• Chapter 03 : Database Design Process

3.1 Overview

Q : 03-01-01 : Define Feasibility Study, Requirements Analysis, Project Planning and Data Analysis ?

Answer :

Feasibility Study : This is also called preliminary investigation of the required database. It involves the area identification and selection i.e. which area or aspect is to be selected to start with. After the project is selected, it is allocated a specific find and a proper planning is chalked out for its practical implementation. Side by side, a proper market analysis is also worked out.

Requirements Analysis : During this activity, the requirements are gathered i.e. the possible inputs for the database and the required functionality out of it. The users precisely narrate their needs of the database and the possible domain and restrictions are also chalked out.

Project Planning : A proper schedule is laid down to accomplish this activity. All the cost factors are taken into consideration i.e., the salaries of team members, theft logistics involved, other trivial expenses (such as marriage gifts, insurances etc) and hardware costs.

Data Analysis : This is an important analysis aspect while designing a database. It involves :

Data Flow Diagrams (DFD)
Decision Tables
Decision Trees

3.2 Data Modeling

Q : 03-02-01 : Describe Ingredients of Data Modeling ?

Q : 03-02-01 : Describe Entities / Objects, Attributes , Relationships, Cardinality, Modality and ERD (Entity Relationship Diagram) ?

Answer :

Ingredients of Data Modeling :

Entities / Objects : [A data entity / object is anything that is participating in the system. It is always properly identifiable] i.e., a TEACHER, a STUDENT, an AEROPLANE.

Attributes : [Attributes define the objects, describe their characteristics and in some cases, make references to other objects(s)] i.e., attributes for a TEACHER could be: Teacher Name, Gender, Last Degree, Appointment Date, Pay Scale, Nationality, Telephone No. etc.

Relationships : [The relationship indicates how the Entities/Objects are Connected or Related to each other]. The Data objects are related / connected to one another in different ways. Important to note are :

All the relationships define the relevant connections between both objects.
All the relationships are bi-directional.
We have to consider only the relevant relationship (in the context of the requirement).

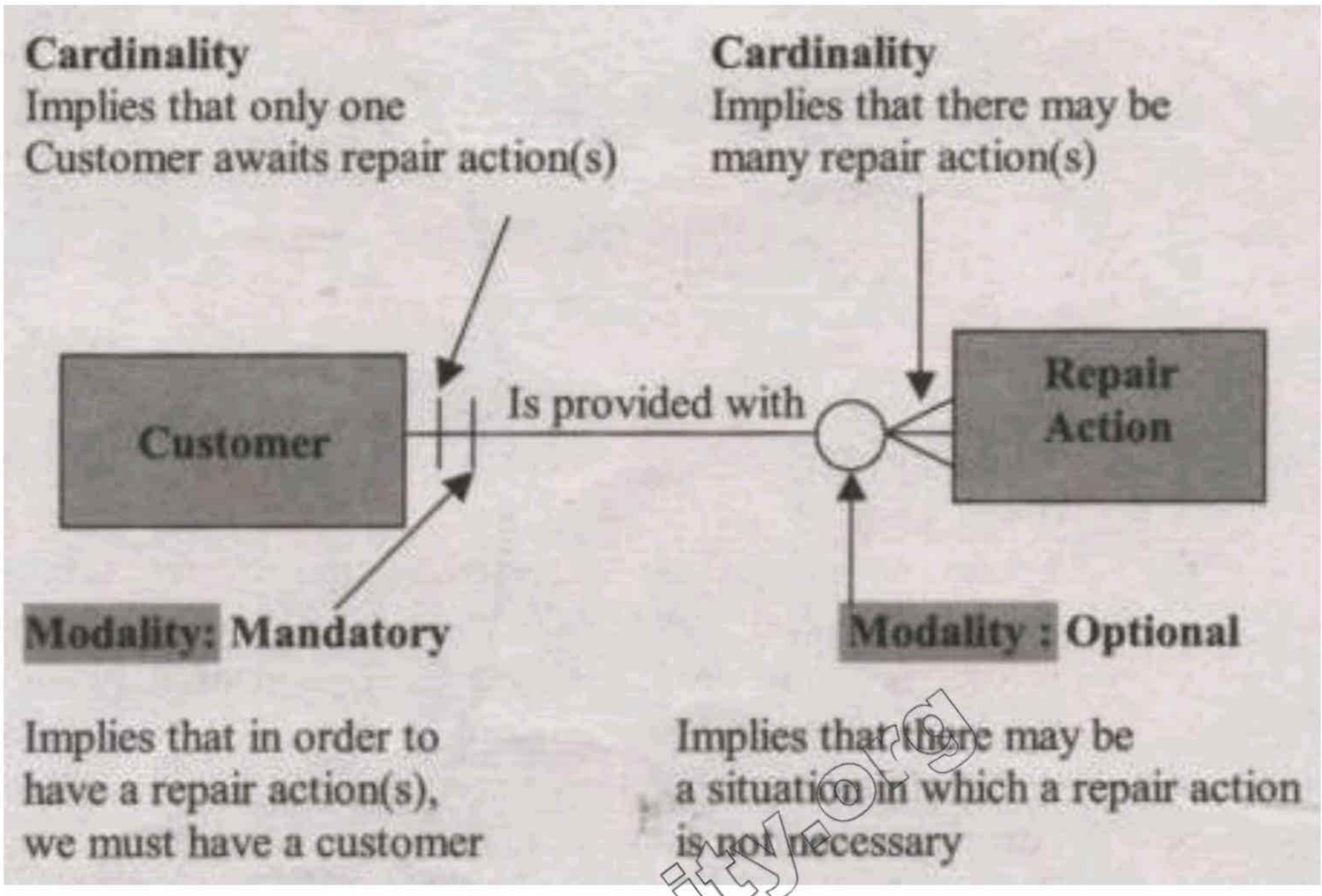
Cardinality : Whether some occurrence(s) of object-1 are related to some occurrence(s) of object-2. It is expressed as one or many. A relationships can be One to One, One to Many, Many to Many, Recursive and None.

