Introduction to Management Information Systems

Data Systems

Data Resource Management

Learning objectives

- understand distributing databases
- understand business intelligence systems
- to be able to explain data warehousing and data mining
- to understand business analysis in respect of data visualization, big data and data security

Distributed Databases

Distributed Database (DDB)

A collection of multiple, logically interrelated databases distributed over a computer network.

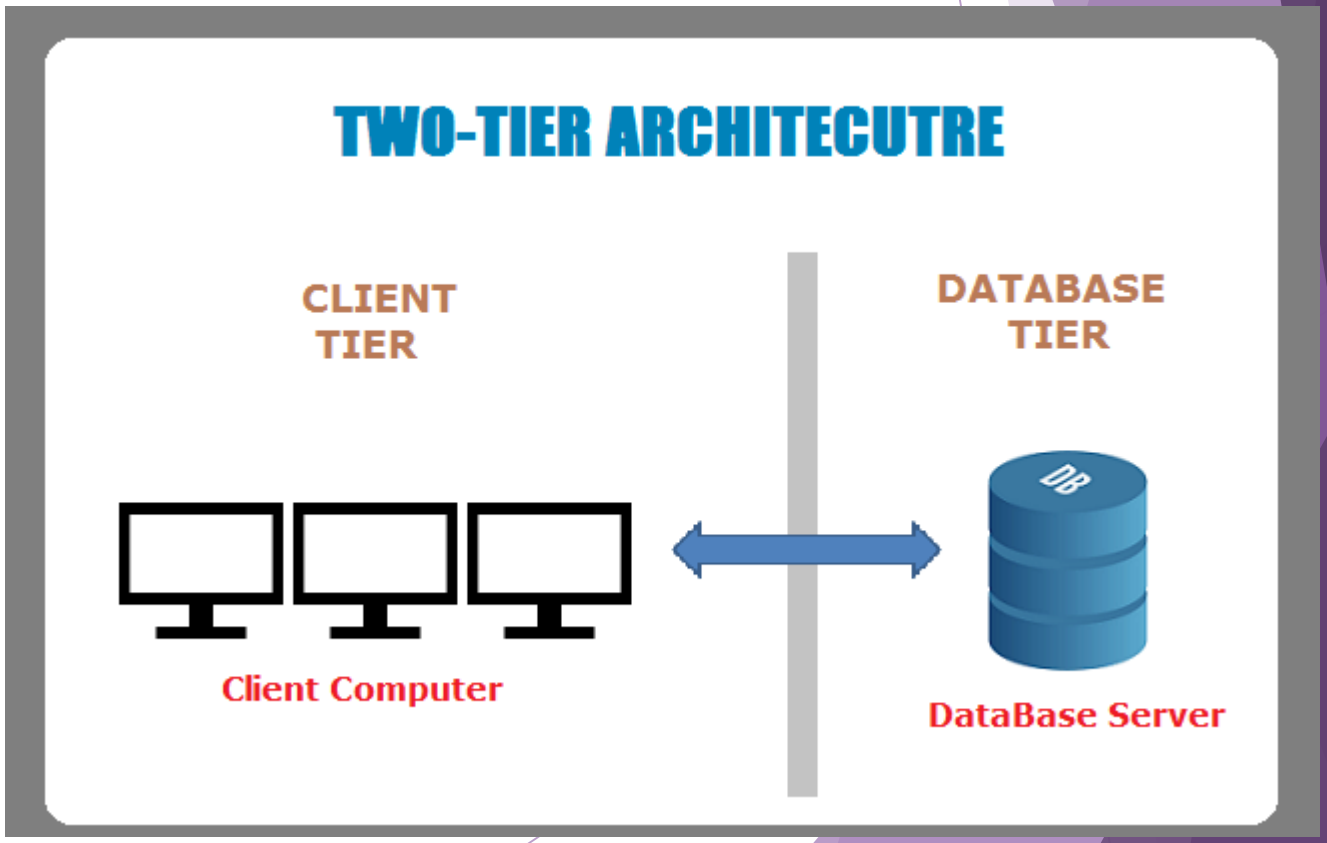
DDBMS manages a DDB such that the distribution is transparent to the user

Distributed Database Concepts

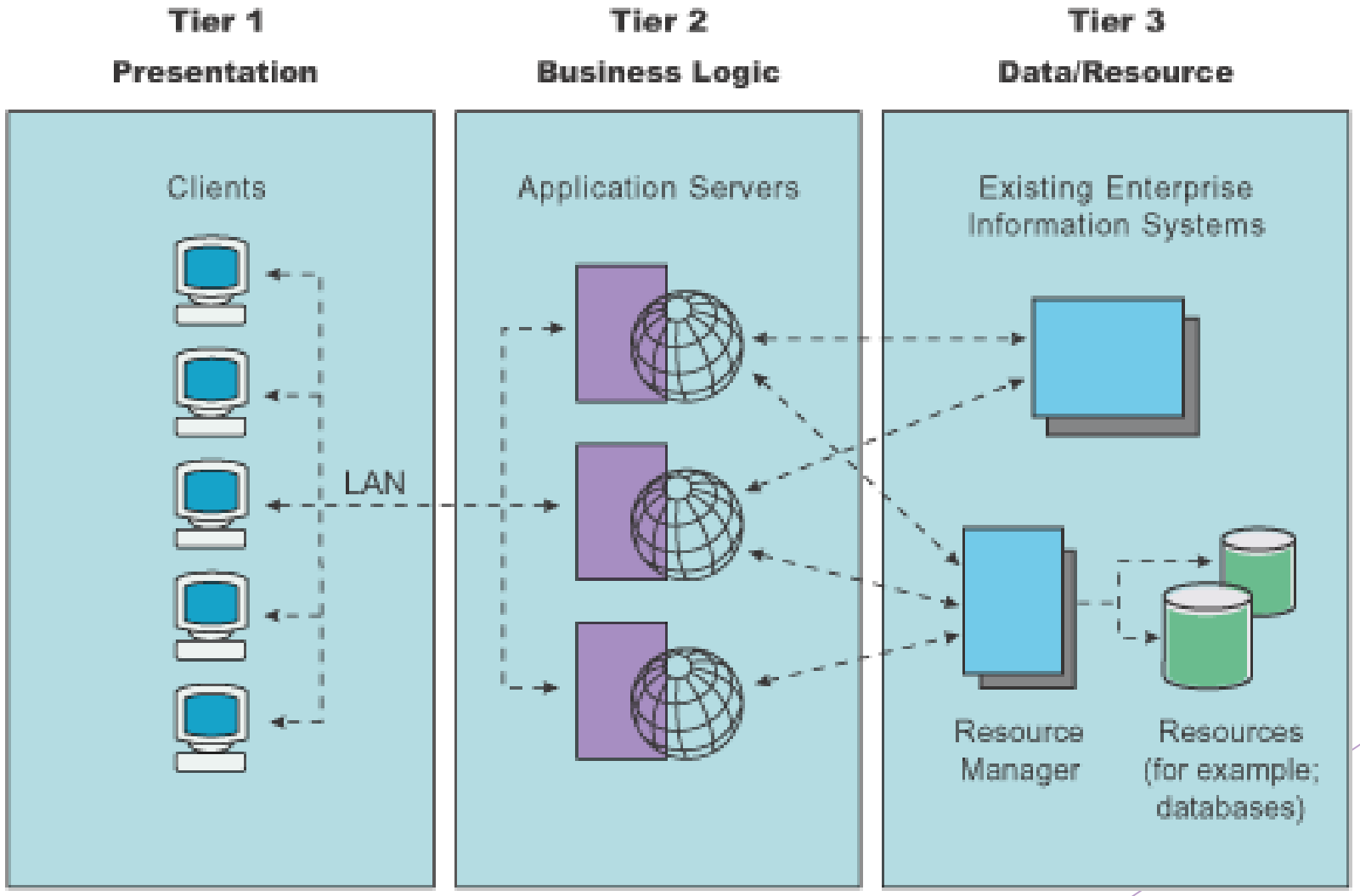
- What constitutes a distributed database?
  - Connection of database nodes over computer network
  - Logical interrelation of the connected databases
  - Possible absence of homogeneity among connected nodes
- Distributed database management system (DDBMS)
  - Software system that manages a distributed database

