Chapter 11

Advanced Database Concepts

Database Management System (DBMS)
4th Semester , BEIT & BCE

Compiled by :

Ankit Bhattarai, Assistant Professor
ankitbhattarai@cosmoscollege.edu.np

Cosmos College of Management & Technology

Syllabus

Full marks: 100

Internal: 50
Final: 50

Internal Marks:

Theory: 30
Practical: 20

Unit	Contents	Hours	Remarks
1.	Introduction	4	
2.	Data Models	4	
3.	Relational Model	4	
4.	Relational Database Query Languages	8	Important
5.	Database Constraints and Relational Database Design	8	Important
6.	Security	3	
7.	Query Processing	3	
8.	File Organization & Indexing	4	
9.	Crash Recovery	3	
10.	Transaction Processing & Concurrency Control	4	
11.	Advanced Database Concepts	3	

Chapter 11

Main Topics

(3 hours)

- Object Oriented Database Concept
- Distributed Database Management System (DDBMS)
- Data Warehouse

  (Homogenous Vs Heterogeneous Database System, Advantages & Disadvantages of DDBMS)

Object Oriented Database Concepts

Concept of Object Oriented

- It is a database management system (DBMS) that supports the modelling and creation of data as objects.
- This includes some kind of support for classes of objects and the inheritance of class properties and methods by subclasses and their objects.
- Criteria to be object oriented database system
    i. It should be DBMS
    ii. It should be object oriented system.

Object Oriented Database Concepts

It should be DBMS.

The first criterion translates into five features: persistence, secondary storage management, concurrency, recovery and an ad hoc query facility.

It should be object oriented system

The second one translates into eight features: complex objects, object identity, encapsulation, types or classes, inheritance, overriding combined with late binding, extensibility and computational completeness.

Object Oriented Database Concepts

Object-Oriented Data Model (OODM): A (logical) data model that captures the semantics of objects supported in objected-oriented programming.

Object-Oriented Database (OODM): A persistent and sharable collection of objects defined by an OODM.

