

# Business Intelligence and Systems

CIS 3730  
Designing and Managing Data

J.G. Zheng  
Fall 2010



# Introduction

## ◆ Types of Information Processing

### ■ Transactional Processing

- ◆ Focus on data insertion, modification, deletion, and transmission

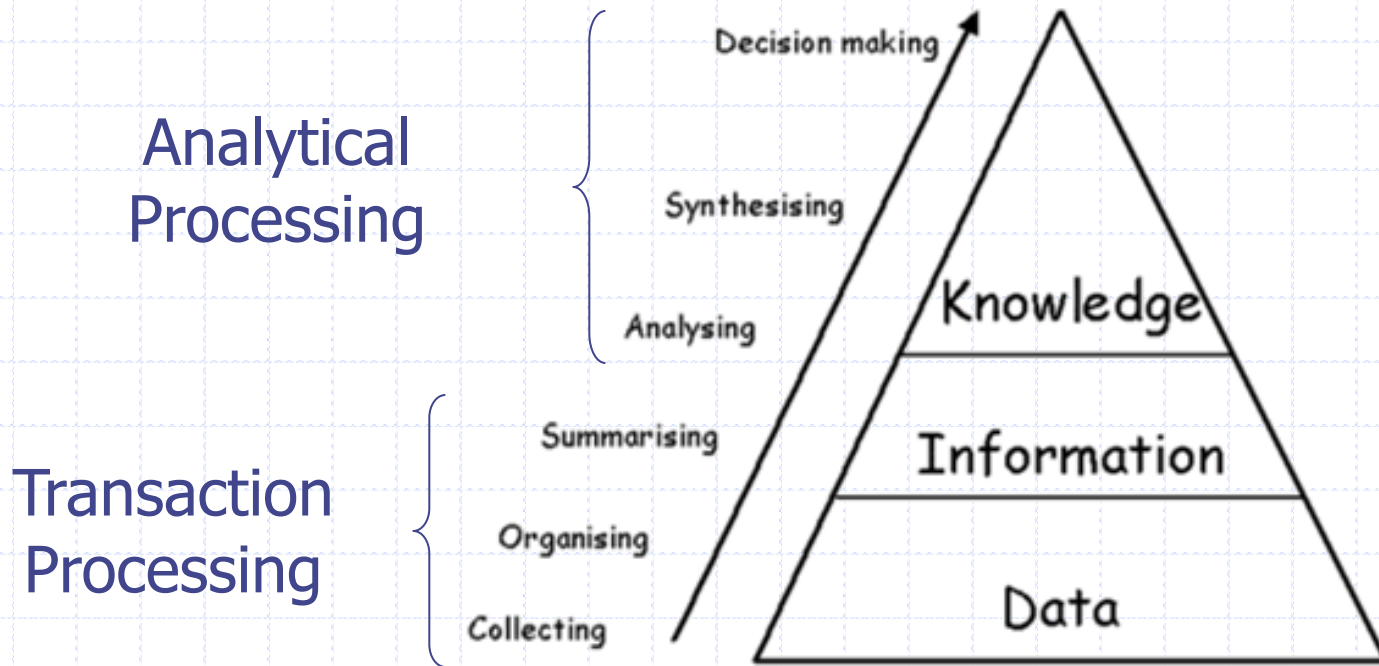
### ■ Analytical Processing

- ◆ Focus on reporting, analysis, transformation, and decision support

# Data Information and Knowledge

## ◆ DIKW

- <http://en.wikipedia.org/wiki/DIKW>



# Business Intelligence (BI)

- ◆ BI is the an umbrella term for concepts, methods, and technologies used to support decision making.
  - Traditionally it has been also understood as Decision Support System (DSS).
  - A brief history of DSS:  
<http://dssresources.com/history/dsshistory.html>
- ◆ Narrowly speaking, intelligence comes from data (facts).
  - In this sense, BI focuses on analytical processing.
- ◆ Broadly speaking, intelligence, or knowledge, also comes from human experience and tacit knowledge.
  - In this sense, BI is also related to knowledge management.

