

# Management Information Systems

## B10. Data Management: Warehousing, Analyzing, Mining, and Visualization



- Code: 166137-01+02
- Course: Management Information Systems
- Period: Spring 2013
- Professor: Sync Sangwon Lee, Ph. D

## Contents

- **Part I: IT in the Organization**
  - 01. Strategic Use of Information Technology in the Digital Economy
  - 02. Information Technologies: Concepts and Management
- **Part II: The Web Revolution**
  - 03. Network Computing, Discovery, Communication, and Collaboration
  - 04. E-Business and E-commerce
  - 05. Mobile, Wireless, and Pervasive Computing
- **Part III: Organizational Applications**
  - 06. Transaction Processing, Functional Applications, CRM, and Integration
  - 07. Enterprise Systems: From Supply Chains to ERP to CRM
  - 08. Interorganizational and Global Information Systems



## Contents

- **Part IV: Managerial and Decision Support Systems**
- 09. Knowledge Management
- **10. Data Management: Warehousing, Analyzing, Mining, and Visualization**
- 11. Management Decision Support and Intelligent Systems
  
- **Part V: Implementing and Managing IT**
- 12. Using Information Technology for Strategic Advantage
- 13. Information Technology Economics
- 14. Acquiring IT Applications and Infrastructure
- 15. Managing Information Resources and IT Security
- 16. The Impacts of IT on Individuals, Organizations, and Society



3

## Learning Objectives

- 01. Recognize the importance of data, their managerial issues, and their life cycle.
- 02. Describe the sources of data, their collection, and quality issues.
- 03. Describe document management systems.
- 04. Explain the operation of data warehousing and its role in decision support.
- 05. Describe information and knowledge discovery and business intelligence.
- 06. Understand the power and benefits of data mining.
- 07. Describe data presentation methods and explain geographical information systems, visual simulations, and virtual reality as decision support tools.
- 08. Discuss the role of marketing databases and provide examples.
- 09. Recognize the role of the Web in data management.



4

## 01. Data Management

- Managing Data
  - Data are an asset, when converted to information and knowledge, give the firm competitive advantages.
  - IT applications cannot be done without using some kind of data which are at the core of management and marketing operations.



<http://chanellewebtech.com>

5

## 01. Data Management

- Managing Data
  - However, managing data is difficult for various reasons.
    - The amount of data increases exponentially with time.
    - Data are scattered throughout organizations.
    - Data are collected by many individuals using several methods.
    - External data needs to be considered in making organizational decisions.
    - Data security, quality, and integrity are critical.
    - Selecting data management tools can be a major problem.

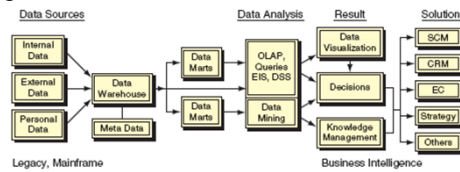


<http://www.marketo.com>

6

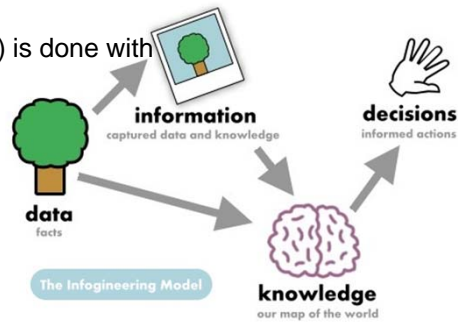
# 01. Data Management

- Data Life Cycle Process
  - Businesses run on data that have been processed to information and knowledge, which managers apply to businesses problems and opportunities.
  - The result of all these activities is the generating of decision support and knowledge.
  - The result -



# 01. Data Management

- Data Life Cycle Process
  - This transformation of data into knowledge and solutions is accomplished in several ways.
    - New data collection occurs from various sources.
    - It is temporarily stored in a database then preprocessed to fit the format of the organizations data warehouse or data marts.
    - Users then access the warehouse or data mart and take a copy of the needed data for analysis.
    - Analysis (looking for patterns) is done with
      - Data analysis tools
      - Data mining tools



## 01. Data Management

- Data Sources
  - The data life cycle begins with the acquisition of data from data sources.
  - Types of data sources
    - 1) Internal data sources
    - 2) Personal data sources
    - 3) External data sources



## 01. Data Management

- Types of Data Sources
  - 1) Internal data source
    - Internal data sources are usually stored in the corporate database and are about people, products, services, and processes.



## 01. Data Management

- Types of Data Sources
  - 2) Personal data source
    - Personal data is documentation on the expertise of corporate employees usually maintained by the employee.
    - It can take the form of:
      - Estimates of sales
      - Opinions about competitors
      - Business rules
      - Procedures
      - Etc.



<http://www.dundas.com>

## 01. Data Management

- Types of Data Sources
  - 3) External data source
    - External data sources range from commercial databases to Government reports.
    - Internet and commercial database services are accessible through the Internet.



