RAMA UNIVERSITY
www.ramauniversity.ac.in

FACULTY OF EGINEERING

DATA MINING & WAREHOUSEING
LECTURE-30
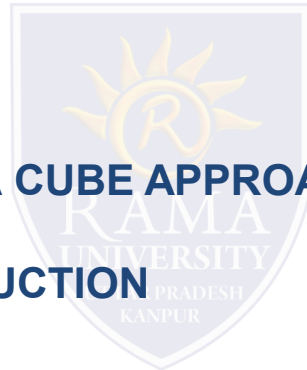
MR. DHIRENDRA
ASSISTANT PROFESSOR
RAMA UNIVERSITY

OUTLINE

❖ **WHAT IS CONCEPT DESCRIPTION**

❖ **CONCEPT DESCRIPTION VS. OLAP**

❖ **DATA GENERALIZATION AND SUMMARIZATION-BASED CHARACTERIZATION**

❖ **CHARACTERIZATION: DATA CUBE APPROACH**

❖ **ATTRIBUTE-ORIENTED INDUCTION**

❖ **MCQ**

❖ **REFERENCES**

Descriptive vs. predictive data mining

– Descriptive mining: describes concepts or task-relevant data

sets in concise, summarative, informative, discriminative

forms

– Predictive Predictive mining: mining: Based on data and analysis analysis, constructs constructs

models for the database, and predicts the trend and
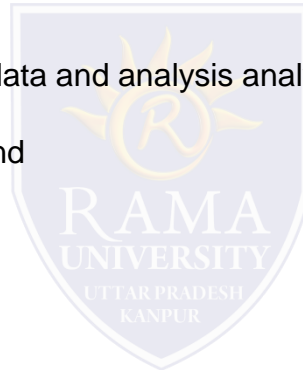
properties of unknown data

• Concept description:

– Characterization: provides a concise and succinct

s mmari ation summarization of the gi en v collection collection of data

– Comparison: provides descriptions comparing two or more

collections of data

Concept Description vs. OLAP

Concept description:

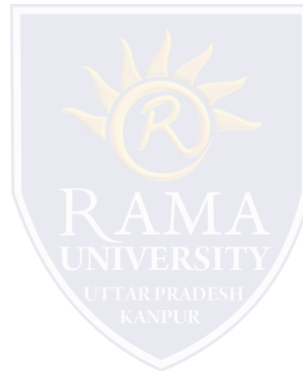– can hadle an e complex data types of the attributes

and their aggregations

– a more aut td oma e process

• **OLAP:**

– restricted to a small number of dimension and

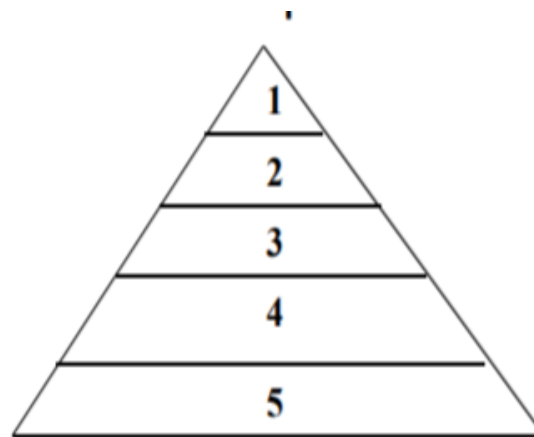measure types

– user-controlled process

D t a a generali ti za on

– A process which abstracts a large set of task-relevant data in

a database database from a low conceptual conceptual levels to higher ones.

– Approaches:

Conceptual levels

• Data cube approach(OLAP approach)

• Attribute-oriented induction approach

Characterization: Data Cube Approach

Perform computations and store results in data cubes

• Strength

– An efficient implementation of data generalization

– Computation of various kinds of measures
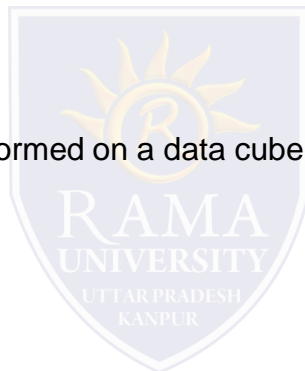
• count( ), sum( ), averag (e), max( )

– Generalization and specialization can be performed on a data cube

by roll-up and drill-down

• Limitations

• handle only dimensions of simple nonnumeric data and measures of simple aggregated numeric values.