# Why BI is Needed

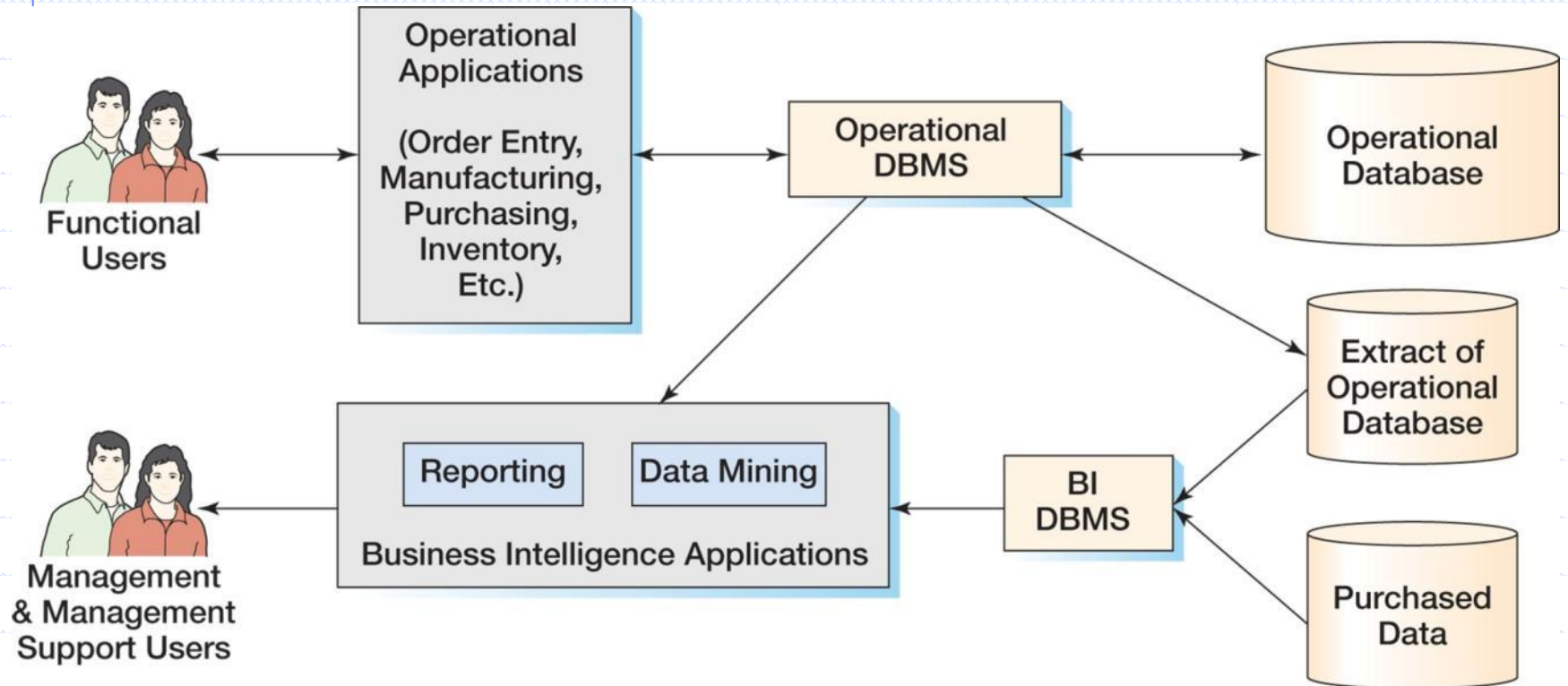
## ◆ Information overloading

- There are too much data and information that we cannot find useful information (knowledge) from them.