<http://www.dundas.com>

## 01. Data Management

- Methods for Collecting Raw Data
  - The task of data collection is fairly complex.
  - Which can create data-quality problem requiring validation and cleansing of data?



<http://www.dundas.com>

13

## 01. Data Management

- Methods for Collecting Raw Data
  - Collection can take place
    - In the field
    - From individuals
    - Via manually methods
      - Time studies
      - Surveys
      - Observations
      - Contributions from experts
    - Using instruments and sensors
    - Transaction processing systems (TPS)
    - Via electronic transfer
    - From a web site (Clickstream)

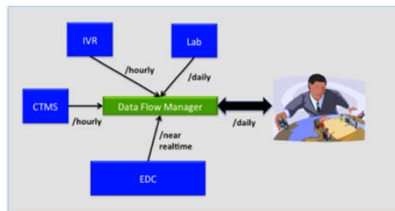


<http://www.nedarc.org>

14

## 01. Data Management

- Data Flow Manager (DFM)
  - One way to improve data collection from multiple external sources
  - DFM takes information from external sources and puts it where it is needed, when it is needed, in a usable form.
  - DFM consists of
    - A decision support system
    - A central data request processor
    - A data integrity component
    - Links to external data suppliers
    - The processes used by the external data suppliers.



<http://www.solitexconsulting.com>

## 02. Data Quality

- Data Quality and Integrity
  - Data quality (DQ) is an extremely important issue since quality determines the data's usefulness as well as the quality of the decisions based on the data.
  - Data quality is the cornerstone of effective business intelligence.
  - Data integrity means that data must be accurate, accessible, and up-to-date.

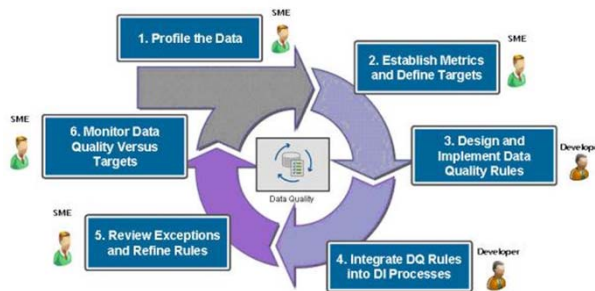


<http://www.enterworks.com>



## 02. Data Quality

- Types of Data Quality
  - Intrinsic DQ: Accuracy, objectivity, believability, and reputation.
  - Accessibility DQ: Accessibility and access security.
  - Contextual DQ: Relevancy, value added, timeliness, completeness, amount of data.
  - Representation DQ: Interpretability, ease of understanding, concise representation, consistent representation.



<http://www.tdwi.org>

17

## 03. Document Management

- Document Management
  - Document management is the automated control of electronic documents, page images, spreadsheets, word processing documents, and other complex documents through their entire life cycle within an organization, from initial creation to final archiving.



<http://www.qis.com>

18

### 03. Document Management

- Document Management
  - Maintaining paper documents, requires that:
    - The documents be distributed to the appropriate individuals in a timely manner
    - Everyone have the current version
    - An update schedule be determined
    - Security be provided for the document

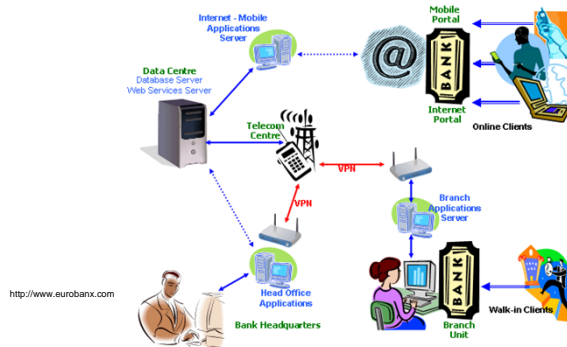


<http://www.huddle.com>

[thisbroken.com](http://thisbroken.com)

### 04. Data Processing

- Transactional vs. Analytical Data Processing
  - Transactional processing takes place in operational systems (TPS) that provide the organization with the capability to perform business transactions and produce transaction reports.
  - The data are organized mainly in a hierarchical structure and are centrally processed.
  - This is done primarily for fast and efficient processing of routine, repetitive data.



<http://www.eurobanx.com>

## 04. Data Processing

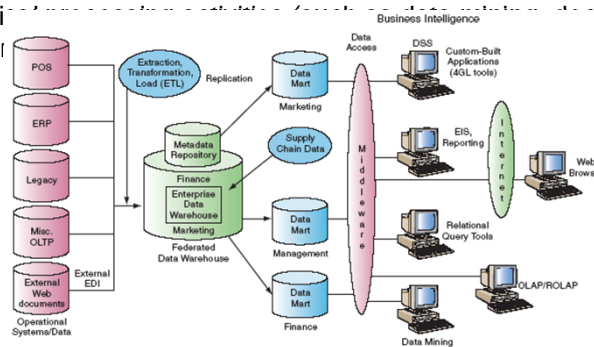
- Transactional vs. Analytical Data Processing
  - A supplementary activity to transaction processing is called analytical processing, which involves the analysis of accumulated data.
  - Analytical processing, sometimes referred to as business intelligence, includes data mining, decision support systems (DSS), querying, and other analysis activities.
  - These analyses place strategic information in the hands of decision makers to enhance productivity and make better decisions, leading to greater competitive advantage.



