Data Warehousing and Data Mining

UNIT-I

DATA WAREHOUSE

Data :- In Computing, data is information that has been translated into a form that is efficient for movement or processing. Relative to today's computers and transmission media, data is information converted into binary digital form. It is acceptable for data to be used as a singular subject or a plural subject. Raw data is a term used describe data in its most basic digital format.

Data Warehouse :- The term "Data warehouse" was first coined by "Bill Inmon in 1990". According to Inmon, a data warehouse is a "subject oriented, integrated time-variant, and non-volatile collection of data." This data helps analysts take informed decisions in an organization.

An operational database undergoes frequent changes on a daily basis on account of the transactions that take place. Suppose a business executive wants to analyze previous feedback on any data such as a product, a supplier, or any consumer data, then the executive will have no data available to analyze because the previous data has been updated due to transactions.

A data warehouses provides us generalized and consolidated data in multidimensional view. Along with generalized and consolidated view of data, a data warehouses also provides us Online Analytical Processing (OLAP) tools.

These tools help us in interactive and effective analysis of data in a multidimensional space. This analysis results in data generalization and data mining.

Data Mining functions such as association, clustering, classification, prediction can be integrated with OLAP operations to enhance the interactive mining of knowledge at multiple level of abstraction. That's why data ware house has now become an important platform for data analysis and online analytical processing.

Understanding a Data Warehouse :

A data Warehouse is a database, which is kept separate from the Organization's Operational database.

* There is no frequent updating done in a DW.
* It possesses consolidated historical data, which helps the organization to analyze its business.
* A data warehouse helps executives to organize, understand, and use their data to take strategic decisions.
* Data warehouse Systems helps in the integration of diversity of application systems.
* A data warehouse System helps in Consolidated historial data analysis.

Why a Data Warehouse is Separated from Operational Databases

1) An Operational database is Constructed for well-known tasks and workloads such as Searching particular records, indexing, etc. In Contract, data warehouse queries are often complex and they present a general form of data.

2) Operational database Support Concurrent processing of multiple transactions. Concurrency Control and Recovery mechanisms are required for Operational database to ensure robustness and consistency of the database.

3) An Operational database query allows to read and modify Operations, while an OLAP query needs only read access of stored data.

4) An Operational database maintains current data On the other hand, a data warehouse maintains historical data.

Data Warehouse Features :

1) **Subject Oriented :** A data warehouse is subject oriented because it provides information around a subject rather than the organization's ongoing operations. These subjects can be product, customers, Suppliers, Sales, revenue, etc. A data warehouse does not focus on the ongoing operations, rather it focuses on modelling and analysis of data for decision making.

2) **Integrated :** A data warehouse is Constructed by integrating data from heterogeneous source. Such as relational database, flat files, etc. This integration enhances the effective analysis of data.

3) **Time Variant :-** The data collected in a data warehouse is identified with a particular time period. The data in a data warehouse provides information from the historical point of view.

4) **Non-volatile :-** Non-volatile means the previous data is not erased when new data is added to it. A data warehouse is kept separate from the operational database and therefore frequent change in ODB is not reflected in DW.

NOTE: A DW does not require transaction processing, recovery, and concurrency controls, because it is physically stored and separate from the operational database.

DW - Applications :-

A datawarehouse serves as a sole part of a plan-execute-assess "closed-loop" feedback system for the enterprise management.

* Financial Services
* Banking Services
* Consumer goods
* Retail Sectors
* Controlled manufacturing

Types of DataWarehouse

Information processing, analytical processing, and data mining are the three type of DW.

1) Information Processing :- A DW allows to process the data stored in it. The data can be processed by means of querying, basic statistical analysis, reporting using crosstabs, tables, charts, or graphs

2) Analytical Processing :- A DW supports analytical processing of the information stored in it. The data can be analyzed by means of basic OLAP operations, including slice-and-dice, drill down, drillup, and pivoting.

3) Data Mining :- DM supports knowledge discovery by finding hidden patterns and associations, constructing analytical models, performing classification and prediction. These mining results can be presented using the visualization tools.

Data Warehouse (OLAP)	Operational Database (OLTP)
1) It involves historical processing of information.	It involves day-to-day processing.
2) OLAP systems are used by knowledge workers such as executives, managers and analysts	OLTP systems are used by clerks, DBAs or database professionals.
3) It is used to analyze the business.	It is used to run the business
4) It focuses on information Out	It focuses on Data in
5) It is based on Star Schema, Snowflake schema, and Fact constellation Schema.	It is based on Entity Relationship Model.
6) It focuses on information Out	It is application oriented
7) It contains historical data	It contains current data
8) It provides summarized and consolidated data.	It provides primitives and highly detailed data.
9) It provides summarized and multidimensional view of data.	It provides detailed and flat relational view of data.
10) The no: of users is in hundreds.	The no: of users is in thousands
11) The no: of records accessed is in millions	The no: of records accessed is in tens
12) The database size is from 100 GB to 100 TB	The database size is from 100 MB to 100 GB
13) These are highly flexible	It provides high performance.

What is Data Warehousing?

A data warehousing is constructed by integrating data from multiple heterogeneous sources that support analytical reporting, structured and/or ad hoc queries, and decision making. Data warehousing involves data cleaning, data integration, and data consolidations.

Functions of Datawarehouse :-

It works as a collection of data and here is organized by various communities that endures the features to recover that data functions. It has stocked backs about the tables which has high transaction levels which are observed so as to define the datawarehousing techniques and major functions which are involved in.

