RAMA UNIVERSITY

www.ramauniversity.ac.in

FACULTY OF EGINEERING

DATA MINING & WAREHOUSEING
LECTURE-13

MR. DHIRENDRA
ASSISTANT PROFESSOR
RAMA UNIVERSITY

OUTLINE

- CUBE OPERATION
- EFFICIENT PROCESSING OLAP QUERIES
- DATA WAREHOUSE USAGE
- OLAM SYSTEM ARCHITECTURE
- DATA WAREHOUSE SECURITY
- SECURITY
- MCQ
- REFERENCES

Cube Operation

Cube definition and computation in DMQL

define cube sales[item, city, year]: sum(sales_in_dollars)

compute cube sales

Transform it into a SQL-like language (with a new operator cube by, introduced by Gray et al.'96)

SELECT item, city, year, SUM (amount)

FROM SALES

CUBE BY item, city, year

Need compute the following Group-Bys

(date, product, customer),

(date,product),(date, customer), (product, customer),

(date), (product), (customer)

()

Efficient Processing OLAP Queries

- Determine which operations should be performed on the available cuboids

  Transform drill, roll, etc. into corresponding SQL and/or OLAP operations,

  e.g., dice = selection + projection

- Determine which materialized cuboid(s) should be selected for OLAP op.
  - Let the query to be processed be on {brand, province_or_state} with the condition "year = 2004", and there are 4 materialized cuboids available:

    1) {year, item_name, city}

    2) {year, brand, country}

    3) {year, brand, province_or_state}

    4) {item_name, province_or_state} where year = 2004

    Which should be selected to process the query?

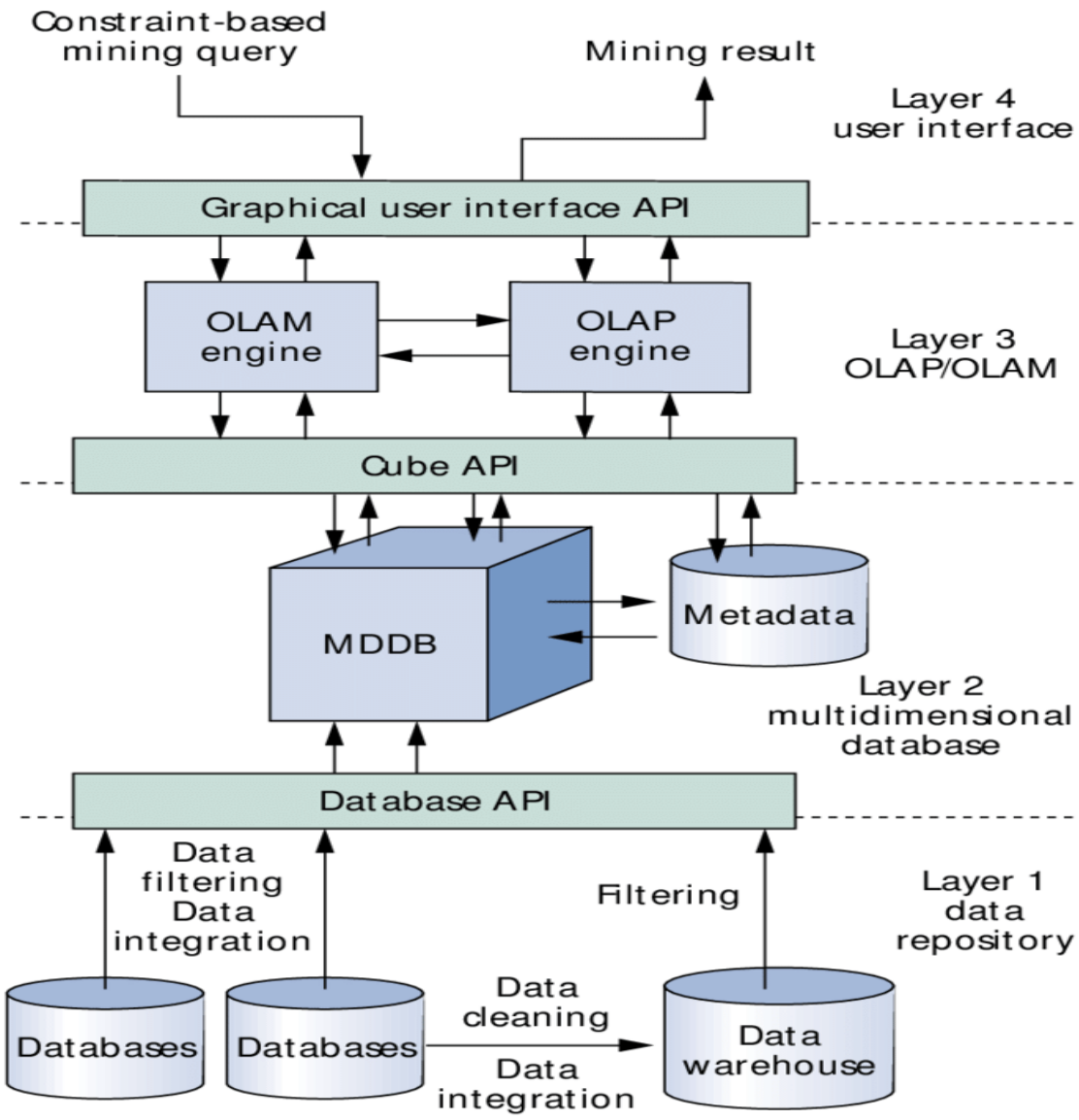
- Explore indexing structures and compressed vs. dense array structs in MOLAP

