DATABASE - 1

3RD CLASS

COMPUTER SCIENCE DEPARTMENT

11th Lecture – File Organization, and Designing Database

Sunday 1st of December 2024

LECTURER :

DR. RAYAN YOUSIF ALKHAYAT

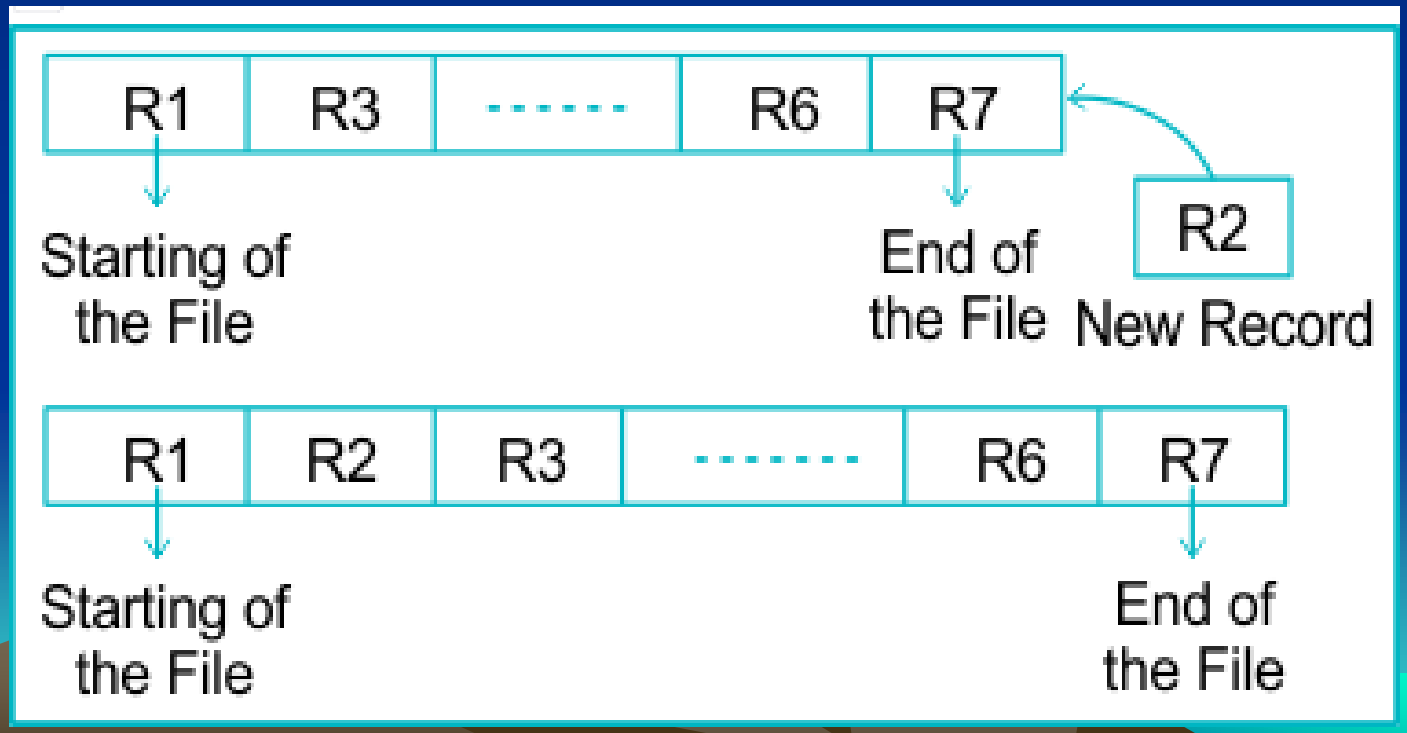
File Organizations

- A file organization is a technique for physically arranging the records of a file on secondary storage devices.
- In choosing a file organization for a particular file in a database, you should consider seven important factors

1. Fast data retrieval
2. High throughput for processing data input and maintenance transactions.
3. Efficient use of storage space .
4. Protection from failures or data loss.
5. Minimizing need for reorganization.
6. Accommodating growth.
7. Security front unauthorized access.