1) Data Consolidation
2) Data cleaning.
3) Data Integration

Feature	OLTP	OLAP
Characteristic	It is a system which is used to manage Opirational Data.	It is a system which is used to manage Informational Data.
Users	Clerks, clients and Information Technology professionals	Knowledge workers including managers, executive and analysts.
System Orientation	OLTP system is a customer oriented. Transaction & query processing are done by clerks clients and information technology professionals	OLAP system is market oriented, knowledge workers including managers, do data analysis executive and analysts.
Data Contents	OLTP system manages current data that too detailed and are used for decision making.	OLAP system manages a large amount of historical data provides facilities for summarization and aggregation and stores and manages data at different levels of granularity.
Data Base	100MB - GB	100GB - TB
Database Design	OLTP system usually uses an entity relationship (ER) data model and application-oriented database design.	OLAP system typically user either a star or snow flake model and subject oriented database design.
Volume of Data	Not very large	Because of their large volume, OLAP data are stored on multiple storage media.
Access Mode	Read/write	Mostly write
Number of Records accessed	Tens	Millions

Data Warehouse - Architecture :-

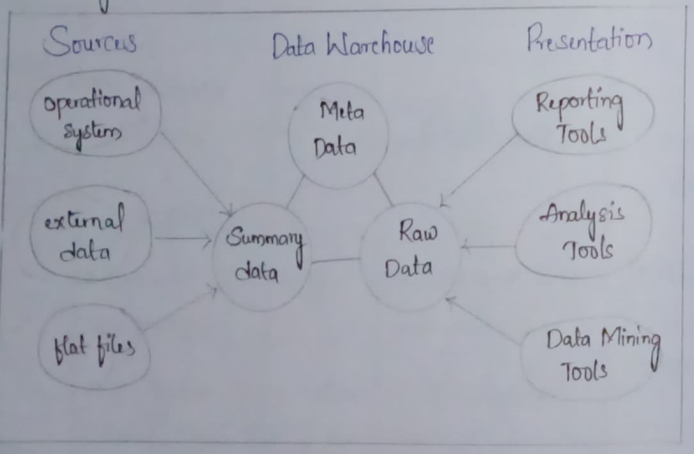
There are 3 ways you can construct a data warehouse System. These approaches are classified by the no: of tiers in the architecture. They are:

1) Single - Tier - Architecture
2) Two - Tier - Architecture
3) Three - Tier - Architecture

Single - Tier Data Warehouse - Architecture :-

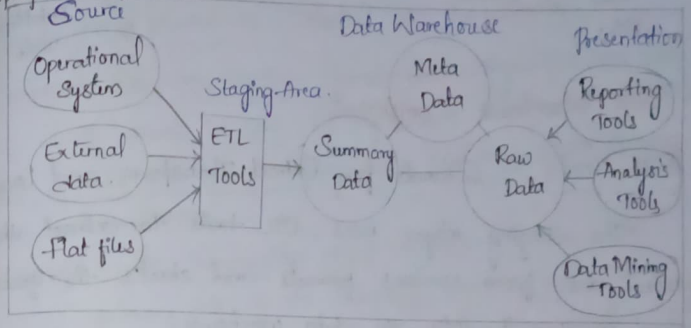
The Single-Tier architecture is not a frequently practiced approach. The main goal of having such an architecture is to remove redundancy by minimizing the amount of data stored.

It's primary dis-advantage is that it doesn't have a component that Spparatis analytical and Transactional processing.



Two - Tier Data Warehouse - Architecture :-

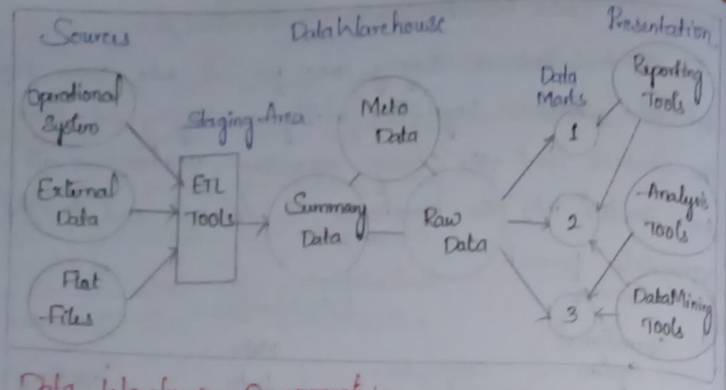
It includes a staging area for all data source, before the data warehouse layer. By adding a staging are between the sources and the storage repository, you ensure all data loaded into the warehouse is cleansed and in the appropriate format.



Three - Tier Data Warehouse - Architecture

It is most widely used architecture for data warehouse Systems.

1) The bottom Tier :- The database of the warehouse, where the cleansed and transformed data is loaded.
2) The middle Tier :- The application layer giving an abstracted view of the database. It arranges the data to make it more suitable for analysis. This is done with an OLAP Server, implimented using the ROLAP or MOLAP model.
3) The Top-Tier :- Is where the user accesses and interacts with the data. It represents the front-end client layer. You can use reporting tools, Query, Analysis or data mining Tools.

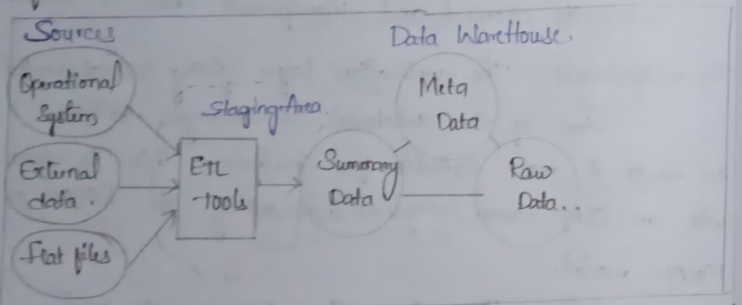


Data Warehouse Components:

ETL Tools :- Stands for "Extract, Transform, and Load."

