OLAP (Online Analytical Processing)

G.Lakshmi Priya & Razia Sultana.A

Assistant Professor/IT

OVERVIEW
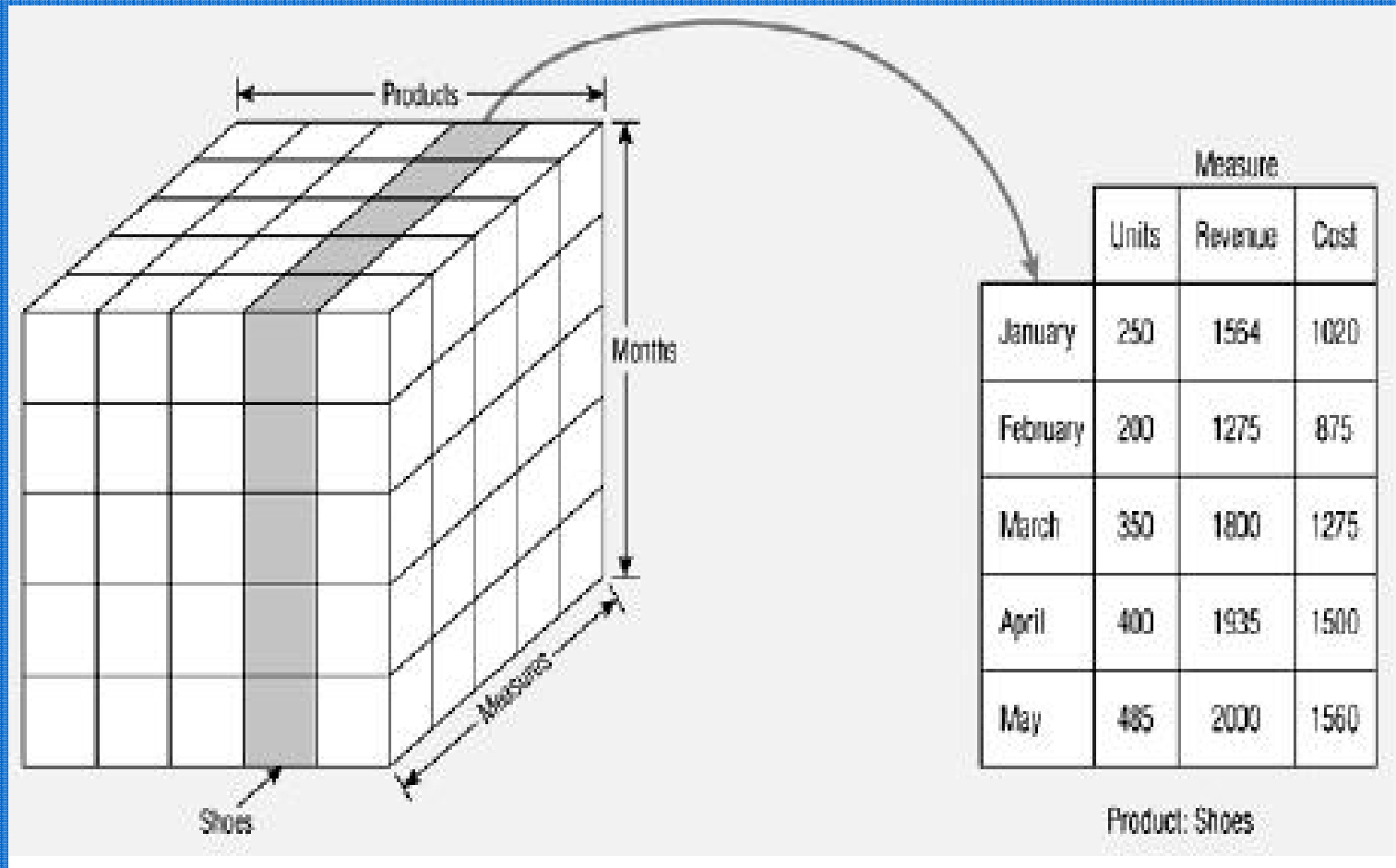
- INTRODUCTION
- OLAP CUBE
- HISTORY OF OLAP
- OLAP OPERATIONS
- DATAWAREHOUSE
- DATAWAREHOUSE ARCHITECHTURE
- DIFFERENCE BETWEEN OLAP & OLTP
- TYPES OF OLAP
- APPLICATIONS OF OLAP

INTRODUCTION TO OLAP

- OLAP (online analytical processing) is **computer processing** that enables a user to easily and selectively extract and view data from different points of view.