Data Warehouse Usage

Three kinds of data warehouse applications

- **Information processing**
  - ❑ supports querying, basic statistical analysis, and reporting
  - ❑ using crosstabs, tables, charts and graphs
- **Analytical processing**
  - ❑ multidimensional analysis of data warehouse data
  - ❑ supports basic OLAP operations, slice-dice, drilling, pivoting
- **Data mining**
  - ❑ knowledge discovery from hidden patterns
  - ❑ supports associations, constructing analytical models, performing classification and prediction, and presenting the mining results using visualization tools

OLAM System Architecture



Data Warehouse Security

Responsibility and Confidentiality :

- The Data Warehouse contains confidential and sensitive University data. In order to use its data, you must have proper authorization. Your authorization means that you have the authority to use the data and the responsibility to share stewardship of the data with the other users of the collection.
- Once authorized, you can access the data that you need to do your job. All authorized users are cautioned, however, that they are entrusted to use the data they retrieve from the Warehouse with care. Confidential data should not be released to others except for those with a "legitimate need to know."
- Please remember that you should never share Business Objects queries with other users with the data intact -- send the query without the data. More information about sending and saving Business Objects documents.

Querying Data with Security Restrictions :

- If you execute a query requesting data that you are not authorized to access, you will get results which may be incomplete because they are missing the data you are not allowed to access.
- If your authorization is limited to a specific set of data, be sure when querying the data that your record selection conditions include your security restrictions. For example, if you are authorized to access just data for a particular department, one of your record selection conditions should state something like "If Organization= 'My Organization'," where My Organization is the code of your department. This will document why the query gets the results it does, and will also help your query run faster.

SECURITY

- A DWH by nature is an open accessible system. The aim of DWH is generally to make large amounts of data easily accessible to the users, thereby enabling the users to extract information about the business as a whole.
- It is important to establish early any security and audit requirements that will be placed on the DWH.
- Clearly, adding security will affect performance because further checks require CPU cycles and time to perform.
- **Requirement :** Security can affect many different parts of the DWH such as :
  - User access – can be done by
    - Data classification
    - By Level of security required
    - By Data sensitivity
    - By job Function
  - User Classification
    - Top-down company hierarchy (department , section, group)
    - By Role
- Data load
- Legal Requirements : it is vital to establish any legal requirements (law) on the
- data being stored.
- Audit Requirements : such as connections, disconnections, data access, data change.
- Network Requirements : (routes)

SECURITY

❑**Data movement** : Type of file to be moved & the manner in which file has to be moved (flat file, encrypted / decrypted, summary generation, results temporary tables)

❑**Documentation :** It is important to get all the security and audit requirements clearly documented as this will be needed as part of any cost justification. This document should contain all the information gathered on:

- ❑ Data classification
- ❑ User classification
- ❑ Network requirement
- ❑ Data movement & storage requirements
- ❑ All auditable actions

❑**Query generation**

Multiple Choice Question

1. _____ are designed to overcome any limitations placed on the warehouse by the nature of the relational data model.
a) Operational database.
b) Relational database.
c) Multidimensional database.
d) Data repository.

2.. MDDB stands for _____.
a) multiple data doubling.
b) multidimensional databases.
c) multiple double dimension.
d) multi-dimension doubling.

3. _____ is data about data.
a) Metadata
b) Microdata
c) Minidata
d) Multidata

4. _____ is an important functional component of the metadata.
a) Digital directory.
b) Repository
c) Information directory.
d) Data dictionary.

5. EIS stands for _____.
a) Extended interface system.
b) Executive interface system.
c) Executive information system.
d) Extendable information system.

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
- 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.