Modality : It defines the nature of the relationship :

Optional …….. represented by 0

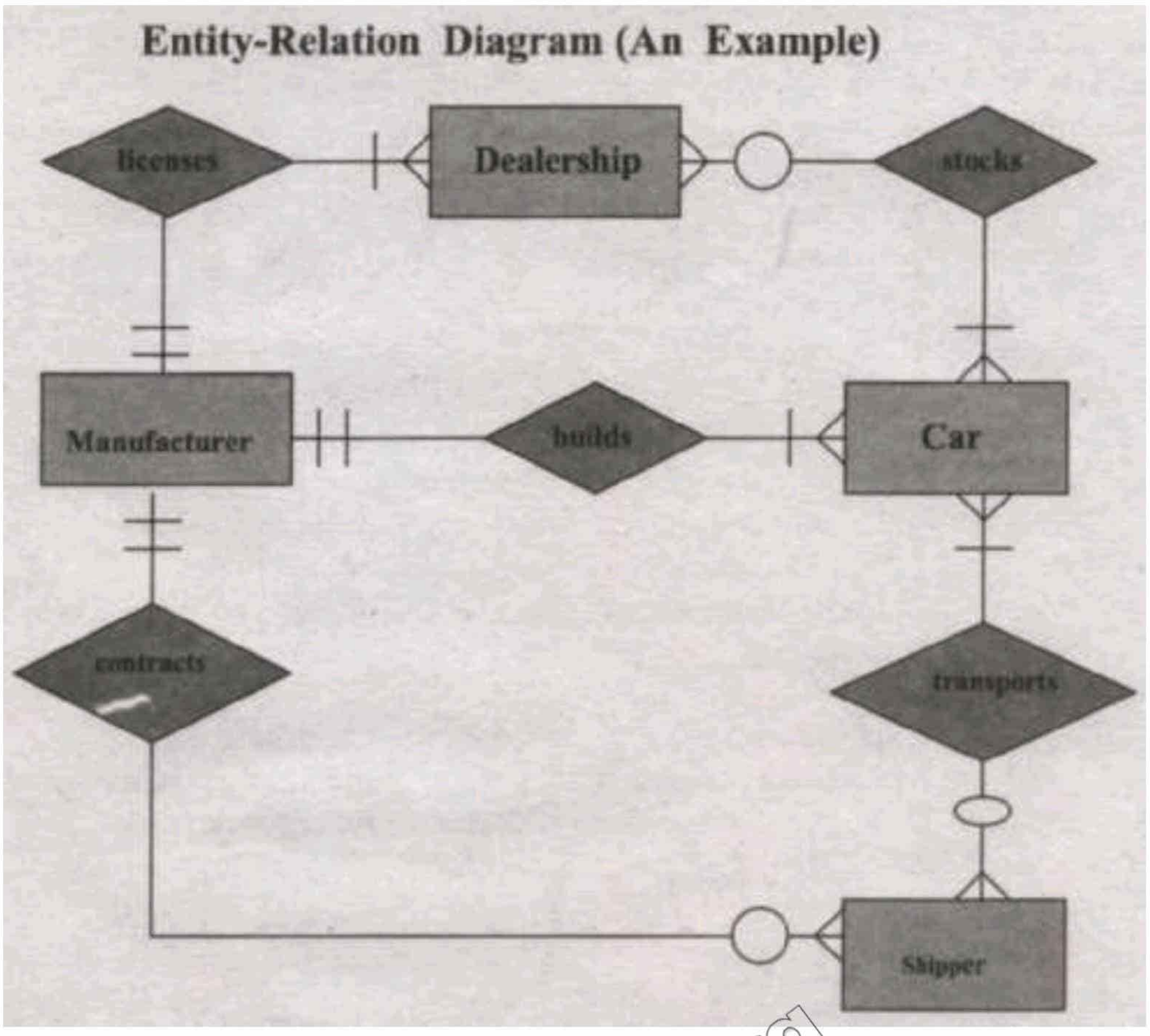
Mandatory …….. represented by 1

Consider two objects Customer and Repair Action in a Workshop environment :



A simple Data Model can be drawn from the above as :

