

UNIT-5 DWDM

- Aggregation [AKTU-21-22]
- Historical Information
- Query Facility
- OLAP Function and Tools [AKTU-21-22, 22-23, 23-24]
- OLAP Servers [AKTU-21-22]
- ROLAP, MOLAP, HOLAP [AKTU-21-22, 22-23, 23-24]
- Security and Backup & Recovery
- Tuning data warehouse [AKTU-22-23]
- Types of Warehousing application [AKTU-22-23]
- Web Mining, Spatial Mining and Temporal Mining [AKTU-21-22, 22-23, 23-24]

SUBSCRIBE 

What is Aggregation? [AKTU-21-22]

Aggregation in data mining means collecting and summarizing data so we can understand it better and use it for analysis.

When we have a lot of data from different sources, it's hard to make sense of it all. So we group and summarize the data to get useful information.

How Does Data Aggregation Work?

Step 1 :- Raw data is often too detailed or messy to be useful by itself.

Step 2 :- So we combine similar piece of data to make it easier to understand.

Step 3 :- We use simple calculations like Sum, Count, Average.

Step 4 :- Aggregated data can be shown in report and dashboards to help people make quick and smart decision.

Historical Information ?

Historical Information means old data collected over time - like past sales, customer behavior, or Product Performance

In a data warehouse, historical data is stored so businesses can look back and see.

- What happened?
- When did it happen?
- How did thing change over time?

Query Facility ?

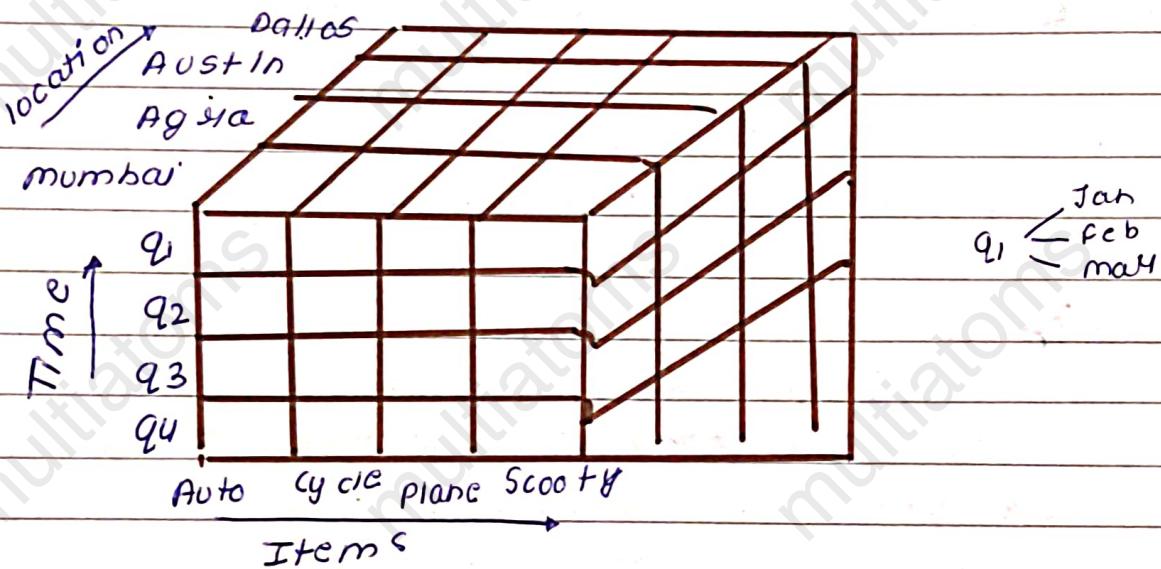
A query facility is a tool or system that lets users ask questions (queries) to retrieve, analyze or summarize data from a data warehouse or database. It's like a search engine for your data, helping you get specific information quickly

What is OLAP?

OLAP (Online Analytical Processing) is a technology used in data warehousing to analyze large amount of data quickly. It allows users to explore and

Summarize data from different angles to
Find insights, trends or patterns

OLAP is especially useful for handling multi-dimensional data, where data is organized into dimensions (ex → time, location, product) and measures (ex → sales, profit)

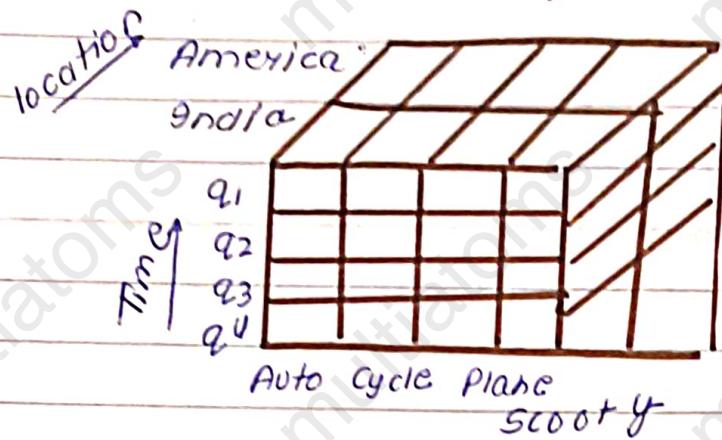


OLAP Functions [AKTU-23-24, 21-22, 22-23]