<http://www.tech-faq.com>

## 05. Data Warehouse

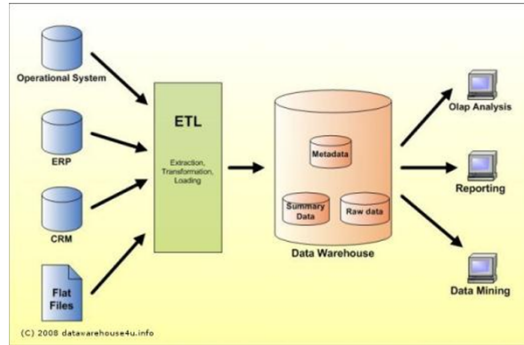
- Data Warehouse (DW)
  - A data warehouse is a repository of subject-oriented historical data that is organized to be accessible in a form readily acceptable for analytical processing activities (querying, data mining, decision support, querying)



Information Technology for Management, Ed. 5, Efraim Turban et al., Wiley

## 05. Data Warehouse

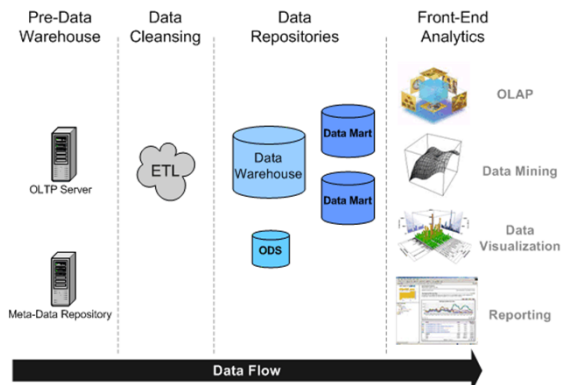
- Benefits of Data Warehouse
  - The ability to reach data quickly, since they are located in one place
  - The ability to reach data easily and frequently by end users with Web browsers.



<http://datawarehouse4u.info>

## 05. Data Warehouse

- Characteristics of Data Warehouse
  - Consistency
    - In the warehouse data will be coded in a consistent manner.
  - Relational
    - Typically the data warehouse uses a relational structure.



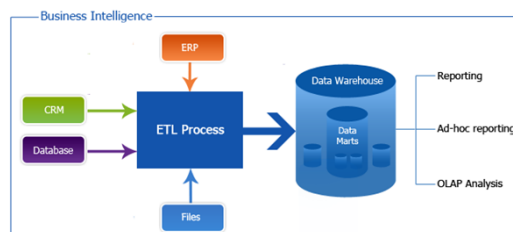
## 05. Data Warehouse

- Characteristics of Data Warehouse
  - Nonvolatile
    - Once entered into the warehouse, data are not updated.
  - Organization
    - Data are organized by subject



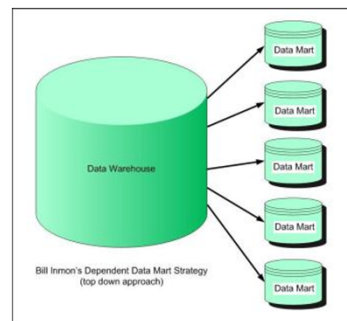
## 05. Data Warehouse

- Characteristics of Data Warehouse
  - Client/server
    - The data warehouse uses the client/server architecture mainly to provide the end user an easy access to its data.
  - Time variant
    - The data are kept for many years so they can be used for trends, forecasting, and comparisons over time.
  - Web-based
    - Data warehouses are designed to provide an efficient computing environment for Web-based applications



## 05. Data Warehouse

- Data Mart
  - A data mart is a small scaled-down version of a data warehouse designed for a strategic business unit (SBU) or a department.
  - Since they contain less information than the data warehouse they provide more rapid response and are more easily navigated than enterprise-wide data warehouses.

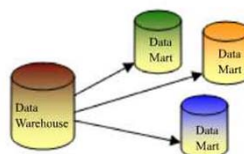


<http://www.datamonkeys.com>

27

## 05. Data Warehouse

- Types of Data Mart
  - Replicated (dependent) data marts
    - Replicated (dependent) data marts are small subsets of the data warehouse.
    - In such cases one replicates some subset of the data warehouse into smaller data marts, each of which is dedicated to a certain functional area.
  - Stand-alone data marts
    - A company can have one or more independent data marts without having a data warehouse.
    - Typical data marts are for marketing, finance, and engineering applications.