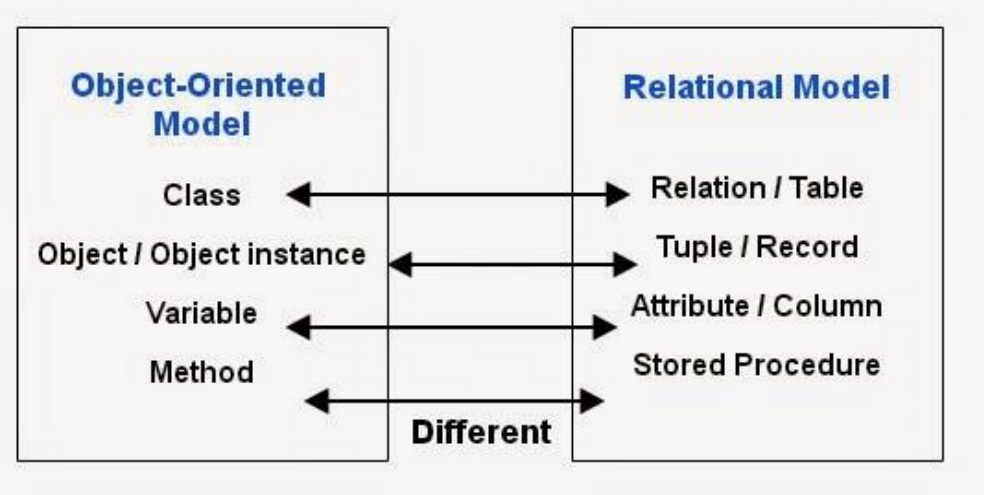


Figure I. Comparison between Object Oriented Model & Relational Model

Object Oriented Database Concepts

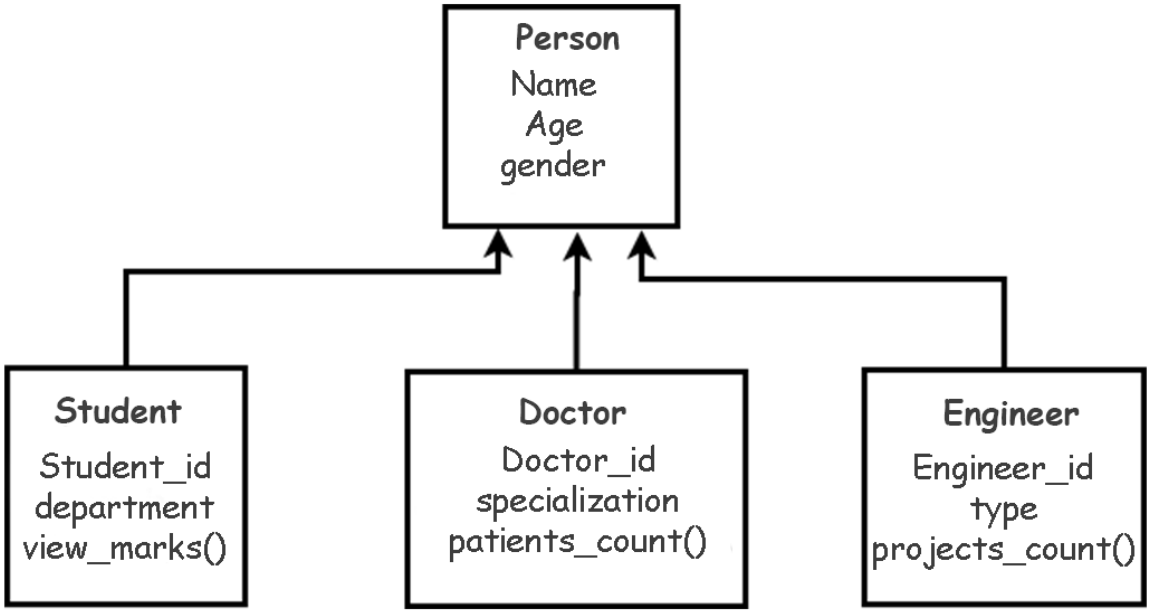


Figure II. Object Oriented Model

Object Oriented Database

Features

- **Persistence** (*that the data survives after the process with which it was created has ended)*
- **Support of transactions**
- **Simple querying of bulk data**
- **Concurrent access**
- **Resilience** *(failure with recovery)*
- **Security**

Why Object-Oriented Database ?

- Industry Trends: Integration and Sharing
- Seamless integration of operating systems, databases, languages, spreadsheets, word processors, AI expert system shells.
- Sharing of data, information, software components, products, computing environments.
- Referential sharing: Multiple applications, products, or objects share common sub-objects.
- (Hypermedia links are then used to navigate from one object to another)
- Object-oriented databases allow referential sharing through the support of object identity and inheritance.

Distributed Database Management System (DDBMS)

- A Distributed Database Management System (DDBMS) consists of a single logical database that is split into a number of fragments.
- Each fragment is stored on one or more computers under the control of a separate DBMS, with the computers connected by a communications network.
- A distributed database system consists of loosely coupled sites that share no physical component but appears to user as a single system.
- **Distributed Database**: A logically interrelated collection of shared data (and a description of this data) physically distributed over a computer network.

Distributed Database Management System (DDBMS)

- **Distributed DBMS**: The software system that permits the management of the distributed database and makes the distribution transparent to users.
- **Reasons for Distributed Database**

i. Business unit autonomy and distribution
ii. Data sharing
iii. Data communication costs
iv. Data communication reliability and costs
v. Multiple application vendors
vi. Database recovery
vii. Transaction and analytic processing

Distributed Database Management System (DDBMS)

- **Distributed DBMS**: The software system that permits the management of the distributed database and makes the distribution transparent to users.
- **Reasons for Distributed Database**

i. Business unit autonomy and distribution
ii. Data sharing
iii. Data communication costs
iv. Data communication reliability and costs
v. Multiple application vendors
vi. Database recovery
vii. Transaction and analytic processing

Distributed Database Management System (DDBMS)

- **Homogeneous Database System**
- Same DBMS at each node
- Easy to manage, difficult to enforce
- All sites have identical software
- They are aware of each other and agree to cooperate in processing user requests.
- It appears to user as a single system

Distributed Database Management System (DDBMS)

- **Heterogeneous Database System**
- Different DBMSs at different nodes
- Difficult to manage, preferred by independent organizations
- In a heterogeneous distributed database system, at least one of the databases uses different schemas and software.
- A database system having different schema may cause a major problem for query processing.

Distributed Database Management System (DDBMS)

- **Advantage:**

➢ Availability: failure of site containing relation r does not result in unavailability of r is replicas exist.

➢ Parallelism: queries on r may be processed by several nodes in parallel.

➢ Reduced data transfer: relation r is available locally at each site containing a replica of r.

