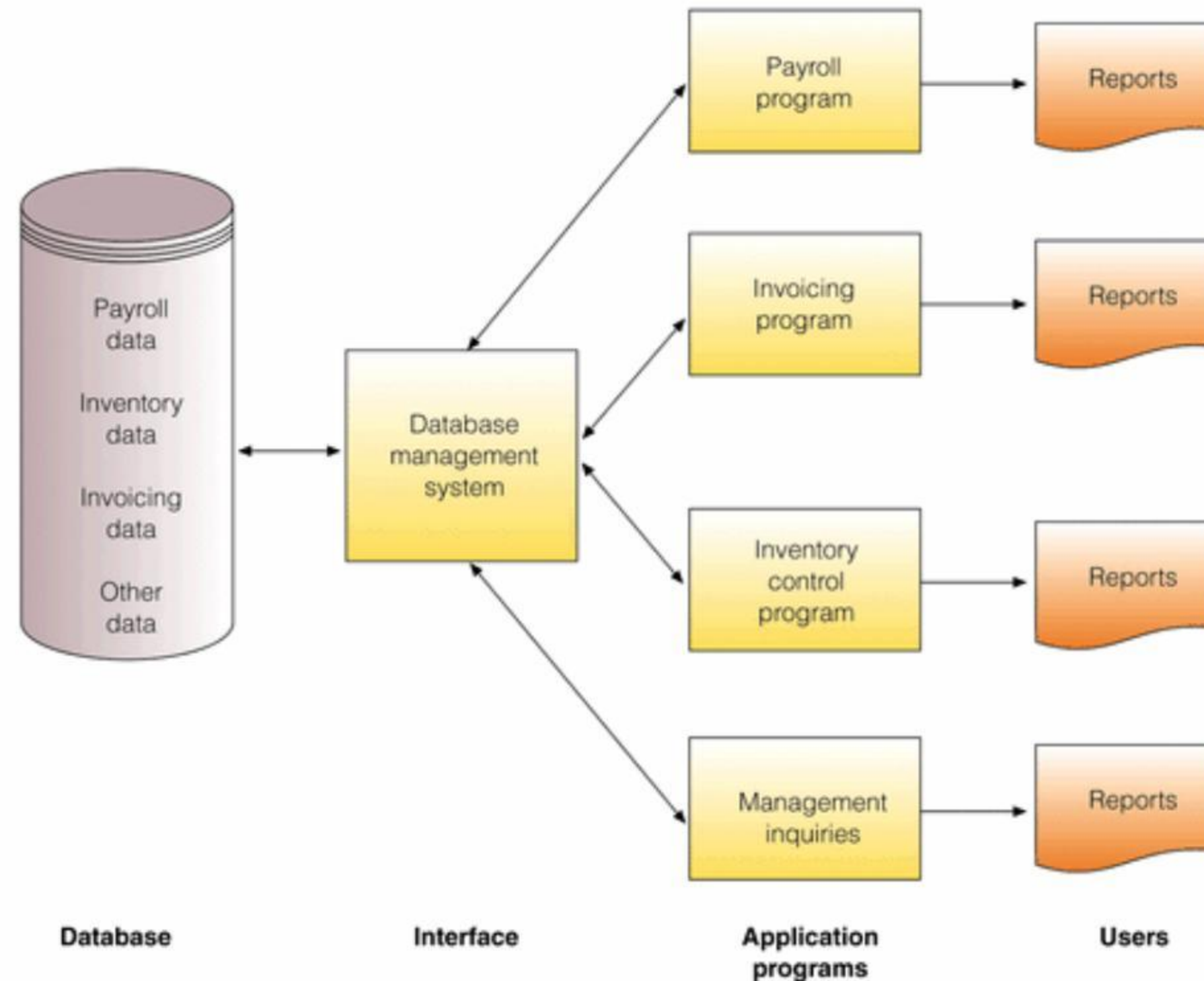
- Because data is located in different files and different departments, difficult to be shared and accessed in timely manner