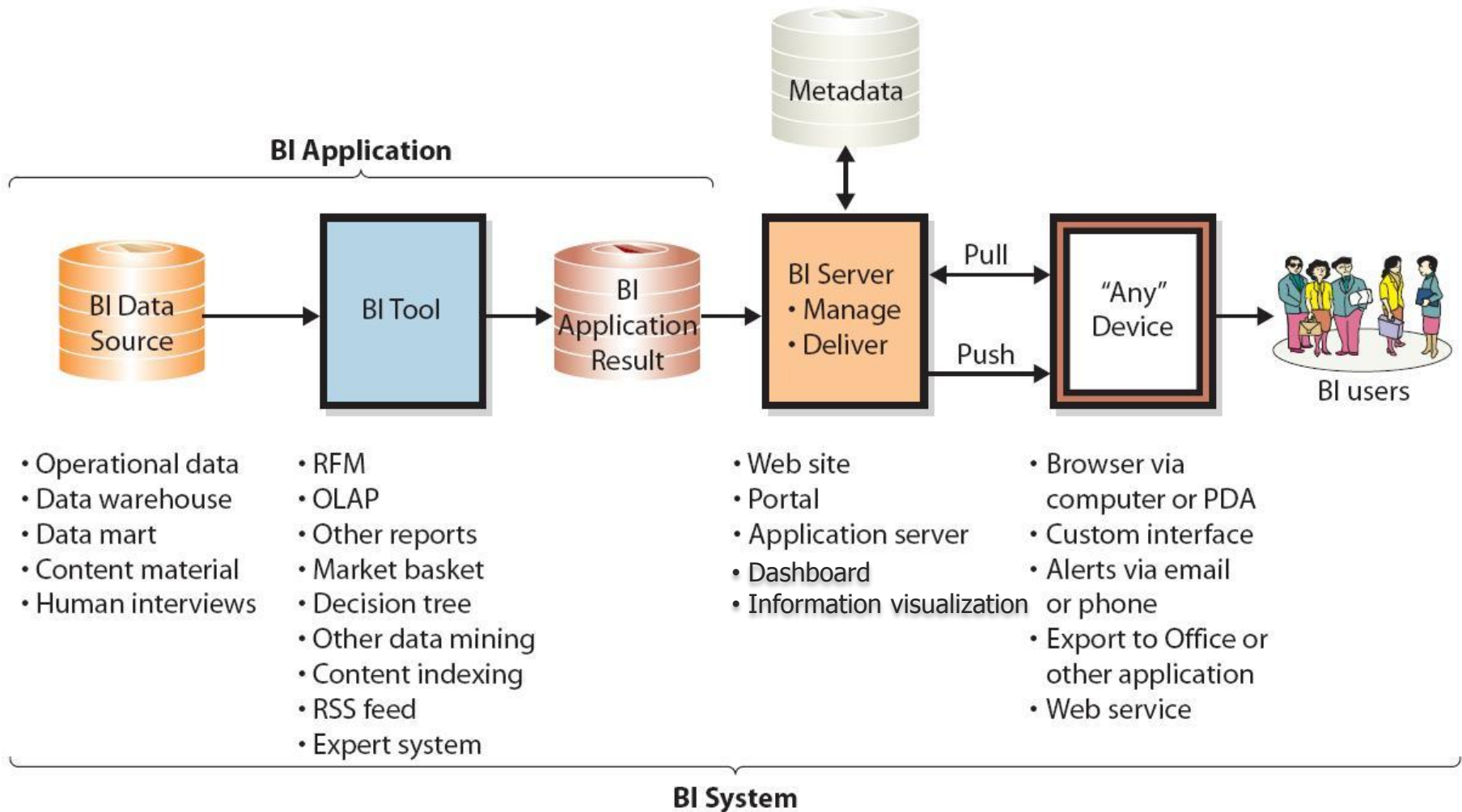
- ◆ A gap between data and knowledge (useful information leading to a decision).