- OLAP allows users to **analyze database information** from multiple database systems at one time.

- OLAP data is stored in **multidimensional databases**.

AN EXAMPLE...

Products

Months

Measures

Shoes

	Measure		
	Units	Revenue	Cost
January	250	1554	1020
February	200	1275	875
March	350	1800	1275
April	400	1935	1500
May	485	2000	1560

Product: Shoes

- Some popular OLAP server software programs include:
  - **Oracle Express Server**
  - **Hyperion Solutions Essbase**

- OLAP processing is often used for **data mining.**

- OLAP products are **typically designed for multiple-user environments**, with the cost of the software based on the number of users.
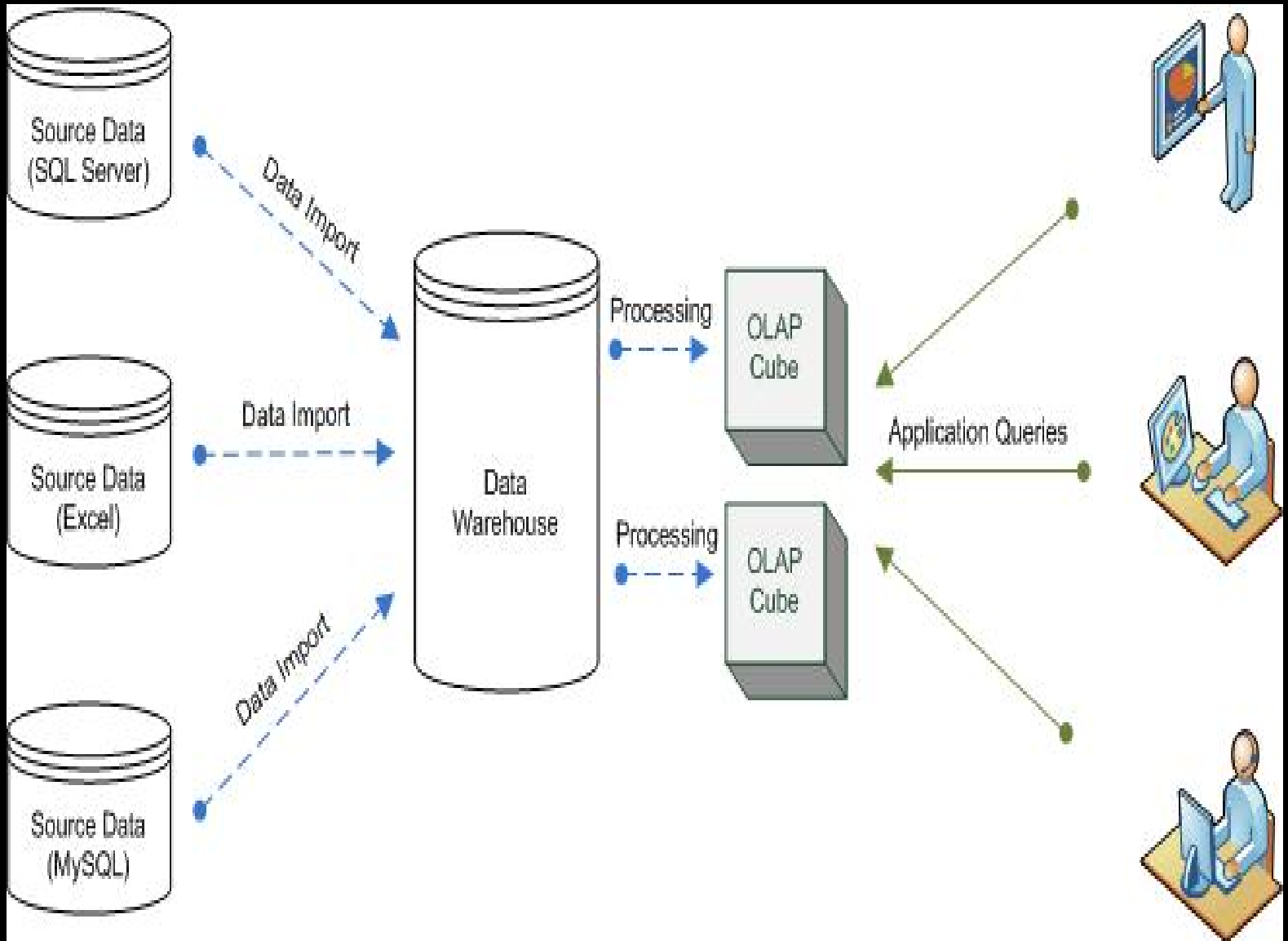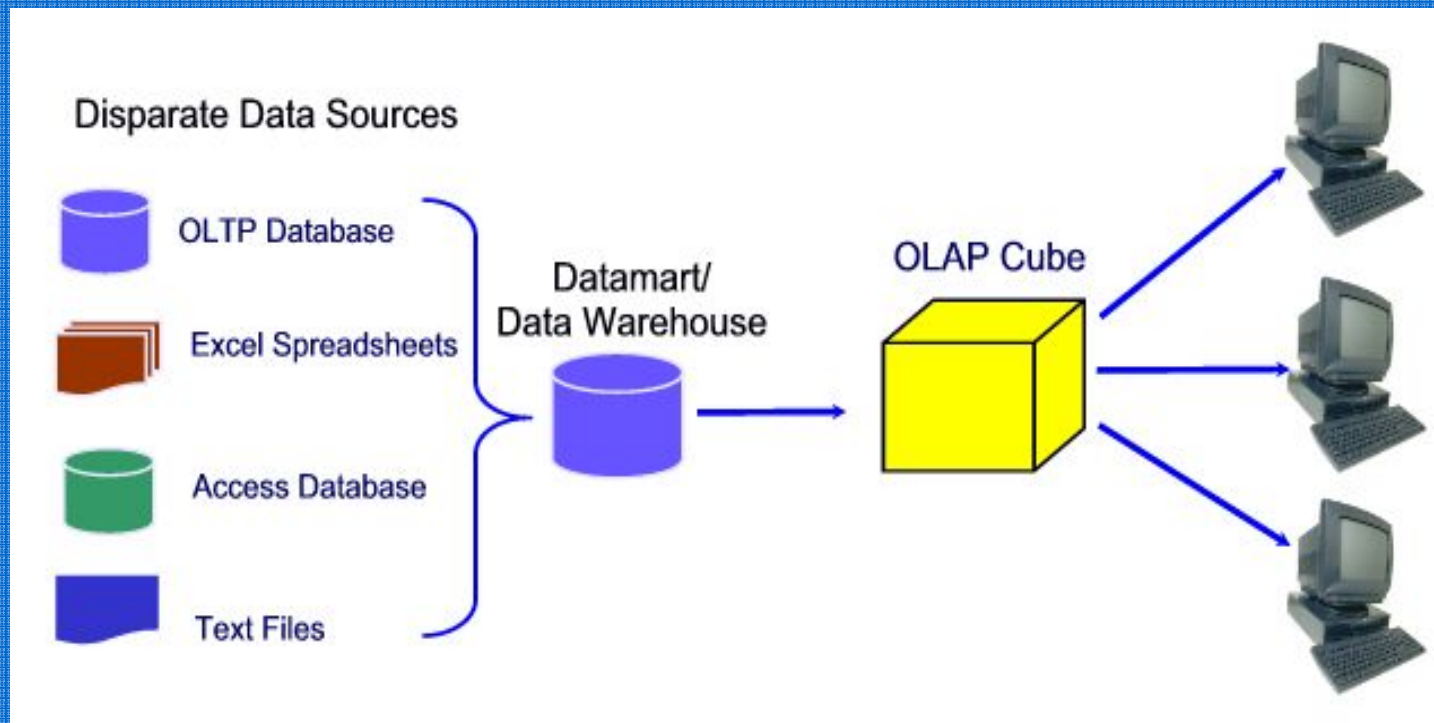
- An OLAP Cube is a data **structure that allows fast analysis of data.**