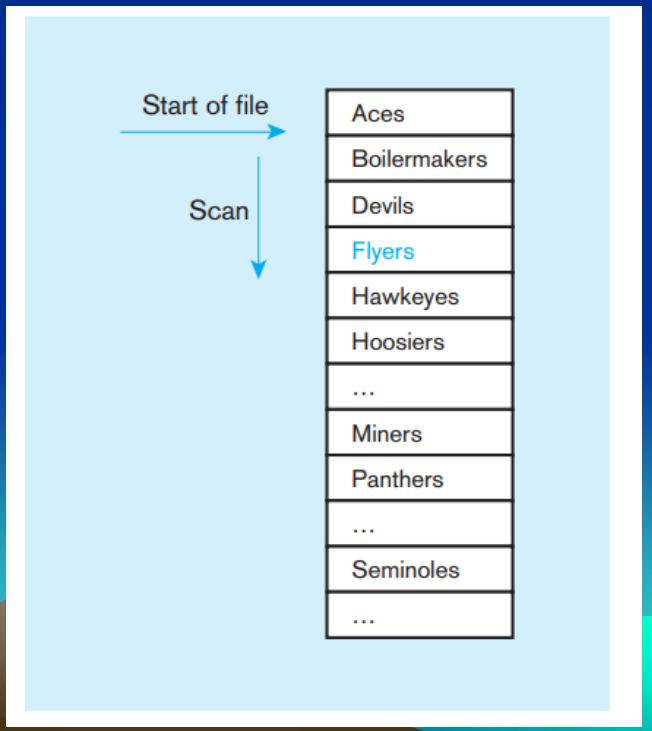
(A). Sequential File Organizations:

- In sequential file organization, the records in the file are stored in sequence according to a primary key value
-



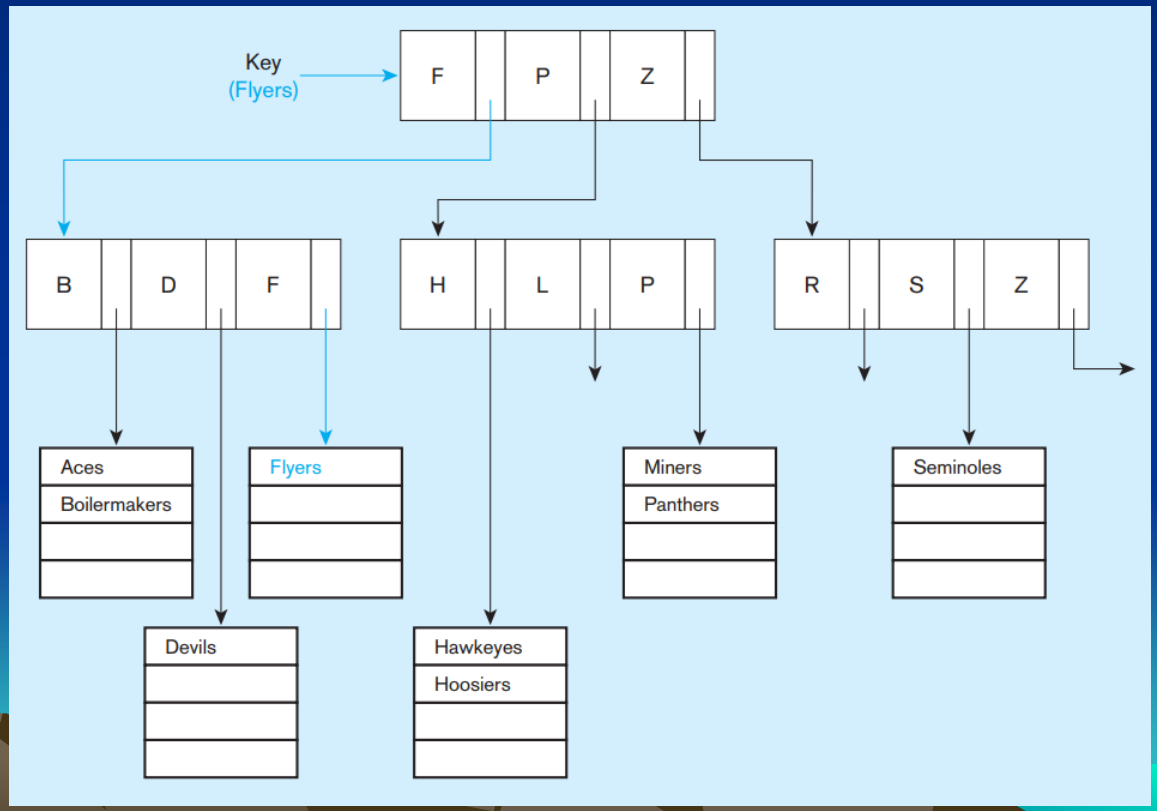
(A). Sequential File Organizations:

- To locate a particular record, a program must normally scan the file from the beginning until the desired record is located.
- A common example of a sequential file is the alphabetical list of persons in the white pages of a telephone directory.

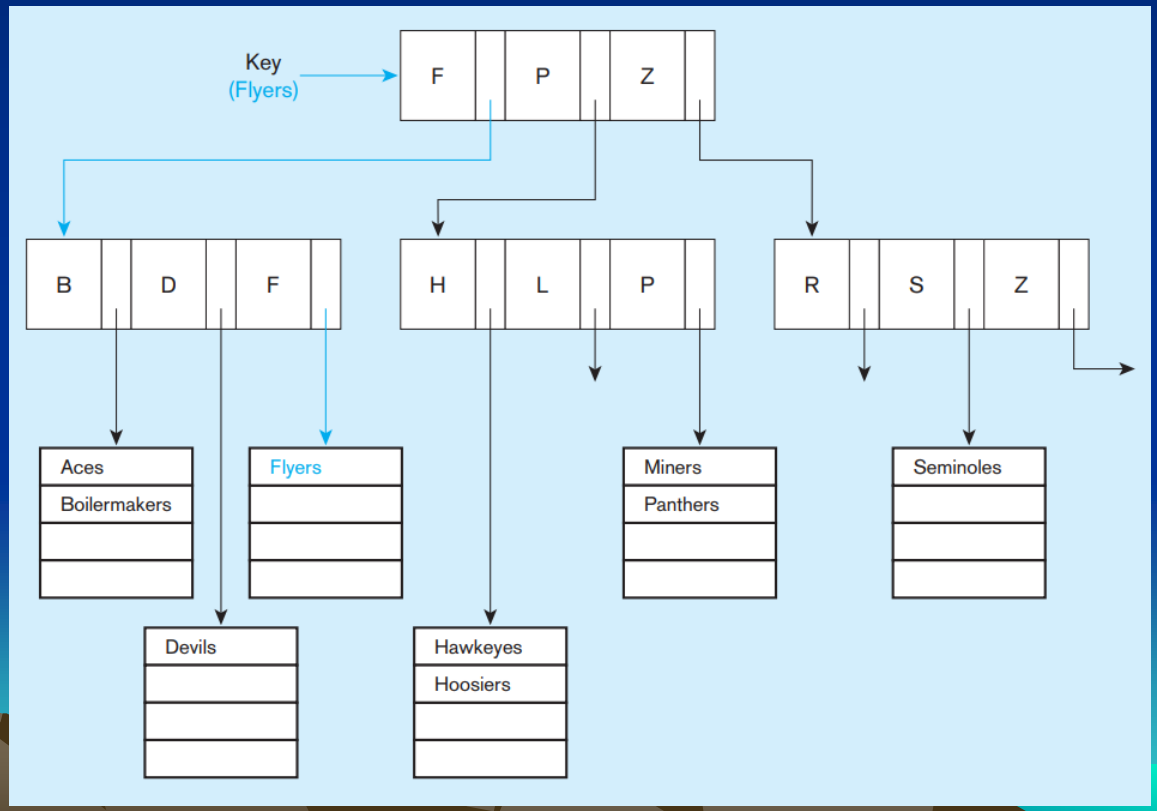


(B). Indexed File Organizations:

- In the indexed file organization, the records are stored either sequentially or nonsequentially and an index is created that allows the application software to locate individual records