- **DATABASE Approach**
- **Database**
 - Organized collection of data, or a collection of related files containing records
- **Entity**
 - Generalized class of people, places or things (objects) for which data is collected, stored and maintained
 - E.g. SUPPLIER, PRODUCT
- **Attribute**
 - Specific characteristics of each entity
 - E.g. SUPPLIER: Name, address
 - PRODUCT: Product ID, Product Price
- **Database Management System (DBMS)**



- **DATABASE Approach**





• Advantages of Database Approach



1. Reduced data redundancy	Data about a person / invoice / product is stored only one time
2. Improved data integrity	Since data is stored only once for each entity, we don't need to worry about updating multiple records for the same entity (i.e. storing home address several times for the same person)
3. Easier updating of data	Again the advantage of one storage location
4. Data and program independence	The data files are separate from the applications (HR, payroll, invoicing) and thus can be used by many applications
5. Improved strategic use of data	Accurate, complete, up-to-date data is used by decision makers
6. Improved security	Backups and access can be better controlled by passwords, ensuring privacy



- **Some more advantages of Database Approach**
 - Standardization of data access
 - Shared data and information resources
- **Disadvantages**
 - More complex
 - DBMS could be difficult to set up and operate
 - More expensive
 - More expensive to purchase, additional personnel and additional hardware required
 - Difficult to recover from failure
 - Failure in DBMS shuts down entire database



- **When building a database, following must be considered**
 - **Content:** What data should be collected and at what cost?
 - **Access:** What data should be provided to which users and when?
 - **Logical Structure:** How should data be arranged?
 - **Physical Organization:** Where should data be physically located?
- **Logical Design**
 - Abstract model of how data should be structured and arranged
 - **Data Model** : diagram of entities and their relationships
- **Physical Design**
 - Fine tunes logical design for performance and cost (improved response time, reduce storage space)



- **Entity Relationship Diagram**

- Use basic graphical symbols to show the organization and relation between data
- One to one
- One to many
- Many to many





- **The Relational Database Model**
- **Relational Database**
 - Organize data into two-dimensional tables (**relations**) with columns and rows
 - One table for each entity
 - **Fields** (columns) store data representing an attribute
 - Rows store data for separate **records**
 - **Key field**: identifies a record
 - **Primary Key**: A field that uniquely identifies a set of records, cannot be duplicated and distinguishes records
 - **Domain** is the allowable values for these attributes. E.g. attribute for pay does not include negative numbers



- **Manipulating Data**

- For inquiries and analyzing data
- **Selecting**: eliminating rows according to certain criteria
- **Projecting**: eliminating columns in a table
- **Joining**; combining two or more tables
- As long as tables share at least one common attribute, tables in a relational database can be **linked** to provide useful information and reports



Project	Description	Dept. No.
155	Payroll	001
498	Widgets	002
226	Sales Manual	003

Project

Dept. No.	Dept. Name	SSN
001	Accounting	10-10
002	Manufacturing	23-25
003	Marketing	10-45

Department

SSN	Name	Hire Date	Dept. No.
10-10	Rasheed Khan	10-07-1997	001
23-25	Haider Ali	02-17- 1998	002
10-45	Safdar Ahmed	01-05-1985	003

Manager



- **Data Cleanup**

- Valuable information: accurate, complete, reliable, economical, flexible, relevant, simple, timely, verifiable, accessible, secure
- Data cleanup is to develop data with these characteristics





- **Group of programs used as an interface between a database and application programs/users, used to create, implement, use and update a database.**
 - Makes physical database available for different logical views required by users
- **Single User DBMS**
 - Databases for personal computer are meant for single users
 - Access, FileMaker Pro, Microsoft InfoPath
- **Multuser DBMS**
 - Used by large mainframe computers
 - Powerful, expensive, allow hundreds of people to access
 - Oracle, Sybase, DB2 by IBM, Teradata database



