

# Data Lake vs Data Warehouse: Know the Difference

## What is Data Warehouse?

A data warehouse is a blend of technologies and components which allows the strategic use of data. It is a technique for collecting and managing data from varied sources to provide meaningful business insights.

It is electronic storage of a large amount of information by a business which is designed for query and analysis instead of transaction processing. It is a process of transforming data into information.

## What is Data Lake?

A Data Lake is a storage repository that can store large amount of structured, semi-structured, and unstructured data. It is a place to store every type of data in its native format with no fixed limits on account size or file. It offers high data quantity to increase analytic performance and native integration.

Data Lake is like a large container which is very similar to real lake and rivers. Just like in a lake you have multiple tributaries coming in, a data lake has structured data, unstructured data, machine to machine, logs flowing through in real-time.

## Data Warehouse Concept:

Data Warehouse stores data in files or folders which helps to organize and use the data to take strategic decisions. This storage system also gives a multi-dimensional view of atomic and summary data. The important functions which are needed to perform are:

1. Data Extraction
2. Data Cleaning
3. Data Transformation
4. Data Loading and Refreshing

## Data Lake Concept:

A Data Lake is a large size storage repository that holds a large amount of raw data in its original format until the time it is needed. Every data element in a Data lake is given a unique identifier and tagged with a set of extended metadata tags. It offers wide varieties of analytic capabilities.

## Key Difference between the Data Lake and Data Warehouse



[.\(/cdn.guru99.com/images/1/022218\\_0517\\_DataLakevsD1.png\)](https://cdn.guru99.com/images/1/022218_0517_DataLakevsD1.png)

Here are key differences between the two data associated terms in the mentioned aspects:

Parameters	Data Lake	Data Warehouse
<b>Storage</b>	In the data lake, all data is kept irrespective of the source and its structure. Data is kept in its raw form. It is only transformed when it is ready to be used.	A data warehouse will consist of data that is extracted from transactional systems or data which consists of quantitative metrics with their attributes. The data is cleaned and transformed
<b>History</b>	Big data technologies used in data lakes is relatively new.	Data warehouse concept, unlike big data, had been used for decades.

<b>Data Capturing</b>	Captures all kinds of data and structures, semi-structured and unstructured in their original form from source systems.	Captures structured information and organizes them in schemas as defined for data warehouse purposes
<b>Data Timeline</b>	Data lakes can retain all data. This includes not only the data that is in use but also data that it might use in the future. Also, data is kept for all time, to go back in time and do an analysis.	In the data warehouse development process, significant time is spent on analyzing various data sources.
<b>Users</b>	Data lake is ideal for the users who indulge in deep analysis. Such users include data scientists who need advanced analytical tools with capabilities such as predictive modeling and statistical analysis.	The data warehouse is ideal for operational users because of being well structured, easy to use and understand.
<b>Storage Costs</b>	Data storing in big data technologies are relatively inexpensive then storing data in a data warehouse.	Storing data in Data warehouse is costlier and time-consuming.
<b>Task</b>	Data lakes can contain all data and data types; it empowers users to access data prior the process of transformed, cleansed and structured.	Data warehouses can provide insights into pre-defined questions for pre-defined data types.
<b>Processing time</b>	Data lakes empower users to access data before it has been transformed, cleansed and structured. Thus, it allows users to get to their result more quickly compares to the traditional data warehouse.	Data warehouses offer insights into pre-defined questions for pre-defined data types. So, any changes to the data warehouse needed more time.
<b>Position of Schema</b>	Typically, the schema is defined after data is stored. This offers high agility and ease of data capture but requires work at the end of the process	Typically schema is defined before data is stored. Requires work at the start of the process, but offers performance, security, and integration.
<b>Data processing</b>	Data Lakes use of the ELT (Extract Load Transform) process.	Data warehouse uses a traditional ETL (Extract Transform Load) process.
<b>Complain</b>	Data is kept in its raw form. It is only transformed when it is ready to be used.	The chief complaint against data warehouses is the inability, or the problem faced when trying to make change in in them.

<b>Key Benefits</b>	They integrate different types of data to come up with entirely new questions as these users not likely to use data warehouses because they may need to go beyond its capabilities.	Most users in an organization are operational. These type of users only care about reports and key performance metrics.
---------------------	---	---

## Summary:

- A Data Warehouse is a blend of technologies and components which allows the strategic use of data.
- A Data Lake is a storage repository that can store a large amount of structured, semi-structured, and unstructured data.
- Data Warehouse stores data in files or folders which helps to organize and use the data to take strategic decisions.
- A Data Lake is a large size storage repository that holds a large amount of raw data in its original format until the time it is needed.
- Data warehouse concept, unlike big data, had been used for decades.
- Big data technologies integrate with the use of data lakes is relatively new.