• Lack of intelligent analysis, can't tell which dimensions should be used and what levels should the generalization reach

Attribute-Oriented Induction

Proposed in 1989 (KDD '89 workh) s op

• Not confined to categorical data nor particular

measures.

• How it is done?

– Collect the task-relevant data(initial relation) g usin a

relational database query

– Perform generalization by attribute removal or attribute

generalization.

– Apply aggregation by merging identical, generalized tuples

and accumulating their respective counts.

– Interactive presentation with users.
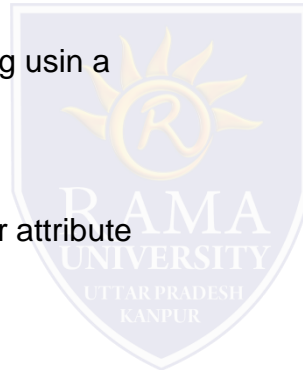
Multiple Choice Question

1. Various visualization techniques are used in _____ step of KDD.
a) selection
b) transformaion
c) data mining.
d) interpretation.

2. Extreme values that occur infrequently are called as _____.
a) outliers
b) rare values.
c) dimensionality reduction.
d) All of the above.

3. Box plot and scatter diagram techniques are _____.
a) Graphical
b) Geometric
c) Icon-based.
d) Pixel-based.

4. _____ is used to proceed from very specific knowledge to more general information.
a) Induction
b) Compression.
c) Approximation.
d) Substitution.

5. Describing some characteristics of a set of data by a general model is viewed as _____
a) Induction
b) Compression
c) Approximation
d) Summarization
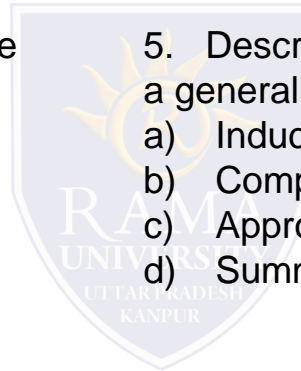
REFERENCES

- https://www.tutorialspoint.com/dwh/dwh_overview.htm

- https://www.geeksforgeeks.org/

- http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf   DATA MINING BOOK WRITTEN BY Micheline Kamber

- https://www.javatpoint.com/three-tier-data-warehouse-architecture

- M.H. Dunham, " Data Mining: Introductory & Advanced Topics" Pearson Education

- Jiawei Han, Micheline Kamber, " Data Mining Concepts & Techniques" Elsevier

- Sam Anahory, Denniss Murray," data warehousing in the Real World: A Practical Guide for Building Decision Support Systems, " Pearson Education

- Mallach," Data Warehousing System", TMH

- R. Agrawal, A. Gupta, and S. Sarawagi. Modeling multidimensional databases. ICDE'97 S. Chaudhuri and U. Dayal. An overview of data warehousing and OLAP technology. ACM SIGMOD Record, 26:65-74, 1997

- S. Agarwal, R. Agrawal, P. M. Deshpande, A. Gupta, J. F. Naughton, R. Ramakrishnan, and S. Sarawagi. On the computation of multidimensional aggregates. VLDB'96 D. Agrawal, A. E. Abbadi, A. Singh, and T. Yurek. Efficient view maintenance in data warehouses. SIGMOD'97

- E. F. Codd, S. B. Codd, and C. T. Salley. Beyond decision support. Computer World, 27, July 1993.

- J. Gray, et al. Data cube: A relational aggregation operator generalizing group-by, cross-tab and sub-totals. Data Mining and Knowledge Discovery, 1:29-54, 1997.