The staging layer uses ETL tools to extract the needed data from various formats and checks the quality before loading it into the data warehouse.

The data coming from the data source layer can come in a variety of formats from multiple sources into a single database, the system must clean and organize the information.



The Database :- The most crucial component and the heart of each architecture is the database. The warehouse is where the data is stored and accessed.

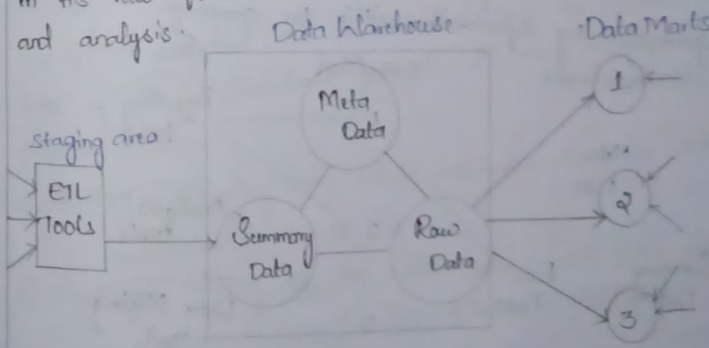
When creating the data warehouse system, you first need to decide what kind of database you want to use.

There are 4 types of database -

1) Relational DataBase :- Row-oriented databases
2) Analytics DataBase :- Developed to sustain and manage analytics.
3) Data Warehouse Applications :- Software for data management and hardware for storing data offered by third-party dealers.
4) Cloud-based Database :- hosted on the cloud.

DATA :- Once the system cleans and organizes the data, it stores in DW. It represents the central repository that stores metadata, summary data, and raw data coming from each source.

1) Metadata :- The information that defines the data. It's primary role is to simplify working with data instances. It allows data analysts to classify, locate, and direct queries to the required data.
2) Summary Data :- Generated by the warehouse manager. It updates as new data loads into the warehouse. This component can includes lightly or highly summarized data. It's main role is to speed up query performance.
3) Raw data :- The actual data loading into the repository, which has not been processed. Having the data in its raw form makes it accessible for further processing and analysis.



Access Tools :-

Users interact with the gathered information through different tools and technologies. They can analyze the data, gather insight, and create reports.

1) Reporting Tools :- They play a crucial role in understanding how your business is doing and what should be done next. Reporting tools include visualizations such as graphs and charts showing how data changes over time.

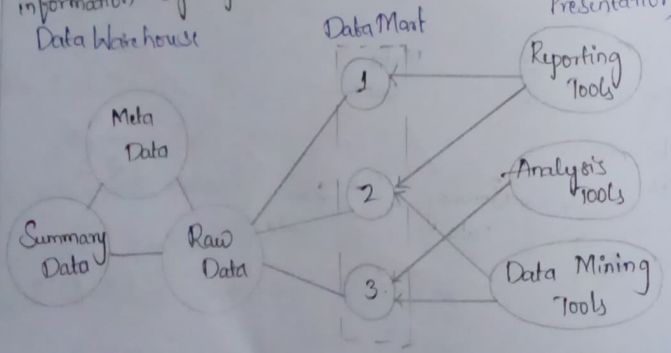
2) OLAP Tools :- Online Analytical processing tools which allow users to analyze multidimensional data from multiple perspectives. These tools provide fast processing and valuable analysis. They extract data from numerous relational data sets and reorganize it into a multidimensional format.

3) Data Mining Tools :- Examine data sets to find patterns within the warehouse and the correlation b/w them. Data mining also helps establish relationships when analyzing multidimensional data.

Data Marts :-

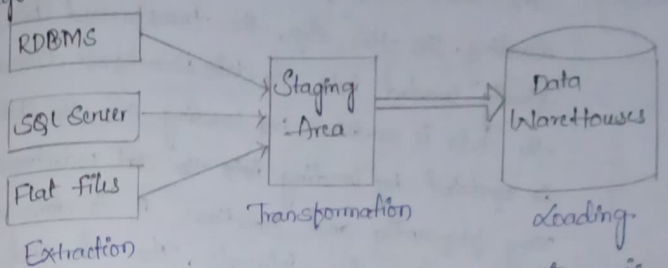
Data Marts allow you to have multiple groups within the system by segmenting the data in the warehouse into categories. It partitions data, producing it for a particular user group.

For instance, you can use data marts to categorize information by departments within the company.



ETL Process in Data Warehouse :

"Extract, Transform and Load." It is a process in which an ETL tool extracts the data from various data source systems, transforms it in the staging area, and then finally, loads it into the Data Warehouse system.



1) Extraction :- Data from various source systems is extracted which can be in various formats like relational databases, No SQL, xml, and flat files into the staging area. It is important to extract the data from various source systems and store it into the staging area first and not directly into data warehouse because the extracted data is various formats and can be corrupted also. Hence loading it directly into DW may damage it and rollback will be much more difficult. Therefore, this is one of the most important step.

2) Transformation :- A set of rules or functions are applied on the extracted data to convert it into a single standard format. It may involves processes / tasks.

a) Filtering — Loading only certain attributes into the data warehouse.

b) Cleaning :- Filling up the NULL values with some default values, mapping U.S.A, United States, and America into USA, etc.

c) Joining :- Joining multiple attributes into one.

b) Splitting :- Splitting a single attributes into multiple attributes.

c) Sorting :- Sorting tuples on the basis of some attributes (generally key-attribute)

3) Loading :- The Third and final step of ETL process is Loading. In this step, the transformed data is finally loaded into the data warehouse. Sometimes the data is updated by loading into the data is updated by loading into the dat warehouse very frequently and sometimes it is done after longer but regular intervals. The rate and period of loading solely depends on the requirements and varies from system to system.

