Unit 5: Managing Data Resources LH 7

- Data resource management
- Types of databases: operational, distributed, external, hypermedia databases
- Data warehousing and data mining
- The database management approach
  - Database management system, database interrogation, database maintenance, application development

Learning Objectives

- Explain the business value of implementing data resource management processes and technologies in an organization
- Outline the advantages of a database management approach to managing the data resources of a business, compared to a file processing approach
- Explain how database management software helps business professionals and supports the operations and management of a business

Learning Objectives

- Provide examples to illustrate the following concepts
  - Major types of databases
  - Data warehouses and data mining
  - Logical data elements
  - Fundamental database structures
  - Database development

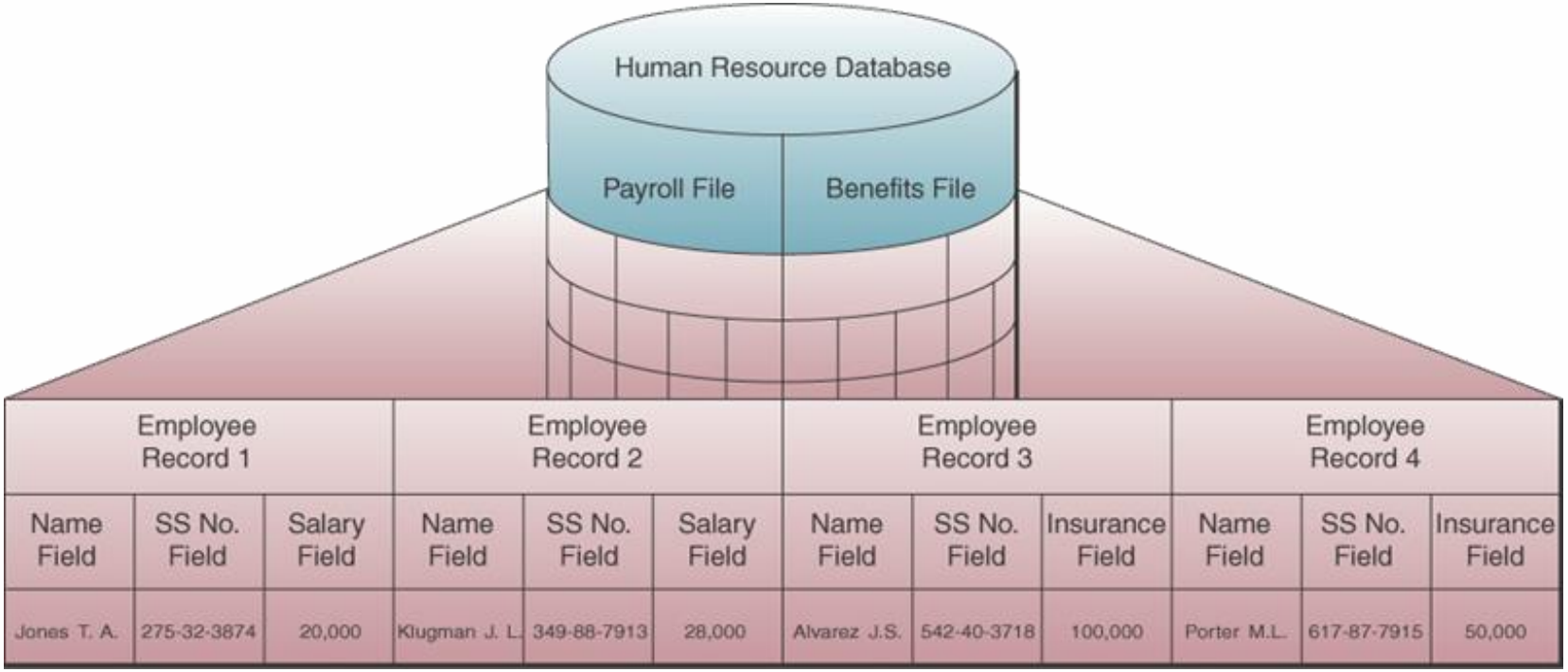
Data resource management

- Data are a vital organizational resource that need to be managed like other important business assets. Today’s business enterprises could not survive or succeed without quality data about their internal operations and external environment.

- Managers need to practice data resource management - a managerial activity that applies information systems technology like database management and other management tools to the task of managing an organization's data resources to meet the information needs of business stakeholders.