- The arrangement of data into cubes overcomes a limitation of relational databases.

- It consists of numeric facts called **measures** which are categorized by dimensions.

- The OLAP cube consists of numeric facts called **measures** which are categorized by **dimensions.**
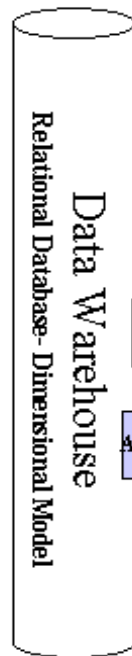
- A multidimensional cube can combine data from disparate data sources and store the information in a fashion that is logical for business users.
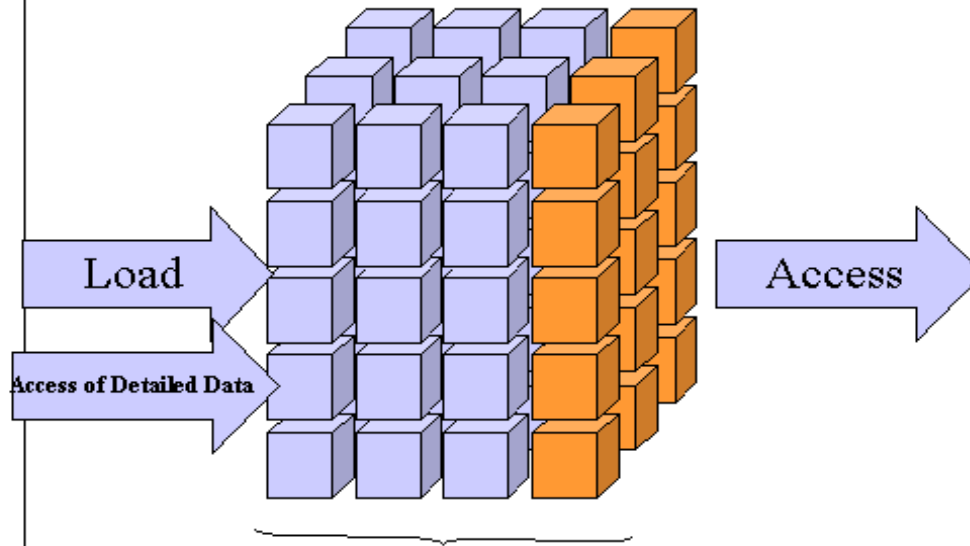

Disparate Data Sources

OLTP Database

Excel Spreadsheets

Access Database

Text Files

Datamart/ Data Warehouse

OLAP Cube

OLAP CUBE

HISTORY OF OLAP

- The term **OLAP** was created as a slight modification of the traditional database term **OLTP (Online Transaction Processing)**.

- Databases configured for OLAP employ a multidimensional data model, allowing for complex analytical and ad-hoc queries with a rapid execution time.

- They borrow aspects of navigational databases and hierarchical databases that are speedier than their relational kind.

- **Nigel Pendse** has suggested that an alternative and perhaps more descriptive term to describe the concept of OLAP is **Fast Analysis of Shared Multidimensional Information (FASMI)**.

- The first product that performed OLAP queries was **Express,** which was **released in 1970** (and **acquired** by **Oracle in 1995 from Information Resources**). However, the term did not appear until 1993 when it was coined by **Ted Codd**, who has been described as **"the father of the relational database"**.