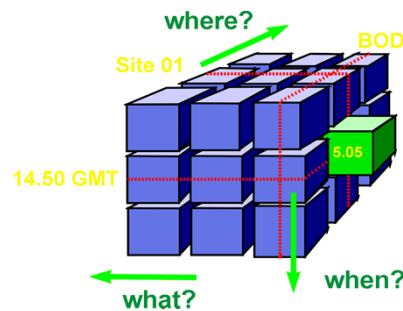


<http://www.exforsys.com>

28

## 05. Data Warehouse

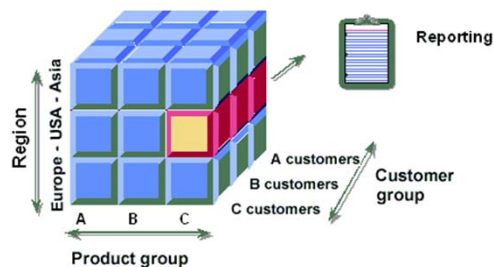
- Data Cube
  - Multidimensional databases (sometimes called OLAP) are specialized data stores that organize facts by dimensions, such as geographical region, product line, salesperson, time.
  - The data in these databases are usually preprocessed and stored in data cubes.



<http://www.nwl.ac.uk>

## 05. Data Warehouse

- Data Cube
  - One intersection might be the quantities of a product sold by specific retail locations during certain time periods.
  - Another matrix might be Sales volume by department, by day, by month, by year for a specific region
  - Cubes provide faster:
    - Slices and dices of the information
    - Queries
    - Rollups
    - Drill downs



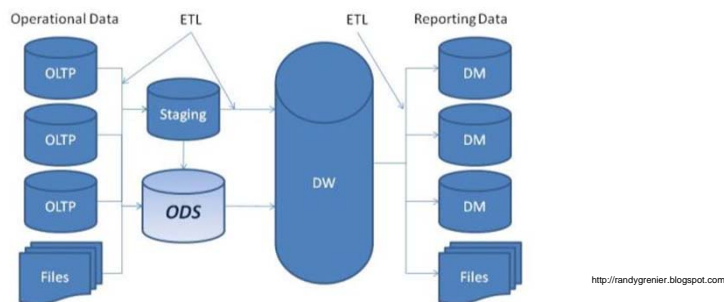
How high is the contribution margin for product group C in the USA region?

<http://en.sap.info>

## 05. Data Warehouse

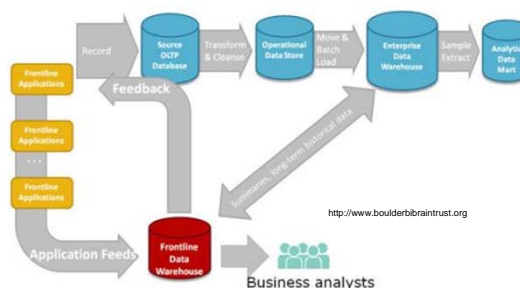
- Operational Data Stores (ODS)
  - Operational data store is a database for transaction processing systems that uses data warehouse concepts to provide clean data to the TPS.
  - It brings the concepts and benefits of a data warehouse to the operational portions of the business.

ODS within Data Warehouse Architecture



## 05. Data Warehouse

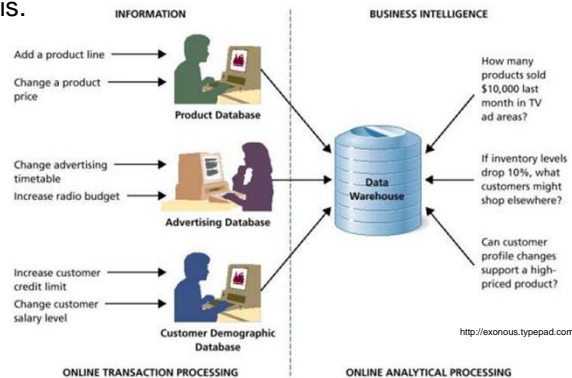
- Operational Data Stores (ODS)
  - It is typically used for short-term decisions that require time sensitive data analysis
  - It logically falls between the operational data in legacy systems and the data warehouse.
  - It provides detail as opposed to summary data.
  - It is optimized for frequent access
  - It provides faster response times.





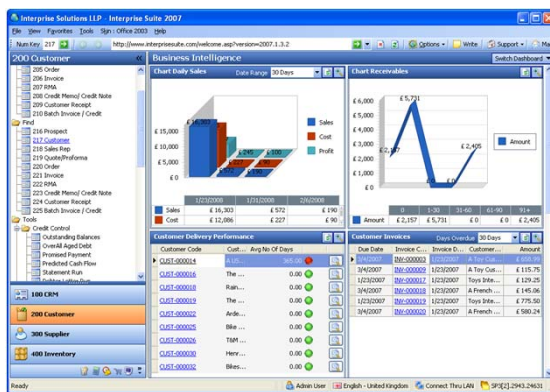
## 06. Business Intelligence

- Business Intelligence (BI)
  - BI is a broad category of applications and techniques for gathering, storing, analyzing and providing access to data.
  - It helps enterprise users make better business and strategic decisions.
  - Major applications include the activities of query and reporting, online analytical processing (OLAP), DSS, data mining, forecasting and statistical analysis.



