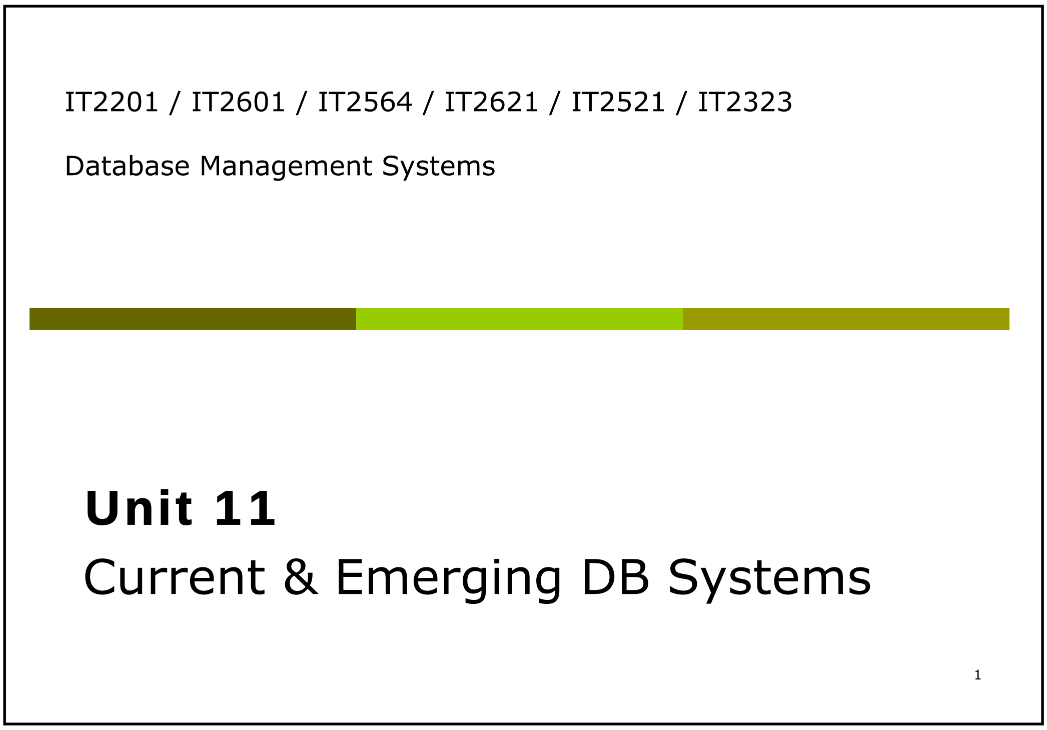
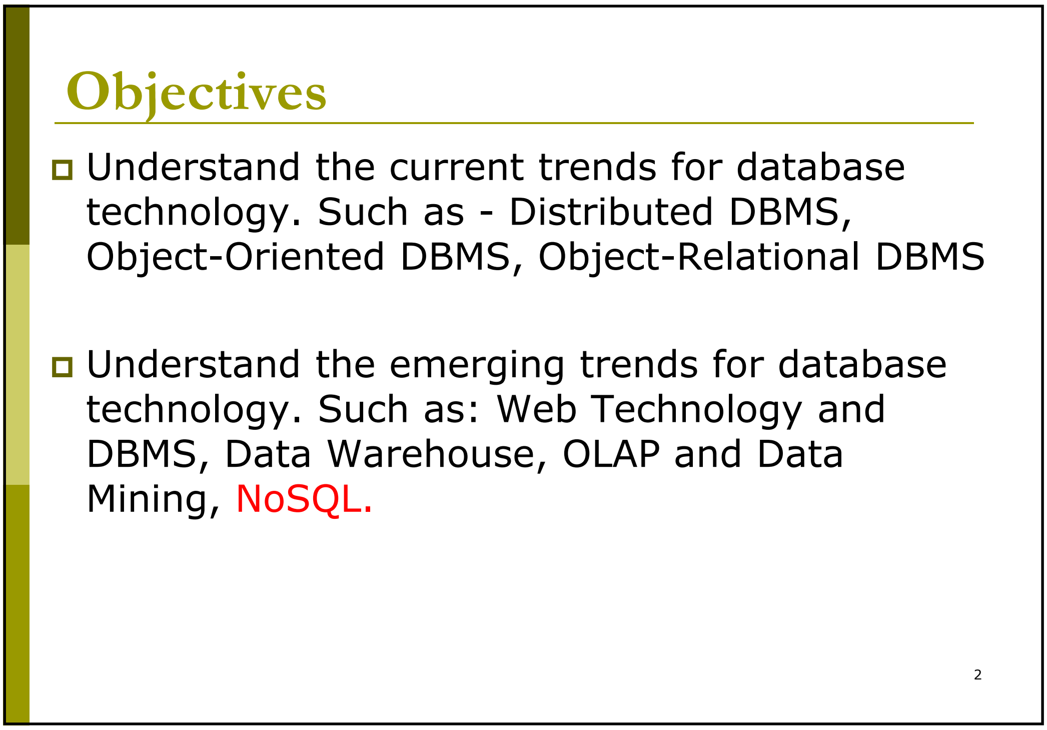


IT2201 / IT2601 / IT2564 / IT2621 / IT2521 / IT2323

Database Management Systems

Unit 11
Current & Emerging DB Systems

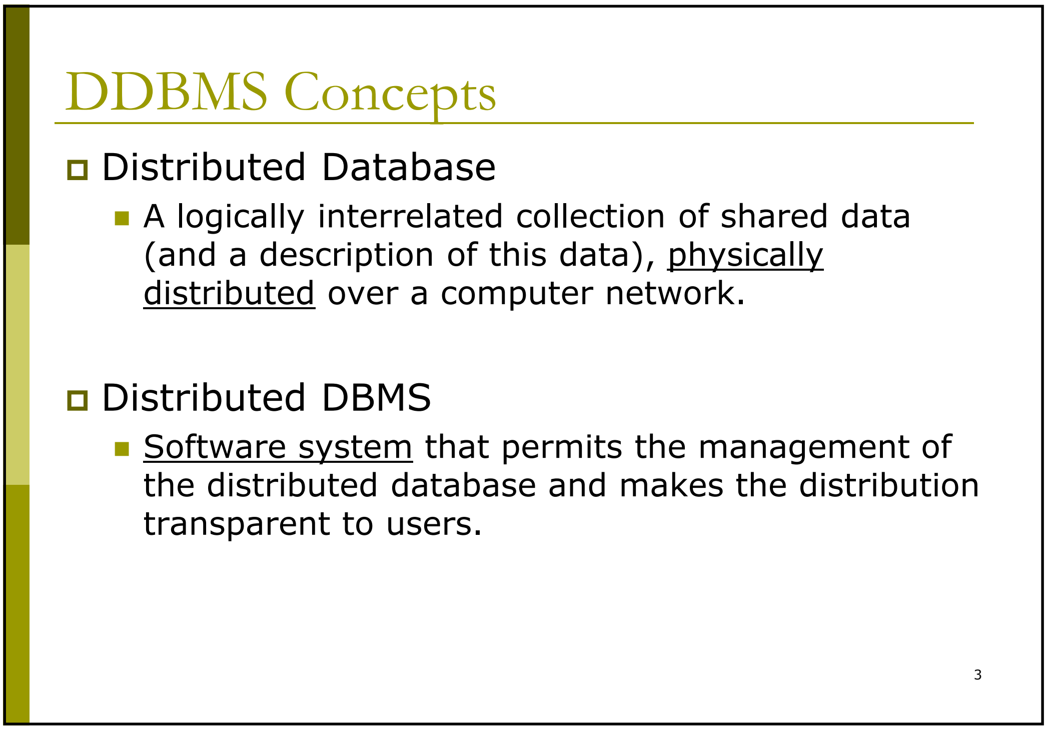
1



Objectives

- Understand the current trends for database technology. Such as - Distributed DBMS, Object-Oriented DBMS, Object-Relational DBMS
- Understand the emerging trends for database technology. Such as: Web Technology and DBMS, Data Warehouse, OLAP and Data Mining, NoSQL.

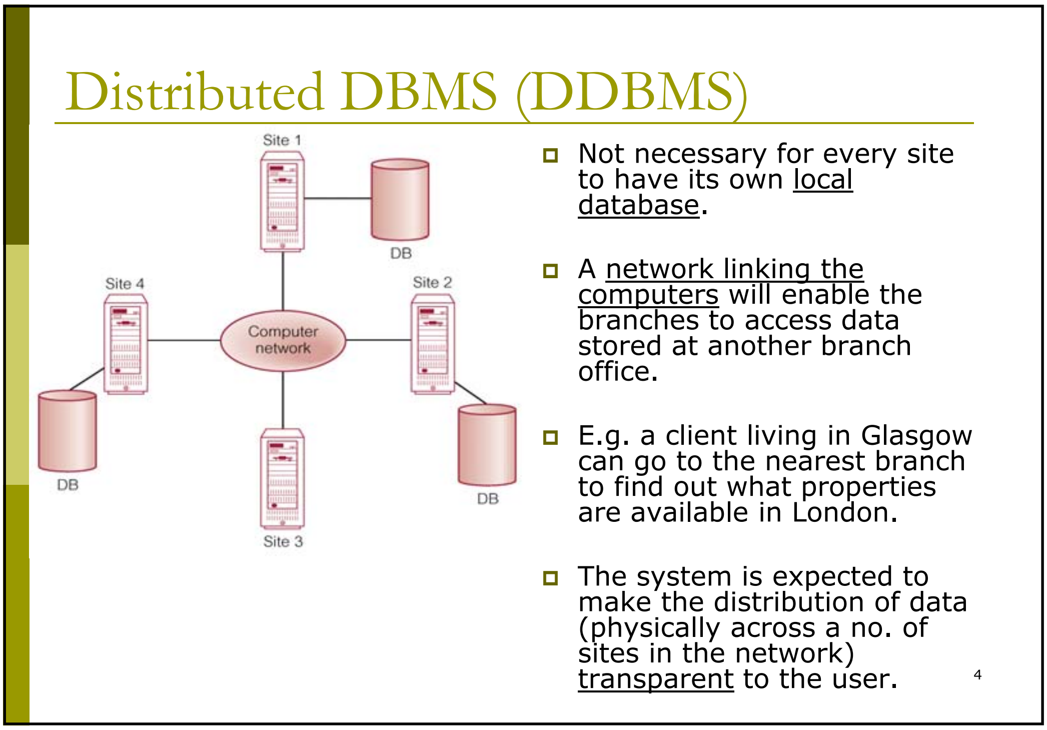
2



DDBMS Concepts

- Distributed Database
  - A logically interrelated collection of shared data (and a description of this data), <u>physically distributed</u> over a computer network.
- Distributed DBMS
  - <u>Software system</u> that permits the management of the distributed database and makes the distribution transparent to users.