- The user-initiated process of navigating by calling for page displays interactively, through the specification of slices via **rotations and drill down/up** is sometimes called "**slice and dice**".

- **Slice**: A slice is a subset of a multi-dimensional array corresponding to a single value for one or more members of the dimensions not in the subset.

- **Dice**: The dice operation is a slice on more than two dimensions of a data cube (or more than two consecutive slices).

- **Drill Down/Up**: Drilling down or up is a specific analytical technique whereby the user navigates among levels of data ranging from the most summarized (up) to the most detailed (down).

- **Roll-up**: A roll-up involves computing all of the data relationships for one or more dimensions. To do this, a computational relationship or formula might be defined.

- **Pivot**: To change the dimensional orientation of a report or page display.

- The output of an OLAP query is typically displayed in a matrix (or pivot) format. The dimensions form the row and column of the matrix; the measures, the values.

DATA WAREHOUSE

- A data warehouse is a repository of an organization's electronically stored data.
- A data warehouse is a
  - subject-oriented,
  - integrated,
  - time-varying,
  - non-volatile

  collection of data that is used primarily in organizational decision making.
- The essential components of a data warehousing system are the means to:
- Retrieve & Analyze data
- Extract, Transform & Load data
- Manage the data dictionary.

- Data warehouse is a collection of data designed to support management decision making.

- Data warehouses contain a wide variety of data that present a coherent picture of business conditions at a single point in time.

- The term data warehousing generally refers to the combination of many different databases across an entire enterprise.

BENEFITS

- ✓ A data warehouse provides a common data model for all data of interest regardless of the data's source.

- ✓ Prior to loading data into the data warehouse, inconsistencies are identified and resolved. This greatly simplifies reporting and analysis.

- ✓ Information in the data warehouse is under the control of data warehouse users so that, even if the source system data is cleared over time, the information in the warehouse can be stored safely for extended periods of time.

✓ Because they are separate from operational systems, data warehouses provide retrieval of data without slowing down operational systems.

✓ Data warehouses facilitate decision support system applications such as trend reports, exception reports, and reports that show actual performance versus goals.