By connecting all the Data Objects along with their Relationships in the above manner, an ERD (Entity Relationship Diagram) is constructed.

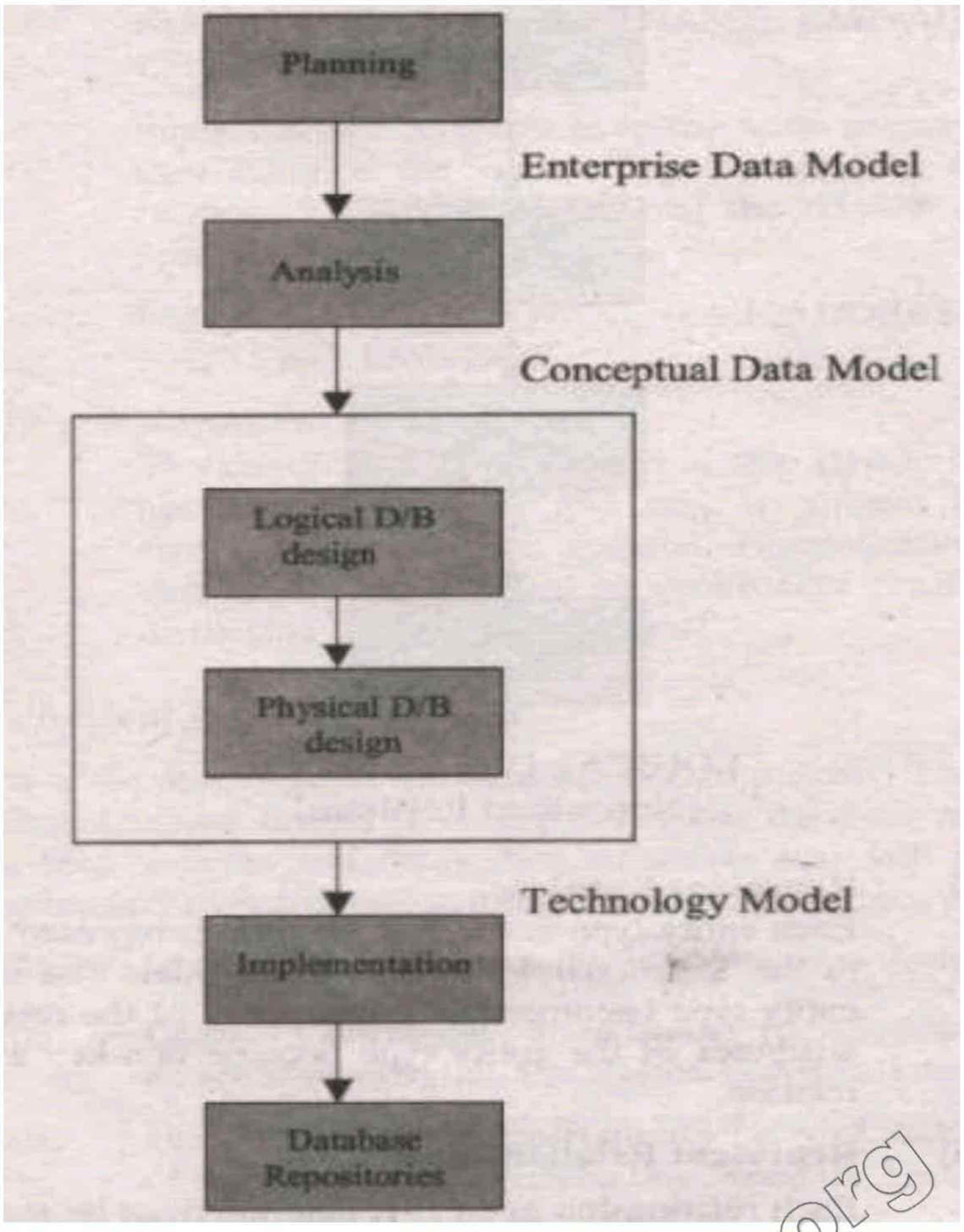


3.3 Database Design

Q : 03-03-01 : Describe major objective of Database Design ? Explain Database Development process with diagram ?