- **Disadvantage:**

➢ Increased cost of updates: each replica of relation r must be updated.

➢ Increased complexity of concurrency control: concurrent updates to distinct replicas may lead to inconsistent data unless special concurrency control mechanisms are implemented.

Distributed Database Management System (DDBMS)

- **Fragmentation**

➢ A relation may be divided into a number of sub-relations, called fragments, which are then distributed.

➢ Data can also be distributed by decomposing a table and storing portions at different sites – called Fragmentation.

➢ There are two types of fragmentation.

i. Horizontal fragmentation

ii. Vertical fragmentation

- **Horizontal Fragmentation**

➢ It consists of a subset of the tuples of a relation.

➢ It groups the tuples in a relation that are collectively used by the important transactions.

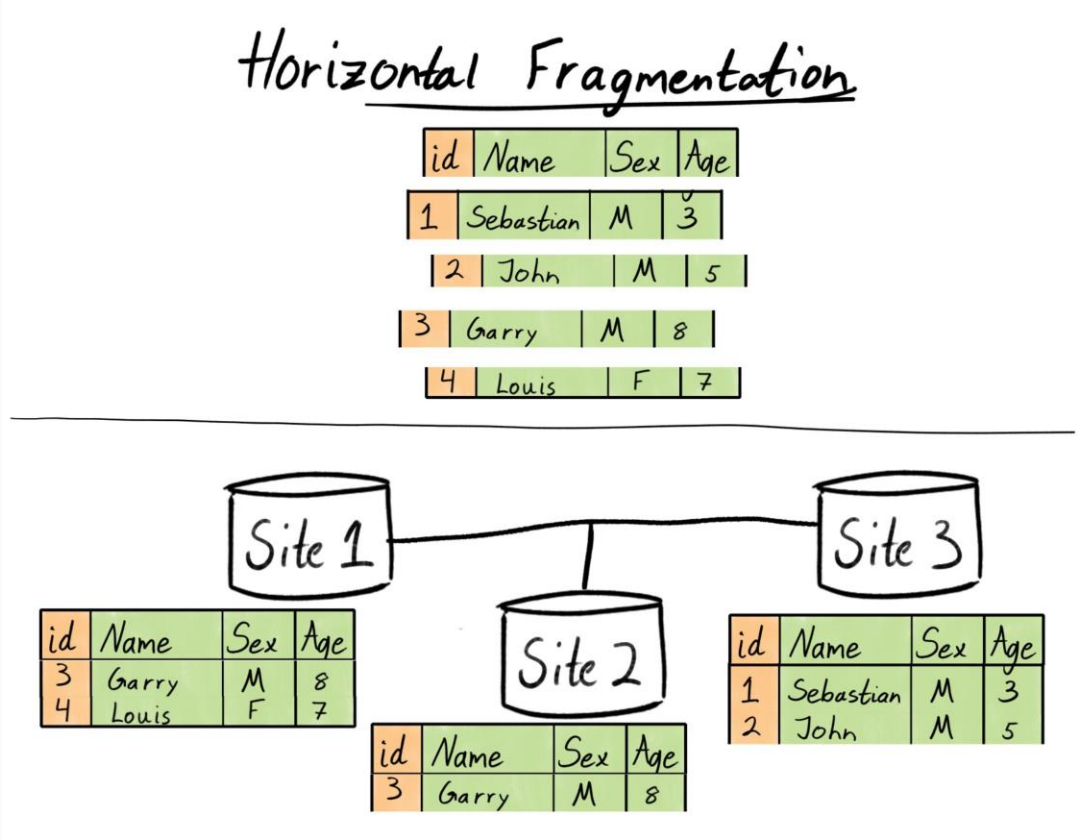
➢ A horizontal fragment is produced by specifying a predicate that performs a restriction on the tuples in the relation.

Example:
EmpAddress = (EmployeeID, Street, City, HouseNumber)

$$EmpAddress1 = \sigma_{City = \text{'Kalimati'}}(EmpAddress)$$

$$EmpAddress2 = \sigma_{City = \text{'Kathmandu'}}(EmpAddress)$$

Horizontal Fragmentation



Distributed Database Management System (DDBMS)

- **Vertical Fragmentation**

➢ Consists of a subset of the attributes of a relation.

➢ It groups together the attributes in a relation that are used jointly by the important transactions.

➢ A vertical fragment is defined using the projection operations of the relation algebra.

