① Designing Data Warehouse :-

A data warehouse is a heterogeneous collection of different data sources organised under unified schema. Some characteristic of Data warehouse are

- Subject Oriented
- Integrated
- Time Varient
- Non-Volatile

Now, some steps that are needed for building any data warehouse are →

i) To extract data from different sources &

For building or designing a data warehouse a data is extracted from various data sources and that data is stored in central storage area.

ii) To transform the transnational data :-

There are various DRMS where many of the companies stores their data. Some of them are : MS Access, MS SQL server, Oracle, Sybase etc. Also these companies stores the data in spreadsheet, flat files, mail system etc. Relating a data from all these sources is done while building a data warehouse.

iii) To load data into dimensional databases &

After building a dimensional model, the data is loaded in the dimensional database. This process combines the several columns together or it may split one field into several column. There are two stages at which transformation of the data column together or it may split one field into the several columns.

ii) To Purchase a front end reporting tool f

There are top notch analytical tools are available in the market. These tools are provided by the several major vendors. A cost effective tool and Data Analyzer is released by Microsoft on it own.

Conversion of the data might be done from object oriented, relational or legacy database to a multidimensional model. One of the largest labor demanding component of data warehouse designing is data cleaning, which is one of the complex process.
The only feasible and better approach for it is incremental updating. Data storage in the data warehouse:
- Refresh the data.
- To provide the time variant data.

2/14

- To store the data as per the data model of the warehouse.
- Purging the data.
- To support the updating of the warehouse data.

Some of the important designs for the data warehouses are:

- Modular Component design.
- Consideration of the parallel architecture.
- Consideration of the distributed architecture.
- Usage protection.
- Characteristic of available sources.
- Design of the metadata Component.
- The fit of the data model.

The major determining characteristic for the design of the warehouse is the architecture of the organization distributed computing enviroment. The distributed warehouse and the federated warehouse are the two basic distributed architecture. There are some benifits from the distributed warehouse, some of them are:

- Improved load balancing

- Scalability of performance

- Higher availability.

② a) Creat function &

```
Create Or Replace Function CUSTM ID (Cust Id
    CUST MASTER (CUST.ID%. Type) IS

C_Id CUST MASTER CUST_ID%. Type;

Begin
    Select Cust ID into C_Id from
    CUST MASTER Where CUST_ID = Cust_Id;
    dbms.output.put line ('Yes!, The Customer
    exists');

    Exception
        When No date found Then
        dbms.output.put line (' The
        Customer doen't exist');
END CUSTM_ID;
```

b) Drop function :-

```
Drop function CUSTM ID;
```

c) Create Package :-

```
Creat Package myCustomer As Function
CUSTM_ID (Cust Id CUST MASTER.Cust_Id% Type);

END myCustomer;
```

Package Body

```sql
Create On Replace Package Body myCustomer
AS
Function CUSTM ID (Cust Id CUST MASTER. Cust_Id%. Type)
IS
C Id   CUST MASTER. Cust_Id%. Type;

Begin
    Select Cust Id Into C.ID From
    CUST MASTER where Cust Id = Cust_Id;
    dbms output put line ('Yes!, The Customer
    exists');

    Exception
        When no data found Then
        dbms output put line (' The Customer
        does not exist');
End CUSTM ID
End my Customer;
```

d) Update Trigger

```sql
Create On Replace Trigger Track Values
Before Update On Cust MASTER
For each row
When (New Cust_Id > 0)
Begin
    Insert Into Cust Old Values Values (Old Cust Id,
    Old Cust Names, Old Phone);
END;
```

(3) I Disadvantages Of Dbms Over Multimedia Dbmsf

i) More Costly :-
Creating and managing a database is quite costly. High cost software and hardware is required for the database. Also highly trained staff is required to handle the database and it also needs continuous maintenance.

ii) High Complexity :-
It is quite complex as it involves creating, modifying and editing a database. Consequently the people who handle a database or work with it need to be quite skilled or valuable data can be lost.

iii) Database Failure :-
All the relavant data for any company is stored in a database. So it is imperative that the database works in optimal condition and there are no failures. A database failure can be catastrophic and can lead to loss or corruption of very important data.

iv) High Hardware Cost :-
A database contains vast amount of data, especially so a large disk storage is required to store all this data.

6/14

v) **Cost of data Conversion :-**

If the database is changed or modified in some manner, all the data needs to be converted to the new form. This cost may even exceed the database creation and management costs sometimes. This is the reason most organisation prefer to work on their old databases rather than upgrade to new ones.