Basic Client-Server Architectures

- Specialized Servers with Specialized functions
- Clients
- DBMS Server

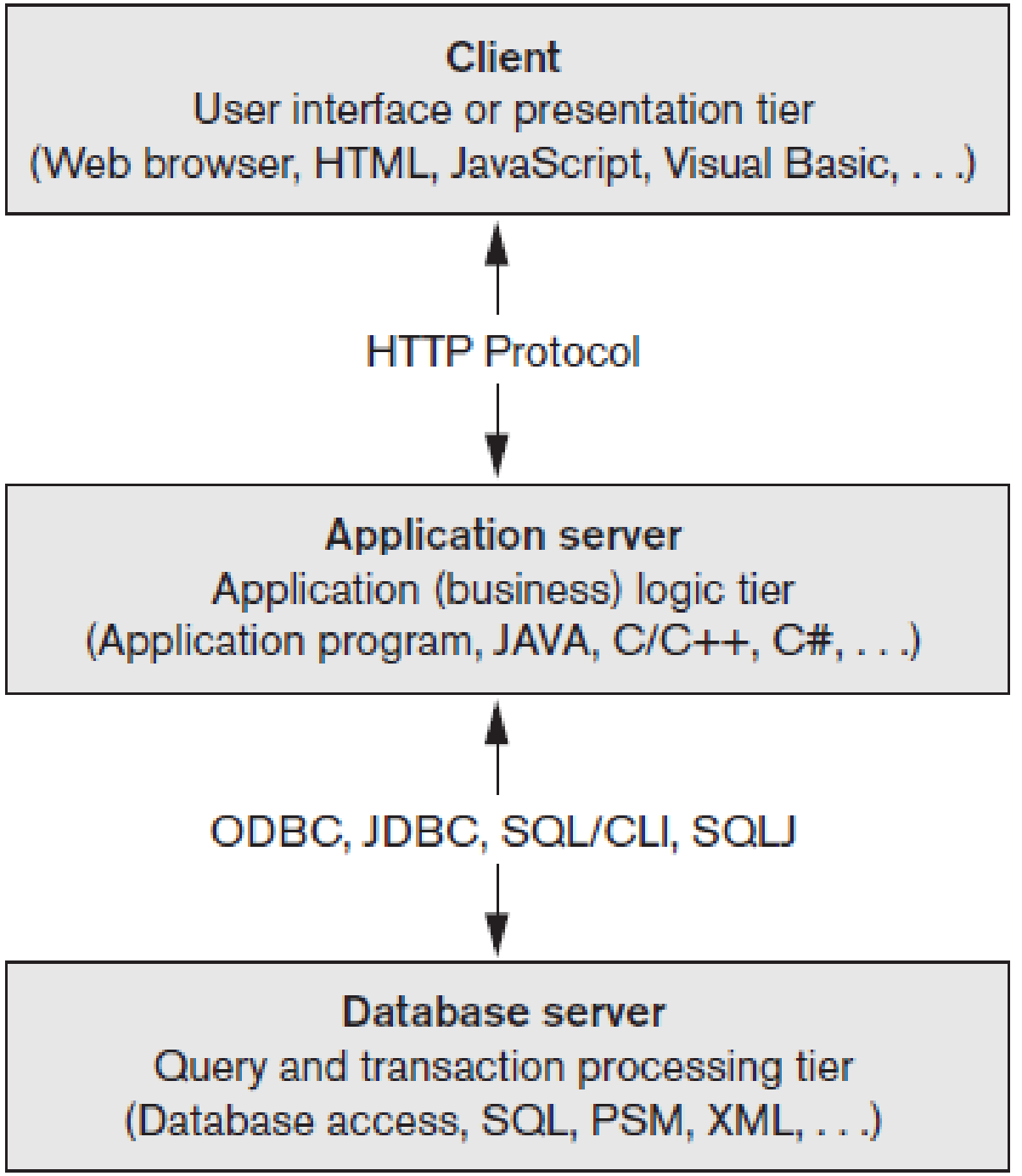


Three Tier Client-Server Architecture



Three-Tier Client/Server Architecture

Division of DBMS functionality among the three tiers can vary



The three-tier client/server architecture

Variations of Distributed Environments

Homogeneous DDBMS

- Same database system

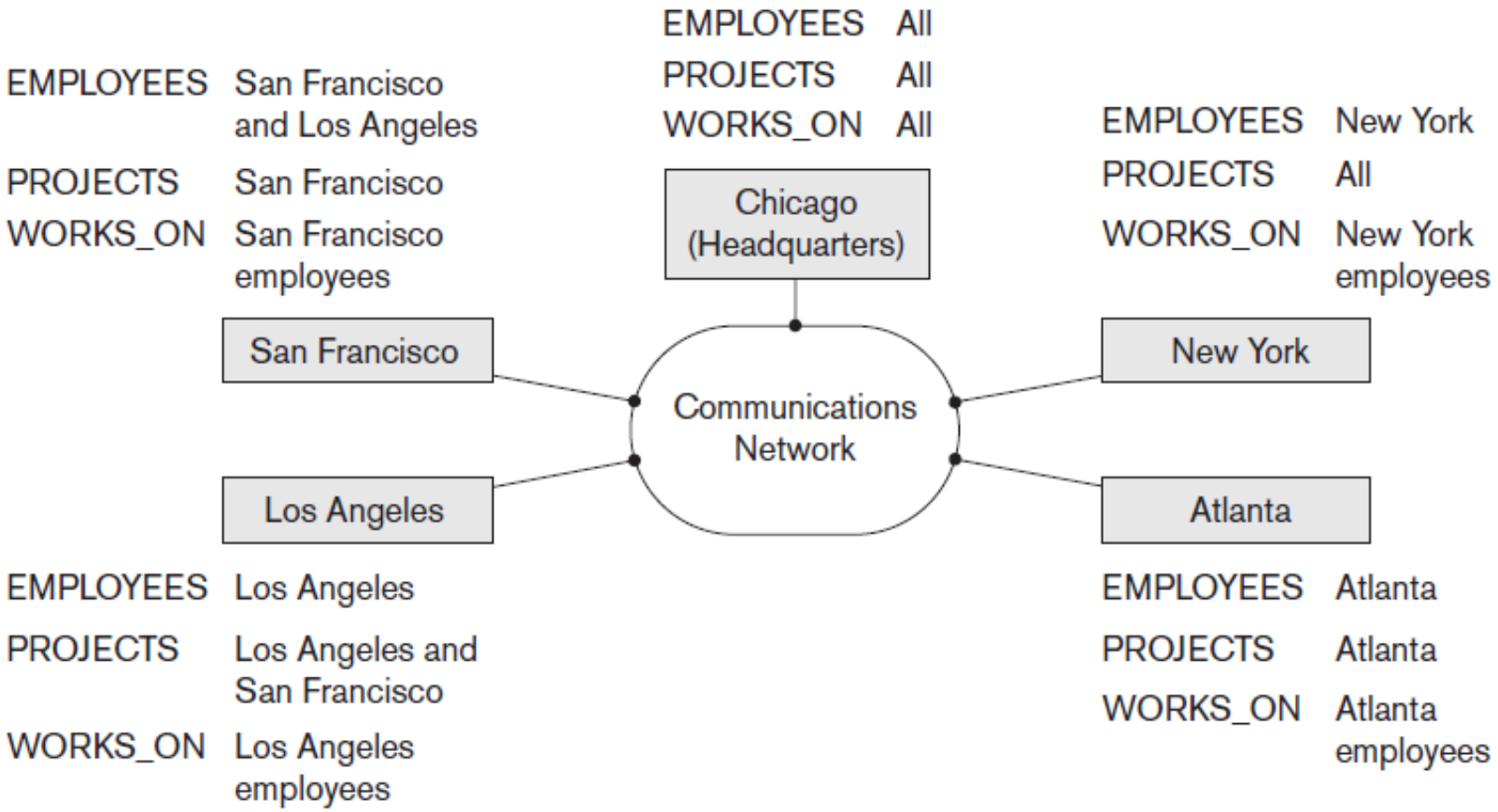
Heterogeneous DDBMS

- Different database systems

Federated or Multidatabase Systems

- Map different systems into one database systems

Distributed Databases