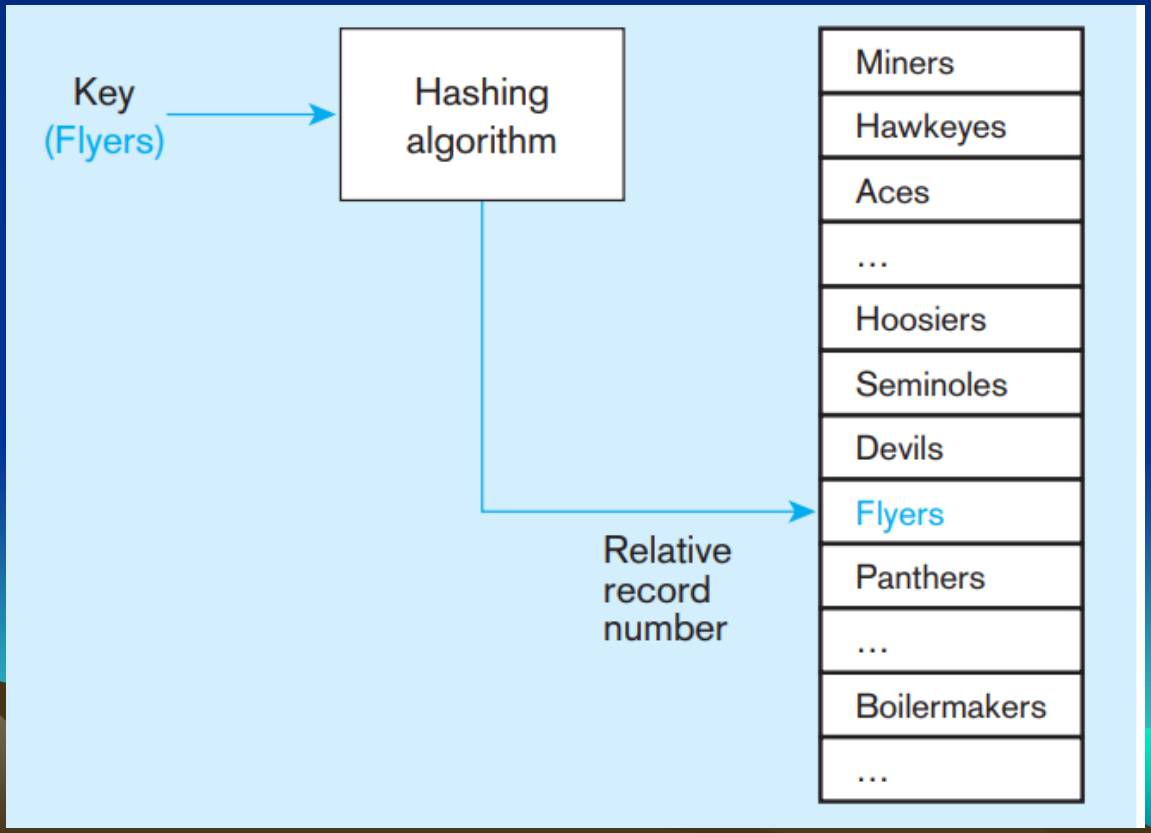


- Like a card catalog in A library in index is a table that used to determine the location of rows in a file that satisfy some condition.
- Each index entry matches a key value with one or more records.
- An index can point to unique records (a primary key index, such as on the Product_ID field of a product record) or to potentially more than one record.
- An index that allows each early to point to more than one record is called a secondary key index.
- Secondary key indexes are important for supporting many reporting requirements and for providing rapid ad hoc data retrieval.