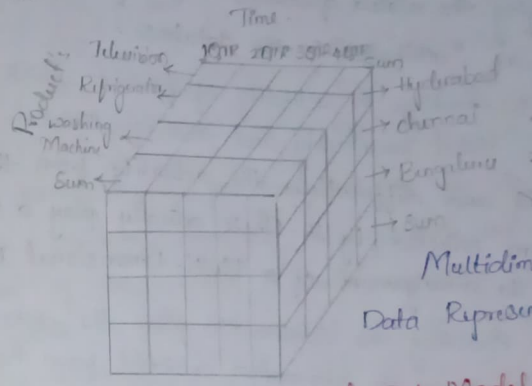
Logical Multi Dimensional Data Model :-

The multi-Dimensional Data Model is a method which is used for ordering data in the database along with good arrangement and assembling of the contents in the database.

It allows customers to interrogate analytical questions associated with market or business trends, unlike relational database which allow customers to access data in the form of Queries. They allow users to rapidly receive answers to the requests which they made by creating and examining the data comparatively fast.

OLAP (Online Analytical processing) and data warehousing uses multi-dimensional database. It is used to show multiple dimensions of data to users.

It represents data in the form of data cubes. Data cubes allows to model and view the data from many dimensions and perspectives. It is defined by dimensions and facts and is represented by a fact table. Facts are numerical measures and fact tables contains measures of the related dimensional tables or names of the facts.



Working on a Multidimensional Data Model :-

Stage-1 :- Assembling data from the client :- In first stage, a Multi-Dimensional Data Model collects correct data from the client. Mostly, software professionals provide simplicity to the client about the range of data which can be gained with the selected technology and collect the complete data in detail.

Stage-2 :- Grouping different segments of the System :- It recognizes and classifies all the data to the respective section they belong to and also builds it problem-free to apply step-by-step.

Stage-3 :- Noticing the different proportions :- It is the basis on which the design of the system is based. The main factors are recognized according to

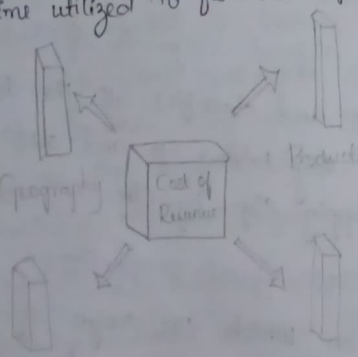
the user's point of view. These factors are also known as "Dimensions".

Stage-4: Preparing the actual-Time factors and their respective Qualities :- The factors which are recognized in the previous step are used further for identifying the related qualities. These qualities are also known as "attributes" in the database.

Stage-5: Finding the actuality of factors which are listed previously and their qualities :- It Separates and differentiates the actuality from the factors which are collected by it. These actually play a significant role in the arrangement of a Multi-Dimensional Data Model.

Stage-6: Building the Schema to place the data, with respect to the information collected from the step above :- In the basis of the data which was collected previously, a schema is built.

for example :- Let us take the example of a firm. The revenue cost of a firm can be recognized on the basis of different factors such as geographical location of firm's workplace, products of the firm, advertisements done, time utilized to flourish a product, etc.



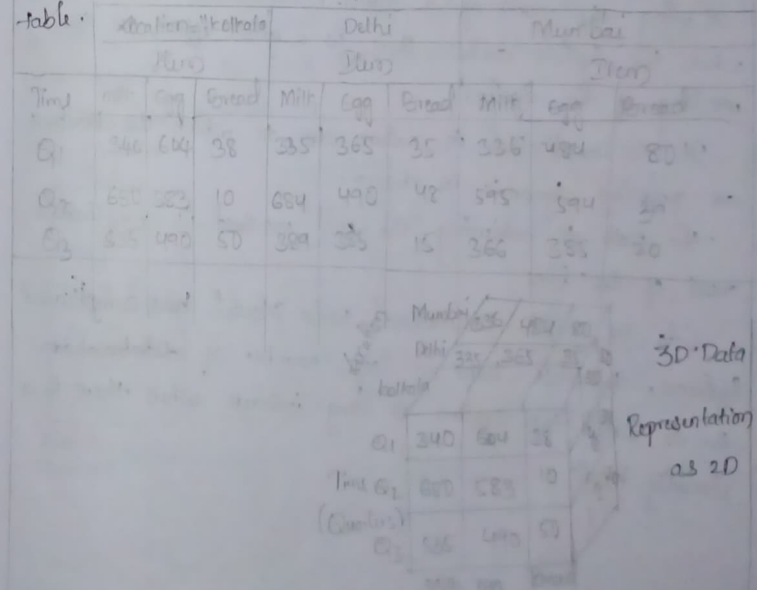
Let us take the example of the data of a factory which sells products per quarter in Bangalore. The data is represented in the table.

Location = "Bangalore"

Time (Quarter)	Type of item			
	Jam	Bread	Sugar	Milk
Q1	350	389	35	50
Q2	260	528	50	90
Q3	483	256	20	60
Q4	436	396	15	40

The factory's sales for Bangalore are for the time dimension, which is organized into quarters and the dimension of items, which is sorted according to the kind of item which is sold. The facts here are represented in rupees (in thousands)

Now, if we desire to view the data of the sales in a 3-dimensional Table, then it is represented in the diagram. Here the data of the sales is represented as a 2-dimensional table.



3D-Data Representation as 2D

Advantages of Multi-Dimensional Data Model.