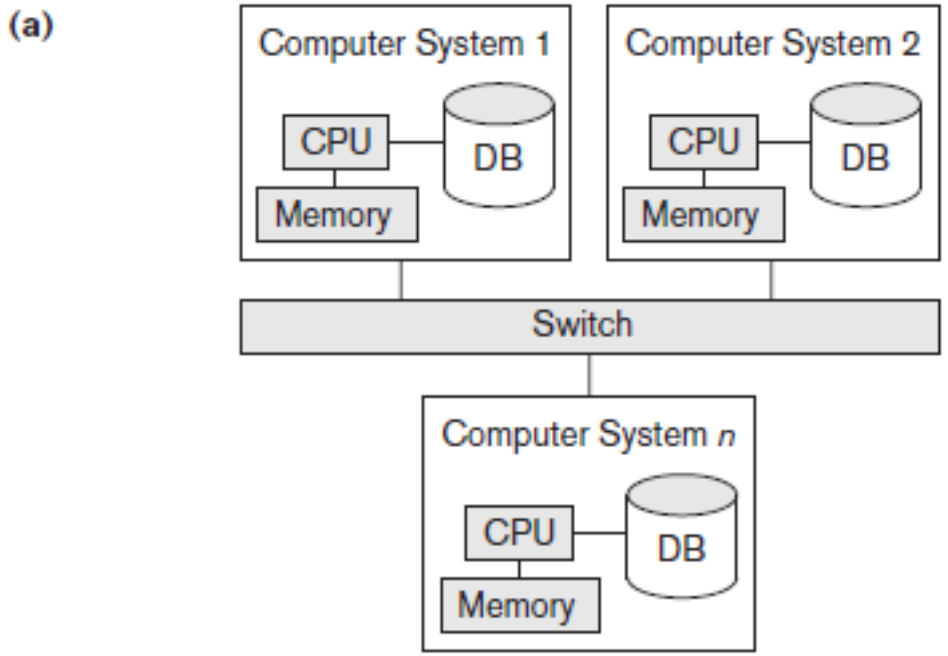
Data distribution and replication among distributed databases

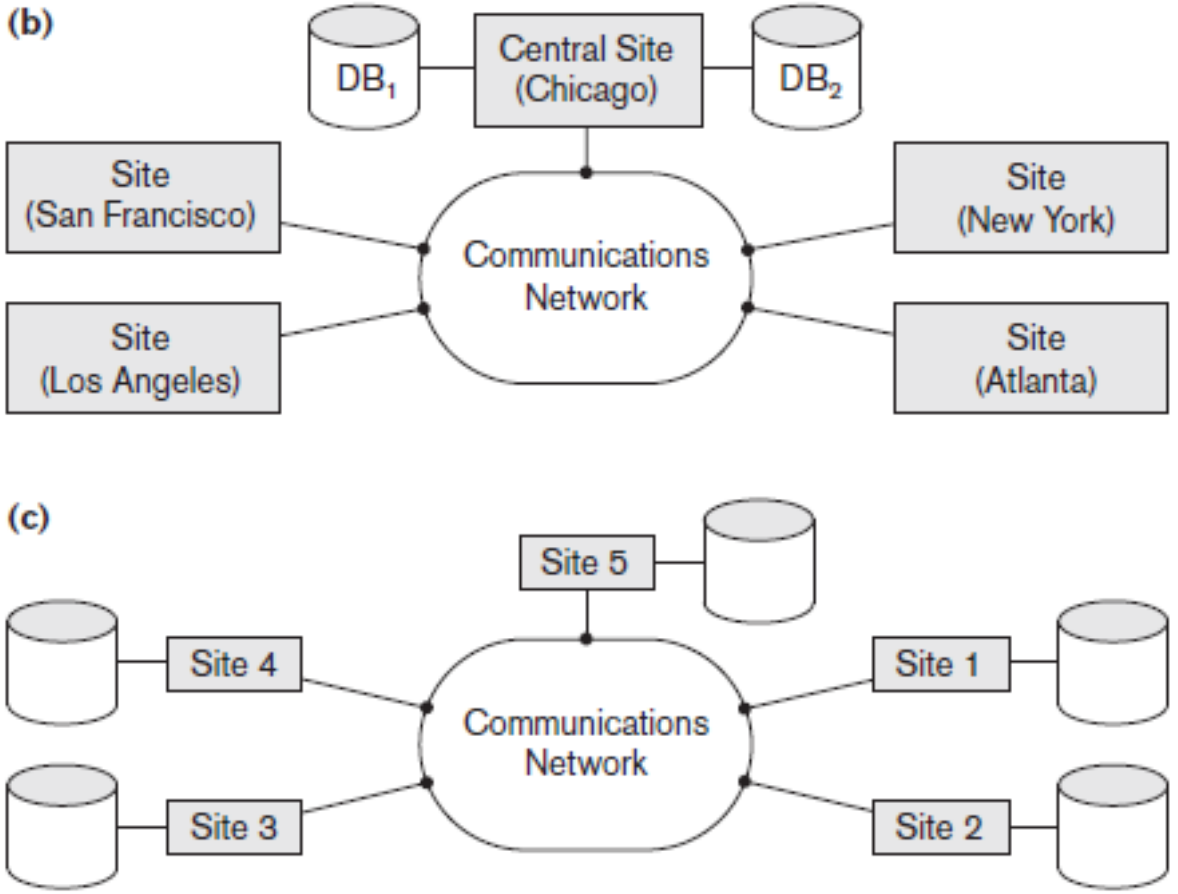
Transparency

Transparency
Hiding implementation details from the end user

- Data organization transparency
  - Location transparency
  - Naming transparency
- Replication transparency
- Fragmentation transparency
  - Horizontal fragmentation
  - Vertical fragmentation
- Design transparency
- Execution transparency

Database System Architectures





(a)Shared-nothing architecture
(b)A networked architecture with a centralized database at one of the sites
(c) A truly distributed database architecture