# BI: A General Process



# BI System Components



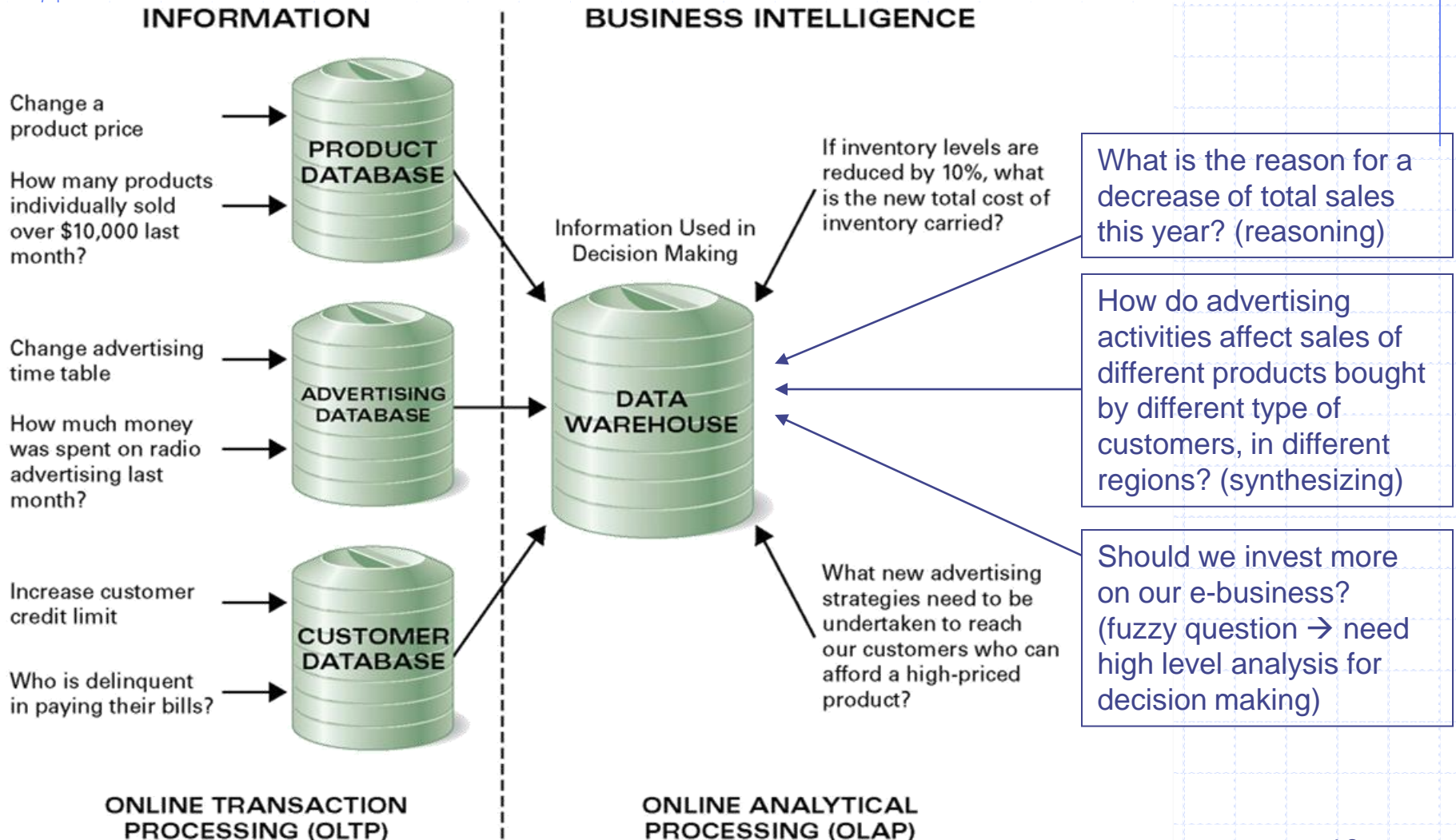
# Major BI Components

- ◆ Data management
  - Data warehouse, data mart, ETL
- ◆ Applications and tools (analytics)
  - Query, OLAP, statistics, data mining
- ◆ Delivery
  - Reporting, user interface, dashboards, information visualization, portal
- ◆ Application domains
  - Performance management, CRM, Portfolio management, etc.