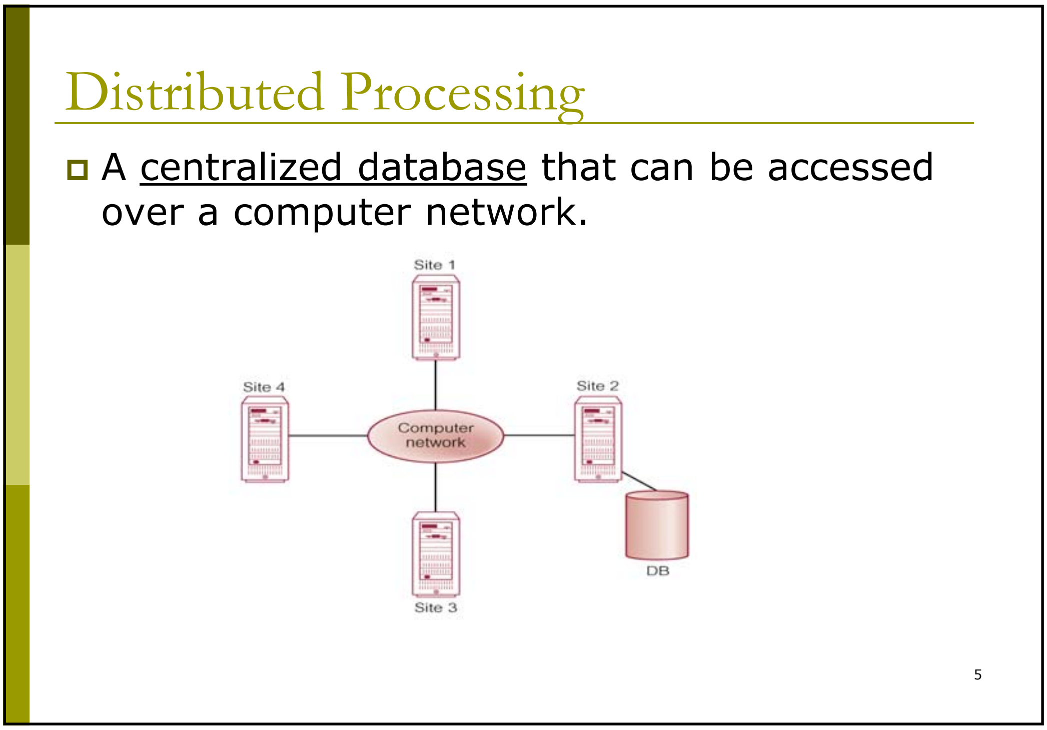
3



Distributed DBMS (DDBMS)

- Not necessary for every site to have its own <u>local database</u>.
- A <u>network linking the computers</u> will enable the branches to access data stored at another branch office.
- E.g. a client living in Glasgow can go to the nearest branch to find out what properties are available in London.
- The system is expected to make the distribution of data (physically across a no. of sites in the network) <u>transparent</u> to the user.

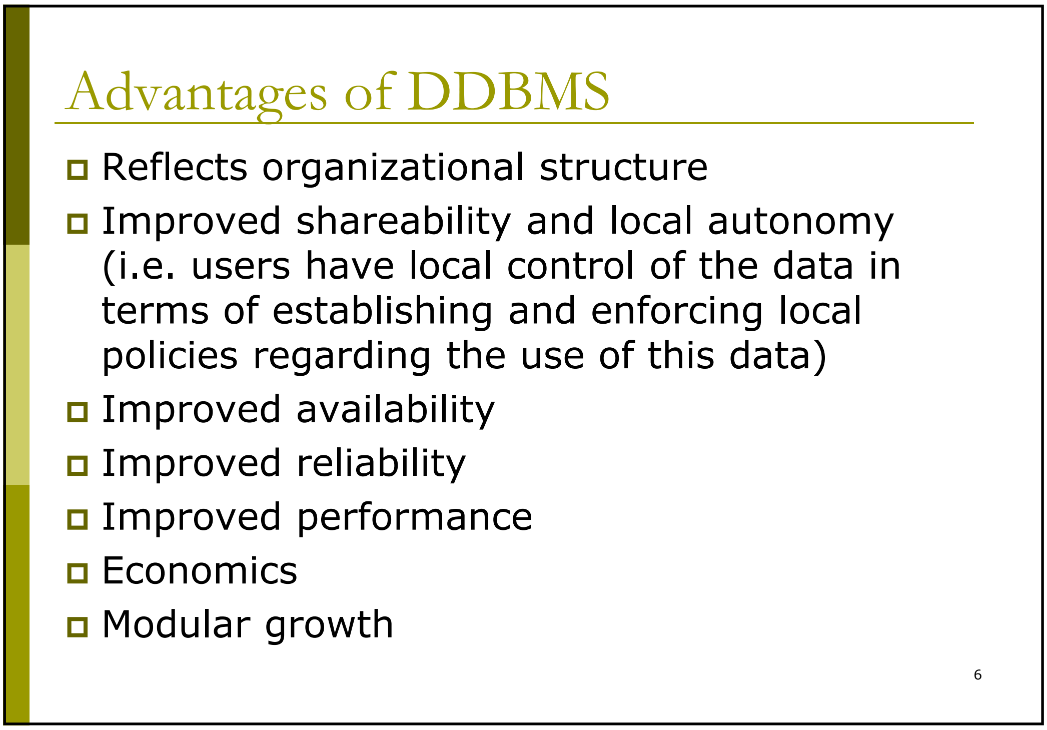
4



Distributed Processing

- A centralized database that can be accessed over a computer network.

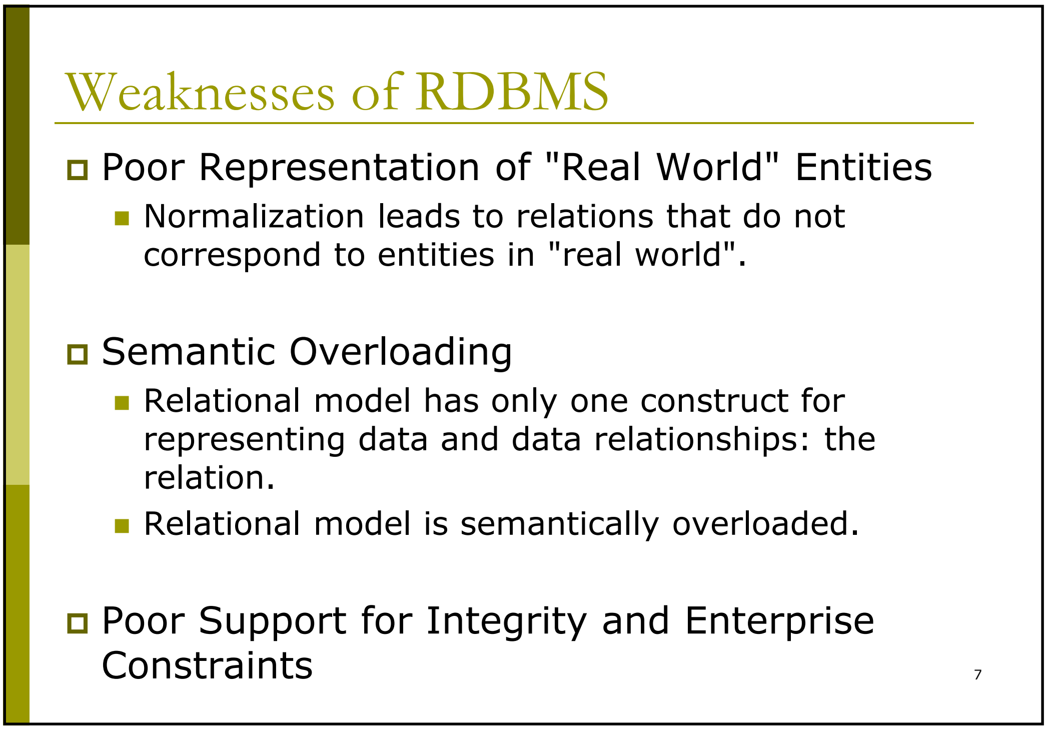
5



Advantages of DDBMS

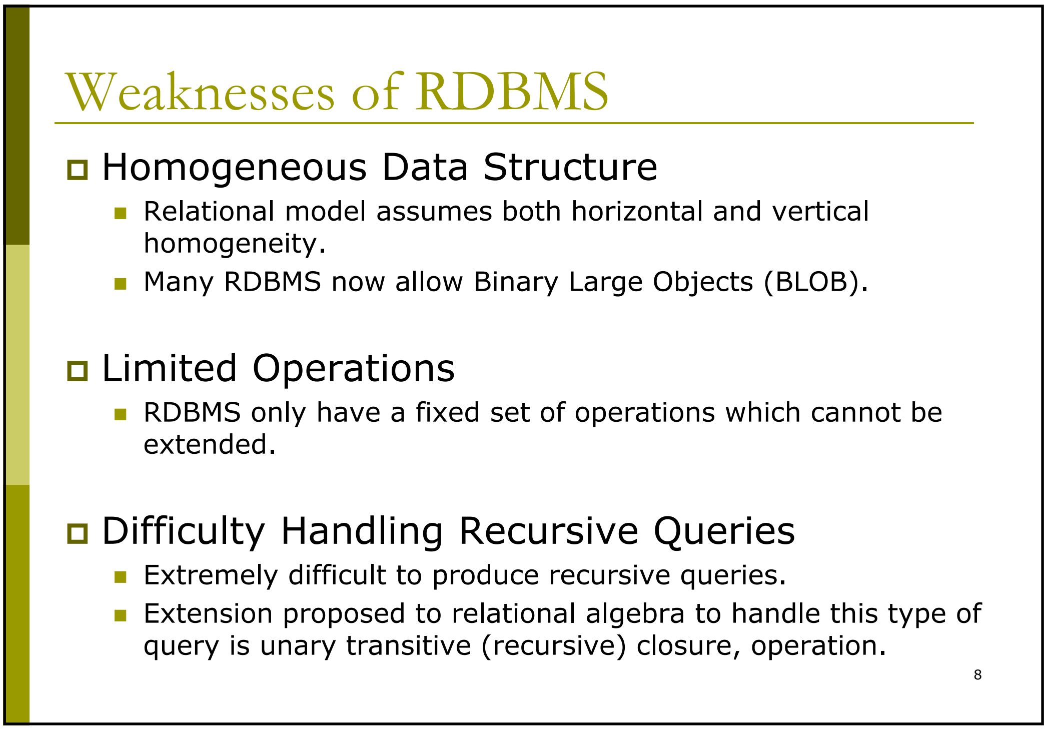
- Reflects organizational structure
- Improved shareability and local autonomy (i.e. users have local control of the data in terms of establishing and enforcing local policies regarding the use of this data)
- Improved availability
- Improved reliability
- Improved performance
- Economics
- Modular growth