1) A multi-dimensional data model is easy to handle.
2) It is easy to maintain
3) Its performance is better than that of normal database eg: Relational database.
4) The representation of data is better than traditional db, That is because the multi-dimensional database are multi-viewed and carry different types of factors.
5) It is workable on complex systems and applications, contrary to the simple one-dimensional db systems.
6) The compatibility in this type of database is an upliftment for projects having lower bandwidth for maintenance staff

Dis-advantages of Multi-Dimensional Data Model.

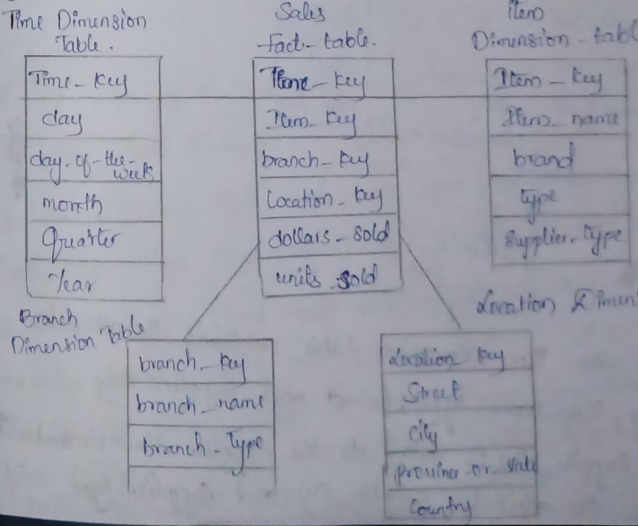
1) The multi-dimensional Data Model is slightly complicated in nature and it requires professionals to recognize and examine the data in the db.
2) During the work of a Multi-Dimensional Data Model, when the system caches, there is a great effect on the working of the system.
3) It is complicated in nature due to which the database are generally dynamic in design.
4) The path to achieving the end product is complicated most of the time
5) As the multi-Dimensional Data Model has complicated systems, database have a large number of ~~databases~~ data due to which the system is very insecure when there is a security breach.

Schema : Multidimensional Schema is defined using Data Mining Query Language (DMQL). The two primitives, cube definition and dimension definition, can be used for defining the data warehouses and data marts.

Schema is a logical description of the entire database. It includes the name and description of records of all record types including all associated data-items and aggregates. Much like a database, a data warehouse also requires to maintain a Schema. A database uses relational model, while a datawarehouse uses Star, Snowflake, and Fact constellation Schema. In this, we will discuss the Schemas.

Star Schema :-

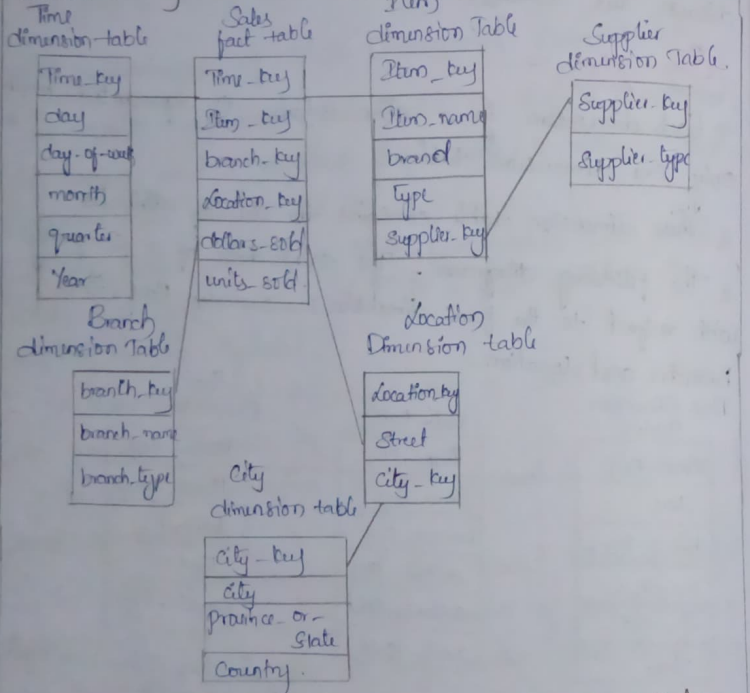
* Each dimension in a star schema is represented with only one-dimension table.
* This dimension table contains the set of attributes
* The following diagram - The sales data of a company with respect to the four dimensions, namely time, item, branch, and location.



* There is a fact table at the center. It contains the keys to each of four dimensions.
* The fact table also contains the attributes, namely dollars sold and units sold.

Snowflake Schema:-

* Some dimension tables in the Snowflake schema are normalized.
* The normalization splits up the data into additional tables.
* Unlike Star schema, the dimensions table in a Snowflake schema are normalized. For eg: The item dimension table in star schema is normalized and split into two dimension tables, namely item and Supplier table