Answer :

Major Objective of Database Design : [To map the conceptual data model to an implementation model that a particular DBMS can process with performance that is acceptable to all users throughout the organization]. Today, database users require information that is complete and up-to-date and they expect to be able to access this information quickly and easily. The Database Development process in diagram :

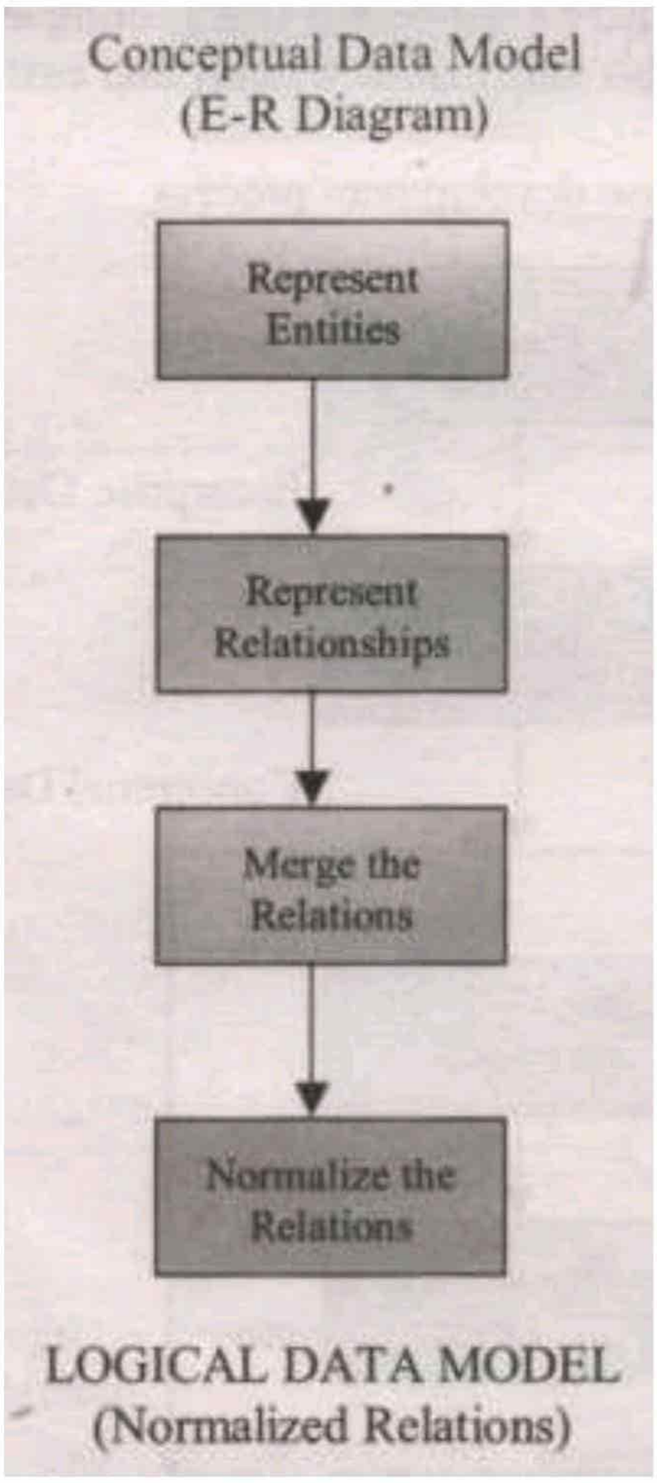


Q : 03-03-02 : Explain Conceptual (Logical) Database Design ?

Answer :

Conceptual (Logical) Database Design : [The process of mapping the conceptual data models (from analysis) to structures that are specific to the target DBMS]. If the target environment is a relational DBMS, then the conceptual data models are mapped to normalized relations. Diagram below presents an overview of logical design process :

Conceptual Data Model
(E-R Diagram)



Represent Entities : Each entity type in the E-R diagram is represented as a relation in the Relational View or Data Model. The identifier of the entity type becomes the Primary key of the relation, and other attributes of the entity type become non-key attributes of the relation.