✓ Data warehouses can work in conjunction with and, hence, enhance the value of operational business applications, notably customer relationship management (CRM) systems.
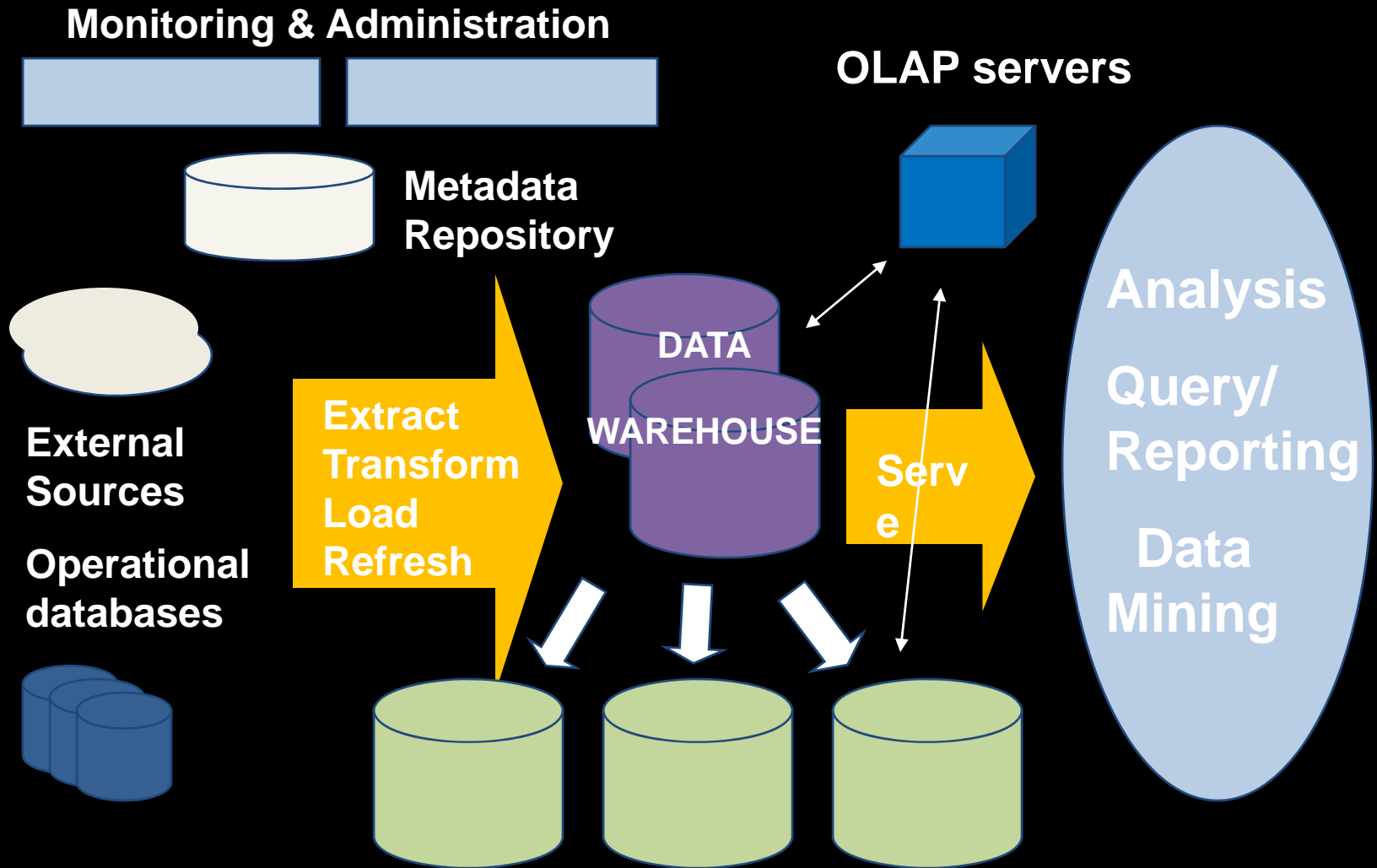
DATA WAREHOUSE ARCHITECHTURE

- Architechture is a conceptualization of how the data warehouse is built.

- One possible simple conceptualization of a data warehouse architecture consists of the following interconnected layers:

- ❖ **Operational database layer:** The source data for the data warehouse - An organization's ERP systems fall into this layer.

- ❖ **Informational access layer:** The data accessed for reporting and analyzing and the tools for reporting and analyzing data - Business intelligence tools fall into this layer. And the Inmon-Kimball differences about design methodology, discussed later in this article, have to do with this layer.

❖ **Data access layer:** The interface between the operational and informational access layer - Tools to extract, transform, load data into the warehouse fall into this layer.

❖ **Metadata layer:** The data directory - This is often usually more detailed than an operational system data directory. There are dictionaries for the entire warehouse and sometimes dictionaries for the data that can be accessed by a particular reporting and analysis tool.

DATA WAREHOUSING ARCHITECHURE

Monitoring & Administration

OLAP servers

Metadata Repository

External Sources

Operational databases

Extract Transform Load Refresh

DATA WAREHOUSE

Serve

Analysis

Query/ Reporting

Data Mining

APPLICATIONS OF
DATA WAREHOUSES

✓Data Mining

✓Web Mining

✓Decision Support Systems (DSS)

TWO TYPES OF
DATABASE ACTIVITY

- **OLTP (Online-Transaction Processing)**

- **OLAP (Online-Analytical Processing)**

- **OLTP**: **O**n-**L**ine **T**ransaction **P**rocessing
- Short Transaction both query and updates
- (e.g., update account balance, enroll is courses)
- Queries are Simple
- (e.g., find account balance, find grade in courses)
- Updates are frequent
- (e.g., Concert tickets, seat reservations, shopping carts)

- **OLAP**: **O**n-**L**ine **A**nalytical **P**rocessing
- Long transactions, usually Complex queries.
- (e.g., all statistics about sales, grouped by department and month)
- "Data mining" operations.
- Infrequent Updates.