Schema Architecture of Distributed Databases

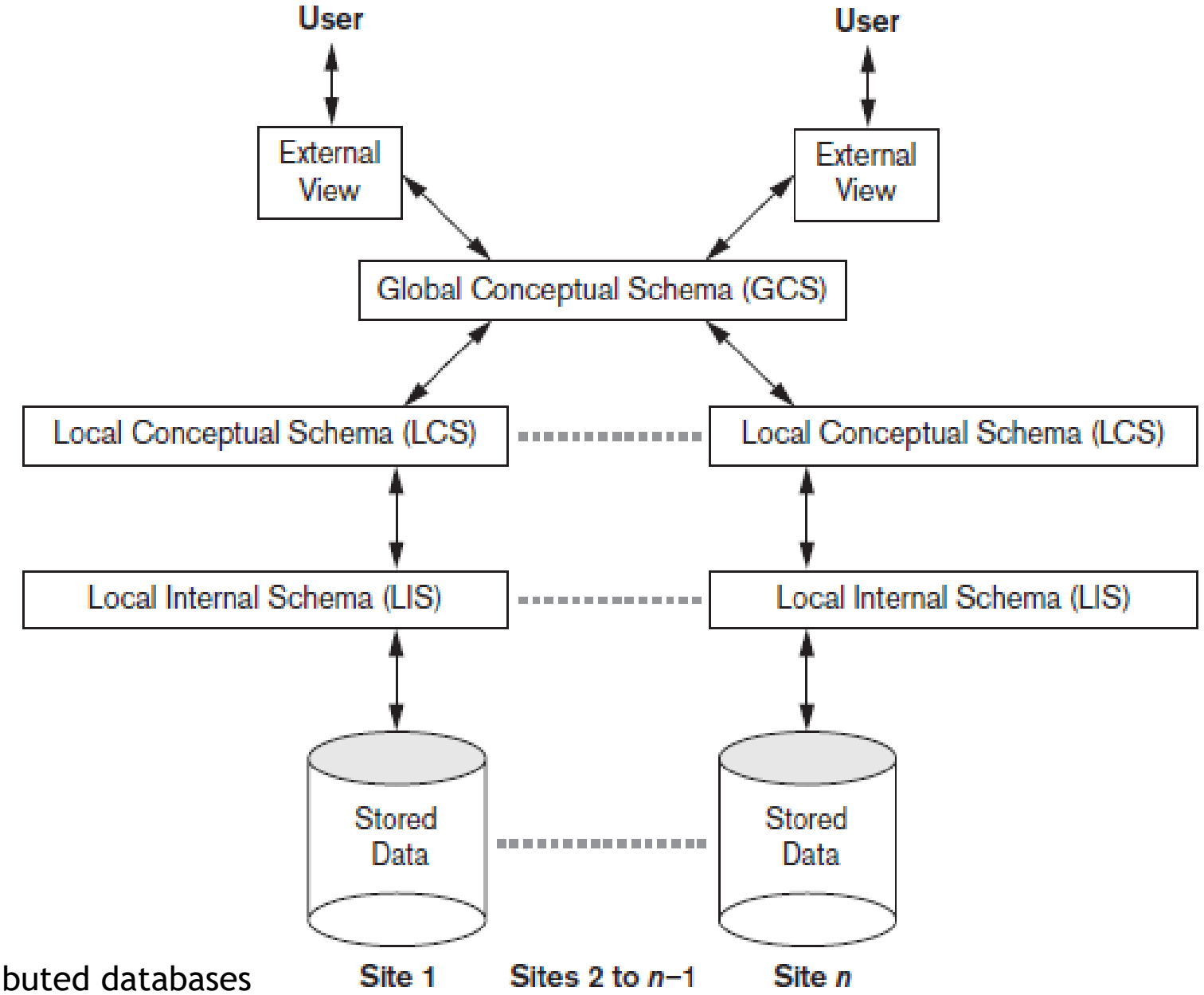


Figure 3 Schema architecture of distributed databases

Advantages of Distributed Databases

Improved ease and flexibility of application development

- Development at geographically dispersed sites

Increased availability

- Isolate faults to their site of origin

Improved performance

- Data localization

Easier expansion via scalability

- Easier than in non-distributed systems

Business Intelligence Systems

business intelligence

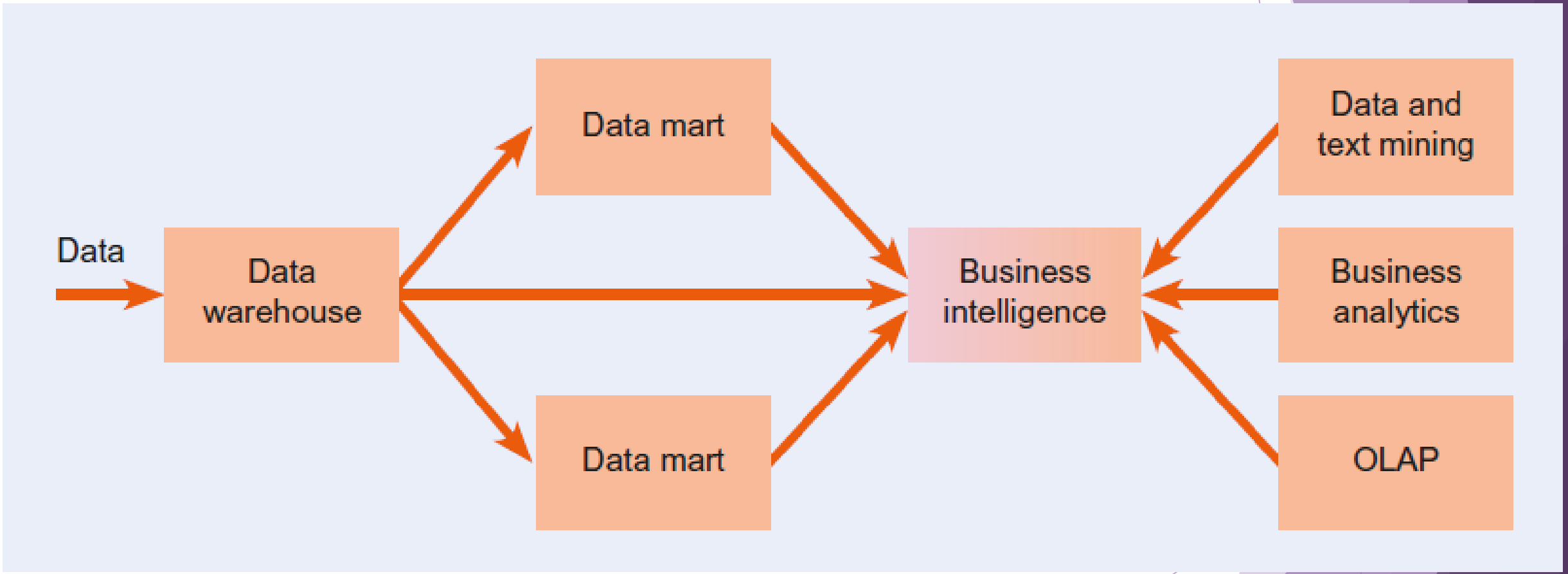
Business intelligence (BI) systems are needed due to

- vast amounts of data in IS
- the need to extract useful information
- in patterns and trends, and
- present the information in an understandable way
- to decision makers

business intelligence (BI) systems

- similarities with MIS (i.e. DSS, EIS)
- provide information at a strategic level
- in large organizations with large data sets
- indirect support for decisions
  - DSS = decision specific orientation
- Data is gathered from various sources
- held in a special database repository called a data warehouse

business intelligence (BI) systems



business intelligence (BI) systems

Data marts

- repositories of data focused on departmental or subject areas

Data mining

- a type of analysis
- aims to identify patterns in the data
- used to predict future behaviour

Business analytic tools

used to conduct analysis of the data warehouse data

using reporting and querying tools

Data Warehousing

data warehouse

large database systems containing

current and historical data

that can be analysed

to produce information

to support organisational decision making

“A subject-oriented, integrated, time variant, and non-volatile collection of data in support of management’s decision-making process.”

William Inmon (2005)

data warehouse

Integrated

- data is collected from diverse sources within an organisation
- enable integrated analysis

Subject-oriented

- e.g. customers and products
- only contain relevant information for decision support for that subject

data warehouse

Non-volatile

- data is transferred from operational IS
  - e.g. sales order processing systems
- into a data warehouse where the information is static – it is not updated.
- old, obsolete data is removed from the database and
- the ‘changed’ data recorded as new data

data warehouse

Support of management's decision-making process

- a data warehouse does not contain 'current data'
- data is not updated
- changes are reported
- a data warehouse holds data over points

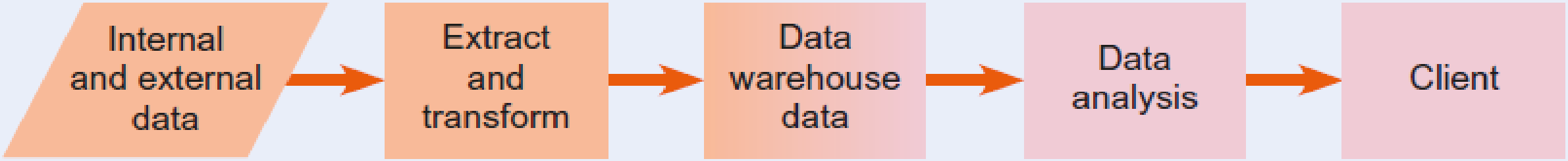
data marts

- a smaller, departmental version of a data warehouse
- easier to manage than a company-scale data warehouse
- focus on one department

A data warehouse can consist of many data marts supporting different (smaller) operations.

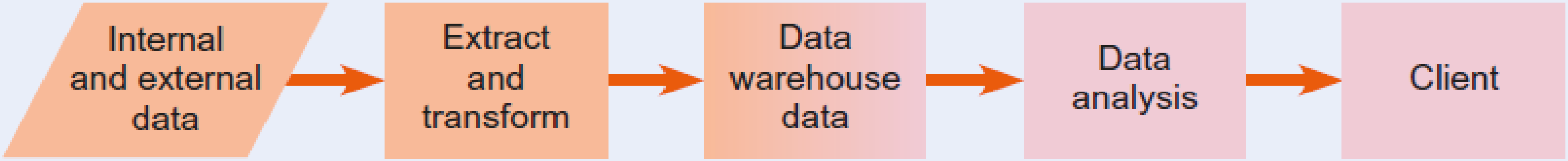
3 main processes

1. information from internal and external sources,
   such as operational systems,
   which record sales or transactions with customers



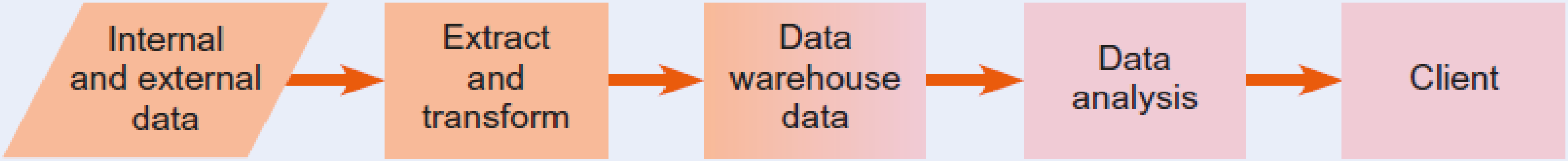
3 main processes

2. data can come from sources such as
   - data from customer transactions
     - EPOS (electronic point-of-sale)
   - ERP systems
   - EDI (electronic data interchange) systems
   - legacy databases
     - holding historical data



3 main processes

3. data is then transformed into a suitable form to the data warehouse ETL software



Extraction, Transformation & Load (ETL)