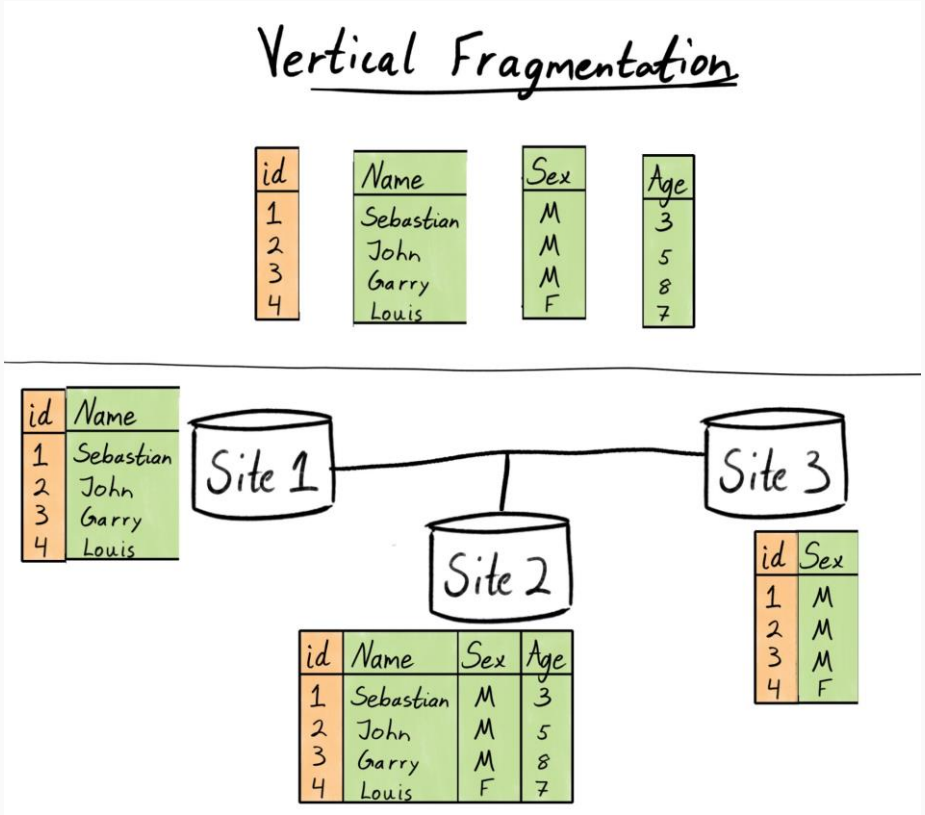
Example:

Customer (CustomerID, CustomerName, CustomerAge, CustomerGender)

$$Customer1 = \pi_{CustomerName,CustomerAge}(Customer)$$

$$Customer1 = \pi_{CustomerID,CustomerAge,\ CustomerGender}(Customer)$$

Vertical Fragmentation



Distributed Database Management System (DDBMS)

➢ **Properties of distributed database**

- A collection of logically related shared data;
- The data is split into a number of fragments;
- Fragments may be replicated;
- Fragments/replicas are allocated to sites;
- The sites are linked by a communications network;
- The data at each site is under the control of a DBMS;
- The DBMS at each site can handle local applications, autonomously;
- Each DBMS participates in at least one global application