Logical Data Elements

- Foundation Data Concepts



Employee Record 1			Employee Record 2			Employee Record 3			Employee Record 4		
Name Field	SS No. Field	Salary Field	Name Field	SS No. Field	Salary Field	Name Field	SS No. Field	Insurance Field	Name Field	SS No. Field	Insurance Field
Jones T. A.	275-32-3874	20,000	Klugman J. L.	349-88-7913	28,000	Alvarez J.S.	542-40-3718	100,000	Porter M.L.	617-87-7915	50,000

Logical Data Elements

- Character
  - A single alphabetic, numeric, or other symbol
- Field or data item
  - Represents an attribute (characteristic or quality) of some entity (object, person, place, event)
    - Examples: salary, job title
- Record
  - Grouping of all the fields used to describe the attributes of an entity
    - Example: payroll record with name, SSN, pay rate

Logical Data Elements

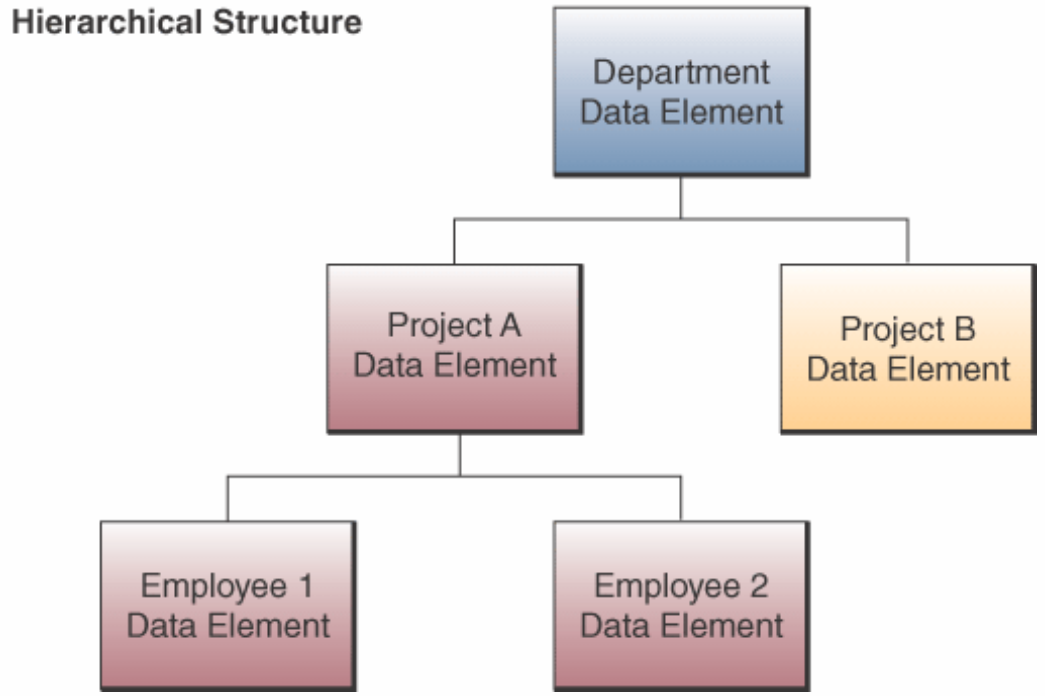
- File or table
  - A group of related records
- Database
  - An integrated collection of logically related data elements

Database Structures

- Common database structures…
  - Hierarchical
  - Network
  - Relational
  - Object-oriented
  - Multi-dimensional