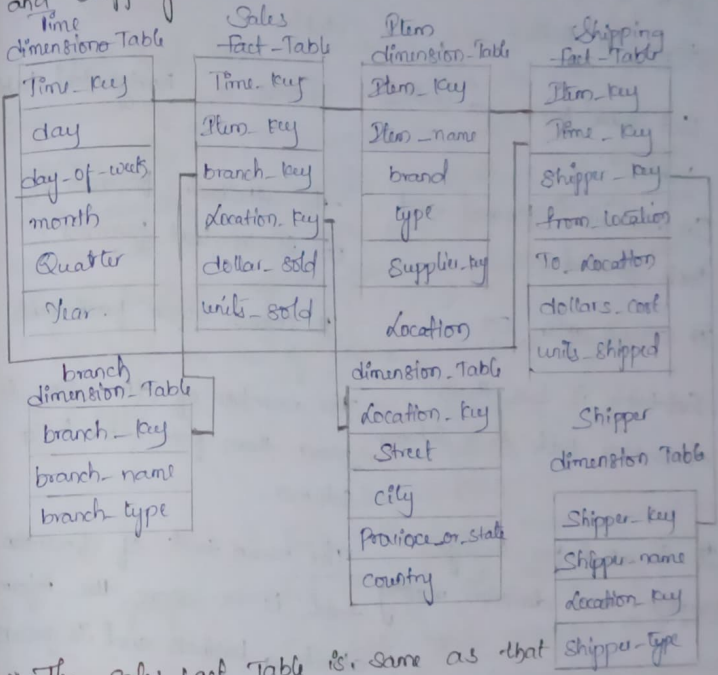


* Now the item dimension table contains the attributes item_key, item_name, type, brand and Supplier_key.
* The Supplier key is linked to the Supplier dimension table. The contains the attributes Supplier_key and Supplier_type.

NOTE: Due to normalization in the Snowflake schema, the redundancy is reduced and therfore, it becomes easy to maintain and the save storage space.

Fact Constellation Schema:-

* A constellation has multiple fact tables. It is also known as galaxy schema.
* The following diagram has 2 fact tables, namely Sales and Shipping.



* The sales fact table is same as that in the Star schema.
* The shipping fact table has the five dimensions, namely item_key, time_key, shipper_key, from_location, to_location
* The shipping fact table also contains two measures, namely dollars sold and units sold.
* It is also possible to share dimension tables b/w fact tables. For eg: Time, Item, location DT are shared b/w the sales and shipping fact table.

Fact Table	Dimension Table
Fact table contains the measuring on the attributes of a dimension table.	It contains the attributes on that truth table calculate the metric.
There is less attributes than dimension Table.	While in DT, there is more attributes than fact table.
There is more records than dimension table.	There is less records than fact table.
Fact table forms a vertical table.	Forms a horizontal table.
In numerical format and text format.	The attribute format of DT is in text format.
It comes after DT	It comes before fact table.
Fact table is less than dimension table in a Schema	The number of dimension is more than fact table in a schema.
It is used for Analysis purpose and decision making	The main task of dimension table is to store the informa about a business and its proce

Fact Table :-

Facts are the numerical measures (or) quantities by which one can analyze relationships b/w dimensions. The relation containing such multi-dimensional data are called Fact Table.

Dimension are the collection of logically related attributes and is viewed as an axis for modelling the data. A dimension table is a table associated with each dimension and helps in describing the dimension further.

eg: Sale of Books

Book-Id	Author	price
B_1	xyz	40
B_2	ABC	20

Dimension - Table

Book Id	Transaction No	
B_1	1	25
B_2	2	26

Fact - Table

Fact Measures :-

Fact table holds the measure data for measuring the performance of your business. It might be sales, purchasing inventory, logistics, banking, telephone data and many more normally one fact table represents a line of one business and it usually tabs more fact, tables to more complex business aspects (purchasing, selling, sales etc).

Based on line of business measures types in Fact table can be : 1) Fully-additive
2) Semi-additive
3) Non-additive.

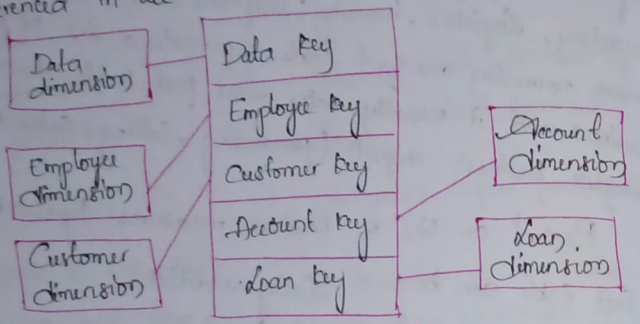
1) Fully-Additive :- Fact table holds measures that are grouped summed through all dimensions

2) Semi-Additive :- Fact table holds measures that can be grouped summed or aggregated through some dimensions but not all the dimensions.

3) Non-Additive :- Non-additive fact table holds measures that can not be grouped, summed or aggregated in any aspect. These measures are normally derived and calculated measures such as percentages, ratio running sums or any similar measures.

Fact-less-facts :-

Simply it mean the key available in the fact that no measures are available fact less fact table are only used to establish relationship between elements of different dimensions and are also useful for describing events and coverage, meaning tables contain information that nothing has happend. It often represents many-to-many relationships. The only-thing they have is an abbreviated key. They still represent a logical phenomenon that is identified by the combination referenced in the dimension table.



Fact-less-Fact Table

Dimension Table Characteristics :-

Dimension Table Key :- Primary key of the dimension table uniquely identifies each row in the table.

Table is wide :- A dimension table has many columns attributes. It is not uncommon for some dimension tables to have more than 50 attributes. Therfore we say that the dimension table is wide. If you lay it out as a table with columns and rows, the table is spread out horizontally.

Dimension table contain textual information that represents the attributes of the business.

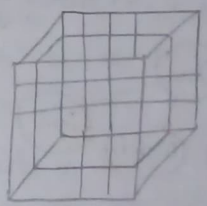
It contain relatively static data.

Are joined to a fact table through a foreign key reference.

OLAP Cube :-

It is also called as Hyper Cube. The OLAP cube is a data structure optimized for very quick data analysis. The olap cube consists of numeric facts called measures which are categorized by dimensions.

Data operations and analysis are performed using single spreed-sheet, where data values are arranged in a row and column format. This is ideal for two-dimensional data. Olap contains multidimension data, with data usually obtain from a different and un-related source. Using a spread sheet is not an optimal option. The cube can store and analyze multipdimensional data in a logical and orderly manner.