- Provides capabilities and tools for organizing, managing and accessing the data in the database
- Data Definition language (DDL)
 - Collection of instructions and commands used to define and describe data and relations in a specific database
 - Basically used to define **schema** (description)
 - Describes logical access paths and logical records in the database
 - SQL: CREATE, DROP, ALTER

```
CREATE TABLE employees (  
    id            INTEGER    PRIMARY KEY,  
    first_name    VARCHAR(50) NULL,  
    last_name     VARCHAR(75) NOT NULL,  
    dateofbirth   DATE       NULL  
);
```



- **Data Dictionary**

- Detailed description of all data used in the database
- Name of data item, range of values used, type of data, amount of storage needed, notation of person who updated it, users who can access it, list of reports that use data item



NORTHWEATERN MANUFACTURING

PREPARED BY:	BORDWELL
DATE:	04 AUGUST 2007
APPROVED BY:	EDWARDS
VERSION:	3.1
PAGE:	1 OF 1
DATA ELEMENT NAME:	PARTNO
DESCRIPTION:	INVENTORY PART NUMBER
OTHER NAMES:	PTNO
VALUE RANGE:	100 TO 5000
DATA TYPE:	NUMERIC
POSITION:	4 POSITIONS OR COLUMNS



- **Data Manipulation Language**

- Specific language used to access, modify, and make **queries**(request for specific data)
- Storing, Retrieving, Manipulating data and Generating reports
- Query By Example (QBE)
 - Visual approach to developing database queries
 - GUI to retrieve data
 - MS Access
- Structured Query Language (SQL)
 - Integral part of relational databases
 - Consists of special keywords and rules
 - Also includes built-in functions AVG, MAX, MIN



- **Data Manipulation Language**



SQL Command	Description
<pre>SELECT ClientName, Debt FROM Client WHERE Debt>1000</pre>	Query displays all clients and amount they owe to the company from database table Client, or client who owe more than \$1000
<pre>SELECT ClientName, ClientNum, OrderNum FROM Client, Order WHERE Client.ClientNum=Order.ClientNum</pre>	Join command that combines data from 2 tables: Client and Order. New table will be created with client name, client number and order number.
<pre>GRANT INSERT ON Client to Guthrie</pre>	Security command, that allows Guthrie to enter values or rows in Client table



- **Require a skilled DBA**

- Expected to have clear understanding of the fundamental business of organizations
- Proficient in the use of selected DBMS
- Stay ahead of emerging technologies and new design approaches
- Role: plan, design, create, operate, secure, monitor, and maintain
- Works with users and programmers
- Database administrator: responsible for defining and implementing consistent principles for a variety of data issues





- **Manipulate content of a database to produce useful information**
 - Searching, filtering, synthesizing, and assimilating the data contained in database
- Businesses use databases, not only for keeping track of employee and customer records, but also to make better decisions and run operations effectively
 - Data warehouse
 - Data mining
 - Business intelligence
 - Web mining and text mining