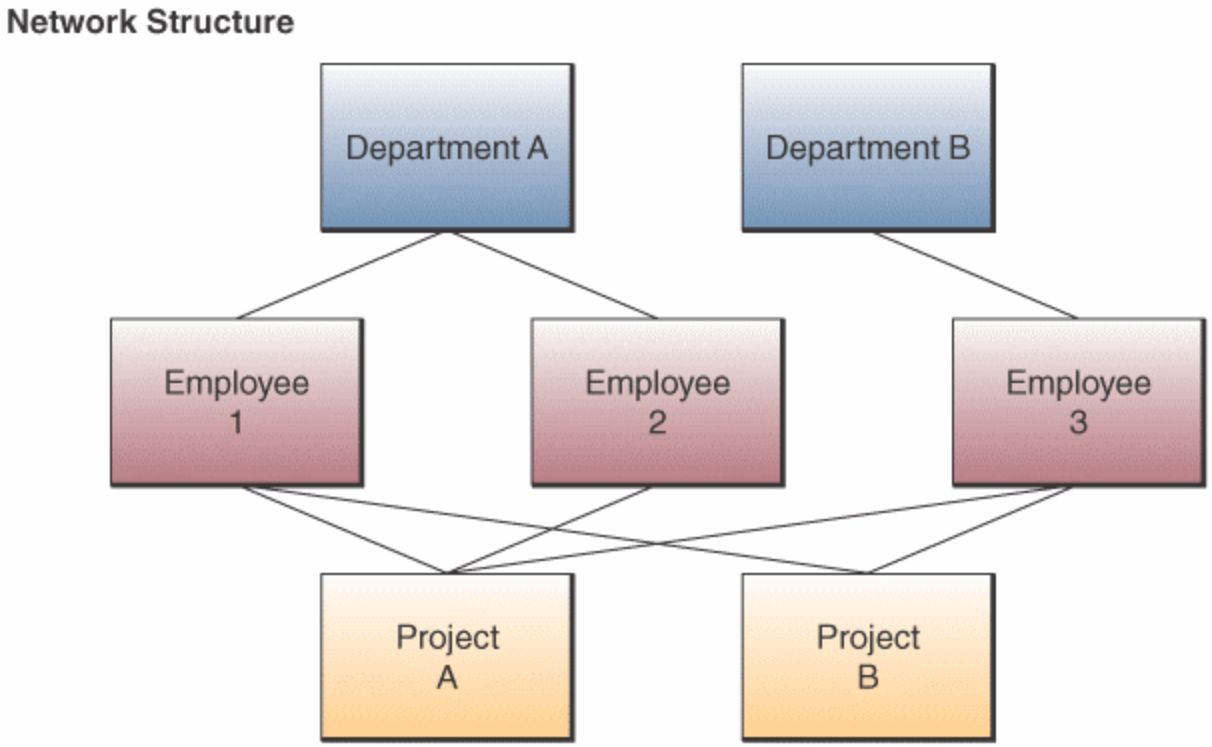
Hierarchical Structure

- Early DBMS structure
- Records arranged in tree-like structure
- Relationships are one-to-many

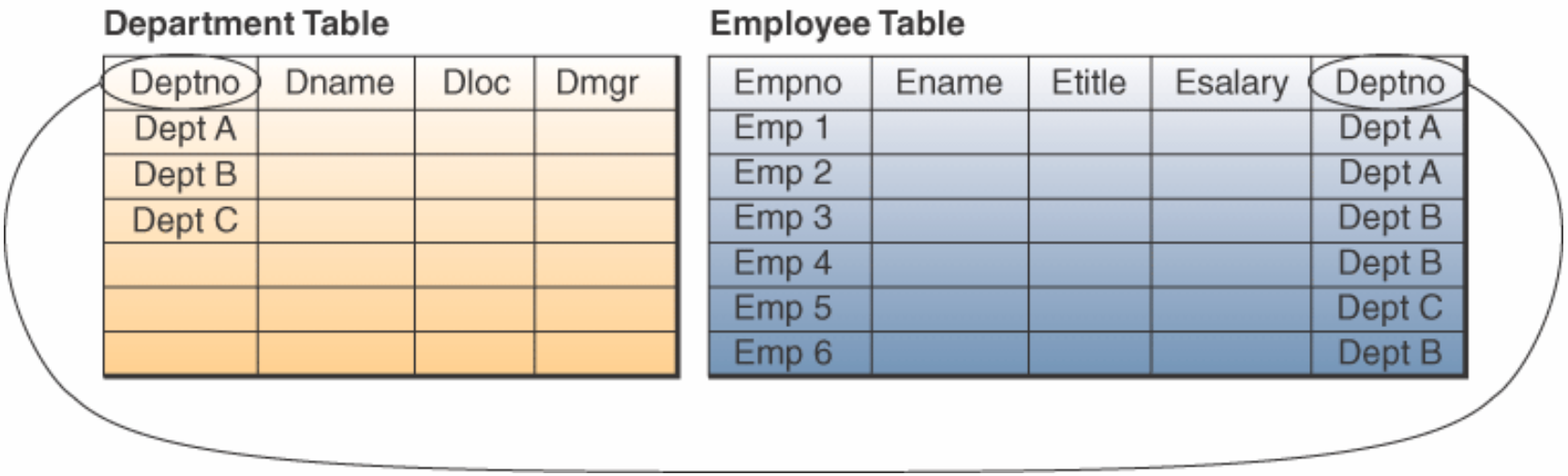


Network Structure

- Used in some mainframe DBMS packages
- Many-to-many relationships



Relational Structure



Most widely used structure

- Data elements are stored in tables
- Row represents a record; column is a field
- Can relate data in one file with data in another, if both files share a common data element