6



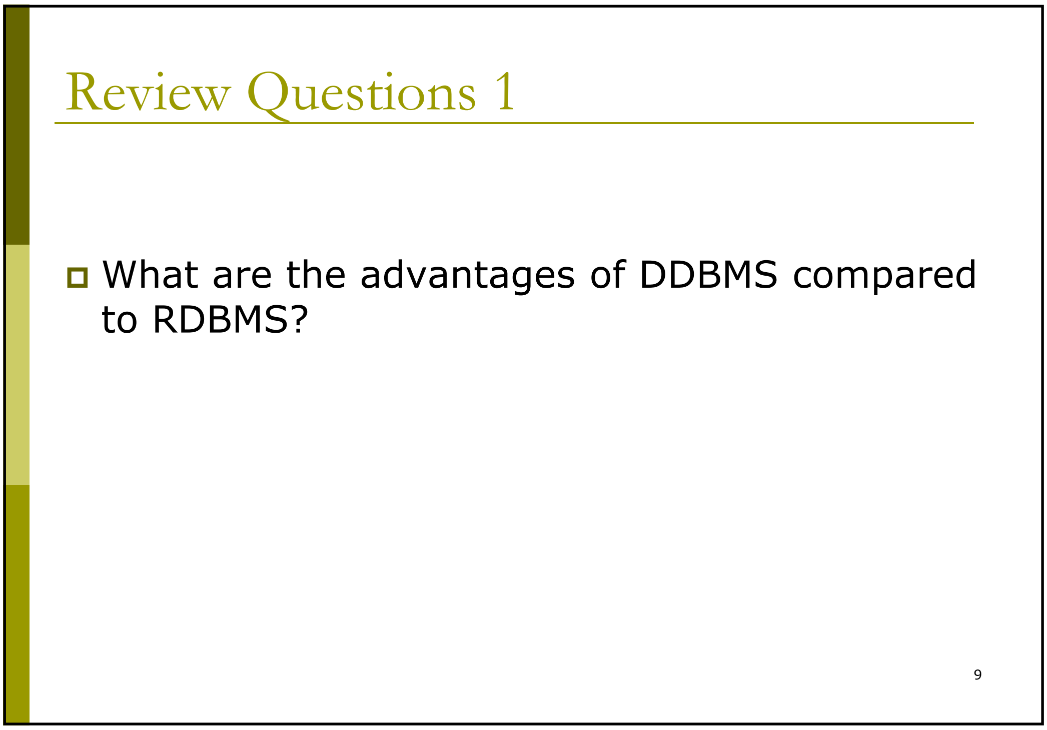
Weaknesses of RDBMS

- Poor Representation of "Real World" Entities
  - Normalization leads to relations that do not correspond to entities in "real world".
- Semantic Overloading
  - Relational model has only one construct for representing data and data relationships: the relation.
  - Relational model is semantically overloaded.
- Poor Support for Integrity and Enterprise Constraints



Weaknesses of RDBMS

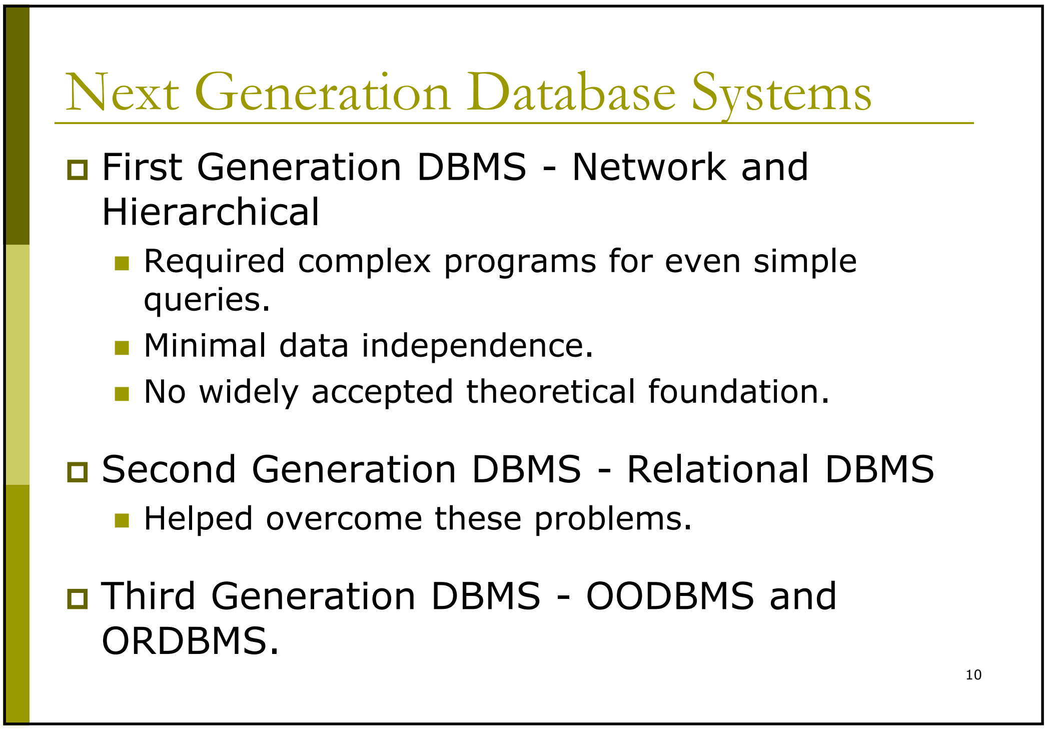
- Homogeneous Data Structure
  - Relational model assumes both horizontal and vertical homogeneity.
  - Many RDBMS now allow Binary Large Objects (BLOB).
- Limited Operations
  - RDBMS only have a fixed set of operations which cannot be extended.
- Difficulty Handling Recursive Queries
  - Extremely difficult to produce recursive queries.
  - Extension proposed to relational algebra to handle this type of query is unary transitive (recursive) closure, operation.



Review Questions 1

- What are the advantages of DDBMS compared to RDBMS?

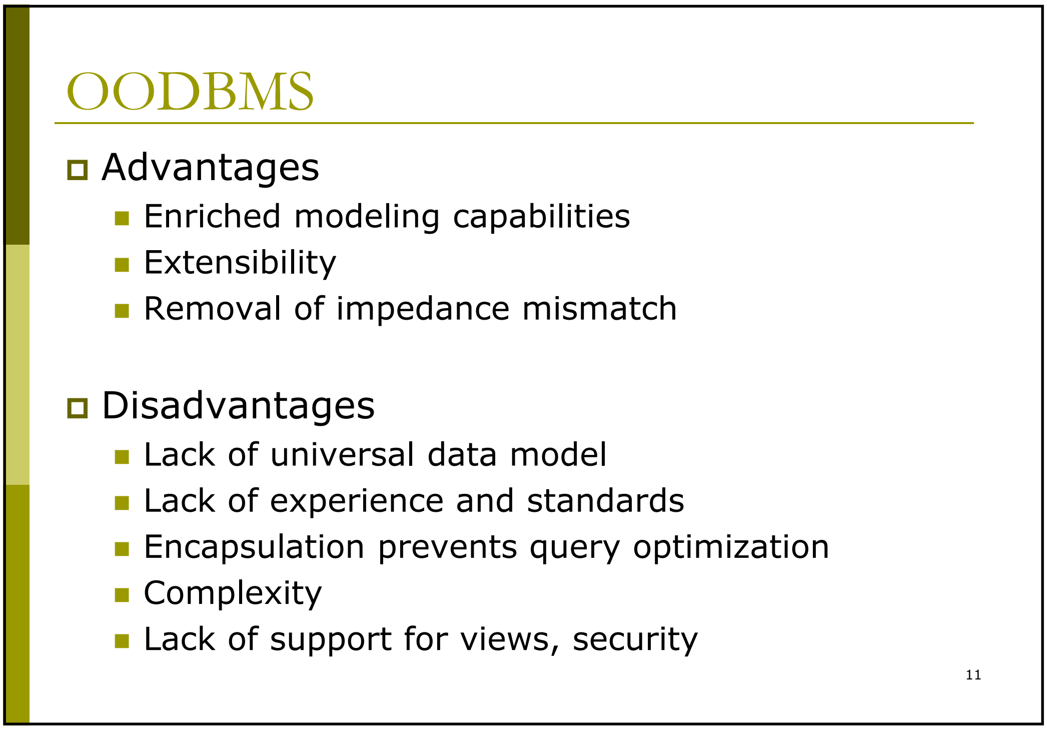
9



Next Generation Database Systems

- First Generation DBMS - Network and Hierarchical
  - Required complex programs for even simple queries.
  - Minimal data independence.
  - No widely accepted theoretical foundation.
- Second Generation DBMS - Relational DBMS
  - Helped overcome these problems.
- Third Generation DBMS - OODBMS and ORDBMS.

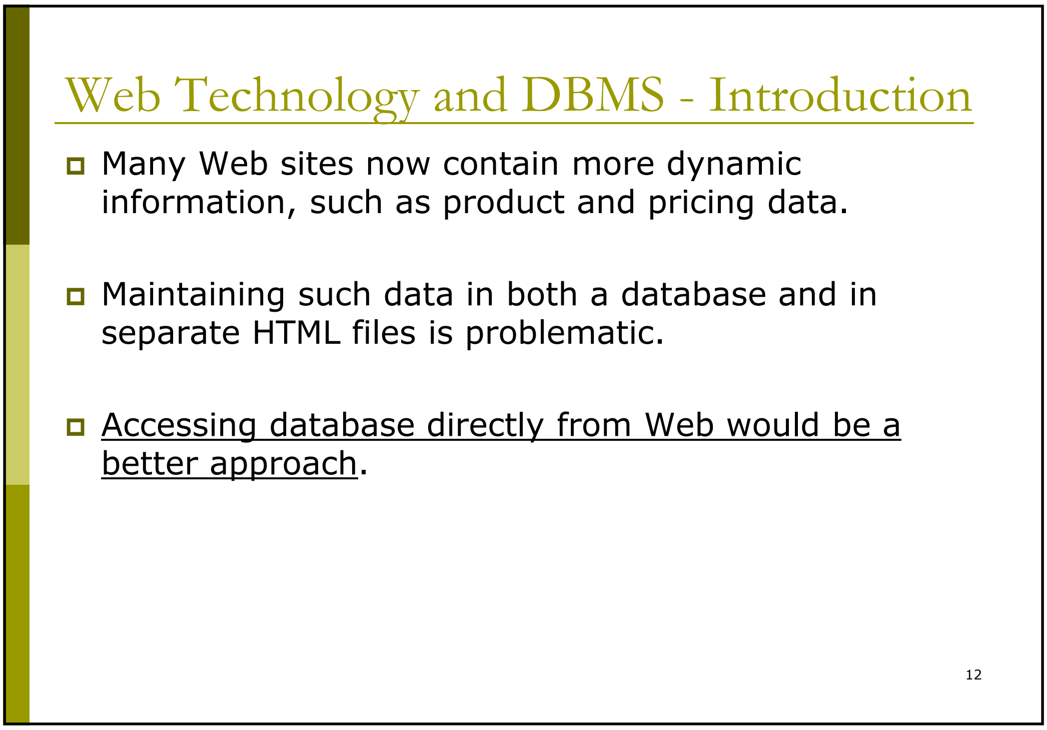
10



OODBMS

- Advantages
  - Enriched modeling capabilities
  - Extensibility
  - Removal of impedance mismatch
- Disadvantages
  - Lack of universal data model
  - Lack of experience and standards
  - Encapsulation prevents query optimization
  - Complexity
  - Lack of support for views, security

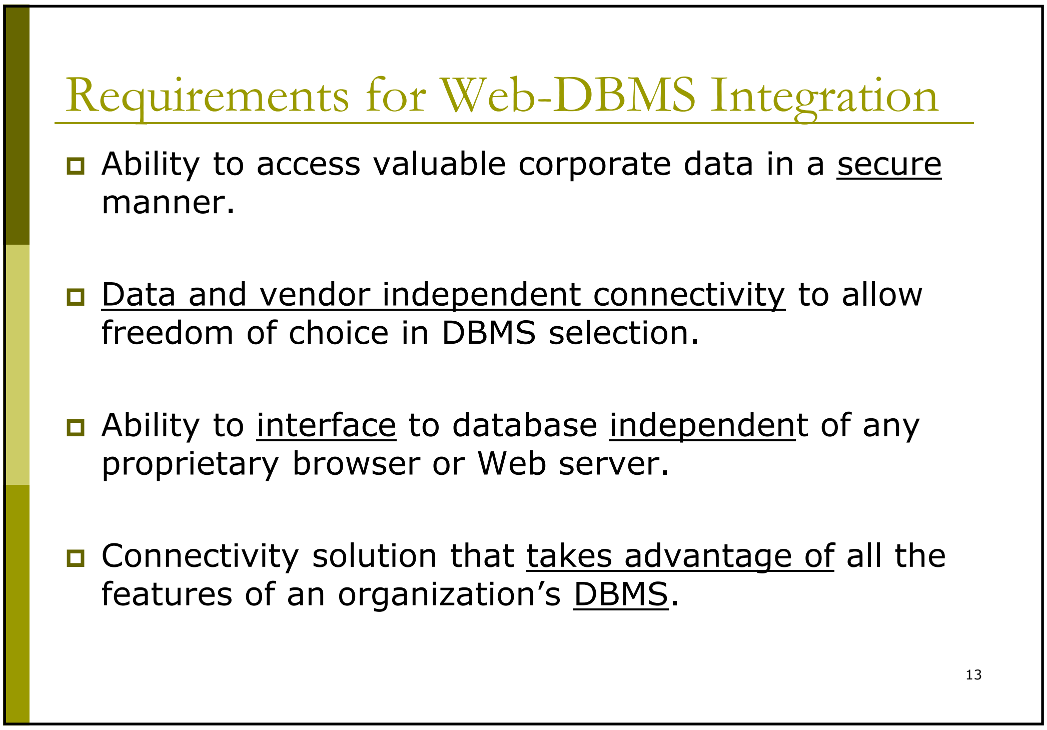
11



Web Technology and DBMS - Introduction

- Many Web sites now contain more dynamic information, such as product and pricing data.
- Maintaining such data in both a database and in separate HTML files is problematic.
- Accessing database directly from Web would be a better approach.