# Data Warehouse

- ◆ Data warehouse is a special kind of database that support data analysis and decision making
- ◆ Traditional (operational) databases facilitate data management and transaction processing. They have two limitations for data analysis and decision support
  - Performance
    - ◆ They are transaction oriented (data insert, update, move, etc.)
    - ◆ Not optimized for complex data analysis
    - ◆ Usually do not hold historical data
  - Heterogeneity
    - ◆ Individual databases usually manage data in very different ways, even in the same organization (not to mention external data sources which may be dramatically different).

# Data Warehousing Supports Analytical Processing



# BI Applications and Tools

## ◆ Query

- Based on simple and direct queries
- Usually involves simple analysis and transformation of data, such as calculating, sorting, filtering, grouping, and formatting

## ◆ OLAP (Online Analytical Processing)

- A multi-dimensional analysis and reporting application
- Great for discovering details from large quantities of data

## ◆ Data mining

- Data mining techniques are a blend of statistics and mathematics, and artificial intelligence and machine-learning.

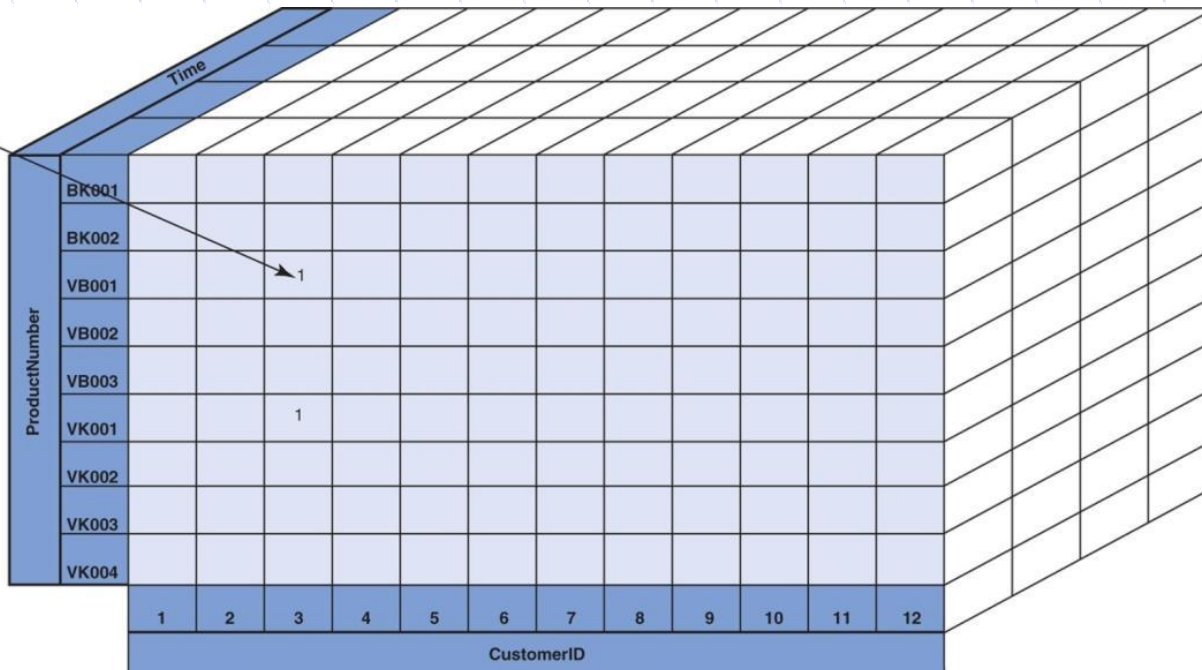
# OLAP

- ◆ The multidimensionality of data warehouse is particularly suitable for *multi-dimensional queries*
  - Such queries are usually arithmetic aggregation operations (sum, average, etc.) on records grouped by multiple dimensions (attributes).