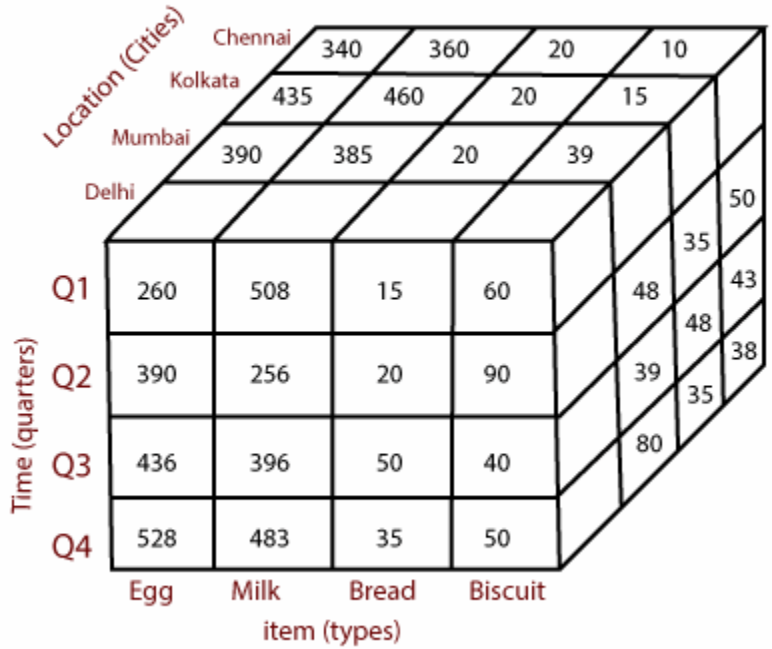
Relational Operations

- Select
  - Create a subset of records that meet a stated criterion
    - Example: employees earning more than $30,000
- Join
  - Combine two or more tables temporarily
  - Looks like one big table
- Project
  - Create a subset of columns in a table

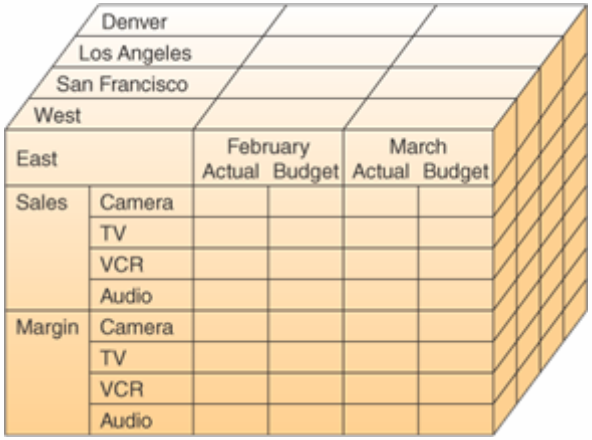
Multidimensional Structure

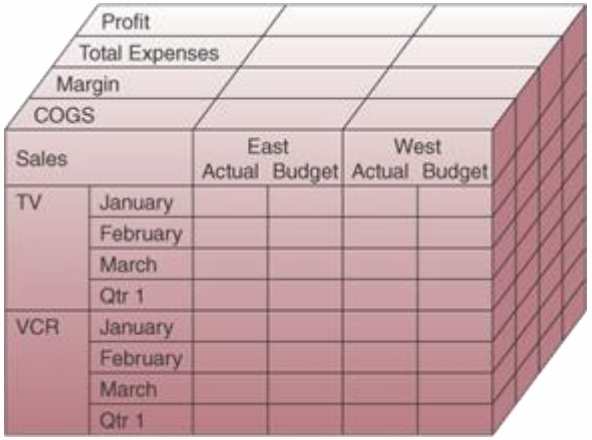
- Variation of relational model
  - Uses multidimensional structures to organize data
  - Data elements are viewed as being in cubes
  - Popular for analytical databases that support Online Analytical Processing (OLAP)