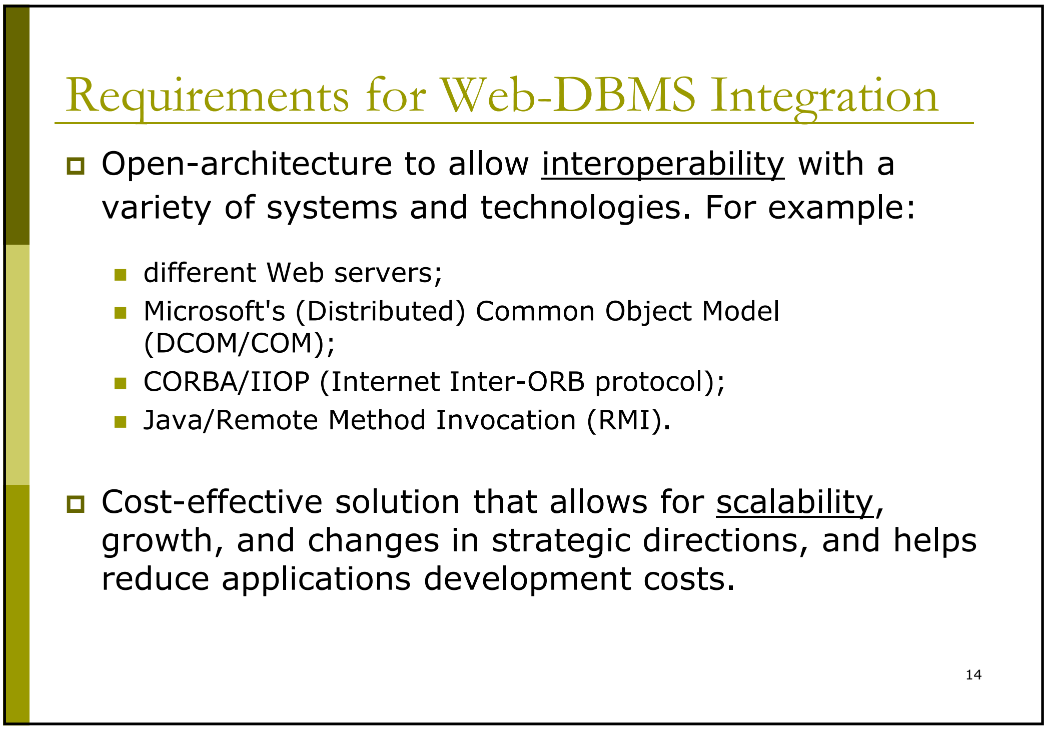
12



Requirements for Web-DBMS Integration

- Ability to access valuable corporate data in a secure manner.
- Data and vendor independent connectivity to allow freedom of choice in DBMS selection.
- Ability to interface to database independent of any proprietary browser or Web server.
- Connectivity solution that takes advantage of all the features of an organization’s DBMS.

13



Requirements for Web-DBMS Integration

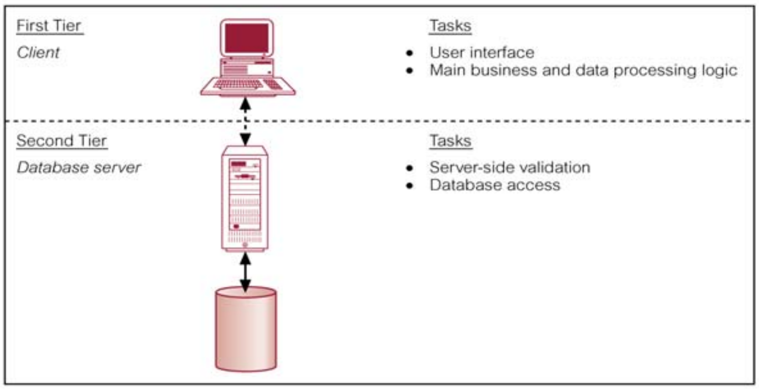
- Open-architecture to allow interoperability with a variety of systems and technologies. For example:
  - different Web servers;
  - Microsoft's (Distributed) Common Object Model (DCOM/COM);
  - CORBA/IIOP (Internet Inter-ORB protocol);
  - Java/Remote Method Invocation (RMI).
- Cost-effective solution that allows for scalability, growth, and changes in strategic directions, and helps reduce applications development costs.

14

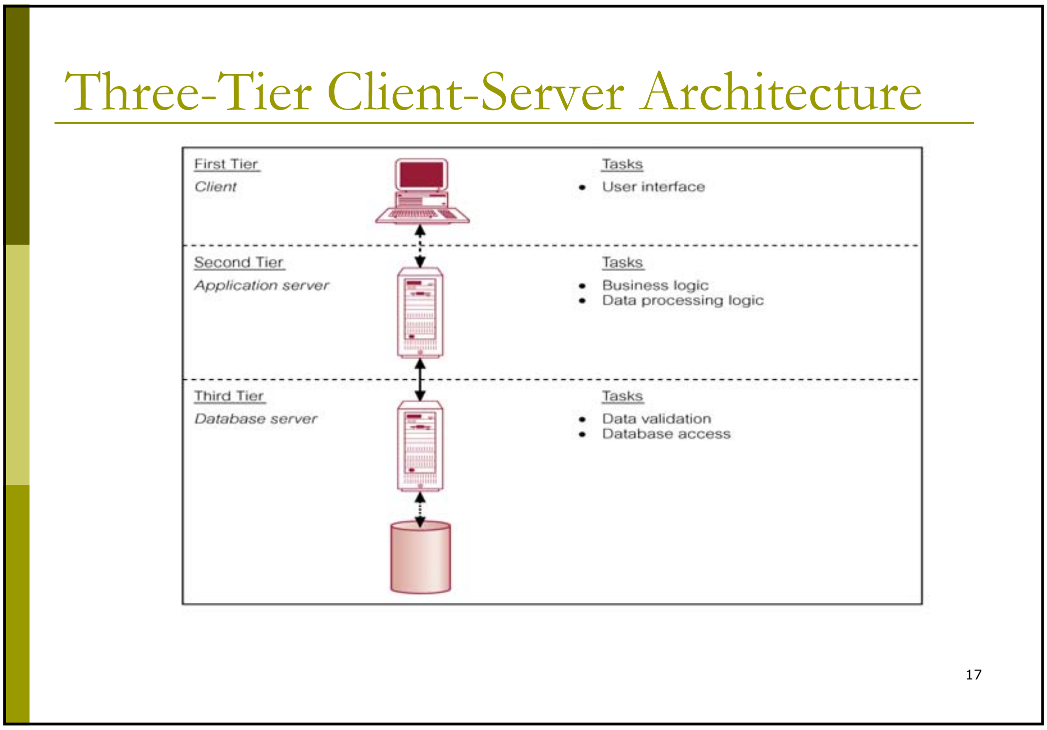
Requirements for Web-DBMS Integration

- Support for transactions that span multiple HTTP requests.
- Support for session- and application-based authentication.
- Acceptable performance.
- Minimal administration overhead.
- Set of high-level productivity tools to allow applications to be developed, maintained, and deployed with relative ease and speed.

Traditional Two-Tier Client-Server Architecture



- Client side presented two problems preventing true scalability:
  - 'Fat' client, requiring considerable resources on client's computer to run effectively.
  - Significant client side administration overhead.