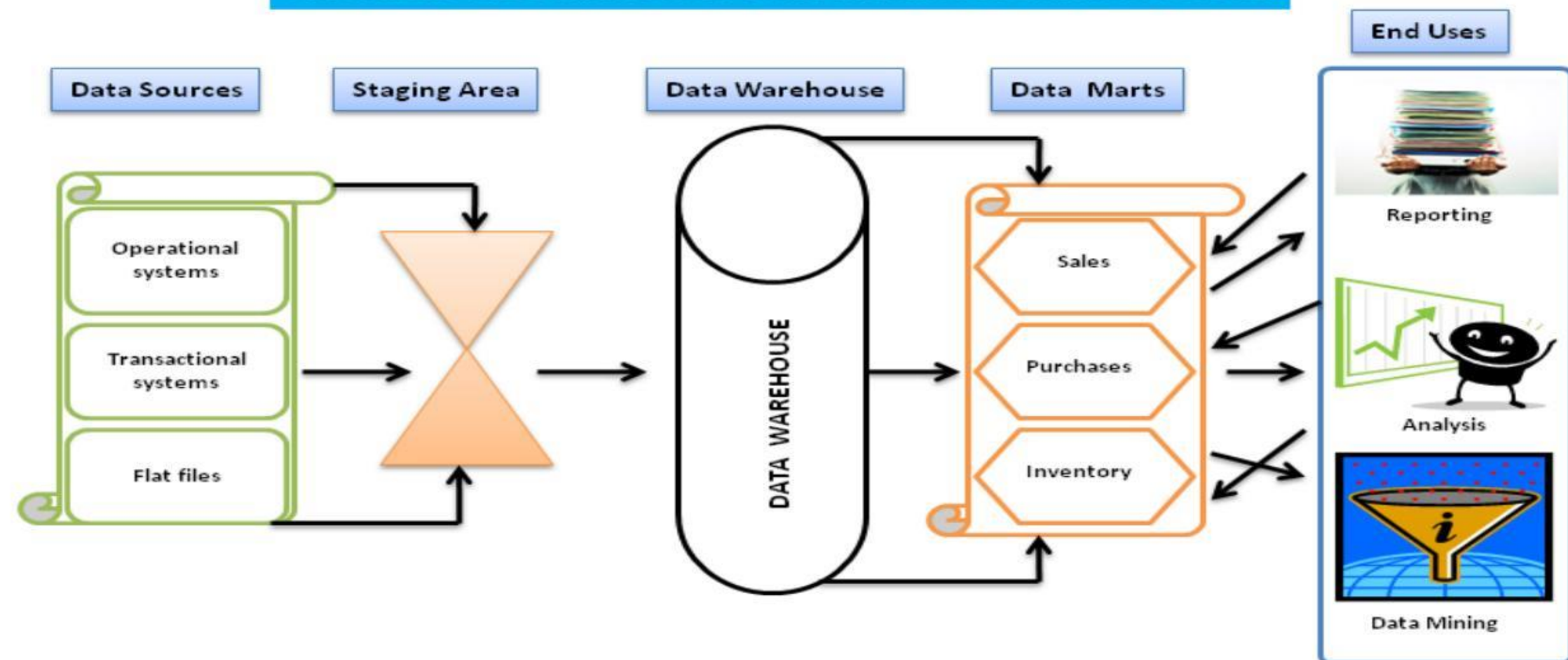
- **Data Warehouse**



- Database that stores current and historical data of potential interest to decision makers throughout the company
- The data is gathered from various operational transaction system, including website transactions
- Consolidates the information from different locations and makes them available for analysis and decisions
- Provides range of standardized query tools, analytical tools and graphical reporting facilities
- Advantage: ability to relate data in innovative ways



A STYLIZED ILLUSTRATION OF DATA WAREHOUSE ARCHITECTURE



Copyright Report Content Writer, 2009. All rights reserved



- **Data Mart**

- Subset of data warehouse
- Contains summarized or highly focused portion of data about a specific area
- E.g. marketing/sales data to deal with customer information
- Useful for smaller groups who want to access detailed data
- Constructed more rapidly, requires less powerful hardware, lower cost





- **Business Intelligence**

- Involves gathering enough of the right information in a timely manner and usable form and analyzing it so that it can have a positive effect on business strategy, tactics or operations.
- Competitive intelligence; information about competitors and the ways that knowledge effects strategy, tactics and operations
 - Beneficial for responding to changing marketplace
- Tools
 - Software for database querying and reporting
 - Multidimensional data analysis (OLAP)
 - data mining