◀ Prev (</data-lake-architecture.html>)

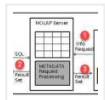
[Report a Bug](#)

Next ▶ (</business-intelligence-definition-example.html>)

## YOU MIGHT LIKE:

### DATA WAREHOUSING

(</multidimensional-online-analytical-processing.html>)



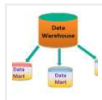
(</multidimensional-online-analytical-processing.html>)

**MOLAP (Multidimensional Online Analytical Processing)**

(</multidimensional-online-analytical-processing.html>)

### DATA WAREHOUSING

(</data-warehouse-vs-data-mart.html>)



(</data-warehouse-vs-data-mart.html>)

**Data Warehouse vs Data Mart: Know the Difference**

(</data-warehouse-vs-data-mart.html>)

### DATA WAREHOUSING



(</data-lake-architecture.html>)

(</data-lake-architecture.html>)

**What is Data Lake? It's Architecture**

(</data-lake-architecture.html>)

## DATA WAREHOUSING

---

(/star-snowflake-data-warehousing.html) (/star-snowflake-data-warehousing.html)

### Star and Snowflake Schema in Data Warehousing

(/star-snowflake-data-warehousing.html)

## DATA WAREHOUSING

---

(/fact-table-vs-dimension-table.html) (/fact-table-vs-dimension-table.html)

### Difference Between Fact Table and Dimension Table

(/fact-table-vs-dimension-table.html)

## DATA WAREHOUSING

---

(/online-analytical-processing.html) (/online-analytical-processing.html)

### What is OLAP (Online Analytical Processing): Cube, Operations & Types

(/online-analytical-processing.html)

# Data Warehousing Tutorial

- 1) What Is Data Warehousing? (/data-warehousing.html)
- 2) Data Warehouse Architecture (/data-warehouse-architecture.html)
- 3) ETL Process (/etl-extract-load-process.html)
- 4) ETL vs ELT (/etl-vs-elt.html)
- 5) What is Data Modelling? (/data-modelling-conceptual-logical.html)
- 6) What is OLAP? (/online-analytical-processing.html)
- 7) MOLAP (/multidimensional-online-analytical-processing.html)
- 8) OLTP vs OLAP (/oltp-vs-olap.html)
- 9) What is Dimensional Model? (/dimensional-model-data-warehouse.html)
- 10) Star and Snowflake Schema (/star-snowflake-data-warehousing.html)
- 11) Data Mart Tutorial (/data-mart-tutorial.html)
- 12) Data Warehouse vs Data Mart (/data-warehouse-vs-data-mart.html)
- 13) What is Data Lake? (/data-lake-architecture.html)

[14\) Data Lake vs Data Warehouse \(/data-lake-vs-data-warehouse.html\)](/data-lake-vs-data-warehouse.html)

[15\) What is Business Intelligence? \(/business-intelligence-definition-example.html\)](/business-intelligence-definition-example.html)

[16\) Data Mining Tutorial \(/data-mining-tutorial.html\)](/data-mining-tutorial.html)

[17\) Data Mining VS Data Warehouse \(/data-mining-vs-data-warehouse.html\)](/data-mining-vs-data-warehouse.html)

[18\) Fact Table VS Dimension Table \(/fact-table-vs-dimension-table.html\)](/fact-table-vs-dimension-table.html)

## **About**

[About US \(/about-us.html\)](/about-us.html)

[Advertise with Us \(/advertise-us.html\)](/advertise-us.html)

[Write For Us \(/become-an-instructor.html\)](/become-an-instructor.html)

[Contact US \(/contact-us.html\)](/contact-us.html)

## **Career Suggestion**

[SAP Career Suggestion Tool \(/best-sap-module.html\)](/best-sap-module.html)

[Software Testing as a Career \(/software-testing-career-complete-guide.html\)](/software-testing-career-complete-guide.html)

[Top Tools List \(/testing-development-tools.html\)](/testing-development-tools.html)

[Certificates \(/certificate-it-professional.html\)](/certificate-it-professional.html)

## **Interesting**

[Books to Read! \(/books.html\)](/books.html)

[Suggest a Tutorial](#)

[Blog \(/blog/\)](/blog/)

[Quiz \(/tests.html\)](/tests.html)

**Execute online**

[Execute Java Online \(/try-java-editor.html\)](/try-java-editor.html)

[Execute Javascript \(/execute-javascript-online.html\)](/execute-javascript-online.html)

[Execute HTML \(/execute-html-online.html\)](/execute-html-online.html)

[Execute Python \(/execute-python-online.html\)](/execute-python-online.html)

© Copyright - Guru99 2018

[Privacy Policy \(/privacy-policy.html\)](/privacy-policy.html)