Three-Tier Client-Server Architecture

First Tier
Client

Tasks
- User interface

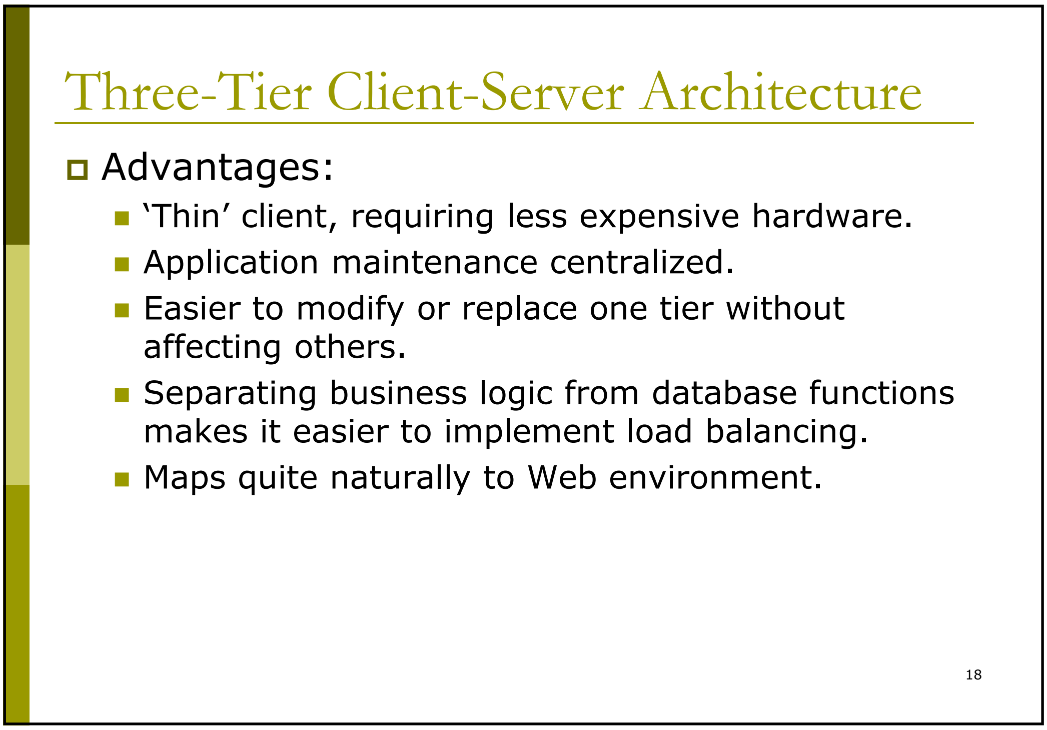
Second Tier
Application server

Tasks
- Business logic
- Data processing logic

Third Tier
Database server

Tasks
- Data validation
- Database access

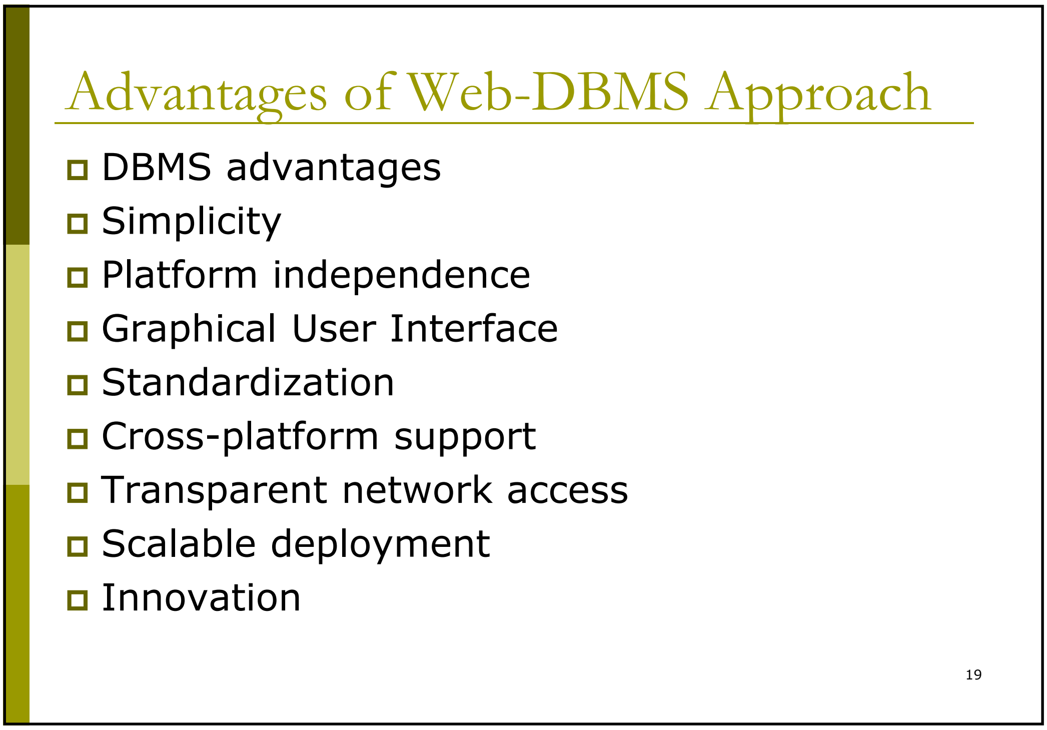
17



Three-Tier Client-Server Architecture

- Advantages:
  - ‘Thin’ client, requiring less expensive hardware.
  - Application maintenance centralized.
  - Easier to modify or replace one tier without affecting others.
  - Separating business logic from database functions makes it easier to implement load balancing.
  - Maps quite naturally to Web environment.

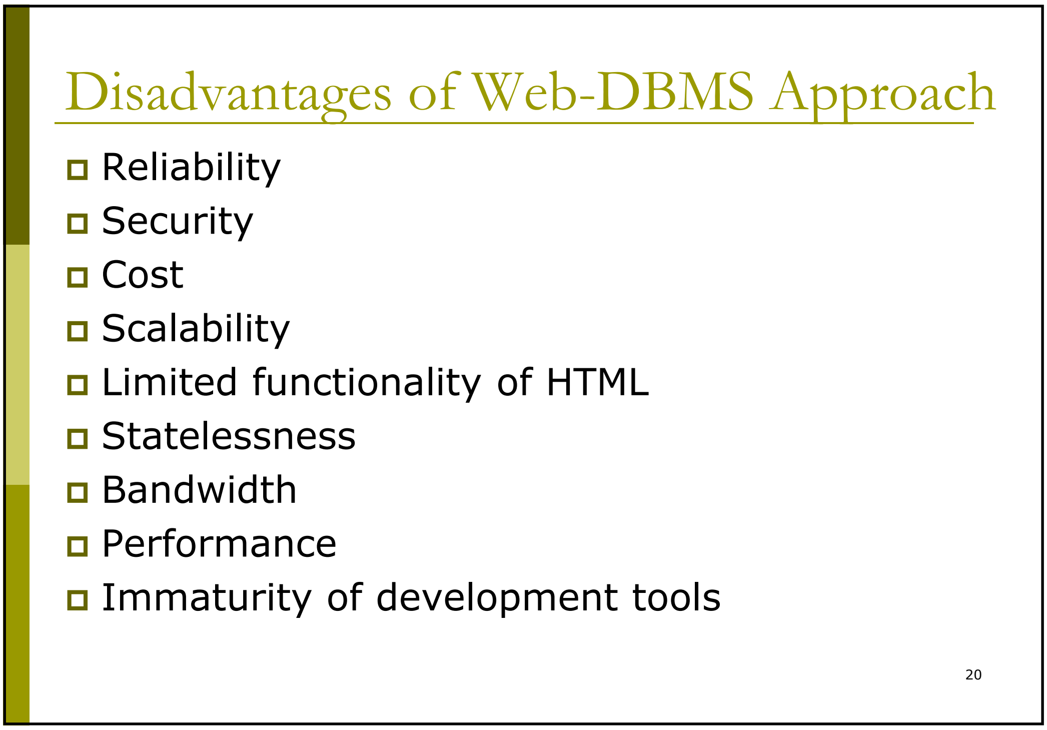
18



Advantages of Web-DBMS Approach

- DBMS advantages
- Simplicity
- Platform independence
- Graphical User Interface
- Standardization
- Cross-platform support
- Transparent network access
- Scalable deployment
- Innovation

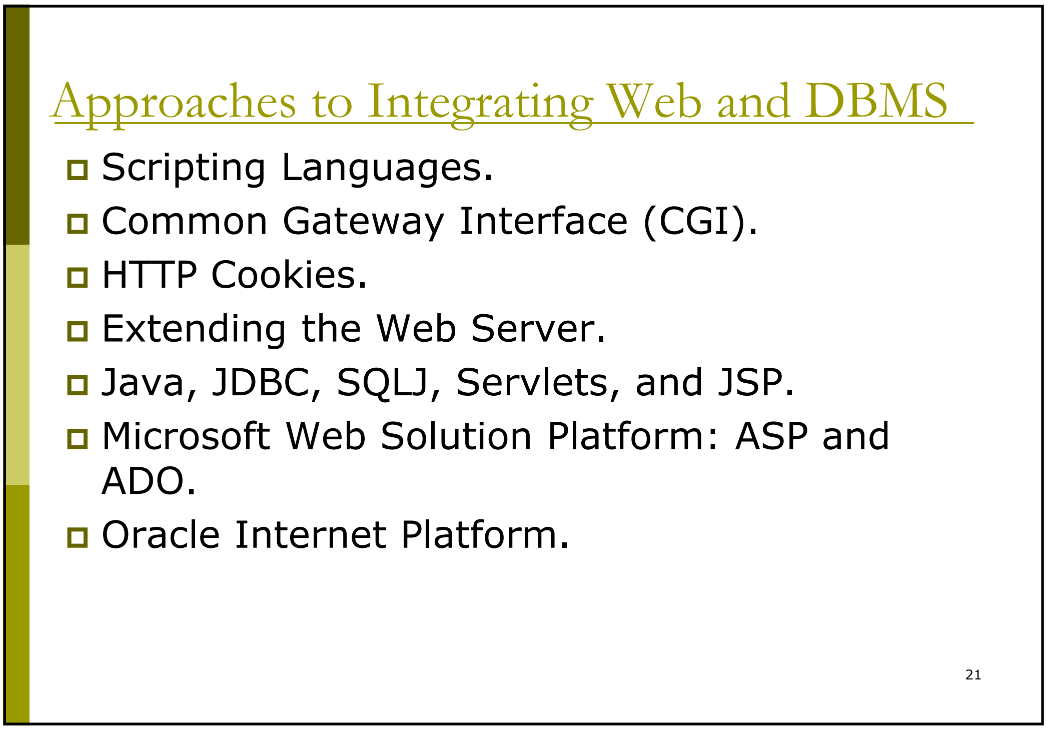
19



Disadvantages of Web-DBMS Approach

- Reliability
- Security
- Cost
- Scalability
- Limited functionality of HTML
- Statelessness
- Bandwidth
- Performance
- Immaturity of development tools

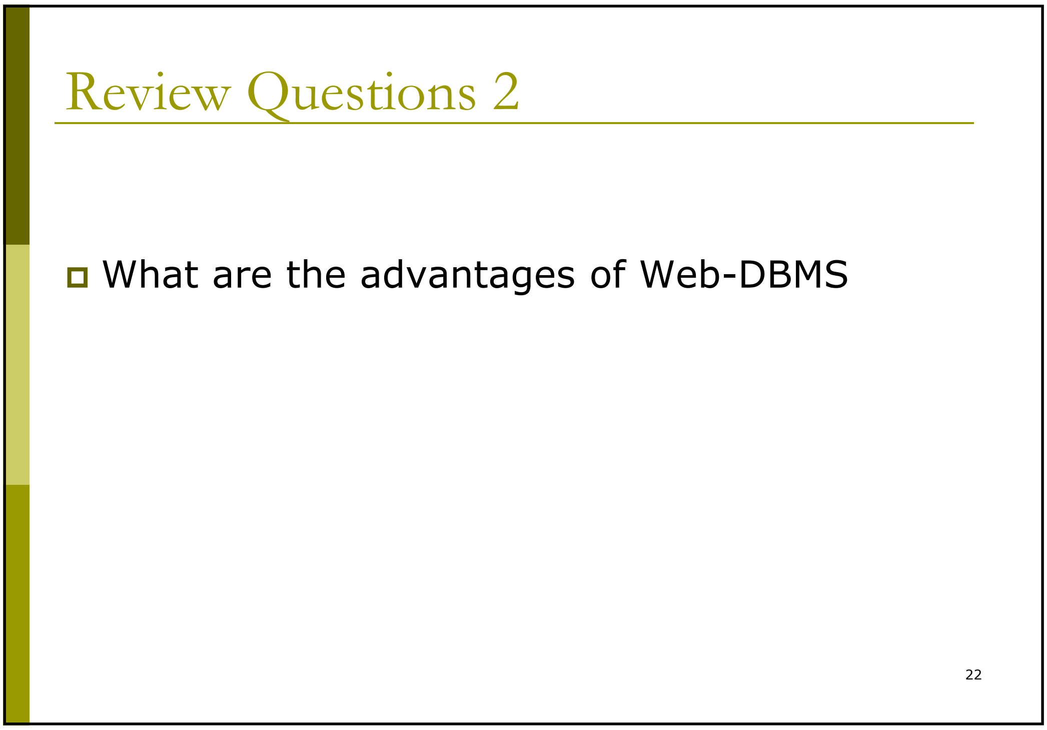
20



Approaches to Integrating Web and DBMS

- Scripting Languages.
- Common Gateway Interface (CGI).
- HTTP Cookies.
- Extending the Web Server.
- Java, JDBC, SQLJ, Servlets, and JSP.
- Microsoft Web Solution Platform: ASP and ADO.
- Oracle Internet Platform.