## 06. Business Intelligence

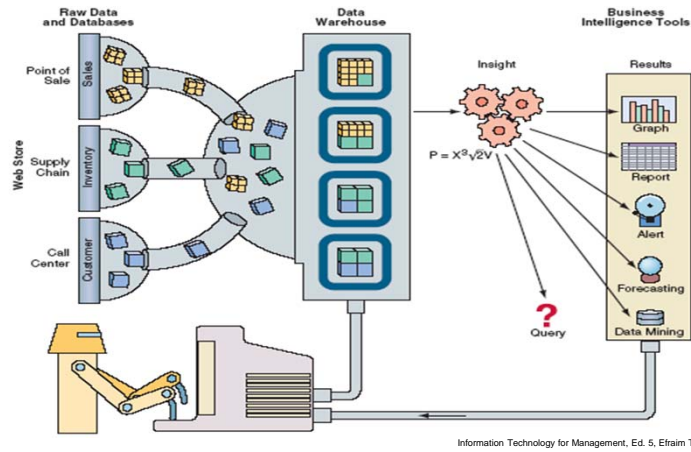
- Components of Business Intelligence
  - Business intelligence includes:
    - Outputs such as financial modeling and budgeting
    - Benchmarking (business performance)
    - Coupons and sales promotions
    - Competitive intelligence
    - Resource allocation
    - Seasonality trends



http://www.publicity4u.eu

## 06. Business Intelligence

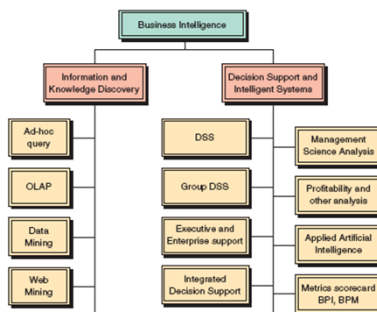
- How Business Intelligence Works.
  - BI starts with knowledge discovery (KD)



Information Technology for Management, Ed. 5, Efraim Turban et al., Wiley

## 06. Business Intelligence

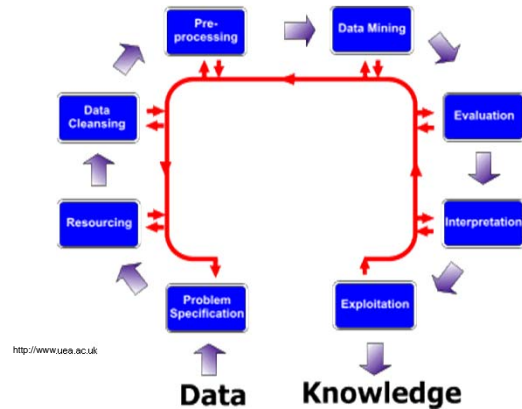
- The Structure of Business Intelligence
  - Discovering useful patterns



Information Technology for Management, Ed. 5, Efraim Turban et al., Wiley

## 06. Business Intelligence

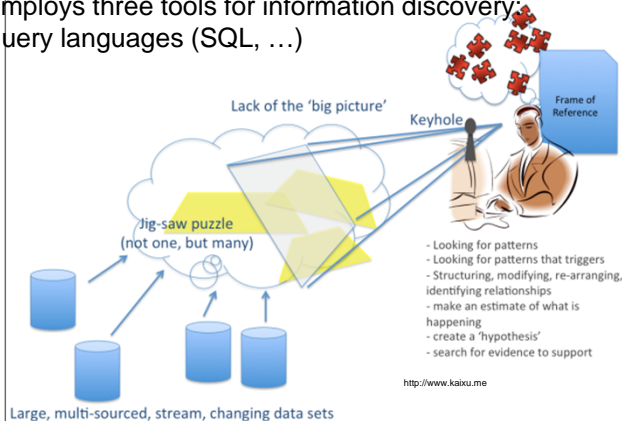
- Knowledge Discovery (KD)
  - Before information can be processed into BI it must be discovered or extracted from the data stores.
  - The major objective of this knowledge discovery in databases (KDD) is to identify valid, novel, potentially useful, and understandable patterns in data.



37

## 06. Business Intelligence

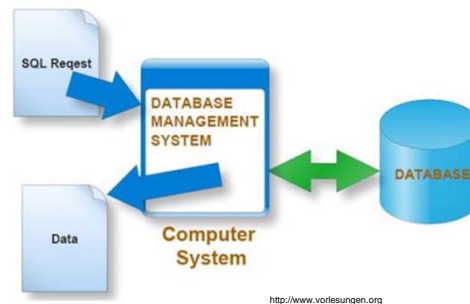
- Knowledge Discovery Technologies
  - KDD supported by three technologies:
    - Massive data collection
    - Powerful multiprocessor computers
    - Data mining and other algorithms.
  - KDD primarily employs three tools for information discovery:
    - Traditional query languages (SQL, ...)
    - OLAP
    - Data mining