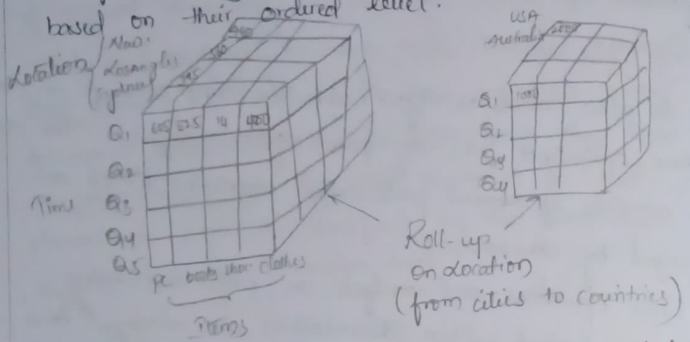
OLAP Operations :-

OLAP operations for multidimensional data is four types :
1) Roll-up
2) Drill-down
3) Slice-and-Dice
4) pivot (rotate)

1) Roll-Up :- It is also called as "consolidation" or "aggregation".

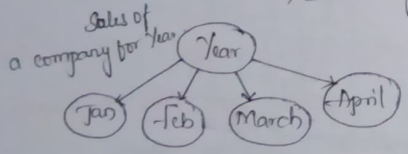
The Roll-up operation can be performed in 2 ways:
1) Reducing dimension.
2) Climbing up concept hierarchy.

Concept hierarchy is a system of grouping things based on their ordered level.



Drill-Down:- In drill down data is fragmented into smaller parts. It is the opposite of the roll-up process.

It can be done via: 1) Moving down the concept hierarchy.
2) Increasing a dimension.



Quarter Q_1 is drilled down to months January, February and March corresponding sales are also registers.

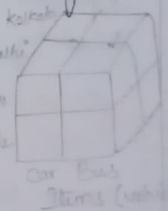
3) **Slice and Dice:-**

Here one dimension is selected and a new sub-cube is created.

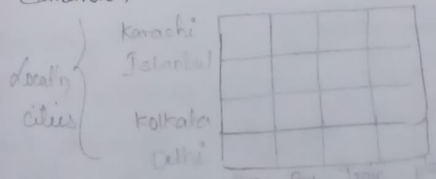
Dice:- This operation similar to a slice. The difference is dice is you select 2 or more dimension. that result in the creation of a sub-cube.

It selects a sub-cube from the OLAP cube by selecting two or more dimensions. In the cube given in the Overview section, a sub-cube is selected by selecting.

Location = "Delhi" or "Kolkata"
Time = "Q_1" or "Q_2"
Item = "Car" or "Bus".



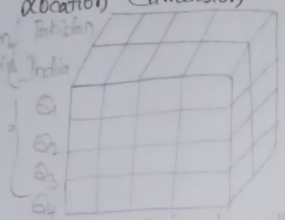
Slice:- It selects a single dimension from the OLAP cube which results in a new sub-cube creation. In the cube given Overview section, Slice is performed on the dimension Time = "Q_1"



1) **Roll-Up:-** It is just opposite of the drill-down operation. It performs aggregation on the OLAP cube:

* Climbing up in the concept hierarchy.
* Reducing the dimensions

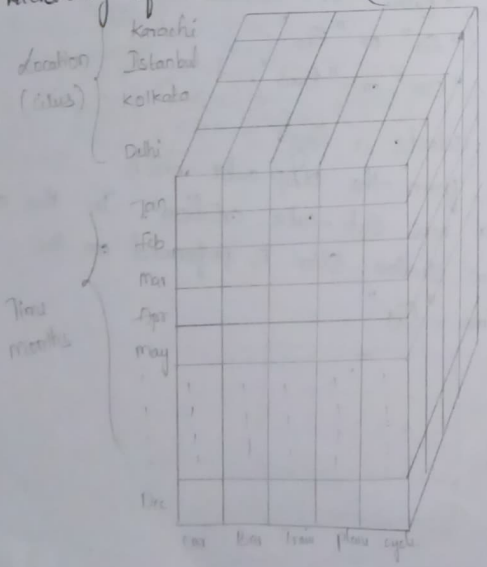
On the cube given in the Overview Section, the roll-up operation is performed by climbing up in the concept hierarchy of location dimension (city → Country)



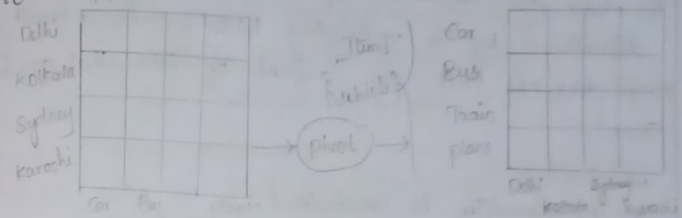
2) **Drill down:-** In drill-down operation, the less detailed data is converted into highly detailed data.

* Moving down in the concept hierarchy.
* Adding a new dimension.

In the cube given in Overview section, the drill-down operation is performed by moving down in the concept hierarchy of Time dimension (Quarter → Month)



3) **Pivot :-** It is also known as rotation operation as it rotates the current view to get a new view of the representation. In the sub-cube obtained after the slice operation, performing pivot operation gives a new view of it.



OLAP Servers: Architecture :-

Online Analytical Processing (OLAP) refers to a set of software tools used for data analysis in order to make business decisions. OLAP provides a platform for gaining insights from database retrieved from multiple database systems at the same time. It is based on a multidimensional data model, which enables users to extract and view data from various perspectives. A multi-dimensional database is used to store OLAP data. Many Business Intelligence (BI) applications rely on OLAP technology.