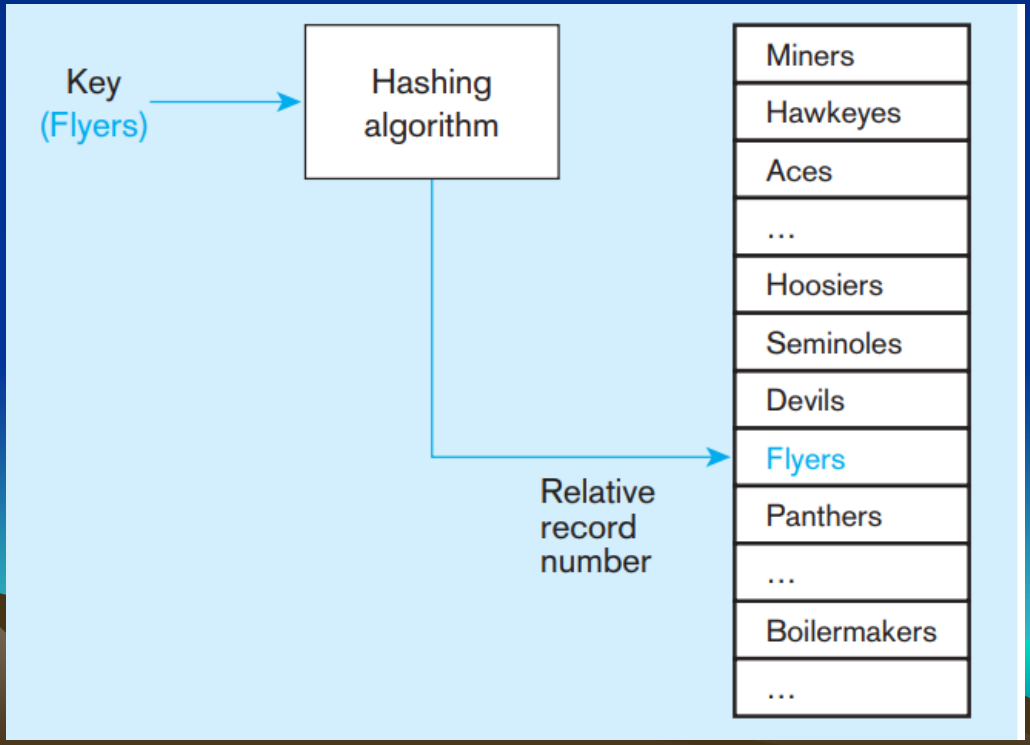


(C). Hashed File Organizations

- In a hashed file organization, the address of each record' is determined using a hashing algorithm .
-



- A hashing algorithm is a routine that converts a primary key value to a record address.
- Although there are several variations of hashed files , in most cases the records are located non sequentially as dictated by the hashing algorithm. Thus, sequential data processing is impractical.
- A typical hashing algorithm uses the technique of dividing each primary key value by a suitable prime number and then using the remainder of the division as the relative storage location.



- For example, suppose that an organization has a set of approximately 1,000 employee records to be stored on drive.
- A suitable prime number would be 997, since it is close to 1,000.
- Now consider the record for employee 12396, when divide this number be 997, the reminder is 432. Thus, this record is stored at location 432 in the file.
- Another technique (not discussed here) must be used to resolve duplicates (or overflow) that can occur with the division/remainder method when two or more keys hash to the same address (known as a "hash clash".)

DESIGNING DATABASE

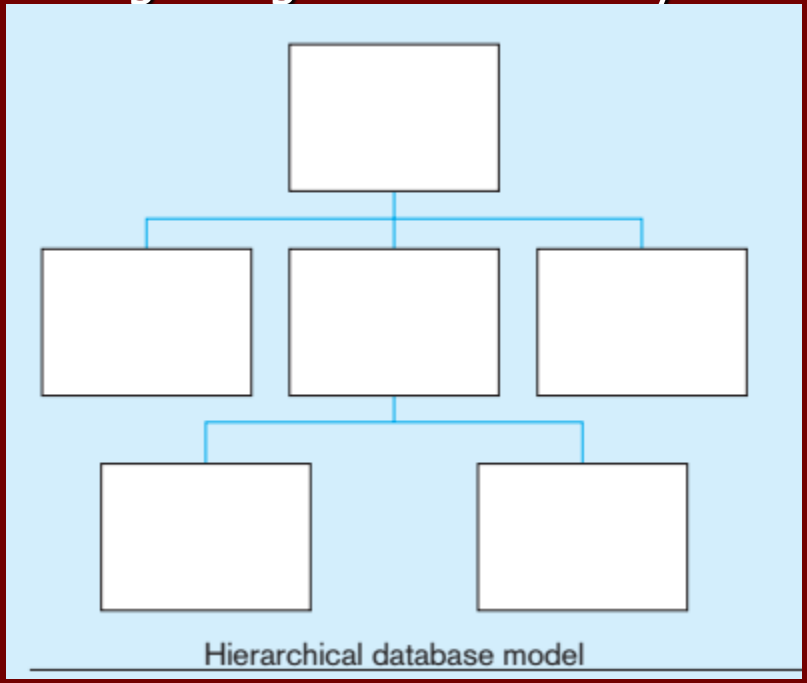
- Most modern information systems utilize database technologies, either database management systems or data warehouse systems, for data storage and retrieval.
- Recall that a database is a collection of logically related data, designed to meet the information needs of multiple users in an organization.
- The relationship between files in a database is due to relationship. Identified in the conceptual and logical data models.