Data mining

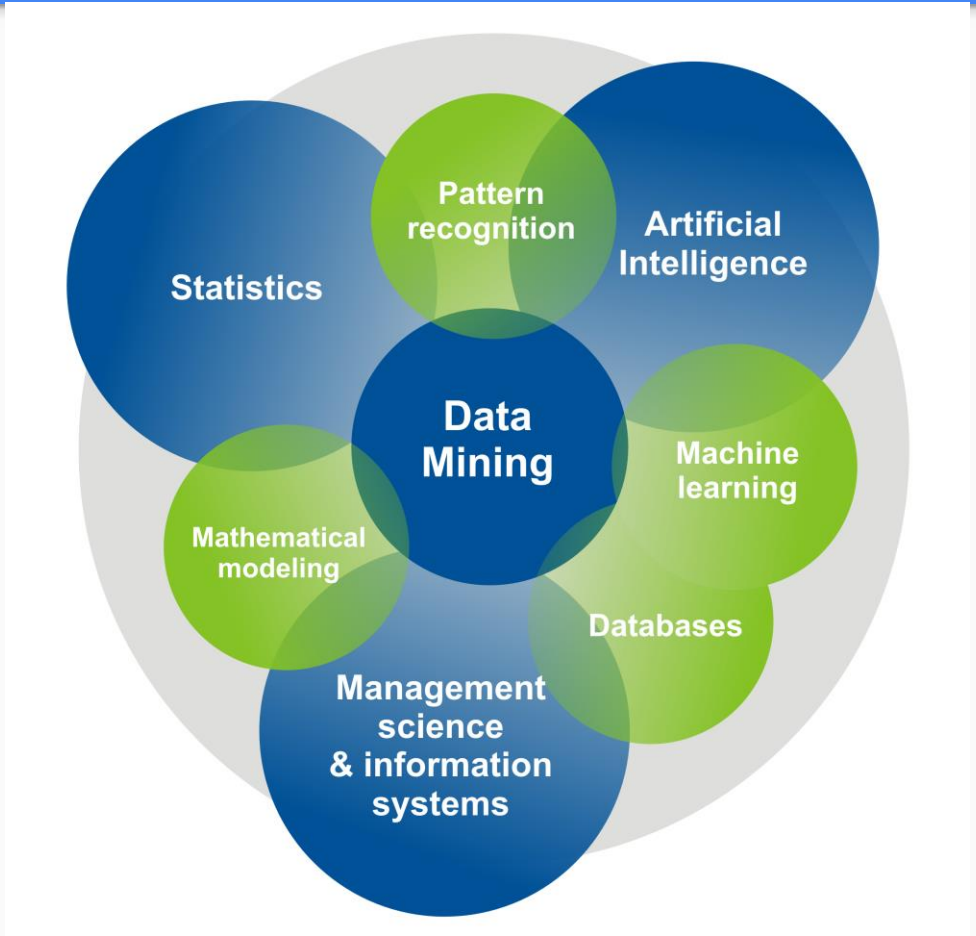
Definition:

“ Discovering meaningful patterns and trends using some *mathematical algorithm* on huge amount of stored data.”

“Extraction of interesting, non-trivial, implicit, previously unknown and potentially useful information or patterns from data.”

“Data mining is basically concerned with the analysis of data and the use of software techniques for finding patterns and regularities in sets of data.”

Data Mining: Confluence of Multiple Disciplines



Application Of Data Mining

- Web page analysis: from web page classification, clustering to PageRank & HITS algorithms
- Collaborative analysis & recommender systems
- Basket data analysis to targeted marketing
- Biological and medical data analysis: classification, cluster analysis (microarray data analysis), biological sequence analysis, biological network analysis
- Data mining and software engineering (e.g., IEEE Computer, Aug. 2009 issue)
- From major dedicated data mining systems/tools (e.g., SAS, MS SQL Server Analysis Manager, Oracle Data Mining Tools) to invisible data mining.

Concept of Data Warehouse

- **“A data warehouse is a subject-oriented, integrated, time variant, and nonvolatile collection of data in support of management’s decision-making process.”** — *W. H. Inmon*
  - Data warehousing emphasizes the capture of data from diverse sources for useful analysis and access, but does not generally start from the point-of-view of the end user who may need access to specialized, sometimes local databases.
  - A data warehouse is a database, which is kept separate from the organization's operational database. There is no frequent updating done in a data warehouse.

Concept of Data Warehouse

- It possesses consolidated historical data, which helps the organization to analyze its business.
- A data warehouse helps executives to organize, understand, and use their data to take strategic decisions.
- Data warehouse systems help in the integration of diversity of application systems.
- A data warehouse system helps in consolidated historical data analysis.

Data Warehouse **should be (Features)**

1. Subject Oriented

- It organized around major subjects such as customer, products, sales, etc.
- Provide a simple/concise view around subject specific issues by excluding data that are not useful in the decision support process.
- Focusing on the modeling/analysis of data for decision makers.

Data Warehouse **should be**

2. Integrated

- Constructed by integrating multiple, heterogeneous data sources (*relational databases*, *flat files*, *on-line transaction records*).
- In a data warehouse, all entities should be integrated and consistent i.e. only one name must exist to describe each individual entity.