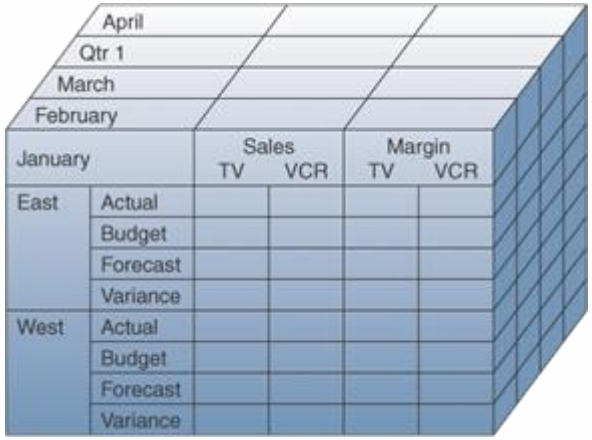
	Location="Chennai"				Location="Kolkata"				Location="Mumbai"				Location="Delhi"			
	item				item				item				item			
Time	Egg	Milk	Bread	Biscuit	Egg	Milk	Bread	Biscuit	Egg	Milk	Bread	Biscuit	Egg	Milk	Bread	Biscuit
Q1	340	360	20	10	435	460	20	15	390	385	20	39	260	508	15	60
Q2	490	490	16	50	389	385	45	35	463	366	25	48	390	256	20	90
Q3	680	583	46	43	684	490	39	48	568	594	36	39	436	396	50	40
Q4	535	694	39	38	335	365	83	35	338	484	48	80	528	483	35	50



Multidimensional Model



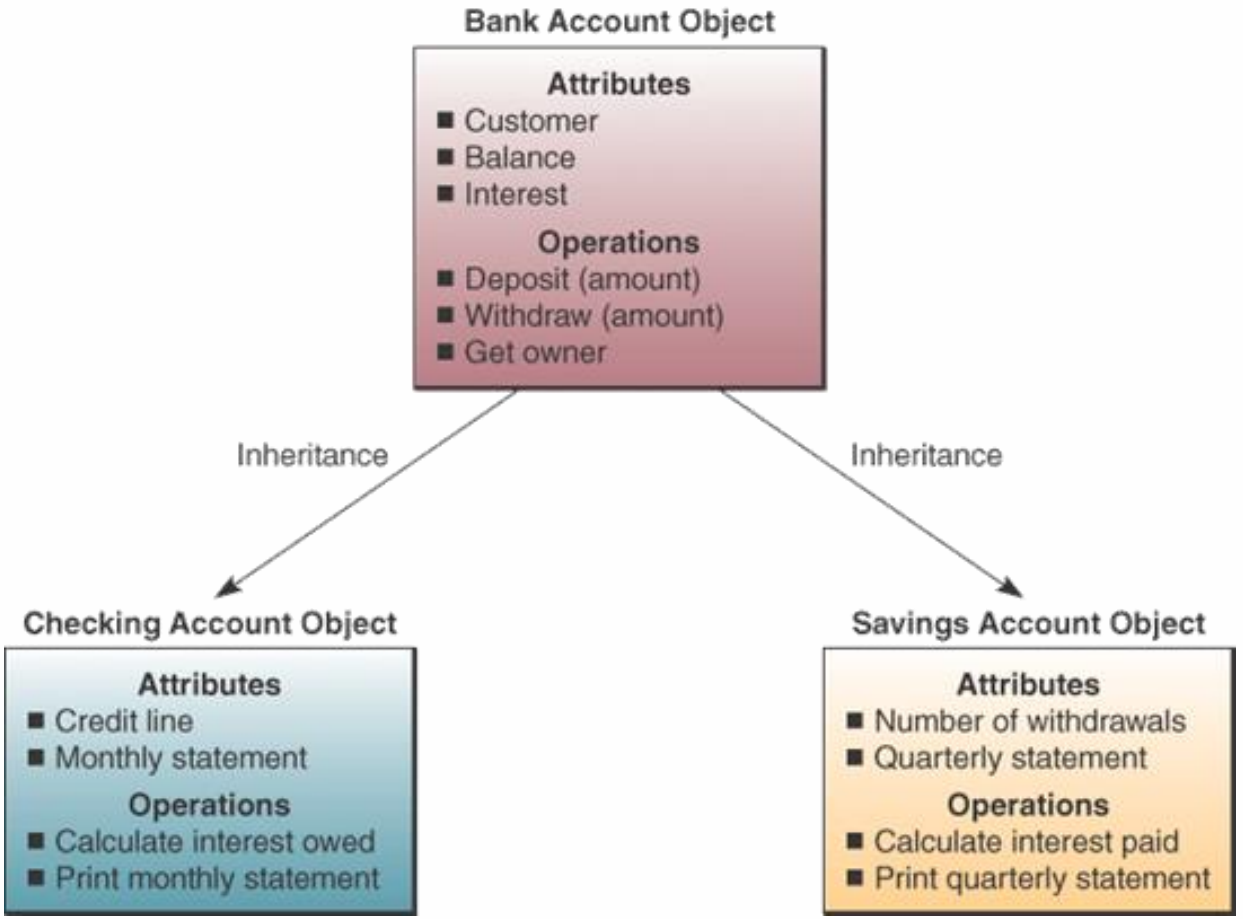




Object-Oriented Structure

- An **object** consists of
  - Data values describing the attributes of an entity
  - Operations that can be performed on the data
- Encapsulation
  - Combine data and operations
- Inheritance
  - New objects can be created by replicating some or all of the characteristics of parent objects