- ◆ Examples
  - “What is the total sales amount grouped by product line (dimension 1), states (dimension 2), years (dimension 3) and ... (other dimensions)?”
  - “What is the total revenue for each store in the last 24 months?”
- ◆ OLAP is a function/operation that is optimized to answer queries that are multi-dimensional in nature

# OLAP Report View

3	Store Sales Net	Store Type					
4	Product Family	Deluxe Supermarket	Gourmet Supermarket	Mid-Size Grocery	Small Grocery	Supermarket	Grand Total
5	Drink	\$8,119.05	\$2,392.83	\$1,409.50	\$685.89	\$16,751.71	\$29,358.98
6	Food	\$70,276.11	\$20,026.18	\$10,392.19	\$6,109.72	\$138,960.67	\$245,764.87
7	Non-Consumable	\$18,884.24	\$5,064.79	\$2,813.73	\$1,534.90	\$36,189.40	\$64,487.05
8	Grand Total	\$97,279.40	\$27,483.80	\$14,615.42	\$8,330.51	\$191,901.77	\$339,610.90

Each cell will show the total quantity of each product that has been purchased by each customer on a specific date



# Data Mining

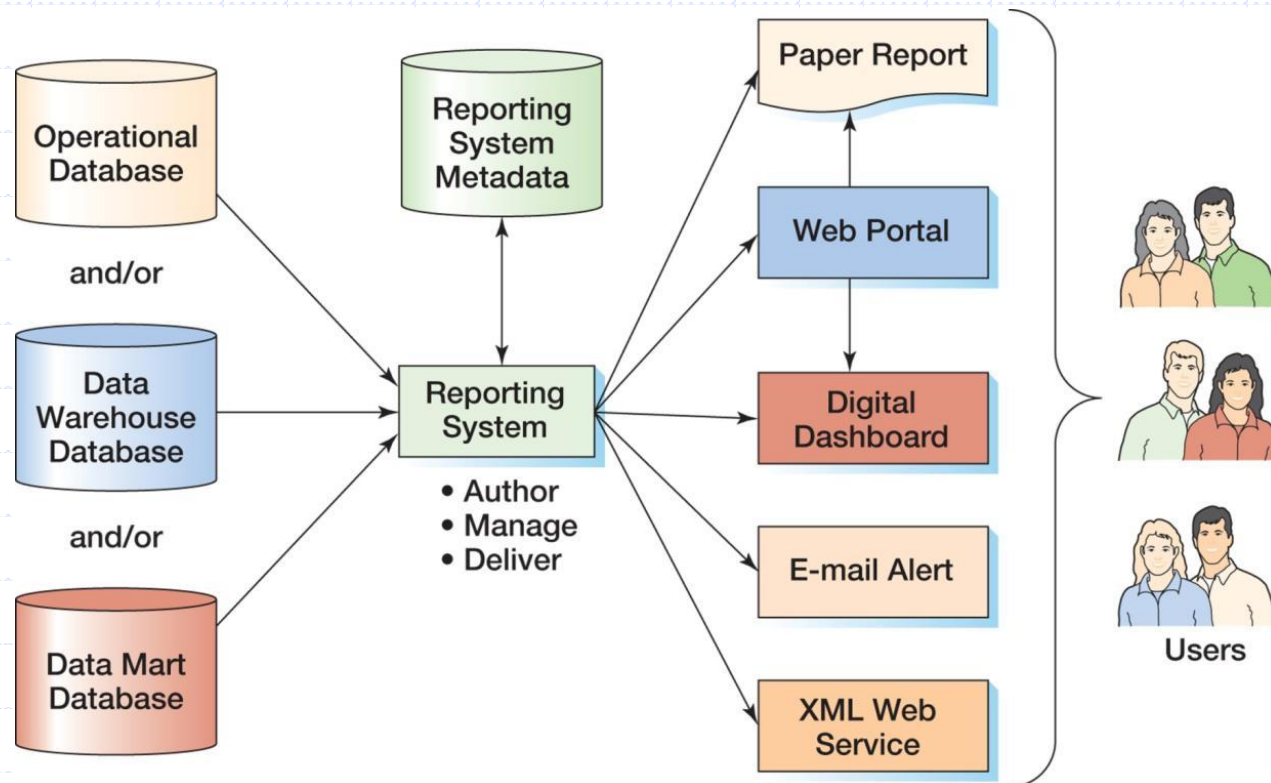
- ◆ Data mining (or, knowledge discovery in database, KDD)
  - Processes and techniques for seeking knowledge (relationship, trends, patterns, etc.) from a large amount of data
  - Non-trivial, non-obvious, and implicit knowledge
  - Extremely large datasets
- ◆ Data mining applications use sophisticated statistical and mathematical techniques to find patterns and relationships among data
  - Classification, clustering, association, estimation, prediction, etc.