DIFFERENCE BETWEEN OLTP & OLAP

Item	OLTP	OLAP
User	IT Professional	Knowledge worker
Functional	Daily task	Decision Making
DB Design	Application oriented	Subject oriented
Data	Up to date, detail, relational	Historical, multidimensional, integrated
Access	Read/write	Read only
DB Size	100 MB-GB	100 GB-TB

TYPES OF OLAP

- **Relational OLAP(ROLAP):**
Extended RDBMS with multidimensional data mapping to standard relational operation.

- **Multidimensional OLAP(MOLAP):** Implemented operation in multidimensional data.

- **Hybrid OnlineAnalytical Processing (HOLAP)** is a hybrid approach to the solution where the aggregated totals are stored in a multidimensional database while thedetail data is stored in the relational database. This is the balance between the data efficiency of the ROLAP model and the performance of the MOLAP model.

Relational OLAP

- Provides functionality by using relational databases and relational query tools to store and analyze multidimensional data.
- Build on existing relational technologies and represent extension to all those companies who already used RDBMS.
- Multidimensional data schema support within the RDBMS.
- Data access language and query performance are optimized for multidimensional data.
- Support for very large databases.

Multidimensional OLAP

- MOLAP extends OLAP functionality to MDBMS.
- Best suited to manage, store and analyze multidimensional data.
- Proprietary techniques used in MDBMS.
- MDBMS and users visualize the stored data as a 3-Dimensional Cube i.e Data Cube.
- MOLAP Databases are known to be much faster than the ROLAP counter parts.
- Data cubes are held in memory called "Cube Cache"

ROLAP v/s MOLAP

Characteristics	ROLAP	MOLAP
SCHEMA	User star Schema •Additional dimensions can be added dynamically.	User Data cubes •Addition dimensions require recreation of data cube.
Database Size	Medium to large	Small to medium
Architecture	Client/Server	Client/Server
Access	Support ad-hoc requests	Limited to pre-defined dimensions

Characteristics	ROLAP	MOLAP
Resources	HIGH	VERY HIGH
Flexibility	HIGH	LOW
Scalability	HIGH	LOW
Speed	• Good with small data sets. • Average for medium to large data set.	• Faster for small to medium data sets. • Average for large data sets.

Implementation of OLAP server

- **ROLAP**:
- Data is stored in tables in relational database or extended relational databases.
- They use an RDBMS to manage the warehouse data and aggregations using often a star schema.
- **Advantage**:
- Scalable
- **Disadvantage**:
- Direct access to cells.

- **MOLAP**:
- Implements the multidimensional view by storing data in special multidimensional data structures.
- **Advantages**:
- Fast indexing to pre-computed aggregations.
- Only values are stored.
- **Disadvantage**:
- Not very Scalable

❖OLE DB for OLAP

OLE DB for OLAP (abbreviated **ODBO**) is a Microsoft published specification and an industry standard for multi-dimensional data processing.

ODBO is the standard **application programming interface (API)** for exchanging metadata and data between an OLAP server and a client on a Windows platform.

ODBO was specifically designed for Online Analytical Processing (OLAP) systems by Microsoft as an extension to Object Linking and Embedding Database (OLE DB).

❖**Marketing and sales analysis**

❖**Consumer goods industries**

❖**Financial services industry** (insurance, banks etc)

❖**Database Marketing**

BENEFITS OF OLAP

- One main benefit of OLAP is **consistency of information and calculations.**

- **"What if" scenarios** are some of the most popular uses of OLAP software and are made eminently more possible by multidimensional processing.

- It allows a manager to **pull down data** from an OLAP database in broad or specific terms.

- OLAP **creates a single platform for all the information and business needs, planning, budgeting, forecasting, reporting and analysis.**