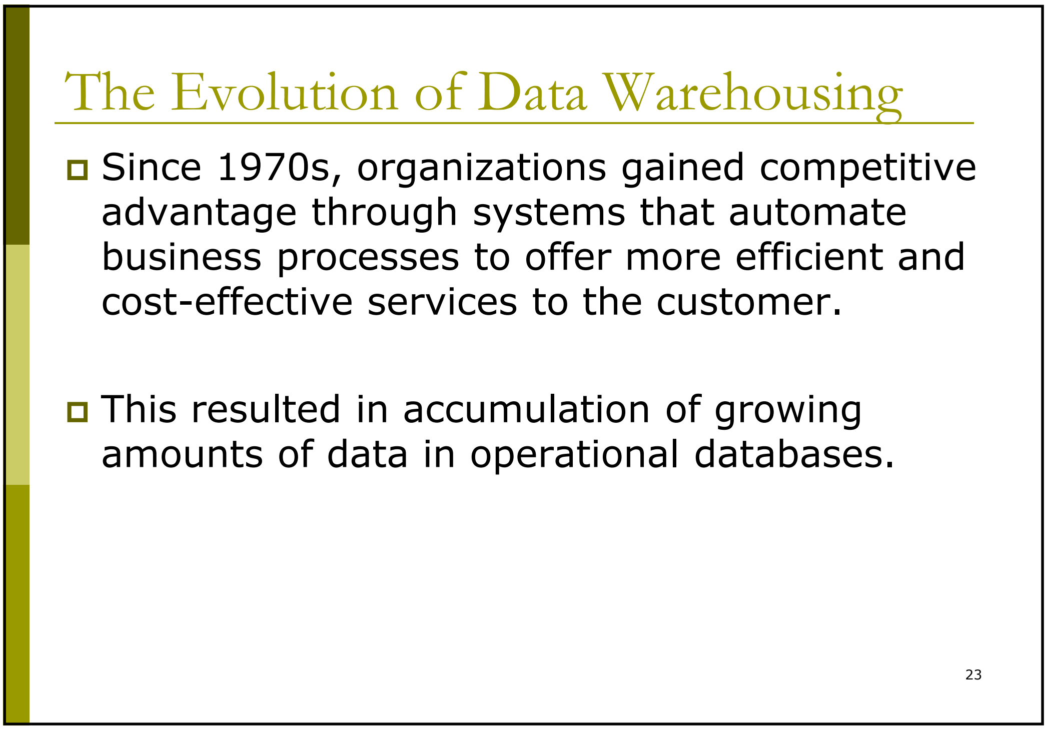
21



Review Questions 2

- What are the advantages of Web-DBMS

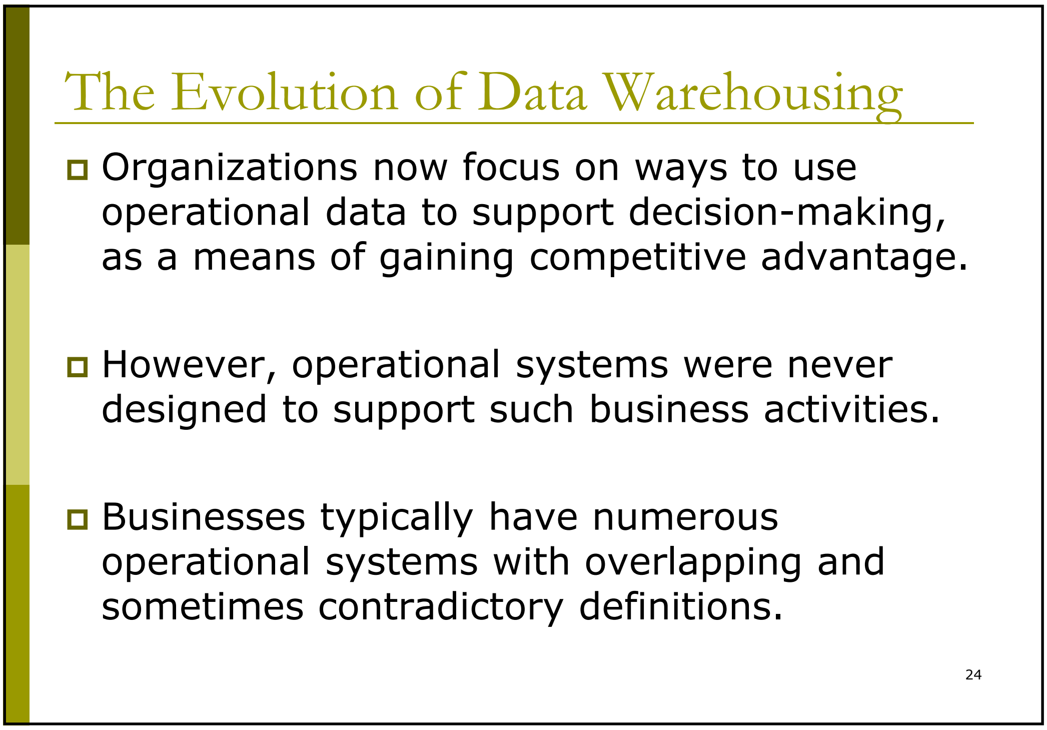
22



The Evolution of Data Warehousing

- Since 1970s, organizations gained competitive advantage through systems that automate business processes to offer more efficient and cost-effective services to the customer.
- This resulted in accumulation of growing amounts of data in operational databases.

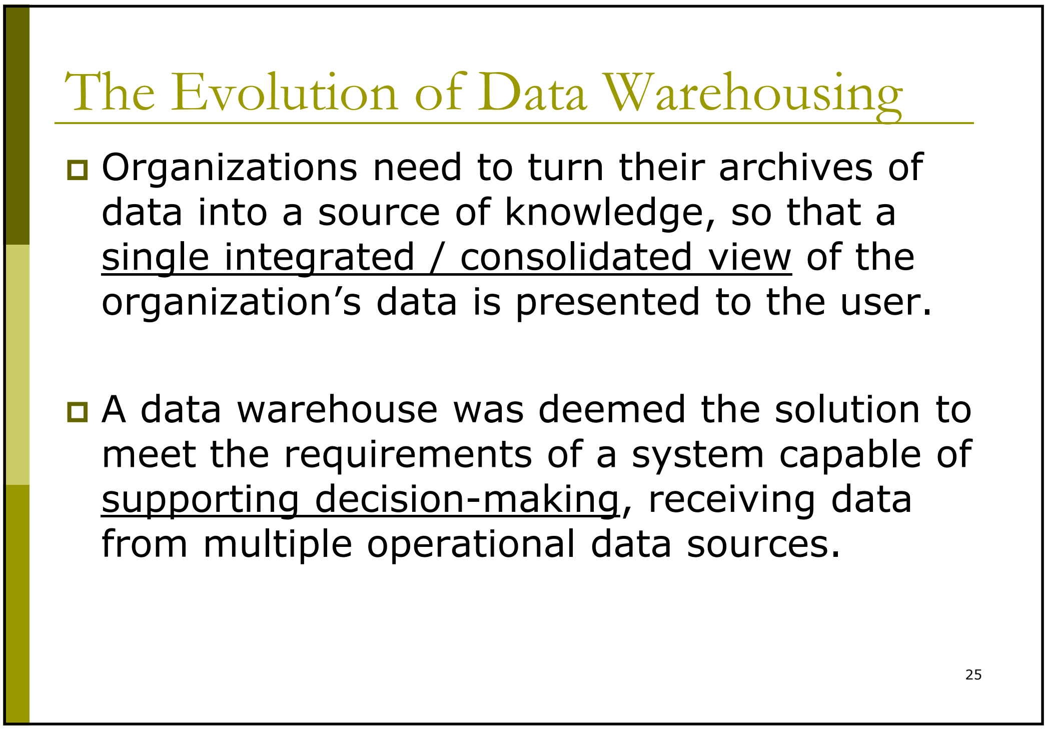
23



The Evolution of Data Warehousing

- Organizations now focus on ways to use operational data to support decision-making, as a means of gaining competitive advantage.
- However, operational systems were never designed to support such business activities.
- Businesses typically have numerous operational systems with overlapping and sometimes contradictory definitions.

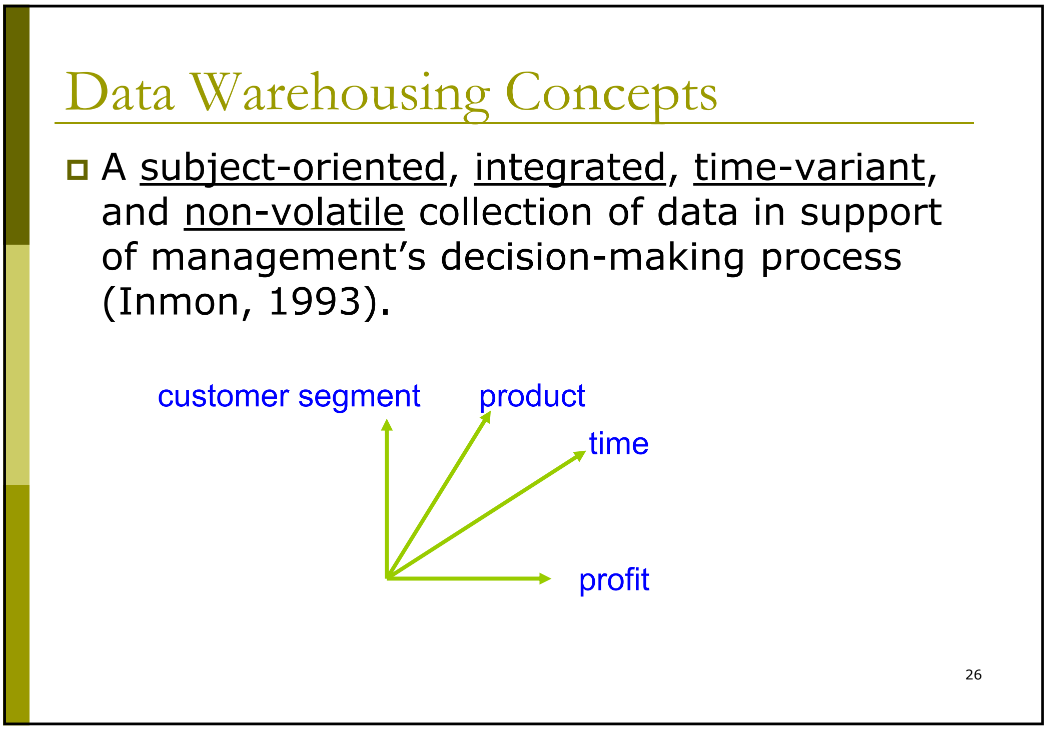
24



The Evolution of Data Warehousing

- Organizations need to turn their archives of data into a source of knowledge, so that a single integrated / consolidated view of the organization's data is presented to the user.
- A data warehouse was deemed the solution to meet the requirements of a system capable of supporting decision-making, receiving data from multiple operational data sources.