ETL software

- extracts data from databases
- transforms data into a suitable format
- loads data into the data warehouse
- processes the source data according to business rules
- rules include definitions of data attributes and calculation methods
- rules can be applied to data in a consistent way

architecture

Data mart centric

Virtual, distributed, federated

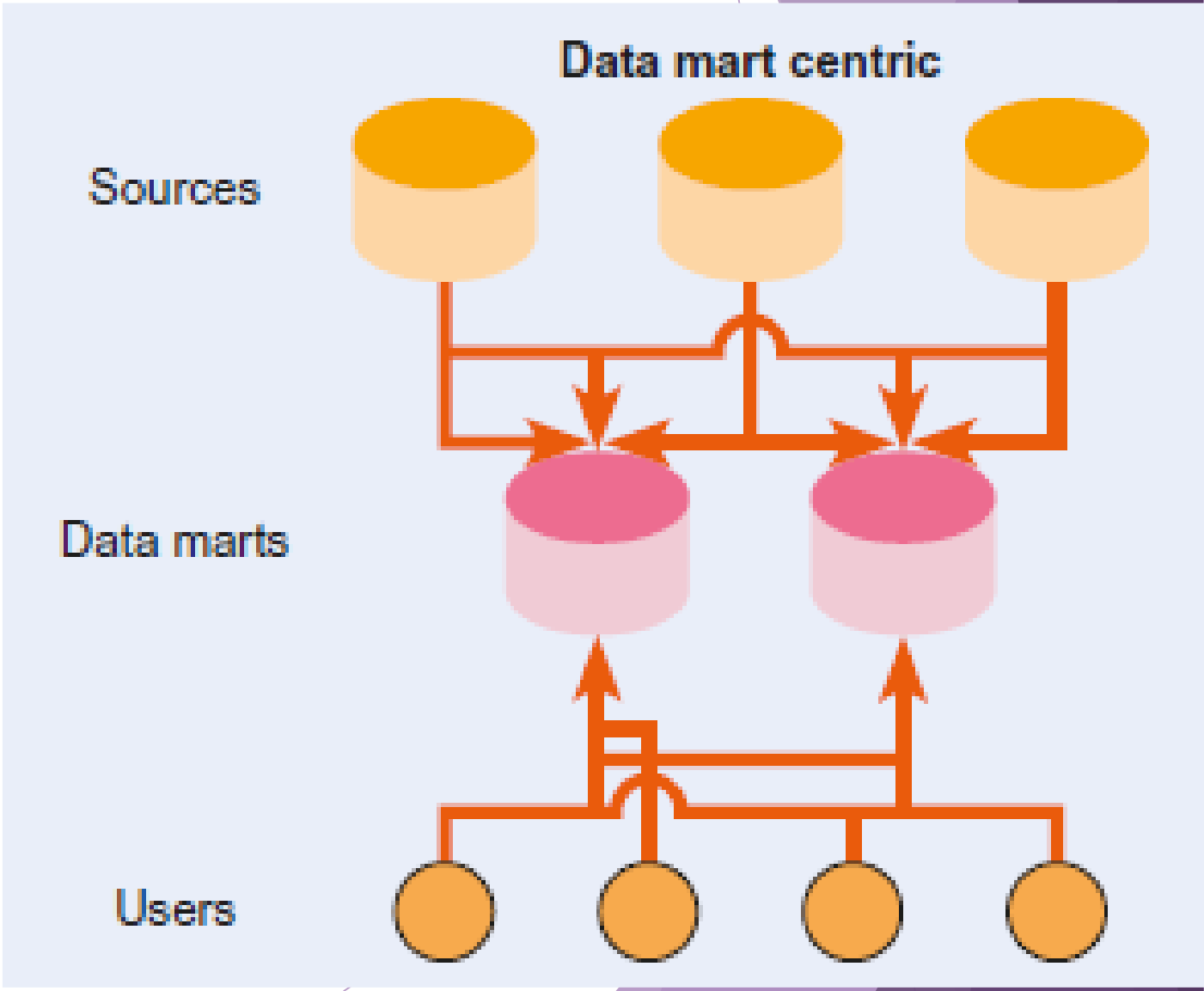
Hub-and-spoke data warehouse

Enterprise data warehouse

architecture

Data mart centric

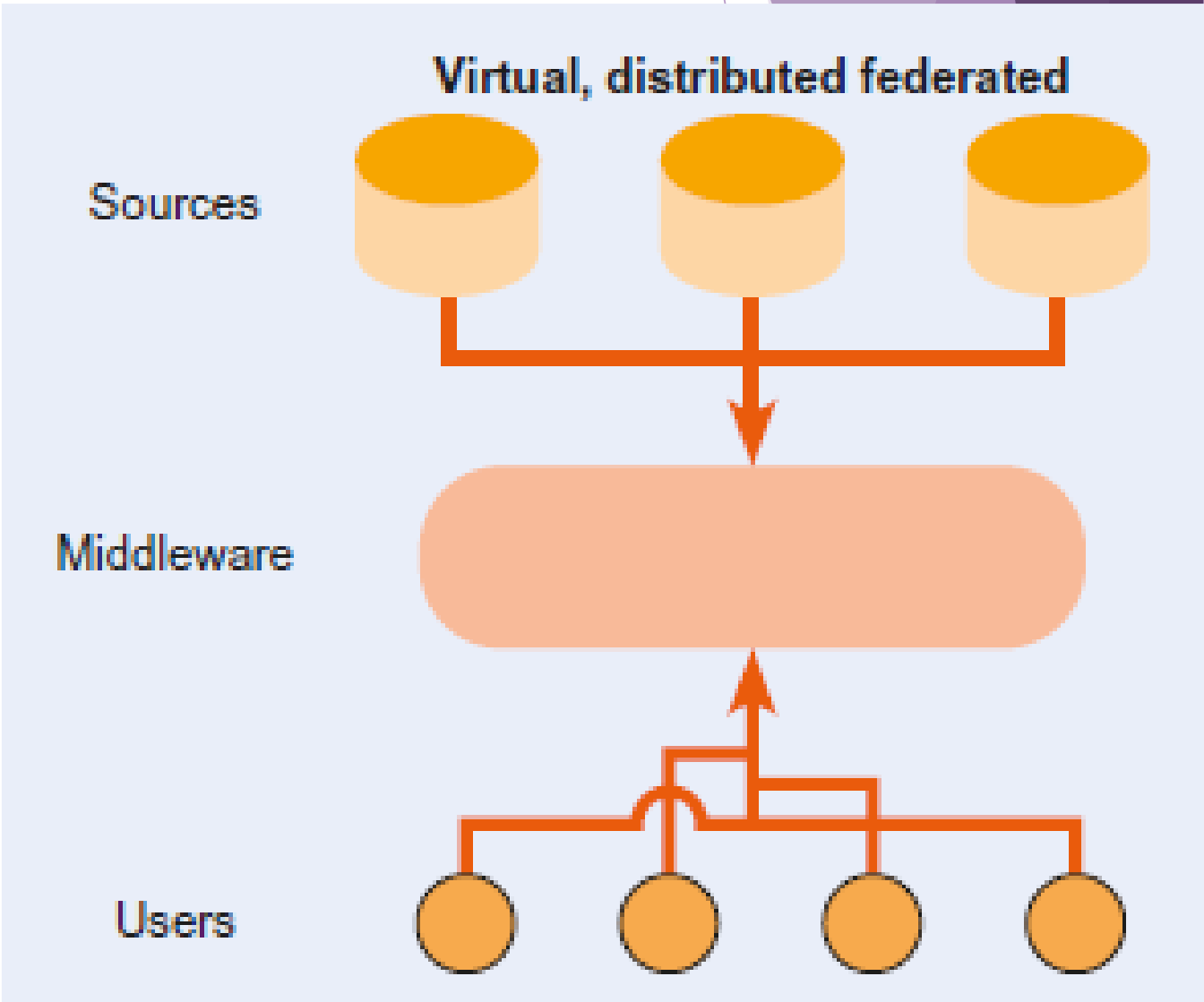
- Data is linked to users through independent data marts
- easy implementation
- lacks an enterprise-wide view of the organisation’s data and
- can lead to inconsistencies of data across data marts



