# ***Lectures in Data Warehouse (4th I-Systems)(Work 1)***

## *Lecture (1)*

**A data warehouse is constructed by integrating data from multiple heterogeneous sources. It supports analytical reporting, structured and/or ad hoc queries and decision making. This tutorial adopts a step-by-step approach to explain all the necessary concepts of data warehousing.**

## Audience

**This tutorial will help computer science graduates to understand the basic-to-advanced concepts related to data warehousing.**

## Prerequisites

**Before proceeding with this tutorial, you should have an understanding of basic database concepts such as schema, ER model, structured query language, etc.**

## Copyright & Disclaimer



**المحاضرات أدناه موزعة على خمسة عشر أسبوعا (كل فصل لكل أسبوع)، والفصل(١٦) تضمن أسئلة وأجوبة لمراجعة للمادة المعطاة..**

**مدرس مادة مستودع البيانات: أ.د. مرتضى محمد حمد**
**الصف الرابع/ قسم نظم المعلومات**
**كلية علوم الحاسوب وتكنلوجيا المعلومات/جامعة الأنبار**

# Table of Contents

# Table of Contents

# Table of Contents

# Table of Contents

# 1. OVERVIEW

The term "Data Warehouse" was first coined by Bill Inmon in 1990. According to Inmon, a data warehouse is a subject-oriented, integrated, time-variant, and non-volatile collection of data. This data helps analysts to take informed decisions in an organization.

An operational database undergoes frequent changes on a daily basis on account of the transactions that take place. Suppose a business executive wants to analyze previous feedback on any data such as a product, a supplier, or any consumer data, then the executive will have no data available to analyze because the previous data has been updated due to transactions.

A data warehouses provides us generalized and consolidated data in multidimensional view. Along with generalized and consolidated view of data, a data warehouses also provides us Online Analytical Processing (OLAP) tools. These tools help us in interactive and effective analysis of data in a multidimensional space. This analysis results in data generalization and data mining.

Data mining functions such as association, clustering, classification, prediction can be integrated with OLAP operations to enhance the interactive mining of knowledge at multiple level of abstraction. That's why data warehouse has now become an important platform for data analysis and online analytical processing.

## Understanding a Data Warehouse

- 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.
- 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.

# Why a Data Warehouse is Separated from Operational Databases

A data warehouses is kept separate from operational databases due to the following reasons:

- An operational database is constructed for well-known tasks and workloads such as searching particular records, indexing, etc. In contrast, data warehouse queries are often complex and they present a general form of data.
- Operational databases support concurrent processing of multiple transactions. Concurrency control and recovery mechanisms are required for operational databases to ensure robustness and consistency of the database.
- An operational database query allows to read and modify operations, while an OLAP query needs only **read only** access of stored data.
- An operational database maintains current data. On the other hand, a data warehouse maintains historical data.

# Data Warehouse Features

The key features of a data warehouse are discussed below:

- **Subject Oriented** - A data warehouse is subject oriented because it provides information around a subject rather than the organization's ongoing operations. These subjects can be product, customers, suppliers, sales, revenue, etc. A data warehouse does not focus on the ongoing operations, rather it focuses on modelling and analysis of data for decision making.
- **Integrated** – A data warehouse is constructed by integrating data from heterogeneous sources such as relational databases, flat files, etc. This integration enhances the effective analysis of data.
- **Time Variant** - The data collected in a data warehouse is identified with a particular time period. The data in a data warehouse provides information from the historical point of view.
- **Non-volatile** - Non-volatile means the previous data is not erased when new data is added to it. A data warehouse is kept separate from the operational database and therefore frequent changes in operational database is not reflected in the data warehouse.

**Note:** A data warehouse does not require transaction processing, recovery, and concurrency controls, because it is physically stored and separate from the operational database.

# Data Warehouse Applications

As discussed before, a data warehouse helps business executives to organize, analyze, and use their data for decision making. A data warehouse serves as a sole part of a plan-execute-assess "closed-loop" feedback system for the enterprise management. Data warehouses are widely used in the following fields:

- Financial services
- Banking services
- Consumer goods
- Retail sectors
- Controlled manufacturing

# Types of Data Warehouse

Information processing, analytical processing, and data mining are the three types of data warehouse applications that are discussed below:

- **Information Processing** – A data warehouse allows to process the data stored in it. The data can be processed by means of querying, basic statistical analysis, reporting using crosstabs, tables, charts, or graphs.
- **Analytical Processing** – A data warehouse supports analytical processing of the information stored in it. The data can be analyzed by means of basic OLAP operations, including slice-and-dice, drill down, drill up, and pivoting.
- **Data Mining** - Data mining supports knowledge discovery by finding hidden patterns and associations, constructing analytical models, performing classification and prediction. These mining results can be presented using visualization tools.

| Data Warehouse (OLAP) | Operational Database(OLTP) |
|---|---|
| It involves historical processing of information. | It involves day-to-day processing. |
| OLAP systems are used by knowledge workers such as executives, managers, and analysts. | OLTP systems are used by clerks, DBAs, or database professionals. |

| | |
|---|---|
| It is used to analyze the business. | It is used to run the business. |
| It focuses on Information out. | It focuses on Data in. |
| It is based on Star Schema, Snowflake Schema, and Fact Constellation Schema. | It is based on Entity Relationship Model. |
| It focuses on Information out. | It is application oriented. |
| It contains historical data. | It contains current data. |
| It provides summarized and consolidated data. | It provides primitive and highly detailed data. |
| It provides summarized and multidimensional view of data. | It provides detailed and flat relational view of data. |
| The number of users is in hundreds. | The number of users is in thousands. |
| The number of records accessed is in millions. | The number of records accessed is in tens. |
| The database size is from 100GB to 100 TB. | The database size is from 100 MB to 100 GB. |
| These are highly flexible. | It provides high performance. |