RAMA UNIVERSITY

www.ramauniversity.ac.in

FACULTY OF EGINEERING

DATA MINING & WAREHOUSEING
LECTURE-03

MR. DHIRENDRA
ASSISTANT PROFESSOR
RAMA UNIVERSITY

OUTLINE

- METADATA
- METADATA REPOSITORY
- DATA CUBE
- DATA MART
- VIRTUAL WAREHOUSE
- MCQ
- REFERENCES

DATA WAREHOUSING

- Process of constructing and using a data warehouse.
- constructed by integrating data from multiple heterogeneous sources
- that support analytical reporting, structured and/or ad hoc queries, and decision making.
- involves data cleaning, data integration, and data consolidations.

METADATA

- simply defined as data about data.
- used to represent other data is known as metadata.
- For example, the index of a book serves as a metadata for the contents in the book.
- metadata is the summarized data that leads us to the detailed data.

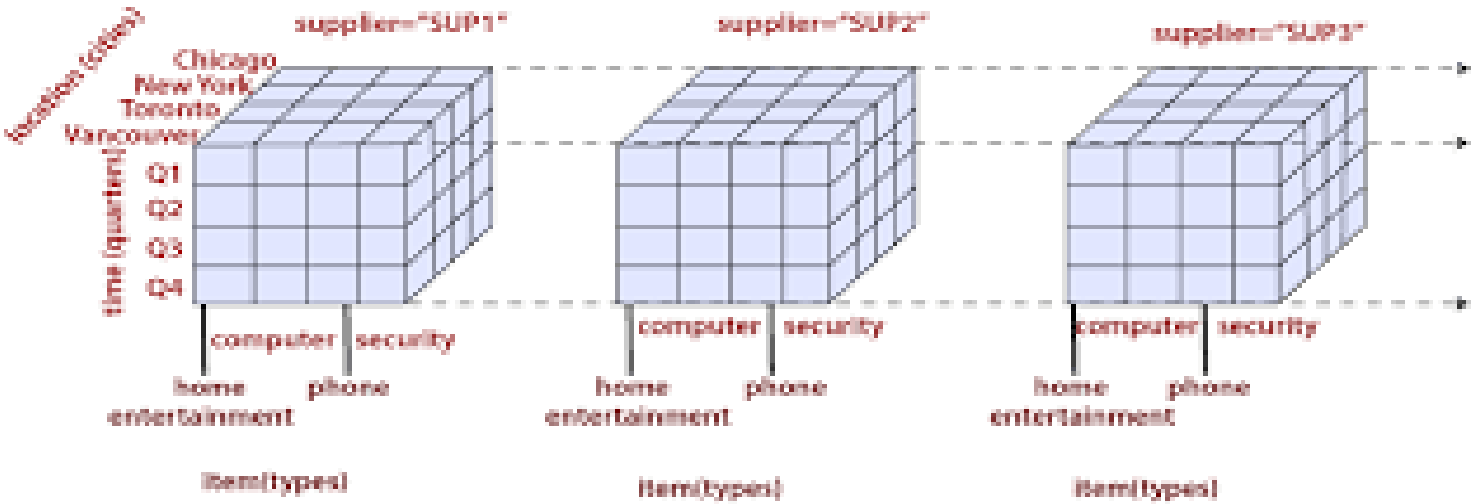
- Metadata is a road-map to data warehouse.
- Metadata in data warehouse defines the warehouse objects.
- Metadata acts as a directory. This directory helps the decision support system to locate the contents of a data warehouse.

METADATA REPOSITORY

- **Business metadata** − It contains the data ownership information, business definition, and changing policies.
- **Operational metadata** − It includes currency of data and data lineage. Currency of data refers to the data being active, archived, or purged. Lineage of data means history of data migrated and transformation applied on it.
- **Data for mapping from operational environment to data warehouse** − It metadata includes source databases and their contents, data extraction, data partition, cleaning, transformation rules, data refresh and purging rules.
- **The algorithms for summarization** − It includes dimension algorithms, data on granularity, aggregation, summarizing, etc.

DATA CUBE

- helps us represent data in multiple dimensions.
- It is defined by dimensions and facts.
- dimensions are the entities with respect to which an enterprise preserves the records.
- Illustration of Data Cube

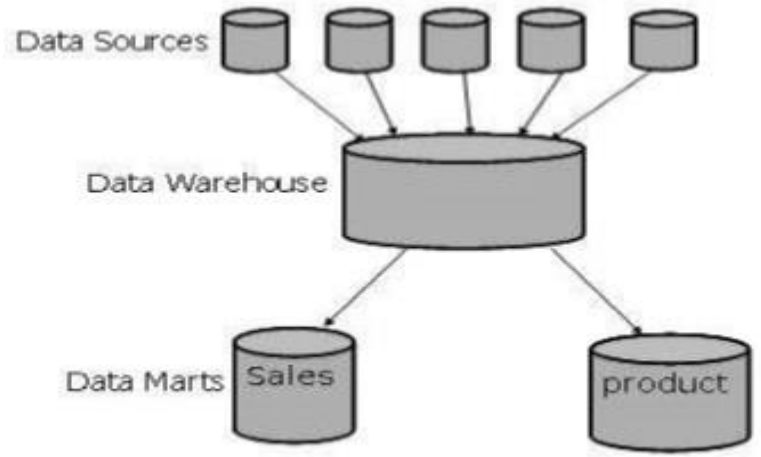


DATA MART

- contain a subset of organization-wide data that is valuable to specific groups of people in an organization.
- data mart contains only those data that is specific to a particular group.
- For example, the marketing data mart may contain only data related to items, customers, and sales. Data marts are confined to subjects.

Points to Remember About Data Marts

- The life cycle of data marts may be complex in the long run, if their planning and design are not organization-wide.
- Data marts are small in size.
- Data marts are customized by department.
- The source of a data mart is departmentally structured data warehouse.
- Data marts are flexible.



VIRTUAL WAREHOUSE

- The view over an operational data warehouse is known as virtual warehouse.
- easy to build a virtual warehouse.
- Building a virtual warehouse requires excess capacity on operational database servers.

Multiple Choice Question

1. _____defines the structure of the data held in operational databases and used by operational applications.
a) User-level metadata.
b) Data warehouse metadata.
c) Operational metadata.
d) Data mining metadata.

2.. _____ is held in the catalog of the warehouse database system.
a) Application level metadata.
b) Algorithmic level metadata.
c) Departmental level metadata.
d) Core warehouse metadata.

3. _____maps the core warehouse metadata to business concepts, familiar and useful to end users.
a) Application level metadata.
b) User level metadata.
c) Enduser level metadata.
d) Core level metadata.

4. _____consists of formal definitions, such as a COBOL layout or a database schema.
a) Classical metadata.
b) Transformation metadata.
c) Historical metadata.
d) Structural metadata.

5. _____consists of information in the enterprise that is not in classical form.
a) Mushy metadata.
b) Differential metadata.
c) Data warehouse.
d) Data mining.

REFERENCES

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

- 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