OLAP provides specific function to analyze data interactively. These functions work with aggregated data are accessed through query facilities

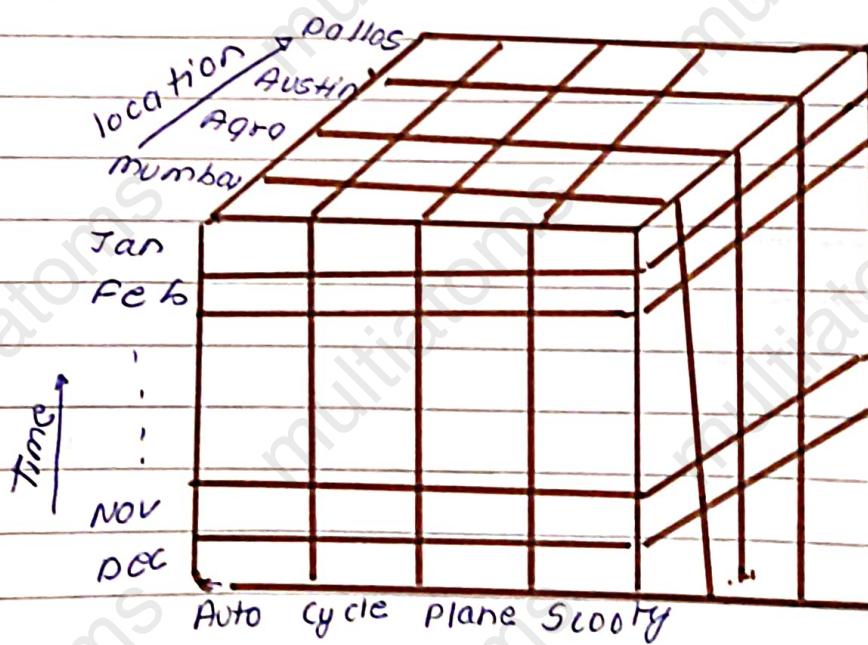
1. ROLUP :-

- Combine data to a higher level of summary
- Ex :- Summarize daily sales into monthly sales



2. Drill-down

- Goes from summarized data to detailed data
- Ex :- From monthly sales, drill down to see daily sales for a specific month



3. Slice

→ Selects a specific subset of data by fixing one dimension

→ E.g. - Look at sales for only one region
(ex "Show sales in New York")

Dollars				
Austin				
Agra				
Mumbai				

cycle Auto Plane Scooty

Items

4. Pivot (Rotate)

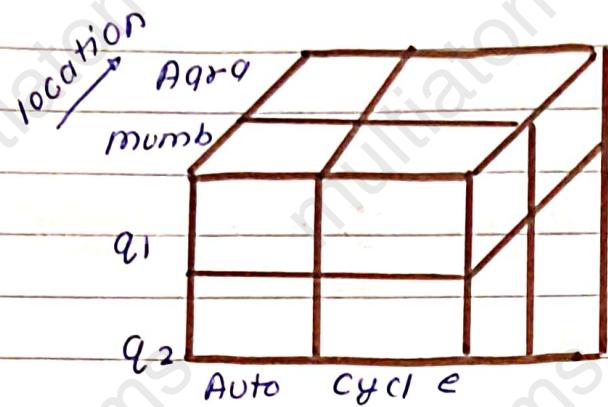
→ Change the view of data by swapping dimensions

Auto				
Cycle				
Plane				
Scooty				

Dollars Austin Agra Mumbai

5. Dice

→ Select a smaller range of data across multiple dimensions



OLAP Tools :- [AKTU-21-22]

OLAP tools are software that make it easy to perform these functions, often with user-friendly interfaces.

1. Microsoft Power BI
2. Tableau
3. Oracle OLAP
4. SAP Business Objects

OLAP Server :- [AKTU-21-22]

An OLAP Server is a specialized software system that powers OLAP in a data warehouse. It organizes and processes large amount of data to support fast multidimensional analysis.

The OLAP Server act as a "brain" that handle queries from users or

OLAP tools and delivers results quickly

Different types of OLAP Servers [AKTU-23-24],
22-23]

1. ROLAP (Relational OLAP)

- ROLAP stores and processes data using relational databases (like SQL, Oracle) which organize data into tables.
- ROLAP Servers are placed between relational backend servers & client frontend tools.
- Queries are written in SQL to perform OLAP Functions.

Advantages

- Handle large datasets
- Uses existing database system, so no need for new storage

Disadvantages

- Slower for complex multidimensional queries because it
- Performance depends on database Optimization

2. MOLAP (multidimensional OLAP)

- MOLAP stores data in a multidimensional cube structure, which is optimized for OLAP analysis. A cube organizes data by dimensions and measures.
- OLAP Functions (slice, dice, pivot) are performed directly on the cube, making queries very fast.