Type of OLAP Servers:- The 3 major OLAP servers are:

1) ROLAP
2) MOLAP
3) HOLAP

1) **Relational OLAP (ROLAP):-** Relational On-line Analytical Processing (ROLAP) is primarily used for data stored in a relational database, where both the base data and dimension tables are stored as relational tables. ROLAP servers are used to bridge the gap between the relational back-end server and the client's front-end tools.

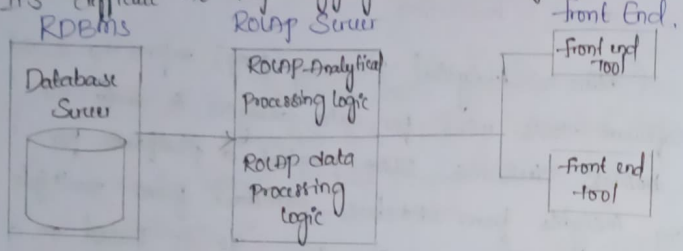
ROLAP Servers store and manage warehouse data using RDBMS and OLAP middleware fills in the gaps.

Benefits:-

1) It is compatible with data warehouses and OLTP systems.
2) The data size limitation of ROLAP technology is determined by the underlying RDBMS. As a result, ROLAP does not limit the amount of data that can be stored.

Limitations:-

1) SQL functionality is constrained.
2) It's difficult to keep aggregated tables up-to date.



II Multi-dimensional OLAP (MOLAP):-

Through array-based multi-dimensional storage engines, Multi-dimensional On-line Analytical processing (MOLAP) supports multi-dimensional views of data.

Storage utilization in multi-dimensional data stores may below if the data set is sparse.

MOLAP stores data on discs in the form of a specialized multi-dimensional array structures. It is used for OLAP, which is based on the array's random access capability. Dimension instances determine array elements, and the data or measured value associated with each cell is typically stored in the corresponding array element. The multi-dimensional array is typically stored in MOLAP in a linear allocation based on nested traversal of the axes in some predetermined order.

However, unlike ROLAP, which stores only records with non-zero facts, all array elements are defined in MOLAP, and as a result, the arrays tend to be sparse, with empty elements occuping a large portion of them.

MOLAP systems typically includes provisions such as advanced indexing and hashing to locate data while performing queries for handling sparse arrays, because both storage and retrieval costs are important when evaluating online performance.

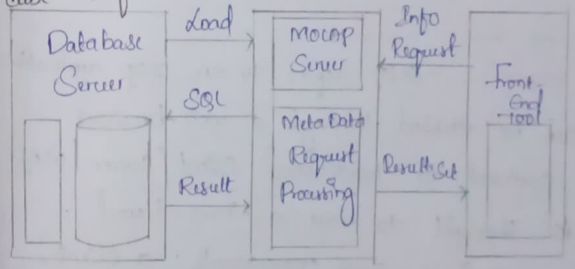
MOLAP cubes are ideal for slicing and dicing data and can perform complex calculations. When the cube is created, all calculations are pre-generated.

Benefits:-

1) Suitable for slicing and dicing operations.
2) Outperforms ROLAP when data is dense.
3) Capable of performing complex calculations.

Limitations:-

1) It is difficult to change the dimensions without re-aggregation.
2) Since all calculations are performed when the cube is built, a large amount of data cannot be stored in the cube itself.



III Hybrid OLAP (HOLAP):-

ROLAP and MOLAP are combined in Hybrid On-line Analytical processing (HOLAP).

It offers greater scalability than ROLAP and faster computation than MOLAP. HOLAP is a hybrid of ROLAP and MOLAP. HOLAP servers are capable of storing large amounts of detailed data. On the one hand, HOLAP benefits from ROLAP's greater scalability.

HOLAP, on the other hand, makes use of cube technology for information. Because detailed data is stored in a relational database, cubes are smaller than MOLAP.

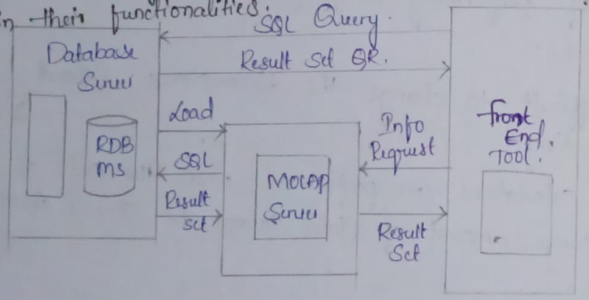
Benefits:-

i) HOLAP combines the benefits of MOLAP and ROLAP
ii) Provide quick access at all aggregation levels.

Limitations:-

i) Because it supports both MOLAP and ROLAP servers, HOLAP architecture is extremely complex.
ii) There is a greater likelihood of overlap, particularly in their functionalities.

II



Other types of OLAP include:-

* Web OLAP (WOLAP):- WOLAP refers to an OLAP application that can be accessed through a web browser. WOLAP, in contrast to traditional client/server OLAP applications, is thought to have a three-tiered architecture consisting of three components: a client, middleware, and a database server.

* Desktop OLAP (DOLAP):- DOLAP is an abbreviation for desktop analytical processing. In that case, the user can download the data from the source and work with it on their desktop or laptop. In comparison to other OLAP applications, functionality is limited. It is less expensive.

* Mobile OLAP (MOLAP):- Wireless functionality or mobile devices are examples of MOLAP. The user is working and accessing data via mobile devices.

* Spatial OLAP (SOLAP):- SOLAP egress combines the capabilities of Geographic Information System (GIS) and OLAP into a single user interface. SOLAP is created because the data can be alphanumeric, image or vector. This allows for the quick and easy exploration of data stored in a spatial database.