Object-Oriented Structure



Source: Adapted from Ivar Jacobsen, Maria Ericsson, and Ageneta Jacobsen, *The Object Advantage: Business Process Reengineering with Object Technology* (New York: ACM Press, 1995), p. 65.

Object-Oriented Structure

- Used in object-oriented database management systems (OODBMS)
- Supports complex data types more efficiently than relational databases
  - Examples: graphic images, video clips, web pages

Case 1: Sharing Business Databases

- Amazon's data vault
  - Product descriptions
  - Prices
  - Sales rankings
  - Customer reviews
  - Inventory figures
  - Countless other layers of content
- Took 10 years and a billion dollars to build

Case 1: Sharing Business Databases

- Amazon opened its data vault in 2002
  - 65,000 developers, businesses, and entrepreneurs have tapped into it
  - Many have become ambitious business partners

- eBay opened its $3 billion databases in 2003
  - 15,000 developers and others have registered to use it and to access software features
  - 1,000 new applications have appeared
  - 41 percent of eBay's listings are uploaded to the site using these resources

Case 1: Sharing Business Databases

- Google recently unlocked access to its desktop and paid-search products
  - Dozens of Google-driven services cropped up
  - Developers can grab 1,000 search results a day for free; anything more requires permission
  - In 2005, the Ad-Words paid-search service was opened to outside applications

Case Study Questions

- What are the business benefits to Amazon and eBay of opening up some of their databases to developers and entrepreneurs?
  - Do you agree with this strategy?
- What business factors are causing Google to move slowly in opening up its databases?
  - Do you agree with its go-slow strategy?

Case Study Questions

- Should other companies follow Amazon and eBay's lead and open up some of their databases to developers and others?
  - Defend your position with an example of the risks and benefits to an actual company

Types of databases: operational, distributed, external, hypermedia databases.

Operational Databases:

- These databases store detailed data needed to support the operations of the entire organization. They are also called subject area databases (SADB), transaction databases, and production databases. Examples are customer databases, personnel databases, inventory databases, and other databases containing data generated by business operations.

External Databases

- Access to external, privately owned online databases or data banks is available for a free to end users and organizations from commercial online services, and with or without charge from many sources on the Internet, especially the Web.