DESIGNING DATABASE

- The relationships imply access paths between data. Each type of database technology allows different types of access paths.
- So, the process of choosing the appropriate type of DBMS or data warehousing technology is one of matching the needed access path, with the capabilities of the database technology.

1.Hierarchical database model :

- In this model, files are arranged in a top-down structure that resembles a tree or genealogy chart.
- Data are related in a nested, one-to-many set of relationships.
- The top file is called the root, the bottom files are called leaves, and intermediate files have one parent, or owner, file and one or several children files.
- Among the oldest of the database architectures, many hierarchical databases exist in larger organizations today.



1.Hierarchical database model :

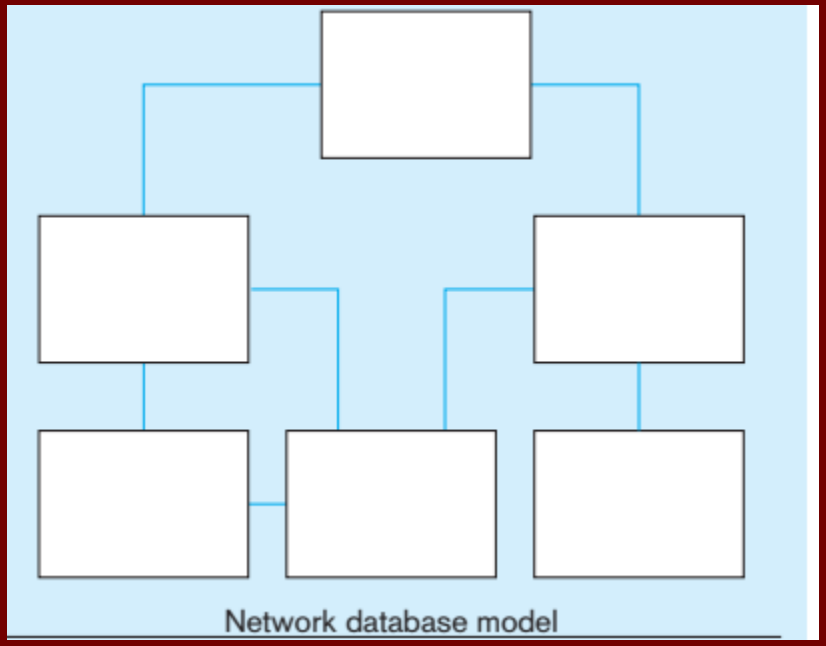
- This technology is best applied when the conceptual data model also resembles a tree and when most data access begins with the same (root) file.
- Hierarchical database technology is used for high-volume transaction processing and MIS applications.
- Few new databases are developed with hierarchical DBMSs since newer applications tend to have broader needs than simply transaction processing or summarization of transaction data. .

2.Network database model:

In this model, each file may be associated with an arbitrary number of files.

Although very flexible because any, relationships can be implemented (a hierarchy is a special case of a network), the form of implementation, usually using pointers; between related records in different files, creates significant overhead in storage space and maintenance time.

Typical network model systems support only one-to-many relationships along each arc in the network, but some support many-to-many relationships.



2.Network database model:

Network model systems are still popular on powerful mainframes and for high-volume transaction processing applications. Since the database designer has such detailed control over data organizations, it is possible to design highly optimized databases with network systems.