25

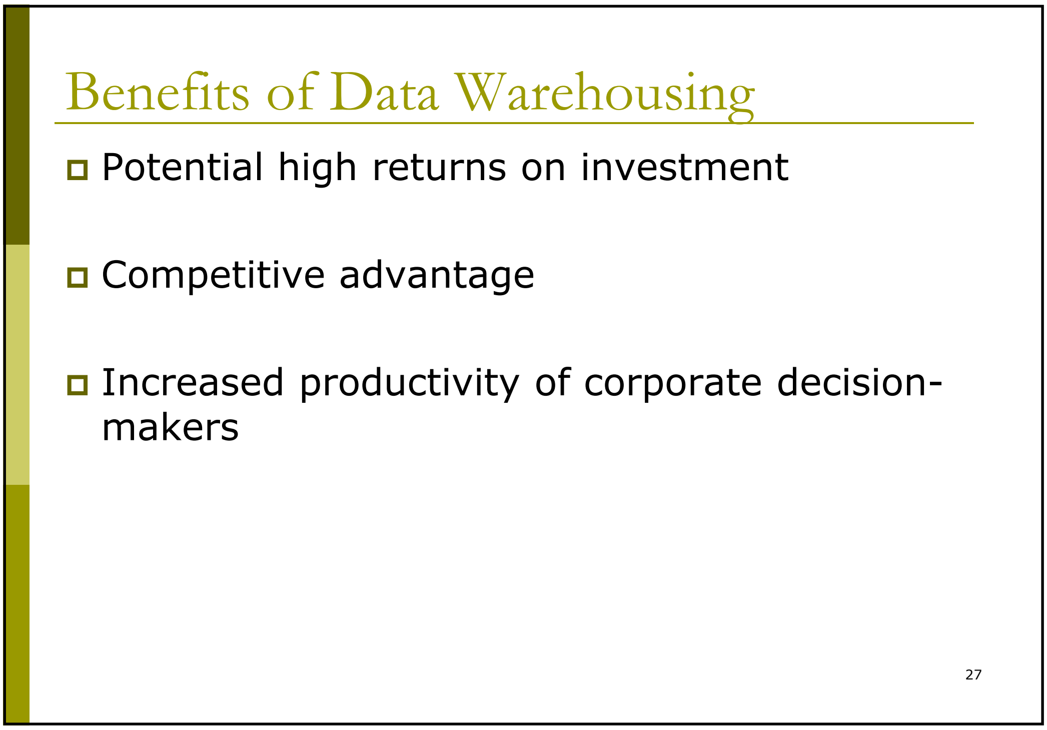


Data Warehousing Concepts

- A subject-oriented, integrated, time-variant, and non-volatile collection of data in support of management's decision-making process (Inmon, 1993).

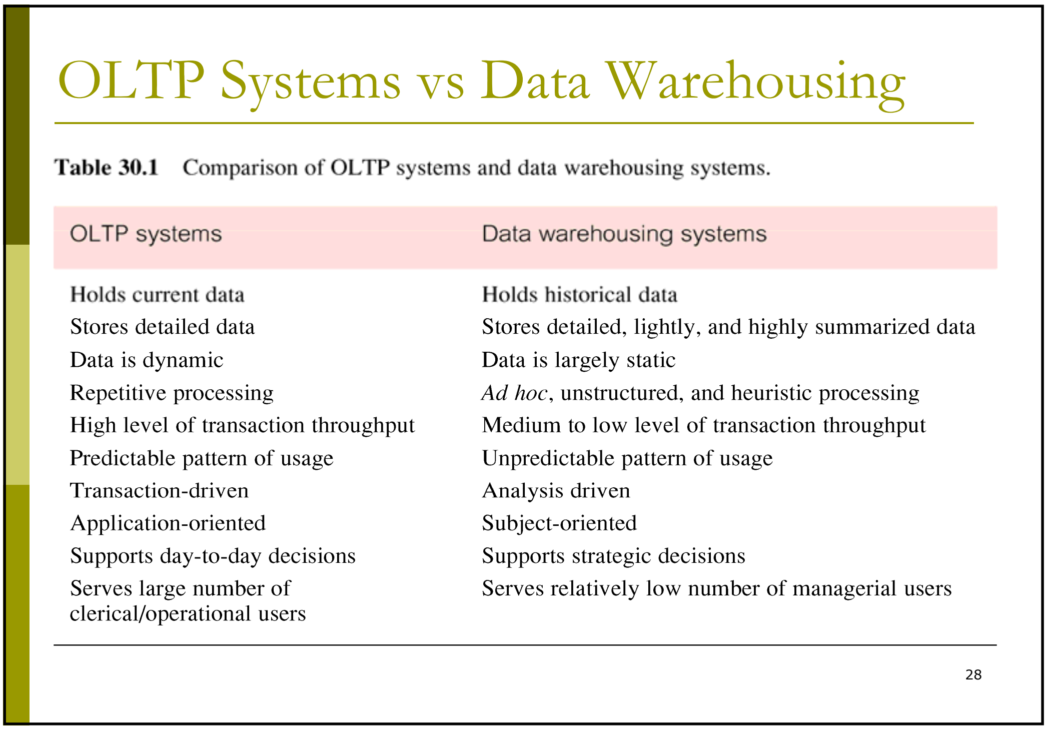
customer segment
product
time
profit

26



Benefits of Data Warehousing

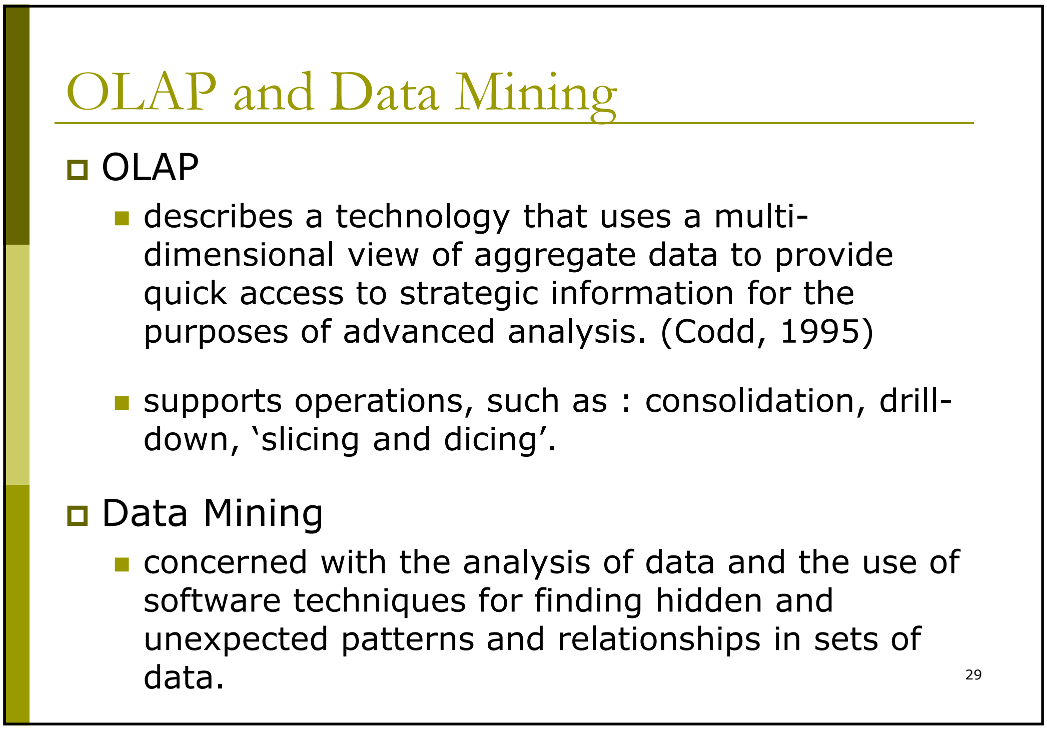
- Potential high returns on investment
- Competitive advantage
- Increased productivity of corporate decision-makers



OLTP Systems vs Data Warehousing

Table 30.1 Comparison of OLTP systems and data warehousing systems.

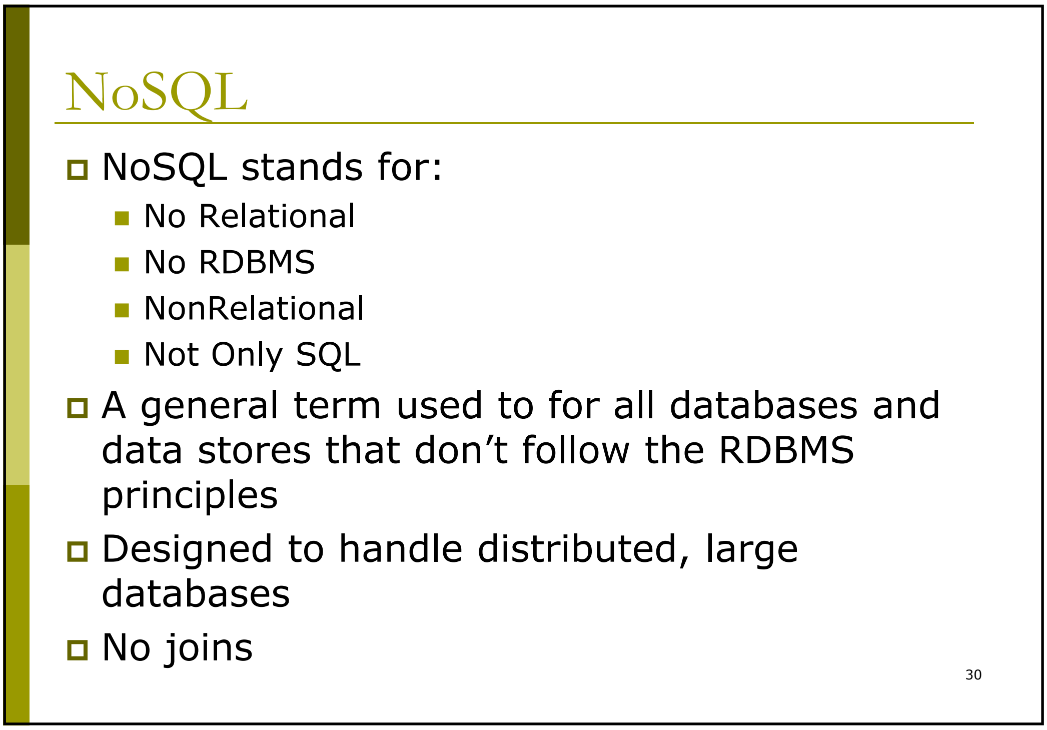
OLTP systems	Data warehousing systems
Holds current data	Holds historical data
Stores detailed data	Stores detailed, lightly, and highly summarized data
Data is dynamic	Data is largely static
Repetitive processing	*Ad hoc*, unstructured, and heuristic processing
High level of transaction throughput	Medium to low level of transaction throughput
Predictable pattern of usage	Unpredictable pattern of usage
Transaction-driven	Analysis driven
Application-oriented	Subject-oriented
Supports day-to-day decisions	Supports strategic decisions
Serves large number of clerical/operational users	Serves relatively low number of managerial users



OLAP and Data Mining

- OLAP
  - describes a technology that uses a multi-dimensional view of aggregate data to provide quick access to strategic information for the purposes of advanced analysis. (Codd, 1995)
  - supports operations, such as : consolidation, drill-down, ‘slicing and dicing’.
- Data Mining
  - concerned with the analysis of data and the use of software techniques for finding hidden and unexpected patterns and relationships in sets of data.

