Name: Pranav Mishra Registration NO: 12114762
Roll No:- RD0C09A55 Course Code: CAP446
Teacher: Punam Maam Course Title: Data Warehouse & Data Mining

Ans:- 1 OLAP Operations with example:-

Online Analytical Processing Server (OLAP) is based on the multidimensional data model. It allows managers, and analysts to get an insight of the information through fast, consistent, and interactive access to information.

OLAP Operations:-

OLAP provides a user-friendly environment for interactive data analysis. One of the most popular front-end applications for OLAP is a PC spreadsheet program.

Here is the list of OLAP operations-

1. Roll-up
2. Drill-down
3. slice and dice
4. Pivot (rotate)

1:- Roll up:- Roll up can be also considered as an aggregation of data. The data which is split up is consolidated and then used further. This can be done by following methods.

→ Reducing dimensions

→ By using concept hierarchy where a system of grouping things is done based on a particular order or level.

Example:

Location	Medal
Delhi	5
New York	2
Pune	3
Los Angeles	5

We roll up on location from cities to countries.

Output:

Location	Medal
India	8
America	7

More detailed data to less detailed data Rollup

2: Drill down:-

Drilling down is nothing but breaking the data further into smaller parts. This dimension can be applied to the data cube.

Drill down is the reverse of roll-up. That means lower level summary to higher level summary.

Drill-down can be performed either by:-

1. Stepping down a concept hierarchy for a dimensio
2. By introducing a new dimension.

Ex:-

Location	Model
India	8
America	7

Drill-down on location from countries to cities.

Location	Medal
Delhi	5
New York	2
Pune	3
Los Angeles	5

Drill down

Less detailed data to More detailed Data.

3:- Slice and dice:-

The slice operation performs a selection on one dimension of the given cube, resulting in a subcube. Reduces the dimensionality of the cubes.

Ex:- If we want to make a select where Medal = 5

Location	Medal
Delhi	5
Los Angeles	5

Slice Operation

The dice operation defines a sub-cube by performing a selection on two or more dimensions.

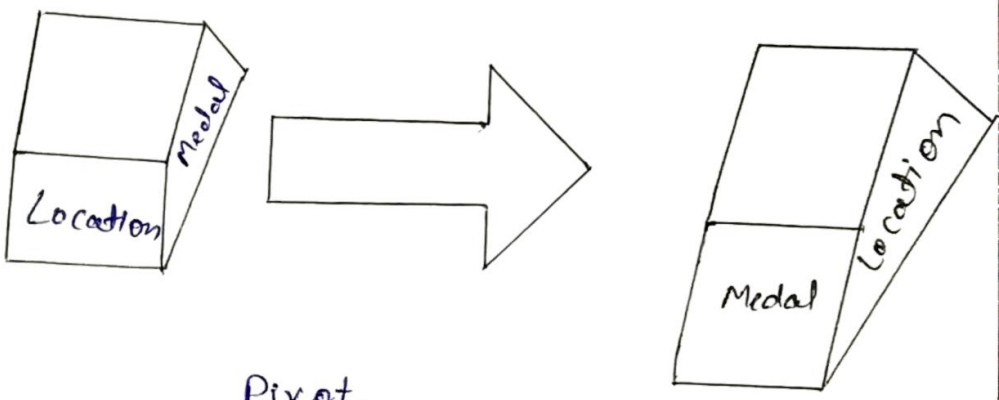
Ex:- If we want to make a select where Medal = 3 or Location = New York

Location	Medal
Pune	3
New York	2

Dice Operation

4:- Pivot:-

The pivot operation is also known as rotation. It rotates the data axes in view in order to provides an alternative presentation of data. Consider the following example:-



Pivot.

Conclusion:-

The core of any OLAP system is numeric facts called measures. These measures can be further divided into dimensions. The measures are then placed at intersection that form the vector space.

The OLAP cube is a matrix interface that will help in doing projection operation like aggregation.

Ans:-2:- Data Mining functionalities:-

Data mining is a technical methodology to detect information from huge data sets.

The main objective of data mining is to identify patterns, trends, or rules that explain data behavior contextually.

Data mining functionalities that are measured to predict the type of patterns in data sets.

There are following data mining functionalities

1. Classification
2. Association Analysis
3. Cluster Analysis
4. Data Characterization
5. Data Discrimination
6. Prediction
7. Outlier Analysis
8. Evolution Analysis

Classification:-

Classification is the technique to categorize element in a collection, basis their predefined functionalities and properties In this, the model can classify new instance whose classification is unknown. These particular instance that are used to create the model called training data. Such a mechanism of classification use methods like if then, decision trees, neural network, or even a set of classification rules.

5/8

Association Analysis :-

Association Analysis is also called Market Basket Analysis. It is a very popular data mining methodology with usage in sales. Association analysis helps to find relations between elements frequently occuring together. It is made up of a series of sets of elements and rules that describe how these are grouped with in the cases. Association rule are used to predict the presence of an element in the database and are based on the manifestation of a specific element identified as important.

Association analysis is based on 2 parts rule-

→ antecedent (if) -

Consequent (then) -

Cluster Analysis :-

It is similar to that of classification. In cluster analysis, similar types of data are grouped, and the only difference is that the class label in unknown. Clustring Algorithm divide the data basis similarities and the grouped data are similar to each other more than the other data in other groups.

Cluster Analysis is used in machine learning, deep learning, image processing, Pattern recognition, and NLP, etc.

Data Characterization :-

The process of data characterization involves summarizing the generic data features, which can result in specific rules to define a target class. To characterize the data without much user intervention or interaction, an attribute oriented

induction technique is used and the resultant characterized data can be visualized in the form of different types of graphs, charts or tables.

Data Discrimination:-

Data Discrimination is a bias, which happens when a data set or source is treated differently than the others, be it intentional or unintentional. This data mining functionality helps to separate peculiar data sets, based on the ambiguity in attribute values.

Prediction:-

Prediction is among the most popular data mining functionalities that determine any missing or unknown element in a data set.

These are two types of predictions:-

Numeric Predictions:- Predict any missing or unknown element in a data set.

Class Predictions:- Predict the class label using a previously built class model.

Outlier Analysis:-

If we are unable to group any data in any class, we use the outlier analysis technique. Outlier analysis helps to learn about data quality. Outlier means data abnormality is most cases. Most outlier in your data sets, low the data quality.

The outlier analysis process helps in checking if there is any data that can be used to analyze

after some clean-up.

It is still important to keep a track of unusual data and activities so that any anomalies can be detected beforehand and any business impact can be detected in advance.

Evolution Analysis :-

Evolution Analysis refers to the study of data sets that may have been through a phase of transformation or change. The evolution analysis models capture evolutionary trends in data, which further contributes to data characterization, classification or discrimination and clustering for multivariate time series.

Conclusion :-

The thing that makes data mining most interesting is that you can get information without asking specific questions.

Data mining also identifies hidden information in addition to future events.

These data mining functionalities contribute towards finding trends in data mining, making it a crucial element of a data scientist's toolbox.