Data Warehouse **should be**

3. Time Variant

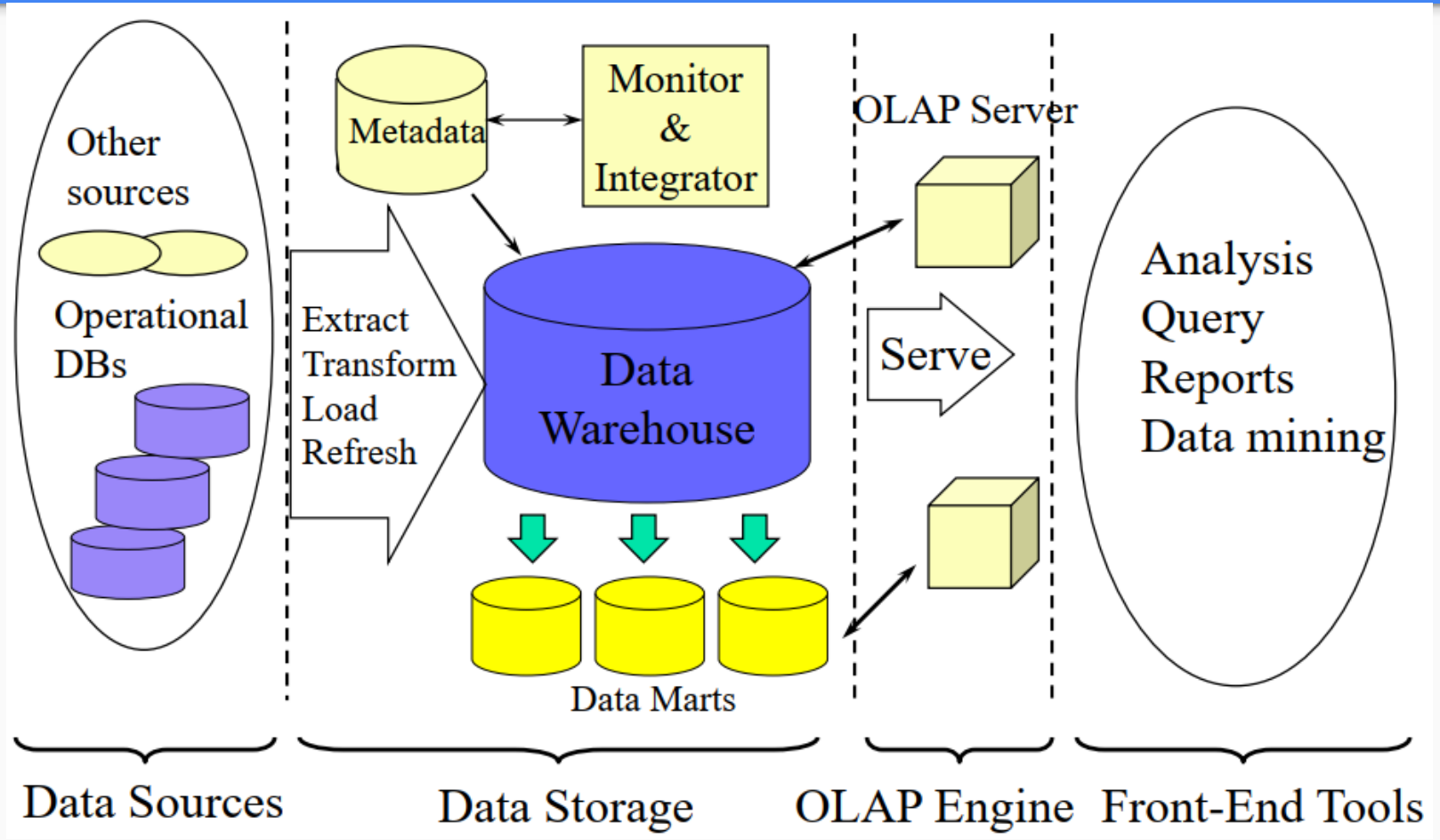
- The time horizon for the data warehouse is significantly longer than that of operational systems.
- Operational database: current value data
  Data warehouse data: provide information from a historical perspective (e.g., past 5-10 years)
- Information can then be sourced according to period.

Data Warehouse **should be**

4. Non volatile

- Operational update of data does not occur in the data warehouse environment.
- End-users who want to update data must use operational database.
- A data warehouse will always be filled with historical data.

Data Warehouse: Multi tier Architecture



Questions

Write short notes on:

i. ORM (Object Relational Model)

ii. Object Oriented DBMS (OODBMS)

iii. Data Mining and data warehouse

2. Compare and Contrast between heterogeneous and homogeneous distributed databases

THANK YOU

Any Queries ?