Represent Relationships : Each relationship in an E-R diagram must be represented in the relational model. It depends on its nature. For example, in some cases, we represent a relationship by making the primary key of one relation a foreign key of another relation. In other cases, we create a separate relation to represent a relationship.

Merge the Relations : In some cases, there may be redundant relations (that is, two or more relations that describe the same entity type). They must be merged to remove the redundancy. This process is also known as View Integration. Suppose we have one relation as :

EMPLOYEE1(EMPNO , NAME, ADDRESS, PHONE)

And another relation as :

EMPLOYEE2(EMPNO, ENAME, EMP-ADDR, EMP-JOBCODE, EMP-DOB)

Since the two relations have the same primary key they describe the same entity and may be merged into one relation. The result of merging the relations is the following relation.

EMPLOYEE(EMPNO , NAME, ADDRESS, PHONE, EMPJOB-CODE, EMP-DOB).

Normalize the Relations : The relations that are created in step (i) and (ii) may have unnecessary redundancy and may be subject to anomalies (or errors) when they are updated. Normalization is the process that refines the relations to avoid these problems.

Q : 03-03-03 : Explain Physical Database Design ?

Answer :

Physical Database Design : [To implement the database as a set of stored records, files, indexes and other data structures that will provide adequate performance and ensure database integrity, security and recoverability].

There are three major inputs to Physical database design :

Logical Database Structures : Developed during logical database design i.e., the Normalized Relations.

User Processing Requirements : The size and frequency of use of the Data-Base, response time, security, backup, recovery etc.

Characteristics of the DBMS and other components of the computer Operating environment.

Q : 03-03-04 : Explain Components of Physical Database Design ?

Answer :

Components of Physical Database Design :

Data Volume and Usage Analysis : To estimate the size or volume and the usage patterns of the database. Estimates of database size are used to select Physical storage devices and estimate the costs of storage. Estimates of usage paths or patterns are used to select the file organization and access methods, to plan for the use of indexes and to plan a strategy for data distribution.

Data Distribution Strategy : Many organizations today have distributed computing networks. For these organizations, a significant problem in physical database design is deciding at which nodes (or sites) in the network to physically locate the data. Basic data Distribution Strategies are :

Centralized : All data are located at a single site. It is fairly easy to do but it has at least three disadvantages :

Data are not readily accessible at remote sites. Data communication costs may be high.

The database system fails totally when the central system fails.

Partitioned : The database is divided into partitions (fragments). Each partition is assigned to a particular site. Major advantage of this is that data is moved closer to local users and so is more accessable.

Replicated : Full copy of database is assigned to more than one site in the network. This approach maximizes local access but creates update problems, since each database change must be reliably processed and synchronized at all of the sites.

Hybrid : In this strategy, the database is partitioned into critical and non-critical fragments. Non-critical fragments are stored at only one site, while critical fragments are stored at multiple sites.

File Organization : A technique for physically arranging the records of a file on secondary storage devices. For selecting a file organization, the system designer must recognize several constraints, including the physical characteristics of the secondary storage devices, available operating systems and file management software, and user needs for storing and accessing data. Following is the criteria for selecting file organizations :

Fact access for retrieval.

High throughput for processing transactions.
Efficient use of storage space.
Protection from failure or data loss.
Minimizing need for re-organization.
Accommodating growth.
Security from un-authorized use.

Indexes : An index is a table that is used to determine the location of rows in a table (or tables) that satis1' some condition. They may be created on primary key, secondary key, foreign key etc.

Integrity Constraints : Database integrity refers to the correctness and consistency of data. It is another form of database protection. While it is related to security and precision, it has some broader implications. Security involves protecting the data from unauthorized operations, while integrity is concerned with the quality of data itself Integrity is usually expressed in terms of certain constraints which are the consistency rules that the database is not permitted to violate.

3.4 Implementation

Q : 03-04-01 : Explain Database Implementation Phase ?

Answer :

Database Implementation Phase : The builder or the database administrator normally requires a server computer which will be linked with hundreds and thousands of computer users who would want to share and interact with the server (database). For this purpose, the DBA might need the services of network administrators to connect the users with the server. The users are normally given the authorizations / permissions defined by their respective managers so that they can perform the authorized tasks while using the database facilities. In distributed computing environment, the database servers and users might be thousands of kilometers apart, so a lot of expensive telecommunication links are required to perform the designated tasks. NADRA and CRICKINFO are some of the typical examples of this type of databases.