38

## 06. Business Intelligence

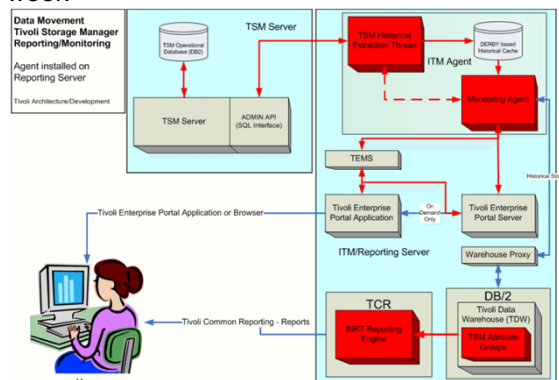
- Queries
  - Queries allow users to request information from the computer that is not available in periodic reports.
  - Query systems are often based on menus or if the data is stored in a database via a structured query language (SQL) or using a query-by-example (QBE) method.



<http://www.vorlesungen.org>

## 06. Business Intelligence

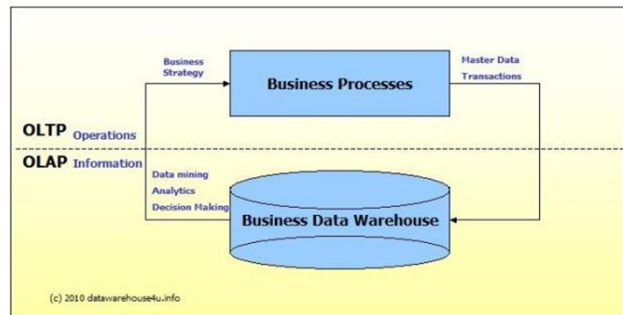
- Queries
  - User requests are stated in a query language and the results are subsets of the relationship
    - Sales by department by customer type for specific period
    - Weather conditions for specific date
    - Sales by day of week
    - ...



<http://www.ibm.com>

## 06. Business Intelligence

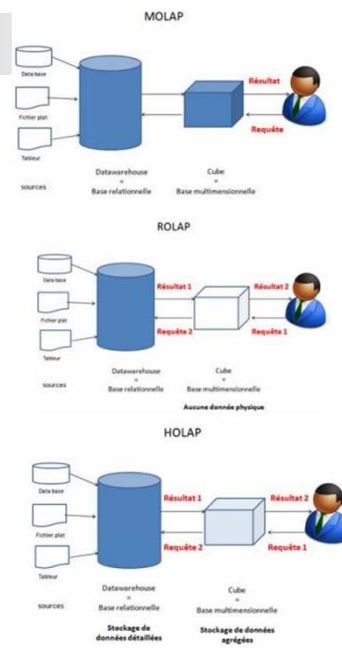
- Online Analytical Processing (OLAP)
  - Online analytical processing is a set of tools that analyze and aggregate data to reflect business needs of the company.
  - These business structures (multidimensional views of data) allow users to quickly answer business questions.
  - OLAP is performed on Data Warehouses and Marts.



<http://datawarehouse4u.info>

## 06. Business Intelligence

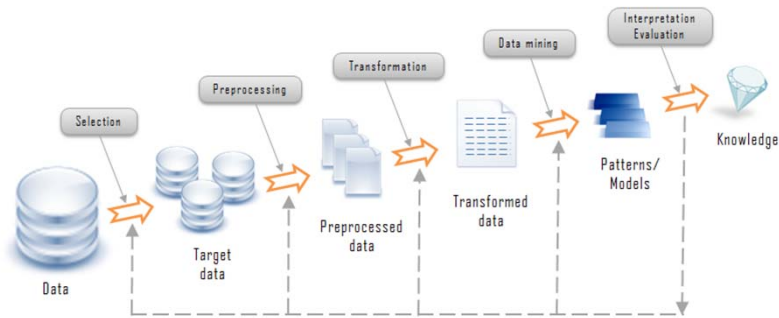
- Types of Online Analytical Processing
  - ROLAP (Relational OLAP)
    - ROLAP (Relational OLAP) is an OLAP database implemented on top of an existing relational database.
    - The multidimensional view is created each time for the user.
  - MOLAP (Multidimensional OLAP)
    - MOLAP (Multidimensional OLAP) is a specialized multidimensional data store such as a Data Cube.
    - The multidimensional view is physically stored in specialize data files.



<http://www.infodcisionnel.com>

## 07. Data Mining

- Data Mining
  - Data mining is a tool for analyzing large amounts of data.
  - It derives its name from the similarities between searching for valuable business information in a large database, and mining a mountain for a vein of valuable ore.



<http://www.rithme.eu>

43

## 07. Data Mining

- Data Mining
  - Data mining technology can generate new business opportunities by providing:
    - Automated prediction of trends and behaviors.
    - Automated discovery of previously unknown or hidden patterns.
  - Data mining tools can be combined with:
    - Spreadsheets
    - Other end-user software development tools
  - Data mining creates a data cube then extracts data.