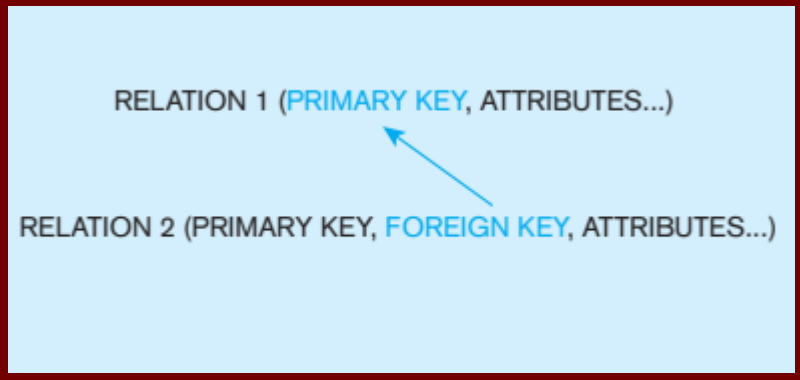
For example, each record type (a record type is shown by each box in the network) can be organized using hashing algorithms or located-near another related record type-an early form of clustering.

2.Network database model:

- Network systems support a wider variety of processing requirements than do hierarchical database systems, but network systems still require significant programming and database design knowledge and time, and hence are used primarily in those organizations with significant expertise with such technologies.

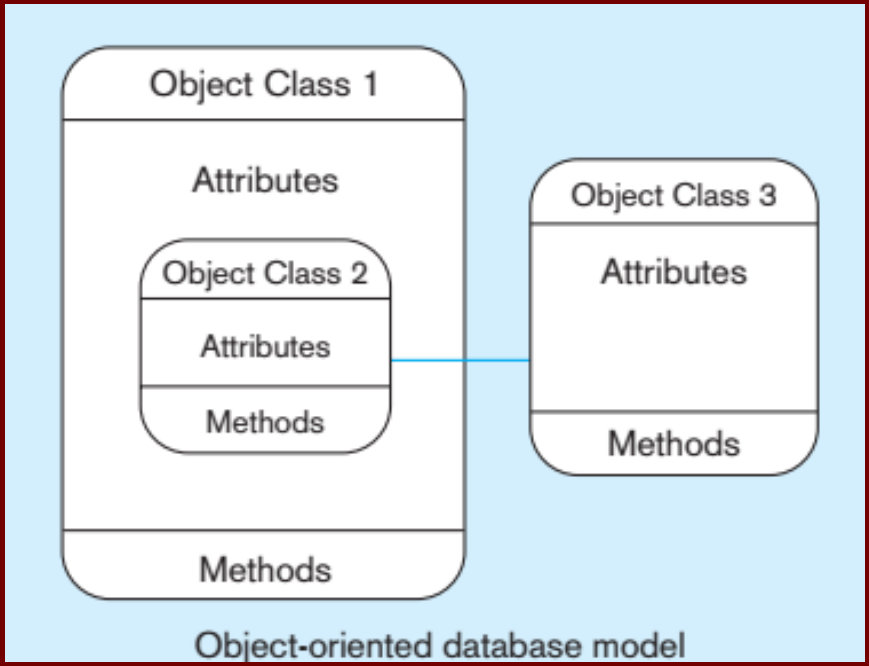
3. Relational database model:

- The most common database model for new systems defines simple tables for each relation and many-to-many relationships.
- Cross-reference keys link the tables together, representing the relationships between entities. Primary and secondary key indexes provide rapid access to data based upon qualifications. Most new applications are built using relational DBMS's, and many relational DBMS;products exist.



4. Object-oriented database model :

- In this model, attributes and methods that operate on those attributes are encapsulated in structures called object classes. Relationships between object classes are shown, in part, be nesting or encapsulating one object class within another object classes are defined from more general object classes.