architecture

Virtual, distributed, federated

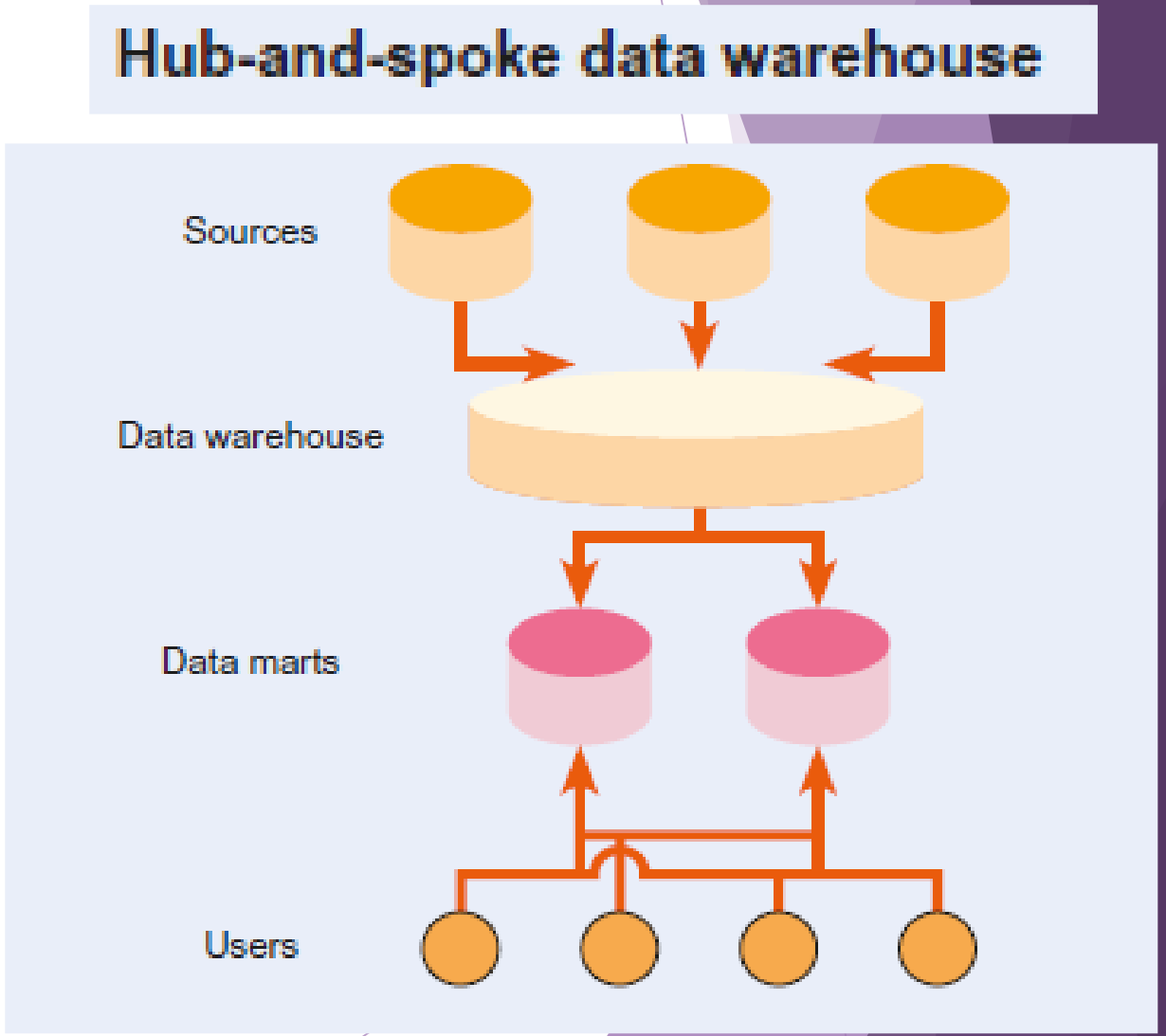
- uses middleware
- links users directly to data sources
- may be performance and data quality issues using this approach



architecture

Hub-and-spoke data warehouse

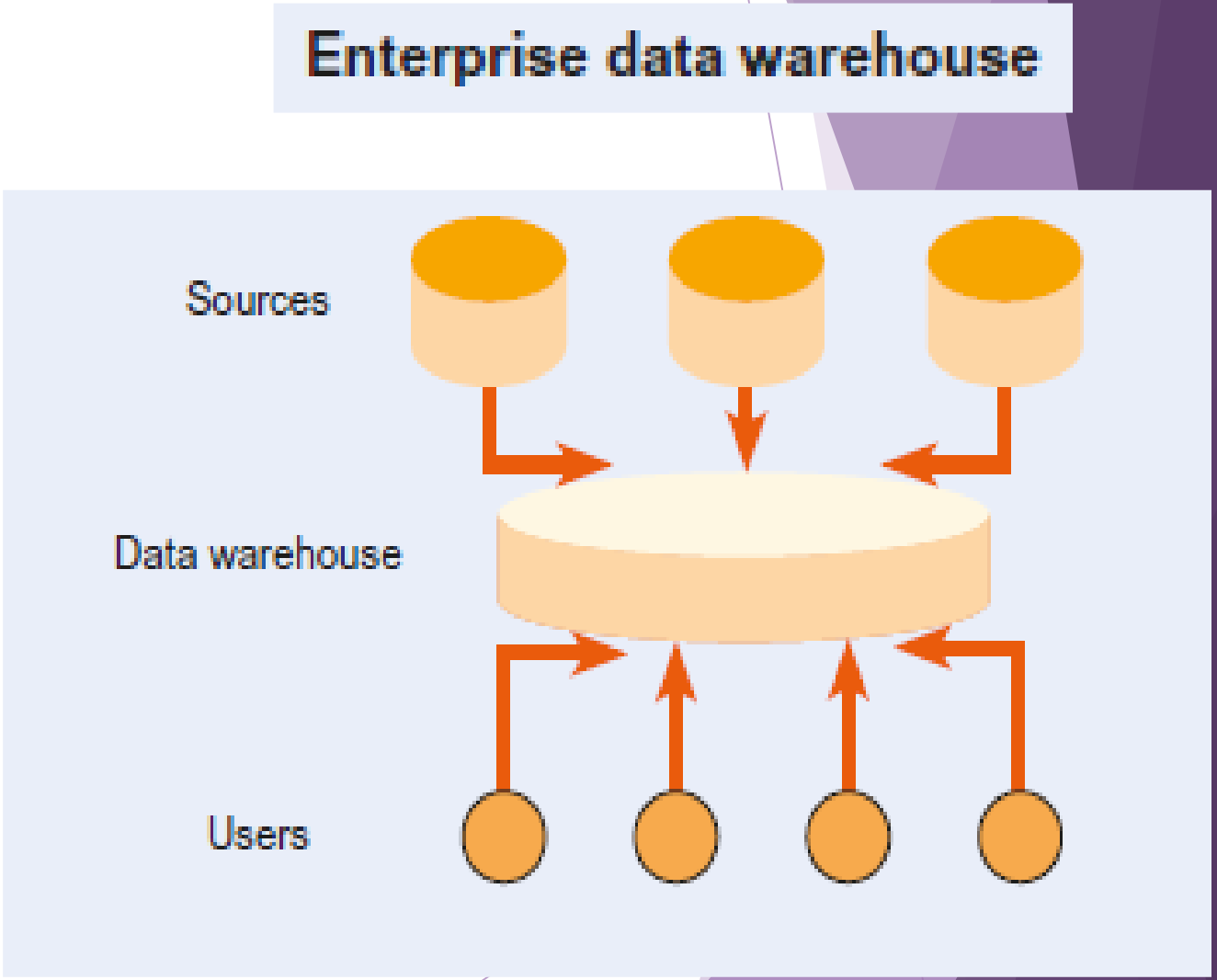
- links users to dependent data marts, linked to an enterprise data warehouse, linked to organizational data sources
- provides the ability for customisation to user needs
- can lead to redundancy of data and
- relatively high operational costs of running both data marts and the data warehouse



architecture

Enterprise data warehouse

- This links users directly to a data warehouse which is linked
- in turn to data sources. This provides a single and thus consistent view of the data across
- the enterprise. It does require leadership from senior management in order to implement
- an enterprise-wide solution.



real-time data warehousing (RDW)

also known as active data warehousing (ADW)