# Common Data Mining Techniques

- ◆ K-means: a clustering technique used to identify groups of entities with similar characteristics
- ◆ Regression analysis: produces equations that offer probabilities that certain events will occur
- ◆ Decision tree analysis: classifies entities into groups based on past history
- ◆ Artificial Neural Network (ANN): non-linear prediction techniques
- ◆ Market Basket Analysis: determines patterns of associated buying behavior
- ◆ Genetic Algorithm (GA): searching optimum solutions in a relative large solution space.

# BI Reporting and Delivery

- ◆ BI reporting is about delivering to users and managing analysis results



# BI Reporting Applications

## ◆ Web portal

- A web portal is a central place for managing and retrieving data and reports
- Example: <https://discoverer.gsu.edu>

## ◆ Dashboard

- Digital dashboard is a visual presentation of data to make it easy to read and understand in a short time
- Why
  - ◆ Quickly understand data and respond quickly
  - ◆ Ability to identify trends
  - ◆ Save time over running multiple reports
  - ◆ Gain total visibility of all systems instantly at one place
- Example: <http://it.usaspending.gov/>

# Summary

## ◆ Key concepts

- Business intelligence
- BI system components
- Data warehouse
- OLAP
- Multidimensional analysis
- Data mining
- BI reporting

# Key Resources

## ◆ DSS Resources

- <http://dssresources.com/>

## ◆ Business intelligence resources

- <http://www.businessintelligence.com/>

# SQL Server BI Resources

## ◆ General

- <http://www.microsoft.com/bi>
- <http://www.microsoft.com/sqlserver/2008/en/us/business-intelligence.aspx>

## ◆ Data warehouse

- <http://www.microsoft.com/SqlServer/2008/en/us/data-warehousing.aspx>
- <http://msftdbprodsamples.codeplex.com>

## ◆ SQL Server Services

- SSAS: <http://www.microsoft.com/sqlserver/2008/en/us/analysis-services.aspx>
- SSIS: <http://www.microsoft.com/sqlserver/2008/en/us/integration.aspx>
- SSRS: <http://www.microsoft.com/sqlserver/2008/en/us/reporting.aspx>

## ◆ Data mining

- <http://www.microsoft.com/sqlserver/2008/en/us/data-mining-addins.aspx>
- <http://www.sqlserverdatamining.com>

## ◆ Business Intelligence w/ Excel and SharePoint

- <http://office.microsoft.com/en-us/products/FX101674131033.aspx>