

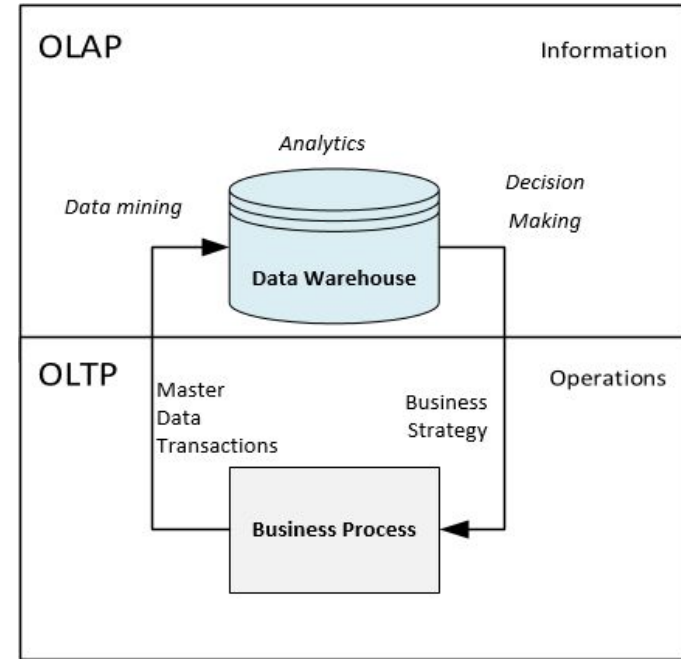
Understand what is Data warehouse



- The concept of data warehousing is pretty simple to create a central location and permanent storage space for the various data sources needed to support a company's analysis, reporting and other BI functions.
- BI :- Gartner says its "A broad category of applications and technologies for gathering, storing, analysing, sharing and providing access to data to help enterprise users make better business decisions."
- Data is extracted on a periodic basis from source systems, which are applications such as ERP systems that contain important company info.
- Data from these systems is moved to a dedicated server that contains a data warehouse.
- When it is moved it is cleaned, formatted, validated, reorganized, summarized, and supplemented with data from many other sources.
- This resulting data warehouse will become the main source of information for report generation and analysis via reporting tools that can be used for such things as ad-hoc queries, canned reports, and dashboards.
- The most important ingredient to a BI solution is that it must include a data warehouse.

OLTP and OLAP Systems

- Before we tackle basic data modeling, we need some background on the systems involved. We can roughly divide systems in two categories: operational and reporting systems.
- Operational systems are often called Online Transaction Processing (OLTP).
- Reporting and analytical systems are referred to as Online Analytical Processing (OLAP).
- OLTP systems support business processes. They work with “live” operational data, are highly normalized, and react very quickly to user actions.
- On the other hand, the primary purpose of the OLAP systems is analytics. These systems use summarized data, which is usually placed in a denormalized data warehousing structure like the star schema.



Let's define a Data Warehouse..

- William H. (Bill) Inmon (born 1945) an American computer scientist, recognized by many as the father of the data warehouse says:
 - A data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process.
- Subject-Oriented: A data warehouse can be used to analyze a particular subject area. For example, "sales" can be a particular subject.
- Integrated: A data warehouse integrates data from multiple data sources. For example, source A and source B may have different ways of identifying a product, but in a data warehouse, there will be only a single way of identifying a product.
- Time-Variant: Historical data is kept in a data warehouse. For example, one can retrieve data from 3 months, 6 months, 12 months, or even older data from a data warehouse. This contrasts with a transactions system, where often only the most recent data is kept. For example, a transaction system may hold the most recent address of a customer, where a data warehouse can hold all addresses associated with a customer.
- Non-volatile: Once data is in the data warehouse, it will not change. So, historical data in a data warehouse should never be altered.

Let's define a Data Warehouse..

- Ralph Kimball (born 1944) is an author on the subject of data warehousing and business intelligence. He is one of the original architects of data warehousing provided a more concise definition of a data warehouse said:
- A data warehouse is a copy of transaction data specifically structured for query and analysis.
- This is a functional view of a data warehouse. Kimball did not address how the data warehouse is built like Inmon did; rather he focused on the functionality of a data warehouse.

- The data needed to provide reports, dashboards, analytic applications and ad-hoc queries all exists within the production applications inside your company, so why not use the BI tools directly against this data?
- You need to integrate many different sources of data in near real-time. This will allow for better business decisions because users will have access to more data.
- You have tons of historical data that you need to gather in one easily accessible place . A Data Warehouse Provides Historical Intelligence.
- You need to keep historical records, even if the source transaction systems does not
- You can restructure the data and rename tables and fields (i.e. numbers without decimal points or tables named F017) so it makes more sense to the users.
- Users are running reports directly against operational systems, causing performance problems. Instead, create a data warehouse so users can run reports off of that.

- The data warehouse can be housed on a server built specifically for a data warehouse, resulting in much quicker access than hardware designed for handling transactions
- There is a risk that BI users might misuse or corrupt the transaction data
- Improve data quality by cleaning up data as it is imported into the data warehouse (providing more accurate data) as well as providing consistent codes and descriptions
- Reports using the data warehouse won't be affected by new releases of application software. Also, trying to maintain reports that summarize data originating within multiple releases of software is very difficult

- A Data Warehouse Saves Time
- Since business users can quickly access critical data from a number of sources—all in one place—they can rapidly make informed decisions on key initiatives. They won't waste precious time retrieving data from multiple sources.
- Not only that but the business execs can query the data themselves with little or no support from IT—saving more time and more money. That means the business users won't have to wait until IT gets around to generating the reports, and those hard working folks in IT can do what they do best—keep the business running.
- A Data Warehouse Enhances Data Quality and Consistency
- A data warehouse implementation includes the conversion of data from numerous source systems into a common format.
- Since each data from the various departments is standardized, each department will produce results that are in line with all the other departments. So you can have more confidence in the accuracy of your data. And accurate data is the basis for strong business decisions.

- A Data Warehouse Provides Historical Intelligence
- A data warehouse stores large amounts of historical data so you can analyze different time periods and trends in order to make future predictions. Such data typically cannot be stored in a transactional database or used to generate reports from a transactional system.
- Data mining Support:
- Once you have built out a data warehouse, there are data mining tools that you can use to help find hidden patterns using automatic methodologies. While reporting tools can tell you where you have been, data mining tools can tell you where you are going

- Ad-hoc reporting and analysis: Since the data warehouse eliminates the need for BI tools to compete with the transactional source systems, users can analyze data faster and generate reports more easily, and slice-and-dice in ways they could never do before.
- Dynamic presentation through dashboards: Managers want access to an interactive display of up-to-date critical management data. That is accomplished via dashboards, which are sophisticated displays that show information in creative and highly graphical forms, much like the instrument panel on an automobile

- **Amazon Redshift**
- **Microsoft Azure**
- **Google BigQuery**
- **Snowflake**
- **Teradata**

Above Mentioned are some popular Data Warehouse Tools , For more information on Data Warehouse Tools, please visit the following [link](#).