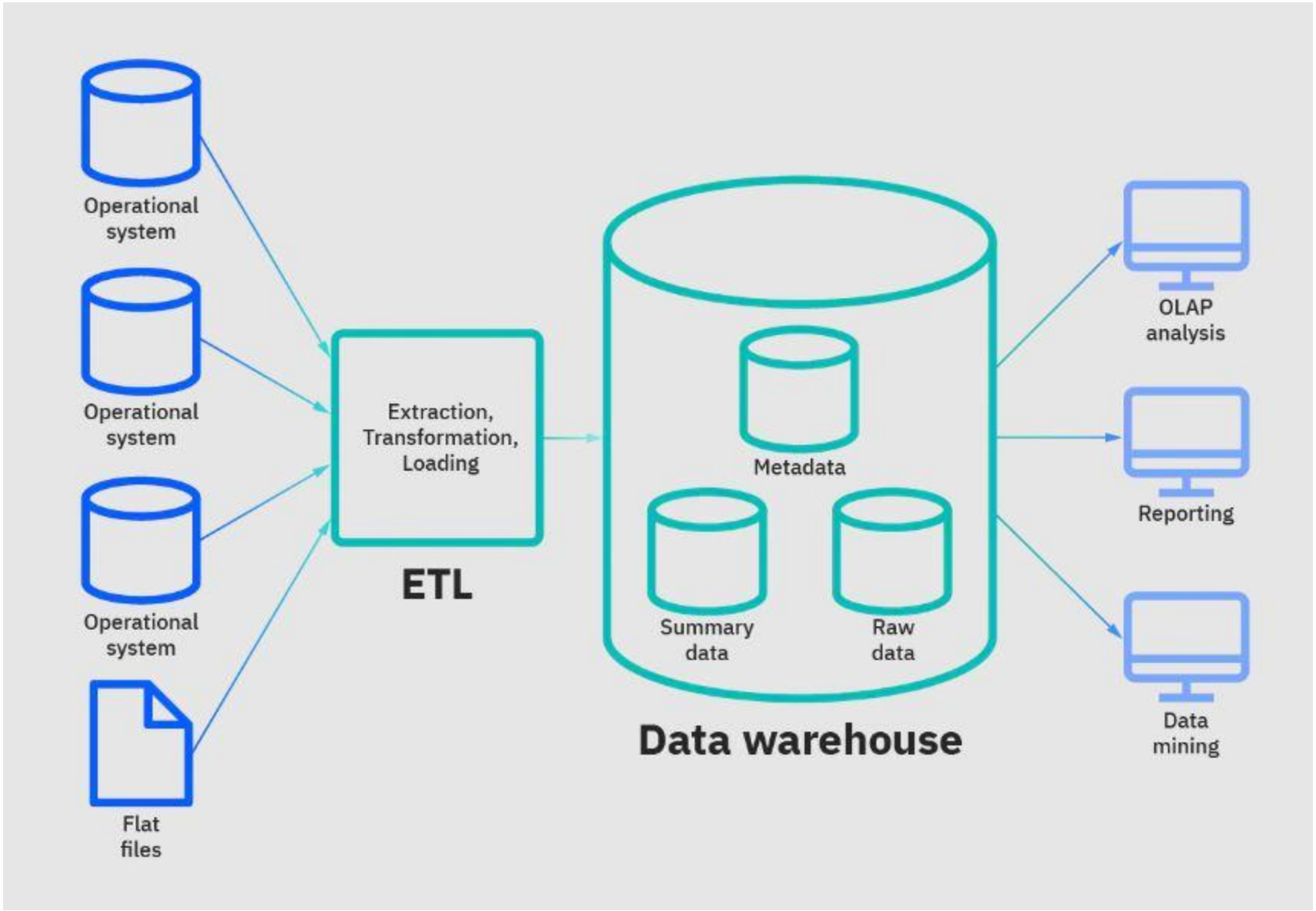
Distributed Databases:

- Many organizations replicate and distribute copies or parts of databases to network servers at a variety of sites. These distributed databases can reside on network servers on the World Wide Web, on corporate Intranets or extranets, or on other company networks. Distributed databases may be copies of operational or analytical databases, hypermedia or discussion databases, or any other type of database. Replication and distribution of databases is done to improve database performance and security.

Hypermedia Databases

- Consists of hyperlinked pages of multimedia (text, graphics, and photographic images, video clips, audio segments, etc.). From a database management point of view, the set of interconnected multimedia pages at a website is a database of interrelated hypermedia page elements, rather than interrelated data records.

Data Warehouse



Data warehousing and data mining:

Data Warehouse

- A data warehouse stores data from current and previous years that has been extracted from the various operational and management databases of an organization. It becomes a central source of data, which has been screened, edited, standardized, and integrated so it can be used by managers and other end user professionals throughout an organization. Data warehouses may be subdivided into data marts, which hold specific subsets of data from the warehouse.

Advantage of Data Warehouse

- **Single source of truth**: A data warehouse applies common standards (formatting processes etc. that are developed by the enterprise) to the data collected from different sources. This ensures that all information is high quality and that there are not duplicates or errors that could impact analysis.

- **Faster operations**: Since the data warehouse consolidates subject area-specific data in a single place with consistency and quality, retrieving data for downstream analytics and reporting becomes simpler, translating into faster decision-making. Without this, enterprise users would be required to log into every individual department system and take the help of IT to consolidate data and generate reports.

- **Historical data**: A data warehouse can keep historical records, eliminating the need to keep hard copies of reports. Enterprise users can easily leverage this to assess changes over time to base present and future decisions.

- **Protection from source system updates**: As the data is consolidated in the warehouse, source system updates will not lead to loss of data or the knowledge it brings. The information would be untouched in the warehouse.

- **Better returns**: The improved efficiency in terms of deriving insights and taking decisions also helps with better business outcomes in the long run. Organizations using a data warehouse can quickly identify trends to capitalize on business opportunities — such as overwhelming demand for a particular product at a particular store/region — and take necessary actions to improve supply, revenue and profits. Similarly, it can also help cut costs.

- **Security and regulatory compliance**: A data warehouse can also help ensure security and regulatory compliance with strong levels of data encryption, such as AES256 and features for access control, user authentication and identity management. Early-stage companies find it hard to set up these layers for different application systems.