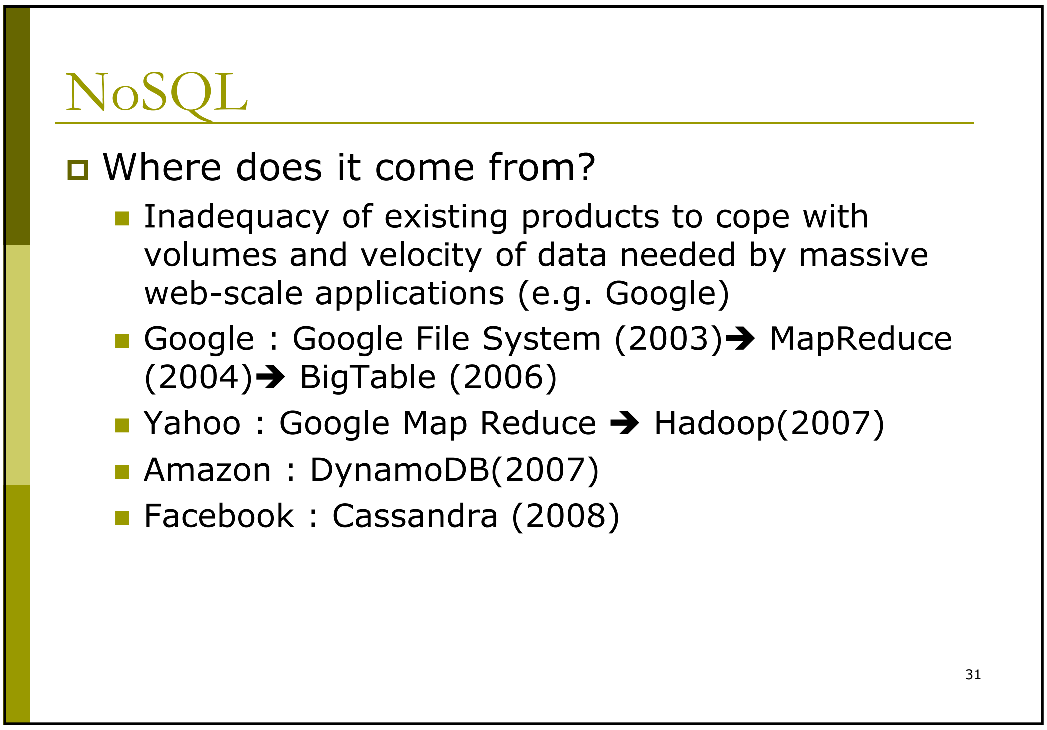
29



NoSQL

- NoSQL stands for:
  - No Relational
  - No RDBMS
  - NonRelational
  - Not Only SQL
- A general term used to for all databases and data stores that don’t follow the RDBMS principles
- Designed to handle distributed, large databases
- No joins

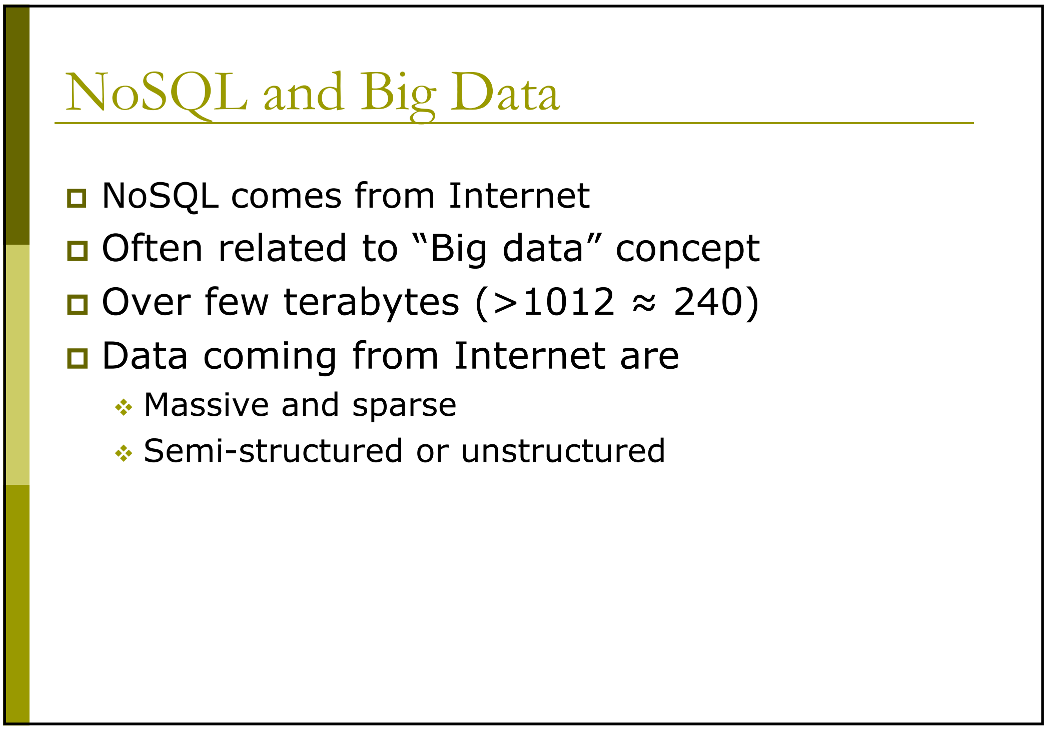
30



NoSQL

- Where does it come from?
  - Inadequacy of existing products to cope with volumes and velocity of data needed by massive web-scale applications (e.g. Google)
  - Google : Google File System (2003)➔ MapReduce (2004)➔ BigTable (2006)
  - Yahoo : Google Map Reduce ➔ Hadoop(2007)
  - Amazon : DynamoDB(2007)
  - Facebook : Cassandra (2008)

31



NoSQL and Big Data

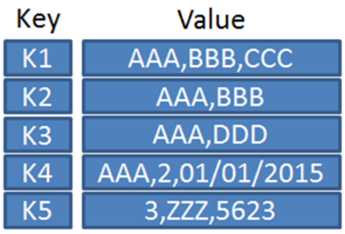
- NoSQL comes from Internet
- Often related to “Big data” concept
- Over few terabytes (>1012 ≈ 240)
- Data coming from Internet are
  - Massive and sparse
  - Semi-structured or unstructured

NoSQL Products

- NoSQL products can be categorized in
  - Key value
  - Document
  - Graph
  - Column Family

Key Value Database

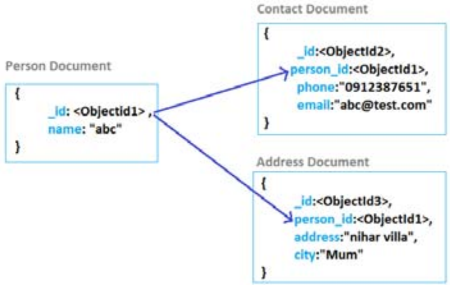
- Simple model, mapping between key and value
- You cannot query without having the key
- Easy to use, very scalable
- Example: Redis (Widely used by Twitter and Flickr)



Key	Value
K1	AAA,BBB,CCC
K2	AAA,BBB
K3	AAA,DDD
K4	AAA,2,01/01/2015
K5	3,ZZZ,5623

Document Database

- Non relational database that stores data as structured document
- Usually in XML or JSON formats
- Usually schema-less
- Usually no join operatic
- Example: MongoDB


Person Document

```
{
        _id: <ObjectId1> ,
        name: "abc"
}
```

Contact Document

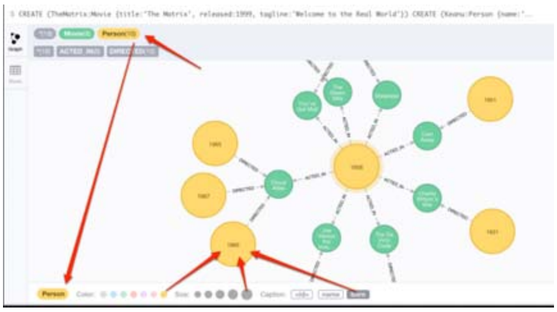
```
{
        _id:<ObjectId2>,
      person_id:<ObjectId1>,
        phone:"0912387651",
        email:"abc@test.com"
}
```

Address Document

```
{
        _id:<ObjectId3>,
        person_id:<ObjectId1>,
        address:"nihar villa",
        city:"Mum"
}
```

Graph Database

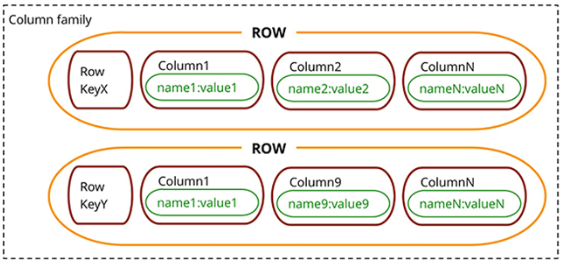
- Useful when objective is to quickly find connections, pattern and relationship between lots of data
- Usage: Recommendations for product, social media
- Example: Neo4J



```
$ CREATE (TheMatrix:Movie {title:'The Matrix', released:1999, tagline:'Welcome to the Real World'}) CREATE (Keanu:Person {name:'...
```

Column Family Database

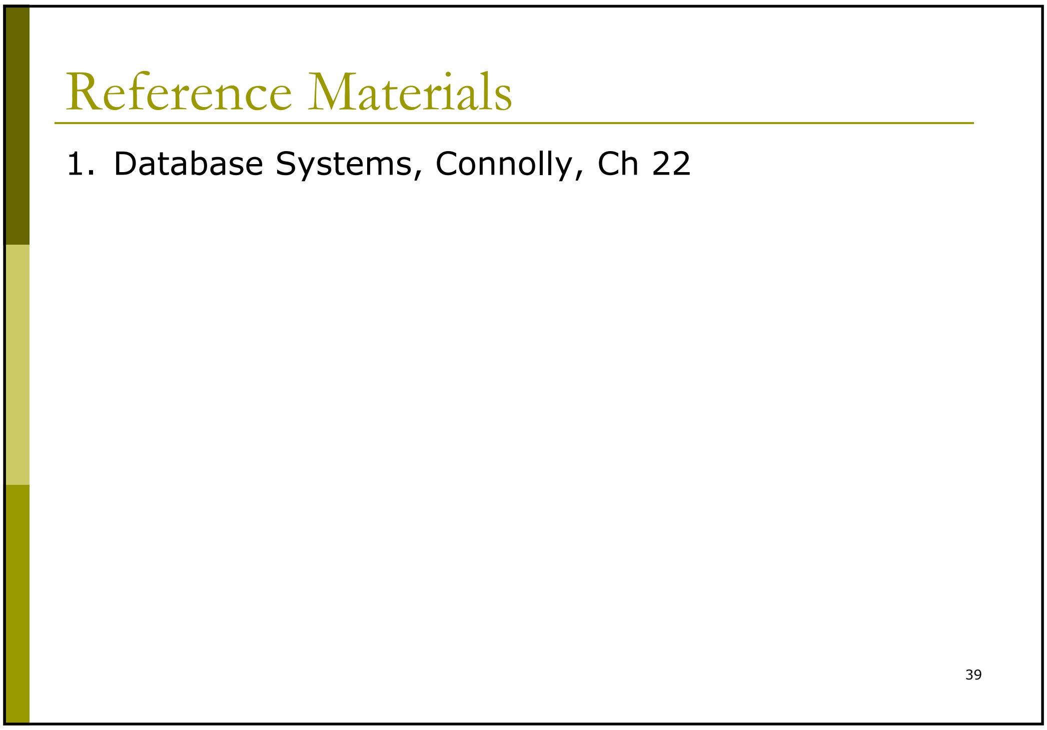
- Store data in column families
  - Many column associated with a row key
- Column families are groups of related data that is often accessed together
- Example: Cassandra,HBase



Review Questions 3

- What are the advantages of Data Warehousing

38



Reference Materials

1. Database Systems, Connolly, Ch 22