II. Data Handling In Multimedia Databases :-

i) **Object Editing &**

Consider editing multimedia objects. For example, two objects may be merged to form a third object. One can project an object to form a smaller object. As an example. Objects may be merged based on time intervals and an object may be projected based on time intervals.

ii) **Browsing :-**

It is essentially carried out by a hypermedia database management system. The multimedia data is presented in terms of nodes and links. One traverses the links to reach the nodes and clicks on the links to get the relevant multimedia data.

(iii) **Filtering :**

It is the process of removing unnecessary material from data. This occurs quite often in video data where material inappropriate for children may be removed from a video clip. This means that video clips have to be filtered and the filtered data displayed to users.

(iv) **Transaction Management :**

This function is also an aspect of data handling as it involves queuing and updating database. There has been some discussion as to whether transaction management is needed in this.

(4) **Algorithm For Select :-**

There are many algorithms for executing a select operation, which is basically a search operation to locate the records in a disk file that satisfy a certain condition.

OP1: $\sigma_{SSN = '123456789'}$ (Employee)

OP2: $\sigma_{DNumber > 5}$ (Department)

OP3: $\sigma_{DNO = 5}$ (Employee)

OP4: $\sigma_{DNO = 5 \text{ AND } Salary > 500000}$ (Employee)

OP5: $\sigma_{SSN = '12345678 9' \text{ AND } PNO = 10}$ (Works On)

Now here are some example of search method-

- **S1- Linear- Search →** Retrieve every record in the file and test whether its attribute values satisfy the selection Condition

- **S2- Binary Search →** If the selection Condition involves an equality comparison on a key attribute on which the file is ordered binary search can be used.

- **S3- Single Record Search →** If the selection Condition involves an equality comparison on a key attribute with a primary index, use the primary index to retrieve the record.

- **S4- Multiple Search →** If the comparision Condition is >, ≥, < or ≤ on a key field which a primary index, use the index to find the record satisfying the corresponding equality Condition then retrieve all subsequent records in the file.

- **S5- Multiple Search (Clustering) →** If the select Condition involves an = on a non-key

attribute with a clustering index, use the clustering index to retrieve all the records satisfying the selection condition.

○ S6- Range Query →
To be used to retrieve records on conditions involving $>, >=, <$ or $<=$. Can also be used for an $=$ for single record search if the indexing field has unique or to retrieve multiple records if the indexing field is not a key.

○ S7- Conjunctive Selection →
If an attribute involved in any single condition in the conjunctive condition has an access path that permits the use of one of the most methods S2 to S6, use that condition to retrieve the records and then check whether each retrieved record satisfies the remaining simple conditions in the conjunctive condition.

⑤ Data Mining :-
It is the process of uncovering patterns and finding anomalies and relationships in large dataset that can be used to make

Predictions about future trends. The main purpose of data mining is extracting valuable information from available data.

⟹ Applications :-

It offers many applications in business For example, the establishment of proper data mining processes can help a company to decrease its costs, increase revenues or derive insights from the behavior and and practice of its customers.

Data mining is also actively utilized in finance. For instance, relevant techniques allow users to determine and access the factors that influence the price fluctuation of financial securities.

⟹ Process :-

i) Define The Problem →

Determine the scope of the business problem and objectives of the data exploration project.

ii) Explore The Data →

This step includes the exploration and collection of data that will help solve the stated business problem.

11/14

iii) Prepare The Data →
Clean and organized Collected data to prepare it for further modeling procedures.

iv) Modeling →
Create a model using data mining techniques that will help solve the stated problem.

v) Interpretation And Evaluation Of Result →
Draw Conclusions from the data model and assess its validity. Translate the result into a bussiness decision.

⇒ Techniques :-

i) Detection Of Anomalies →
Identifying unusual values in a dataset.

ii) Dependency Modeling → Discovering existing relationships within a dataset. This frequently involves regression analysis.

iii) Clustering →
Identifying structures in unstructured data.

iv) Classification → Generalizing the known structure and applying to data

(6) I Fragmentation :-

It is the task of dividing a table into a set of smaller tables. The subsets of tables are called fragments.

Fragmentation should be done in a way so that the original table can be reconstructed from the fragments. This is needed so that the original table can be reconstructed from the fragments whenever required. This requirement is called reconstructive ness.

Advantages :-

o Since data is stored closed to the site of usage, efficiency of the database system is increased.

o Local query optimization techniques are sufficient for most queries since data is locally available.

o Since irrelevent data is not available at the sites, security and privacy of the database system can be maintained.

Disadvantages :-

o When data from different fragments are required, the access speeds may be high.

- In case of recursive fragmentations, the job of reconstruction will need expensive techniques.

- Lack of back-up copies of data in different sites may render the database ineffective in case of failure of a site.

Thus, this is the main reason that fragmentation and Data Allocation are considered crucial while designing a Distributed Dbms.