- **Cloud benefits**: Initially, data warehouses were built on-premise, which required massive investments in hardware, IT teams and training on all aspects related to the repository. Now, companies have the option of cloud-based data warehouse and switch to a pay-as-you-go model where they would be charged only for the storage and compute used. Everything else would be handled by the cloud provider.

Data Mining:


Business Goal
Present Data
Data Mining
Identify Data
Model & Evaluate Data
Prepare Data

- A major use of data warehouse databases is data mining. In data mining, the data in a data warehouse are processed to identify key factors and trends in historical patterns of business activity that can be used to help managers make decisions about strategic changes in business operations to gain competitive advantages in the marketplace.

BASIS FOR COMPARISON	OLTP	OLAP
Basic	It is an online transactional system and manages database modification.	It is an online data retrieving and data analysis system.
Focus	Insert, Update, Delete information from the database.	Extract data for analyzing that helps in decision making.
Data	OLTP and its transactions are the original source of data.	Different OLTPs database becomes the source of data for OLAP.
Transaction	OLTP has short transactions.	OLAP has long transactions.
Time	The processing time of a transaction is comparatively less in OLTP.	The processing time of a transaction is comparatively more in OLAP.
Queries	Simpler queries.	Complex queries.
Normalization	Tables in OLTP database are normalized (3NF).	Tables in OLAP database are not normalized.
Integrity	OLTP database must maintain data integrity constraint.	OLAP database does not get frequently modified. Hence, data integrity is not affected.

- The database management approach

  The development of databases and database management software is the foundation of modern methods of managing organizational data.

  Database Management Approach - Is a method whereby data records and objects are consolidated into databases that can be accessed by many different application programs.

Database management system:

- Database Management Software - (DBMS) serves as a software interface between users and databases. Thus, database management involves the use of database management software to control how databases are created, interrogated, and maintained to provide information needed by end users and their organizations.

The database management approach involves three basic activities:

- **Updating and maintaining** common databases to reflect new business transactions and other events requiring changes to an organization's records.

- **Providing information needed for each end user's** application by using application programs that share the data in common databases.

- **Providing an inquiry/response and reporting capability** through a DBMS package so that end users can easily interrogate databases, generate reports, and receive quick responses to their ad hoc requests for information.

A database management system (DBMS) is a set of computer programs that controls the creation, maintenance, and use of the databases of an organization and its end users.

The four major uses of a DBMS include:

- Database development
- Database interrogation
- Database maintenance
- Application development

Database development

- Database management packages allow end users to develop their own databases.

- Large organizations with client/server or mainframe-based systems usually place control of enterprise database development with database administrators (DBAs) and other database specialists. This improves the integrity and security of organizational databases.

- In database development a data definition language (DDL) is used to develop and specify the data contents, relationships, and structures of each database, and to modify these database specifications when necessary. Such information is catalogued and stored in a database of data definitions and specifications called a data dictionary, which is maintained by the DBA.

Database interrogation:

- Database interrogation means for selecting a database entry for which data is stored within the database in a plurality of fields. End users can use a DBMS by asking for information from a database using a query language or a report generator.

Features of a query language/ report generator:

- Users receive an immediate response in the form of video displays or printed reports
- No difficult program is required
- Users can obtain immediate responses to as needed data requests.

SQL Queries - SQL, or Structured Query Language, is a query language found in many database management packages. It is used to obtaining immediate responses to ad hoc inquiries.

Basic Form of an SQL query is: SELECT..... FROM.....WHERE

- QBE, or Query by Example, is another form of query language found in some database management packages. The QBE method displays boxes for each of the data fields in one or more files. The end user simply “points-and-clicks” to indicate which information they want.

- Graphical and Natural Queries - Most end user database management packages offer GUI (graphical user interface) point-and-click methods to query a database. These methods are easy to use and are translated by the software into SQL commands. Other packages are available that use natural language query statements similar to conversational English

Database maintenance:

- Managers need accurate information in order to make effective decisions. The more accurate, relevant, and timely the information, the better-informed management will be when making decisions. Thus, the databases of an organization need to be updated continually to reflect new business transactions and other events. This database maintenance process is accomplished by transaction processing programs and other end user application packages with the support of the DBMS.

Application development:

- DBMS packages play a major role in application development.
- Application development is made easier by data manipulation language (DML) statements, which can be included in application programs to let the DBMS perform the necessary data handling activities.
- Programmers can also use the internal programming language provided by many DBMS packages or a built-in application generator to develop complex application programs.