<http://www.jumpthecurve.net>

44

## 07. Data Mining

- Data Mining Techniques
  - [Case-based reasoning] uses historical cases to recognize patterns
  - [Neural computing] is a machine learning approach which examines historical data for patterns.
  - [Intelligent agents] retrieving information from the Internet or from intranet-based databases .
  - [Association analysis] uses a specialized set of algorithms that sort through large data sets and express statistical rules among items.
  - [Nearest-neighbor method]
  - [Genetic algorithms]
  - [Decision trees]

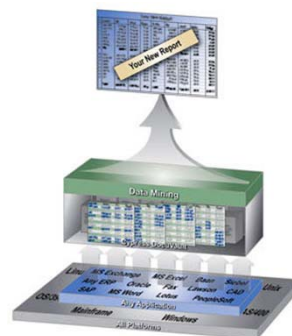


<http://www.rainbowskill.com>

45

## 07. Data Mining

- Data Mining Tasks
  - Classification
    - Infers the defining characteristics of a certain group.
  - Clustering
    - Identifies groups of items that share a particular characteristic. Clustering differs from classification in that no predefining characteristic is given.



<http://www.armsoftware.com>

46

## 07. Data Mining

- Data Mining Tasks
  - Association
    - Identifies relationships between events that occur at one time.
  - Sequencing
    - Identifies relationships that exist over a period of time.
  - Forecasting
    - Estimates future values based on patterns within large sets of data.



<http://www.geek.com>

47

## 07. Data Mining

- Data Mining Tasks
  - Regression
    - Maps a data item to a prediction variable.
  - Time Series
    - Analysis examines a value as it varies over time.



<http://techpubs.sgi.com>

48



## 07. Data Mining

- Other Mining Environments
  - In addition to data stored in traditional databases there are other “structures” that can be mined for patterns.
    - Text Mining is the application of data mining to non-structured or less-structured text files
    - Web Mining is the application of data mining techniques to data related to the World Wide Web. The data may be present in web pages or related to Web activity.
    - Spatial Mining is the application of data mining techniques to data that have a location component.
    - Temporal Mining is the application of data mining techniques to data that are maintained for multiple points in time.



<http://www.thomaslarock.com>

49

## 08. Data Visualization

- Data Visualization
  - Data visualization refers to presentation of data by technologies such as digital images, geographical information systems, graphical user interfaces, multidimensional tables and graphs, virtual reality, three-dimensional presentations, videos and animation.



<http://www.designyoutrust.com>

50

## 08. Data Visualization

- Multidimensionality Visualization
  - Modern data and information may have several dimensions.
  - Dimensions:
    - Products
    - Salespeople
    - Market segments
    - Business units
    - Geographical locations
    - Distribution channels
    - Countries
    - Industries



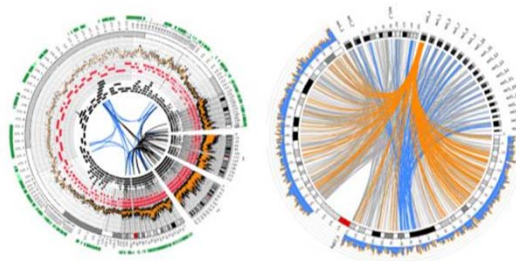
<http://www.panopticon.com>

51

## 08. Data Visualization

- Multidimensionality Visualization
  - Measures:
    - Money
    - Sales volume
    - Head count
    - Inventory profit
    - Actual versus forecasted results

circos is flexible

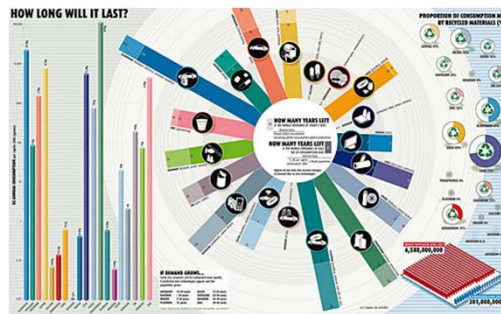


<http://mikeweb.bcgsic.ca>

52

## 08. Data Visualization

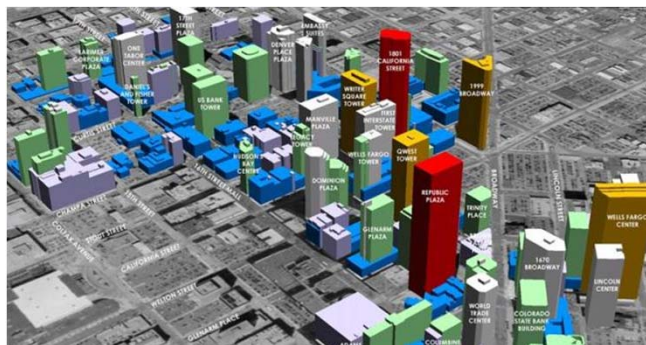
- Multidimensionality Visualization
  - Time:
    - Daily
    - Weekly
    - Monthly
    - Quarterly
    - Yearly



<http://www.noupe.com>

## 08. Data Visualization

- Geographical Information System (GIS)
  - A geographical information system is a computer-based system for capturing, storing, checking, integrating, manipulating, and displaying data using digitized maps.
  - Every record or digital object has an identified geographical location. It employs spatially oriented databases.



