Page No.:
Date: / /

Aggregation

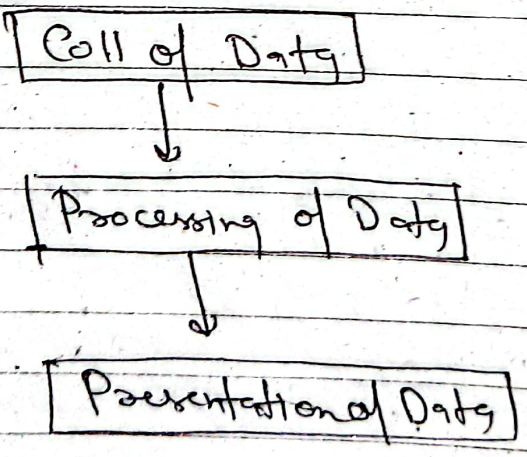
is the process of finding, collecting & presenting in a summarized format to perform statistical analysis of business schemes or analysis of human patterns

It helps in taking prudent decisions in marketing, finance, pricing the product.

Data Aggregators

They are a system in data mining that collects data from numerous sources, then processes the data & repackages them into useful data packages

Working of Data Aggr



Types of Data Aggr

1. Time Aggregation: It provides data for single resources for a defined time period.

Page No.:
Date: / /

② Spatial Aggr : It provides data for group of resources for a defined time period.

Appl of Data Aggr

- Data Aggr is used in many fields where a large number of datasets are involved.
- It helps in making prudent decisions in marketing or finance management.
- It helps in the creation of marketing schemes.
- It plays a major role in e-commerce industries & travel industry.

OLAP function & Tools

OLAP Server

- OLAP Server is based on the multidimensional data model
- It allows managers & analysts to get insight of the information through fast, consistent & interactive access to information

Types of OLAP Servers

- Relational OLAP (ROLAP)

- ROLAP Servers are placed b/w back-end server and client front end tools.
- To store & manage warehouse data ROLAP uses relational or extended-relational DBMS

ROLAP includes following -

- Implementation of aggr navigation logic
- Optimization for each DBMS backend
- Additional tools & services

Multidimensional OLAP (MOLAP)

- MOLAP uses array based multidim storage engines for multidim views of data.
- With multidim data stores the storage utilization may be low if the data set is sparse.
- MOLAP Server use two levels of data storage representation to handle dense & sparse data.

Hybrid OLAP (HOLAP)

- HOLAP is combination of ROLAP & MOLAP
- It offers higher scalibility of ROLAP and faster computation of MOLAP.
- It allows to store the large data of volumes of detailed information

OLAP operations

Roll-up
Drill-down
Slice & Dice
Pivot (Rotate)

Page No.:
Date: / /

Data Mining Interface

Data Mining is the process of finding useful new correlations, patterns & trends by using statistical & mathematical techniques.

Data Mining Interface provides the medium that allows users to communicate with data mining processes.

It is difficult to use data mining query lang. A graphical user interface can be used to comm^{n} with data mining systems.

Backup

Backup is the process of creating a copy of data to protect against accidental or malicious deletion, corruption, hardware failure, attacks & other types of data loss.

Physical Backups

are backups of the physical files used in storing & recovering of database.

Logical Backups

Logical backups contain logical data exported from database for later re-importing.

Page No.:
Date: / /

Recovery

Restore is the process of retrieving data from backup

Backup & Recovery are among the most important tasks for an administrator and data warehouses.

Tuning Data warehouse

It is unpredictable what query the user is going to post in future. Therefore it becomes more difficult to tune a data warehouse system

Difficulties in Data Ware Tuning due to following reasons-

- Date warehouse is dynamic: It never remains constant
- It is very diff to predict what query the user is going to post in the future
- Business requirements change with time.
- Users & their profile keep changing
- The data load on warehouse also changes with time

Tuning Queries

1. Fixed Queries
2. Ad hoc Query

Page No.:
Date: / /

Testing Data Warehouse

Testing is very important for data warehouse systems for data validation and to make them work correctly & efficiently.

There are 3 basic levels of testing performed in DW

- **Unit Testing**

- This type of testing is performed at the developer's end.
- In unit testing, each component of modules is separately tested.

Integration Testing

On this type of testing various individual units or modules of the appl are combined and then tested

System Testing

- It is the form of testing that validates & tests the whole data ware appl.
performed by technical testing team.