- normal data warehouses are updated periodically, e.g. weekly
- RDW load and process data as events happen
- used for realtime operational decisions
  - e.g. process flow performance

disadvantages

- technical difficulty in extracting and transforming real-time data.
- inconsistency in results of queries (constantly updated)
  - e.g. report statistics differ for same day

no lesson Thursday

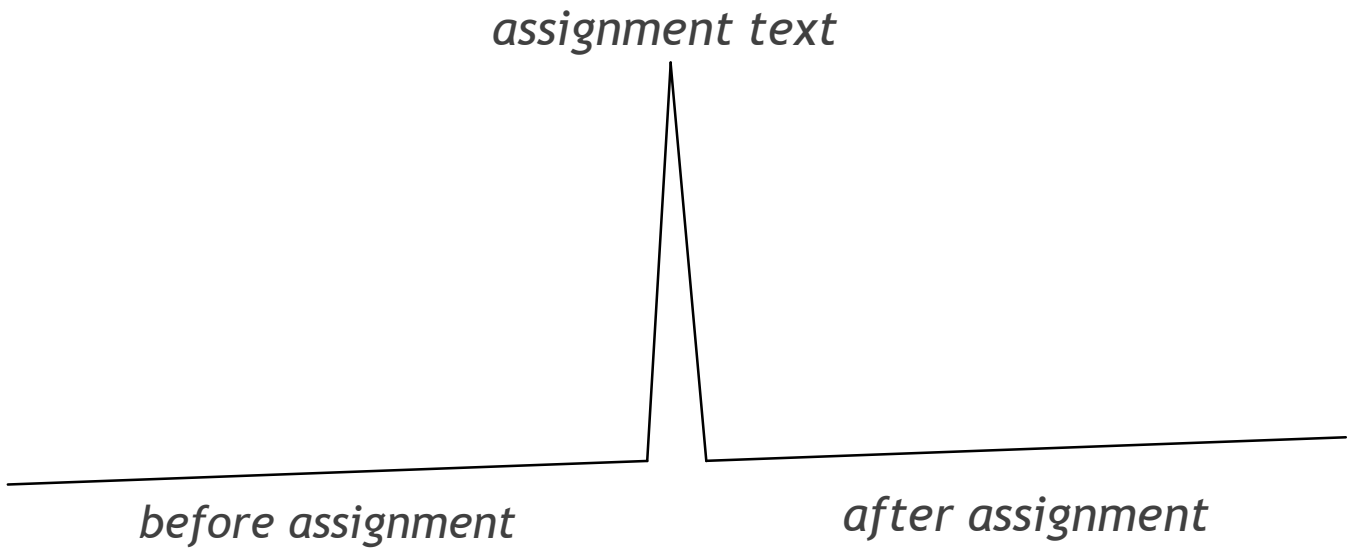
all lessons on Thursday are cancelled

In Thailand, university students need to attend at least **25** *1.5 hour* lessons
(you can't miss 6 lessons including holidays)
or they can't be deemed to have passed a course

THE LAST DAY OF WITHDRAWAL COURSE
WITH RECEIVING GRADE 'W' is FEB 7th, 2025.

Friday, week 11 (this is week 8)

student's level of English



can you see the issue?

Assignment #1

- Who wrote the text?
  - If a students didn't write it then it is not the students work.
- Marks are awarded for evidence of understanding
- If you do not understand what is written then it is zero
- AI is not the correct answer
  - it proves a lack of understanding
  - it is a form of plagiarism
  - therefore it is cheating (if text copied)

Assignment #1 – if you did the work

- Who wrote the text?
  - not students – therefore zero

but,

some students have done some good work

- if all the group has done work and not cheated then let me know
- you will be interviewed to check **you** wrote & understand it

Assignment #1 – if you have made a mistake

- If you do not understand the text
  - your score is zero (0%)

but,

you may have a second chance

- you can receive a lower mark
- for a revised assignment

part 2

Data Mining

Data Mining

the computer-assisted process of sifting through and analyzing vast amounts of data to extract hidden patterns and meaning and to discover new knowledge

Data Mining

- a process that uses statistical, mathematical, artificial intelligence and other techniques
- to extract useful information from large databases.
- open questions such as ‘What are the characteristics of the top 20 per cent of our customers?’
- understanding customers better

Identifying associations

- establishing relationships about items at a particular time,

Identifying sequences

- This involves showing the sequence in which actions occur
- e.g. path or click-stream analysis of a web site

Data Mining

Regression analysis

- develops a formula to fit patterns in the data
- formula is applied to data sets to predict future trends

Modelling

- involves using forecasting and regression analysis to predict sales
  - (e.g. using sales histories to forecast future sales

Clustering

- finding groups of facts that were previously unknown,
  - e.g. identifying new market segments of customers
  - detecting e-commerce fraud.

Cluster analysis

- sorts attributes such as people or events into groups (i.e. clusters)
- degree of association between items is strong and across clusters is weak
- using methods such as statistical techniques, neural networks and genetic algorithms.

division: one cluster and breaking this into separate clusters

agglomerative: separate clusters and joining the clusters together to make new clusters

Classification analysis

- a statistical pattern recognition process
  - e.g. can use neural networks or decision trees
- applied to data sets with more than just numerical data
- analysing historical data into patterns to predict future behaviour,
  - e.g. identifying groups of web site users with similar visitor patterns.
- often used to classify new data into previously defined classes
- by learning the pattern of the data

Classification is distinct from clustering in that at least some of the classes are previously known.

Data Mining

Example applications:

- Retail stores use it to predict future purchase patterns to help them choose which products to stock for the future
- A phone company identifying customers with large bills, who were really small businesses trying to pay the cheaper residential rate
- A coach in the Gymnastics Federation used it to discover what long-term factors contributed to athletes' performance

Text Mining

Text mining

the application of data mining to text files

- normally unstructured content
- finds previously hidden patterns in text within and between documents

text mining system

- queries and finds text within a variety of documents
- variety of document formats e.g. text, pdf
- variety of document platforms such as emails, web pages

information extraction system

- then analyses and processes the text
- extracts information from unstructured data into a structured format
- allows data mining methods such as cluster analysis

Web Mining

data mining information from the web

web content mining

- extraction of information specifically from web pages
- reading and analysing data from web pages

web structure mining

- to analyse information from the links within the web documents
  - e.g. number of external links to a document
  - used to rank web pages for search engines

Web Mining

web usage mining

extracts information from usage data

web page visits and transactions

clickstream analysis

- analysis of user behaviour visiting web pages
- used to target advertisements and marketing campaigns
- to provide information for cross-marketing of alternative products.
- use information gathered on previous customer behaviour
- to recommend products and services

Business Analysis

Business analytics (BA)

describe various approaches to data driven analysis

BA tools cover a wide range of techniques

- includes data mining, text mining and web mining
- including reporting tools (OLAP) and visualisation tools (dashboards)

- BA reporting tools include
  - OLAP and cube analysis
- BA visualisation tools include
  - dashboards and scorecards
- see following details:

Example: Statistical Analytics software (SAS) covers :

Statistics

- use statistical data analysis to help fact-based decisions.

Data and text mining

- builds models over whole enterprise.

Data visualisation

- allows users to interact with graphs

Content categorisation –

- categorises content, triggers business processes

Forecasting and econometrics –

- analyse and predict outcomes based on historical patterns and
- apply statistical methods to economic data, problems and trends.

Operations research

- applies techniques such as optimisation, scheduling and simulation

Model management and deployment

- streamline the process of creating, managing and deploying analytical models.

Quality improvement

- identifies, monitors and measures quality processes over time.

OLAP

Online analytical processing (OLAP)

- refers to the ability to analyse in real time
- large data sets stored in data warehouses
- ‘Online’
  - indicates that users can formulate their own queries

Dr E. Codd (originator of OLAP), defines it as

- the dynamic synthesis, analysis and consolidation of large volumes of multidimensional data

visualization tools

Applications such as spreadsheets, dashboards, scorecards and geographical information systems can be utilised as visualization tools.