Advantage

- Very Fast for multidimensional queries due to Pre-aggregated data
- Great for Complex analysis

Disadvantage

- Limited Scalability
- Cube building takes time and storage

3. Hybrid OLAP (HOLAP)

- HOLAP is combination of MOLAP & ROLAP. It balances Speed and Flexibility. It uses both relational databases and multidimensional cubes.
- Detailed data stays in relational database

Advantages

- Fast for summarized queries
- Flexible for detailed queries

Disadvantages

- More complex to set up and manage
- Performance depends on how well the cube and databases are integrated.

Data Mining Interface

A data mining interface is a user-friendly tool or system that allows users to interact with data mining processes.

It helps users explore, analyze and extract patterns, trends from large datasets without needing deep technical expertise.

Role of Data Mining Interface :-

1. Data Selection
2. Data Preprocessing
3. Algorithm Selection
4. Result Visualization
5. Evaluation

Security :-

Security refers to Protecting the data, Systems and processes in a ~~down~~ environment from unauthorized access, theft and damage.

Key - Security Concept :

1. Access - Control :- Restrict who can access the data warehouse tools
2. Data - encryption :- Protect data by encoding it so only authorized user can read
3. Network - Security Protect the Server From external threats

Backup :

Backup is the process of creating copies of the data to prevent data loss to hardware failure, human errors

Key - Backup Practices :

1. Full Backups :- Copy the entire data warehouse including raw data, aggregated data and metadata.
2. Incremental Backups :- copy only the data that has changed since the last back up
3. Backup Storage :- Store backup in secure, off-site locations.

Tuning Data Warehouse

A Data Warehouse is like a big storage room where a company keeps all its data and its important data in one organized place

Tuning means making the data warehouse faster and more efficient like cleaning and organizing a messy room so you can find things quickly

How Do You Tune a Data Warehouse?

1. Speed Up Slow Queries

- look at which queries are too long
- Try to rewrite them in a simpler way

2. Use Indexes

- Indexes are like a table of content in a book - they help you find info faster

3. Partition Data

- This means splitting big tables into smaller pieces.

4. Avoid Unused Data

- Don't pull in data you don't need.

Ques Name any Five Challenges in Data Warehouse Testing? [AKTU-22-23]

Sol There are five common challenges

1. Huge Data Volumes :- Data warehouse often hold millions or billions records. Testing all that data takes a lot of time and resources.

2. Data Quality Issues :- Data comes from many sources, and inconsistency can cause errors during testing.

3. Complex Transformation: Data is often transformed before storage. Testing these transformations to ensure they work correctly is tricky.

4. Performance Testing: Ensuring the warehouse responds quickly to queries, especially with large datasets is challenging and requires careful testing.

5. Data Integration: Combining data from different systems can lead to mismatches or errors that are hard to detect or fix.

Qn Explain Warehousing application and recent trends? [AKTU-22-23]

Sol'n A Warehouse application is a software system designed to manage and optimize the operations of a warehouse or data warehouse.

In the context of data, it refers to tools or platforms that handle the storage, organization, processing and analysis of large volumes of data in a

data warehouse.

- It gathers data from different places
- Helps users quickly.
- Amazon Redshift, Snowflake, Google BigQuery
Oracle Autonomous

Recent Trends in Warehousing Application

1. Cloud-Based Data Warehouses

- Instead of storing data on physical servers in a company office, data warehouses are now hosted in the cloud

2. Real-Time Data Processing

- Warehousing application now process and analyze data as it arrives instead of waiting for batch updates

3. AI and machine Learning Integration

- Warehousing applications are adding AI and machine learning to automatically find patterns or optimize queries.

Types of Warehousing Applications

1. Enterprise data warehouse (EDW)

→ A central warehouse for all data across a company

2. Cloud - Data warehouse

→ Hosted on cloud platform like Amazon Redshift etc

3. Virtual data warehouse

→ Data stays in source system but appear combined through a virtual layer

Qn discuss about the Web mining, Spatial mining and Temporal mining under the Data Visualization? [AKTU-21-22, 23-24 J 22-23]

So, what is Web-mining:-

→ Web-mining is about extracting useful information from the web - such as websites, search engines or social-media.

How It Relates to Visualization

→ Web mining is about extracting useful

- Helps visualize Online behavior
- useful visualis:- click heatmaps, flow-diagram

2. Spatial mining:

- Spatial mining deals with geographic or location based data like maps, GPS coordinates or satellite images

How It Relates to Visualization

- Shows where events happen and uncovers location-based patterns
- Finding relationship like how close stores are to customers or how weather affects sales in different areas

3. Temporal mining

- Temporal mining focuses on data that changes over time - like logs, transactions. It's like looking at a calendar to see when things happen or repeat such as sales.

How It Relates to Visualization :-

- Track when something happens and reveals trends or changes
- Useful visuals - like - line-graph, timelines

SUBSCRIBE □

JOIN TELEGRAM @