Object-oriented database model

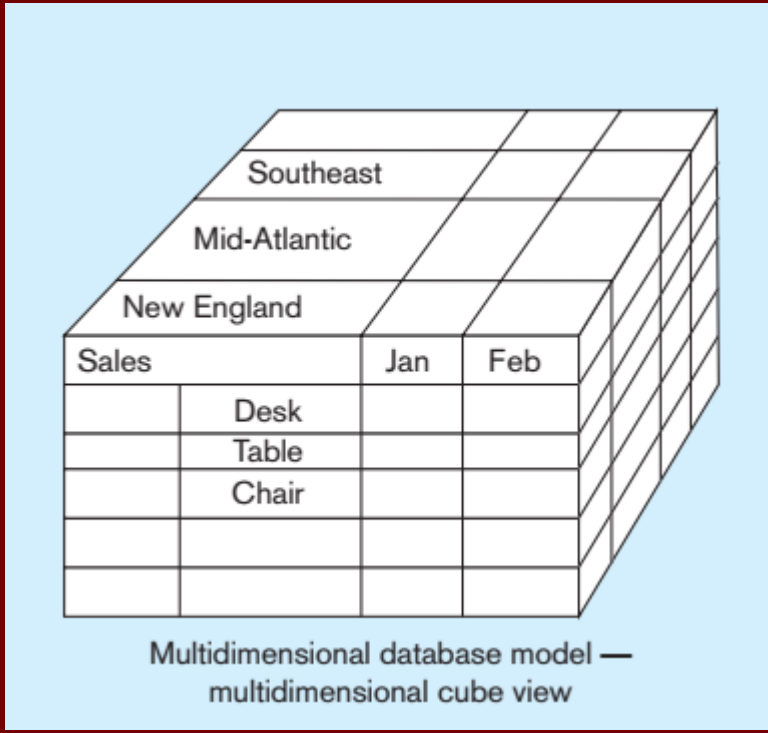
4. Object-oriented database model :

- A major advantage of this data model is that complex data types like graphics, video, and sound are supported as easily a; simpler data types.
- This is the newest DBMS technology and larger organizations are gaining experience with it by selectively using it when complex data or event-driven programming is appropriate for the application.

5. Multidimensional database model:

- This database model is used in data warehousing applications.
- Two ways of viewing this model exist:

- The first views data as a multidimensional table in which each cell contains one or more simple attributes and the dimensions are ways to categorize the raw data.
- These categories, or dimensions, are the factors on which users want to summarize or segment the data, such as time period, geography, lines of business, or people



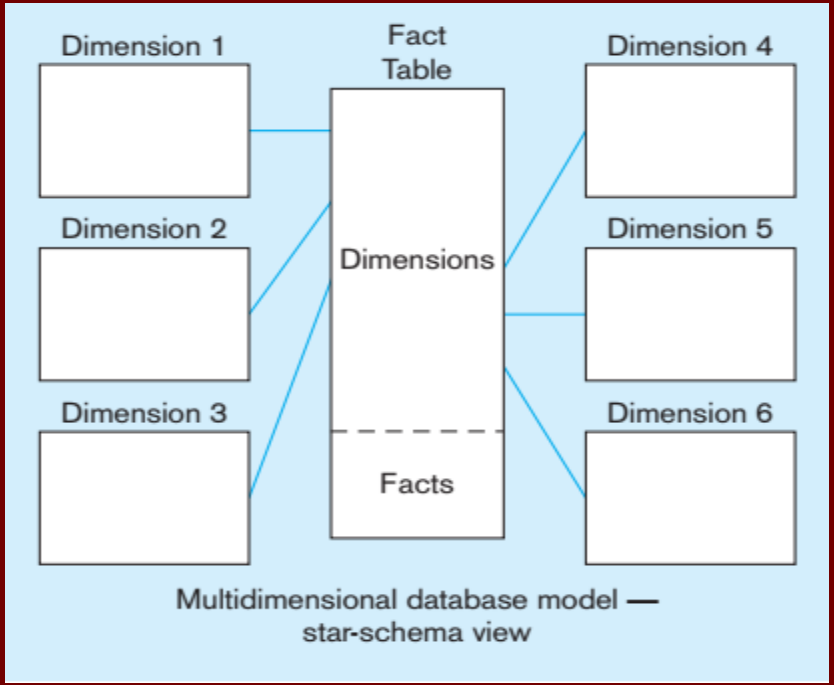
Multidimensional database model — multidimensional cube view

- The second (equivalent) view is called a star schema.

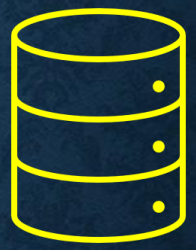
At the center is a fact table, equivalent to the cell in the multidimensional view.

This table contains all the raw attributes and a composite key made up of the primary keys of all the surrounding dimension tables.

The surrounding dimension tables define each of the ways to categorize data, such as all the description data about each salesperson.



Multidimensional database model — star-schema view



END OF LECTURE 11