- **Data Mining**

- Provides insight into corporate data that cannot be obtained with OLAP
- Hidden patterns and relationships are found in large databases by inferring rules
- Type of information retrieved is: association, sequences, classifications, clusters, forecasts
- Extensive use in marketing to improve customer retention, cross-selling opportunities, campaign management, one-to-one marketing
- **Predictive Analysis:** combines historic data with assumptions about future conditions to predict outcome of events such as future product sales or such probabilities
 - Find new market segments that could be profitable
- Oracle, Sybase etc. incorporate data mining functionality



• Online Analytical Processing (OLAP)

- To answer question such as: how many washers sold in each of your sales regions and compare actual results with projected sales?
- Supports multidimensional data analysis, so users can use same data in different ways/dimensions (product, pricing, region, time period)
- Enables users to obtain online answers to ad-hoc questions

OLAP	Data Mining
Used for data analysis and decision making	Used for data analysis and decision making
Top-down, query driven analysis	Bottom up, discovery driven analysis
Users must be knowledgeable of data and its business context	Users trust in data mining tools to uncover valid hypotheses



- **Text Mining and Web Mining**

- Unstructured data is also part of firms useful information
 - E-mails, memos, survey responses, legal cases, service reports are also valuable
- Text Mining tools help businesses analyze this data
 - Extract key elements from unstructured data set
 - Discover patterns and relationships
 - Summarize the information
- Web Mining: Discovery and analysis of useful patterns and info from WWW
 - Understanding customer behaviour
 - Evaluating effectiveness of customer website
 - Quantify success of marketing campaign
 - Content Mining, Structure Mining, Usage Mining



- **Firms use the web to make information from their internal databases available to customers and partners**
- **Middleware and other software make this possible**
 - Web server
 - Application server or CGI
 - Database Server
- **Advantages**
 - Web browser software easier to use than query tools
 - Web interfaces require few or no change to internal databases
 - Less costly



- **Establishing Information Policy**
 - Organizations rules for sharing, disseminating, acquiring, standardizing, classifying and inventorying information
 - Which units share info, where info can be distributed, who can maintain and update it
- **Data administration** in large organizations is responsible for defining policies and procedures for managing organizational resources
- Database administration: design and management group performs the following functions:
 - Establishing physical database
 - Logical relations among elements
 - Access rules and security procedures



- **Ensuring Data Quality**

- Poor data quality is major obstacle to successful customer relationship management
- Data Quality problems
 - Redundant and inconsistent data produced by multiple systems
 - Data input errors
- **Data Quality Audit:** structured survey of accuracy and completeness of data
- **Data cleansing:** detects and corrects incorrect, incomplete, improperly formatted and redundant data



- **What can businesses learn from Text Mining?**
- **Text Mining**
 - Discovery of patterns and relationships from large sets of unstructured data
 - Mobile digital platform has amplified the explosion in digital info
 - Consumer collaboration and sharing offers insights into customer behaviour and attitudes
- **Problem with JetBlue**
 - Receiving large volume of e-mails, no simple way to read everything
 - Used text analysis tools to identify facts, opinions, requests from text of survey responses, e-mails, blog entries, news article etc.
 - Used it with another tool: classifying customers into groups
- **Clarabridge text analytics solution**
 - Delivered as software service



- **Challenges of unstructured data**
 - Lots of digital information generated that has no distinct form
 - Difficult to analyze if there are loads of e-mails (with customer sentiments, preferences, requests etc.)
 - Analyzing customer surveys takes weeks
 - Use slow manual approaches
- **Improving decision making**
 - Spot and address problems quickly
 - Identify facts, opinions, trends etc. to act quickly on customer demands
 - Categorizing comments to reveal less obvious insights
 - Also used to make building improvements



- **Kind of Businesses**

- Airlines
- Hotel chains
- Restaurants
- Also used by location managers





- Data Management and Data Modeling are key aspects of organizing data and information
- Relational model reduce my problems of data inconsistency, easier to control, more flexible
- DBMS produce wide variety of documents, reports, useful for orgs
- Data quality is important to be maintained
- Business Intelligence tools have positive effects on business strategy

Thank You