↙↗ *What is the Data Warehouse?*

A data warehouse is a relational database that is designed for query and analysis rather than for transaction processing. It usually contains historical data derived from transaction data, but it can include data from other sources. In addition to a relational database, a data warehouse environment can include an Extraction, Transportation, Transformation, and Loading (ETL) solution, Online Analytical Processing (OLAP) and Data Mining capabilities, Client Analysis Tools, and other applications that manage the process of gathering data and delivering it to business users.

Formal Definition: “A data warehouse is a subject-oriented, integrated, time variant and non-volatile collection of data in support of management decision making process.”

It means:

• **Subject-Oriented**: *Stored data targets specific subjects.*
Example: It may store data regarding total Sales, Number of Customers, etc. and not general data on everyday operations.
Data warehouses are designed to help you analyze data. For example, to learn more about your company's sales data, you can build a data warehouse that concentrates on sales. Using this data warehouse, you can answer questions such as "Who was our best customer for this item last year?”.

• **Integrated**: *Data may be distributed across heterogeneous sources which have to be integrated*. Example: Sales data may be on RDB, Customer information on Flat files, etc.
Integration is the most important. Data is fed from multiple disparate sources into the data warehouse. As the data is fed, it is converted, reformatted, re- sequenced, summarized, and so forth. The result is that data once it resides in the data warehouse has a single physical corporate image.

• **Time Variant***: Data stored may not be current but varies with time and data have an element of time*. Example: Data of sales in last 5 years, etc.
A data warehouse's focus on change over time is what is meant by the term time variant. In order to discover trends in business, analysts need large amounts of data.

- **Non-Volatile**: *The data warehouse is read-only; it generally has only 2 operations performed on it: Loading of data and Access of data*.

Nonvolatile means that, once entered into the data warehouse, data should not change. This is logical because the purpose of a data warehouse is to enable you to analyze what has occurred.

Most companies have realized that collecting transactional data is useful. In fact, it is tough to find any company that does not record their transactions. The data that has been collected for a number of years reside in various data sources—some in the mainframes, some in proprietary systems, and some in client-server applications. Also, each of these systems was probably built and is being maintained by different people.

The typical dilemma of today's IT managers is not how to collect the data, but how to use the data accumulated over the years. The answer might sound simple: Put everything in one place and run reports against that database. Well, the programmer who built the mainframe system left the company 10 years ago. The consultants that were hired to build the proprietary system have since moved on to other jobs as well. Finally, you're already running the reports against the client server system you use for daily data collection, but those reports are fairly rigid— after they're printed, you can't really change or customize them. Each time you need a specific report, you have to pay a premium rate for a week or two to the outside consultant or to your own programmer. What can you do?

The goal of a data warehouse is to provide your company with an easy and quick look at its historical data. Advanced OLAP (on-line analytical processing) tools let DW users generate reports at a click of a mouse and look at the company's performance from various angles. How much data you need to examine depends on the nature of your business.

Suppose you have a manufacturing plant that produces thousands of parts per hour. The type of information you might be interested in includes the number of defects per hour or per day. Although you might want to examine the number of defective parts this year

against the same number five years ago, such a ratio probably wouldn't provide the best picture of the company's performance. On the other hand, if you're in a car rental business, you might want to examine the number of customers this month against the same number six months ago. If you need to analyze the purchasing trends for customers with various demographic backgrounds, you might wish to examine data collected for a number of years. In short, if you need to make use of the data residing in some or all of your systems, you need to build a data warehouse.

↙ *What is the different between operations of Data base systems and data warehouse?*

	data warehouse	**Operational Database**
User	Knowledge worker	Clerk
Function	Decision support	Day to day operation
Data	Historical ,Summarized Multidimensional, Integrated	Current, up-to-date, detailed
Unit of Work	Complex query	Short, Simple transaction

The major distinguishing features between OLTP and OLAP are summarized as follows:

- **OLTP (on-line Transaction processing)**
  - Major task of traditional relational DBMS.
  - Day-to-day operations: purchasing, inventory, banking, payroll, registration, accounting, etc.
- **OLAP (on-line Analytical processing)**
  - Data analysis and decision making (major task of data warehouse system).

Distinct features	OLTP	OLAP
User and system orientation	customer	market
Data contents	current, detailed	historical
Database design	**ER**(entity-relationship)data module+ **application**	star + subject
View	current, local	evolutionary, integrated
Access patterns	update	read-only but complex querie

- ***OLTP Systems are used to "run" a business.***
- ***The Data Warehouse helps to "optimize" the business***