Spreadsheets

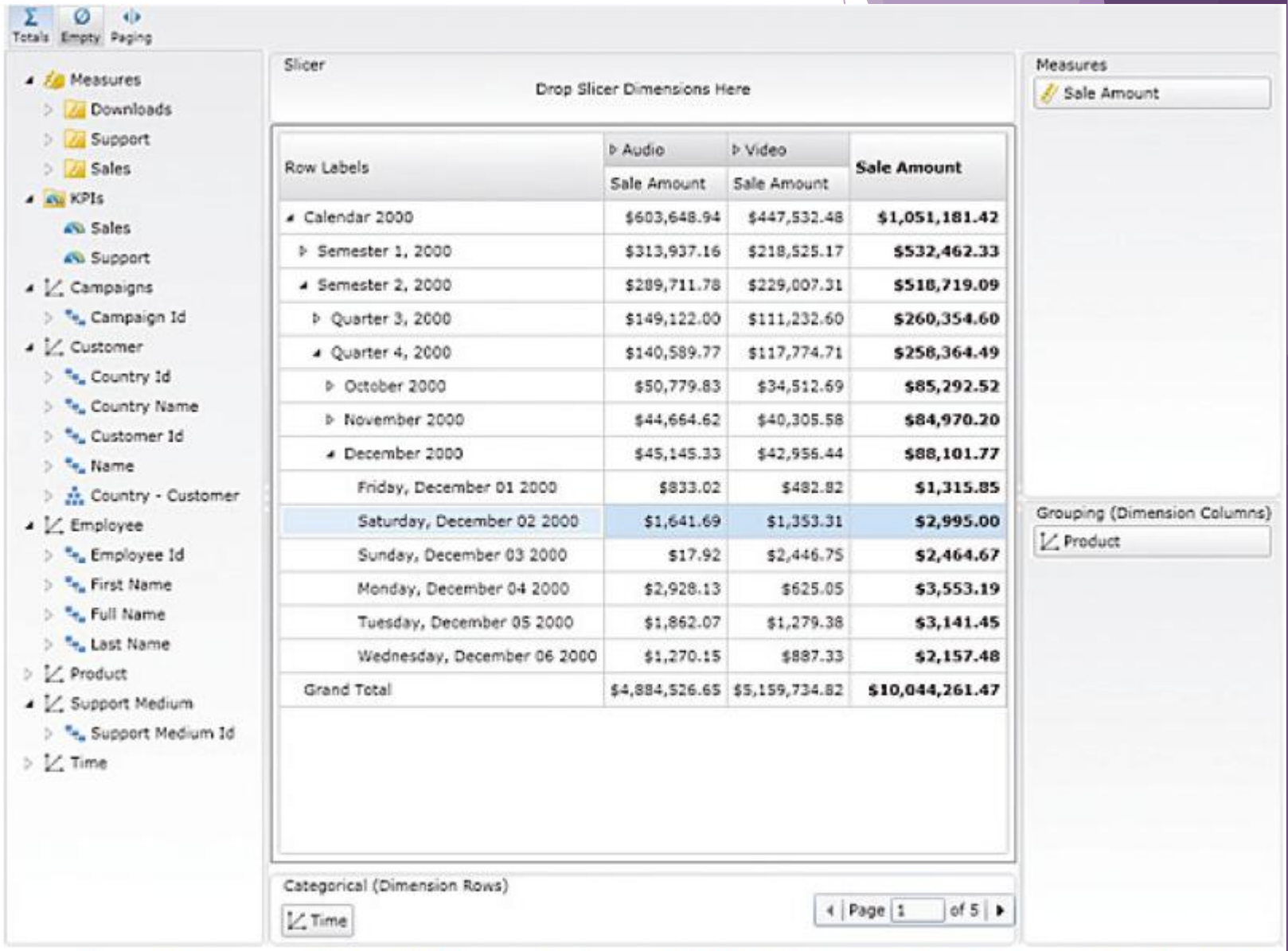
- create different charts
- updated automatically in response to changes in data
- statistical and forecasting capabilities
- useful for providing graphical displays of trends
  - e.g. sales

Dashboards

- a graphical display on the computer
- shows a range of statistics from enterprise-wide software applications
- real-time information.
- includes graphical images such as meters, bar graphs, trace plots, text fields

Scorecards

- provide a summary of performance over a period of time
- examine data from the balanced scorecard perspectives of
  - financial,
  - customer,
  - business process and
  - learning and growth



Source: http://www.dashboardinsight.com/articles/digital-dashboards/fundamentals/a-closer-look-at-scorecards-and-dashboards.aspx

Business activity monitoring (BAM)

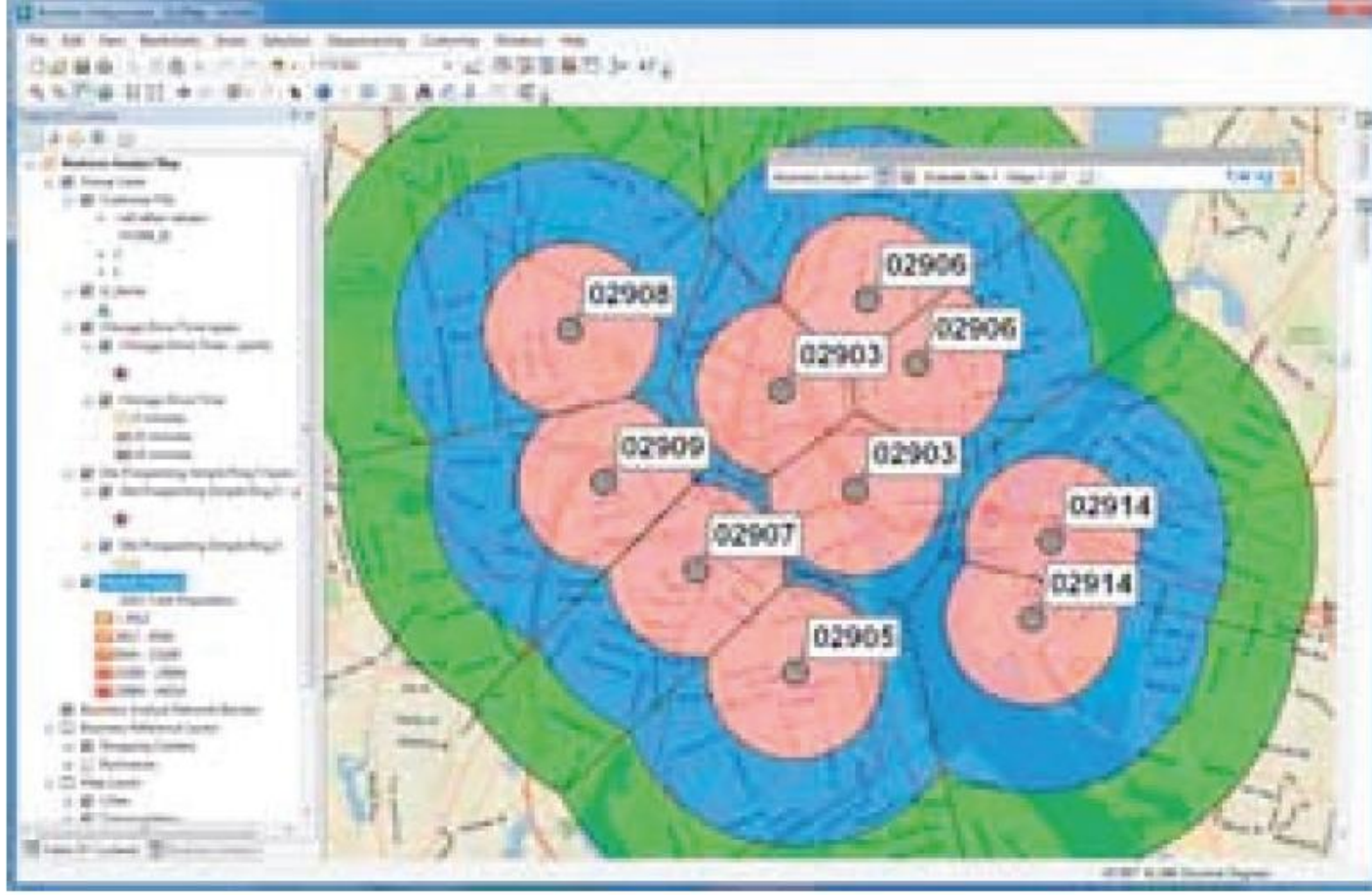
software designed to

- monitor, capture and analyse business performance data
- in real time
- present them visually
- in order that rapid and effective decisions can be taken

Source: http://www.oracle.com/technetwork/middleware/bam/loanflow-098007.html

geographical information system (GIS)

- use maps to display information
- used for performance analysis by marketing staff
- Performance shown by colour-coding
- show demand of customers
- different demographics e.g. average disposable income



Source: http://www.esri.com/news/arcuser/0611/localization-not-location.html

Marketing analysts can review to correct problems in areas of underperformance

- e.g. opportunity in the south of the area to open a new branch

Big Data

big data

- a term created in 2001 by Doug Laney
- describes the growth and availability of data
- structured and unstructured data
- made analyses of data become much easier

3 V's of big date (SAS.com)

- volume - amount of data
- velocity - how quickly data can be transferred
- variety - many types of format

big data

help the store with

- item locations,
- the ordering process, and
- knowing when to expect items to sell

small things add up to large gains

big data makes

- analysis more robust
- but more complicated

Thank you!

any questions?

Preparation for next Thursday

Database

- https://sqlitebrowser.org/dl/
- DB Browser for SQLite - Standard installer for 64-bit Windows