<http://www.denvergov.org>

## 08. Data Visualization

- Visual Interactive Modeling (VIM)
  - Visual interactive modeling uses computer graphic displays to represent the impact of different management or operational decisions on objectives such as profit or market share.



<http://www.org.umu.se>

55

## 08. Data Visualization

- Virtual Reality (VR)
  - Virtual reality is interactive, computer-generated, three-dimensional graphics delivered to the user.
  - These artificial sensory cues cause the user to “believe” that what they are doing is real.



<http://www.infoq.com>

56

## 09. Specialized Databases

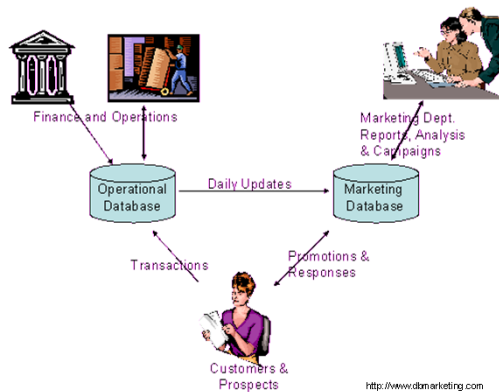
- Specialized Databases
  - Data warehouses and data marts serve end users in all functional areas.
  - Most current databases are static: They simply gather and store information.
  - Today's business environment also requires specialized databases.



<http://www.dbamatrix.com>

## 09. Specialized Databases

- Marketing Transaction Database (MTD)
  - Marketing transaction database combines many of the characteristics of the current databases and marketing data sources into a new database that allows marketers to engage in real-time personalization and target every interaction with customers.



<http://www.dbmarketing.com>

## 09. Specialized Databases

- Interactive Capability
  - An interactive capability an interactive transaction occurs with the customer exchanging information and updating the database in real time, as opposed to the periodic (weekly, monthly, or quarterly) updates of classical warehouses and marts.



<http://www.sita.aero>

## 10. Web-based Data Management Systems

- Web-based Data Management Systems
  - Data management and business intelligence activities—from data acquisition to mining—are often performed with Web tools, or are interrelated with Web technologies and e-business.
  - This is done through intranets, and for outsiders via extranets.





## 10. Web-based Data Management Systems

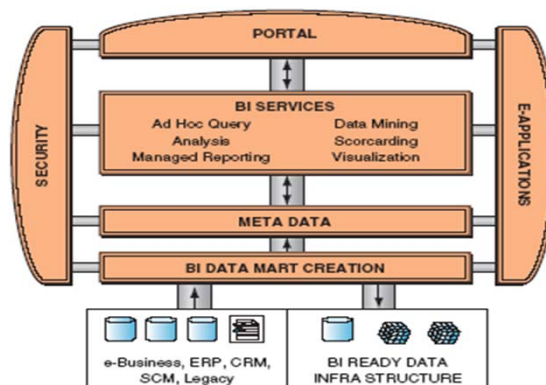
- Features of Web-based Data Management Systems
  - Enterprise BI suites and corporate portals
    - Integrate query, reporting, OLAP, and other tools.
  - Intelligent data warehouse web-based systems
    - Employ a search engine for specific applications which can improve the operation of a data warehouse
  - Clickstream data warehouse
    - occurs inside the Web environment, when customers visit a Web site.



<http://www.ecolabwew.com>

## 10. Web-based Data Management Systems

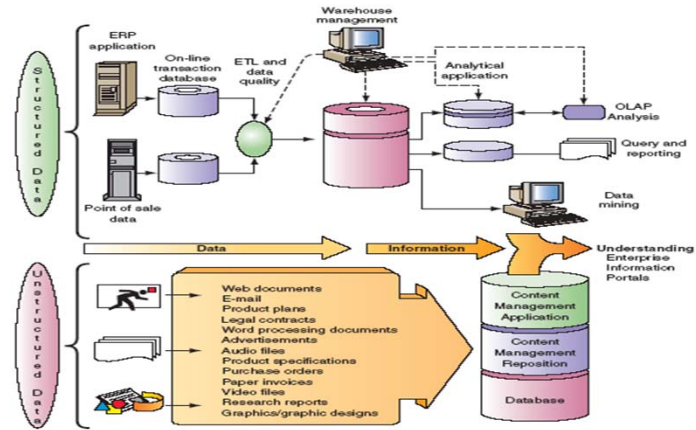
- Structure of Web-based Data Management Systems



Information Technology for Management, Ed. 5, Efraim Turban et al., Wiley

## 10. Web-based Data Management Systems

- Flow of Web-based Data Management Systems



Information Technology for Management, Ed. 5, Efraim Turban et al., Wiley