**Data lake :**

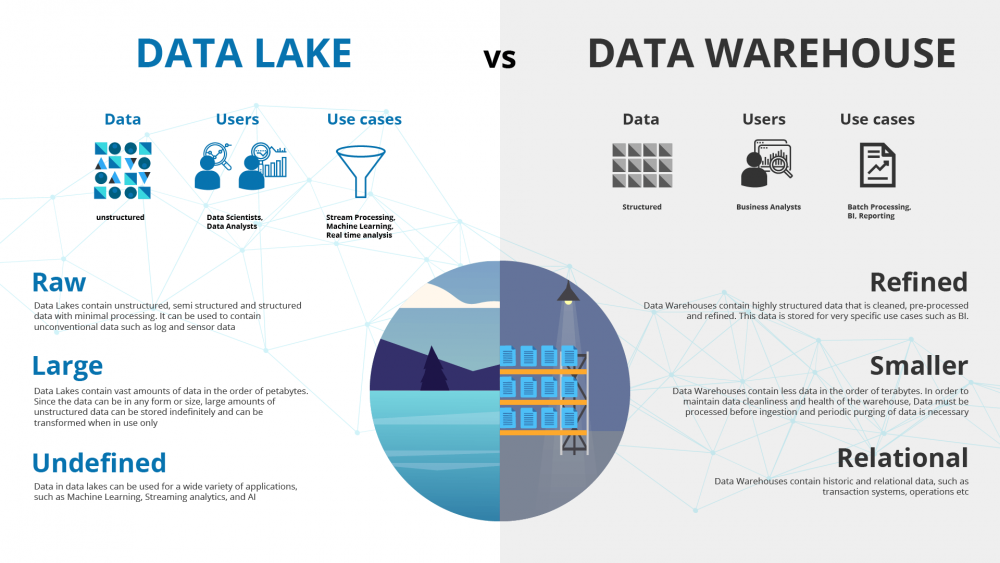
A **[data lake](https://www.qubole.com/blog/data-lake-essentials-part-1-storage-and-data-processing/)** is a central repository that allows you to store all your data – structured and unstructured – in volume

* The **[data lake](https://www.qubole.com/blog/data-lake-essentials-part-3-data-catalog-and-data-mining/)** is created with data streaming in from various sources, and then, multiple users can come to the lake to examine it and take samples.
* Data lakes support native streaming, where streams of data are processed and made available for analytics as it arrives.

**Data Warehouses :**

* It brings all your data together and stores it in a structured manner. It is typically used to connect and analyze data from heterogeneous sources.
* The data warehouse architecture relies on the data structure to support highly performant SQL (Structured Query Language) operations.
* Data lakes support native streaming, where streams of data are processed and made available for analytics as it arrives.
* Data warehouses require sequential ETL to ingest and transform the data before its usage for analytics, and hence they are inefficient for streaming analytics.
* Data warehouse stores the data in a proprietary format. Once the data is stored in the data warehouse, access to this data is limited to SQL and any custom drivers provided by the data warehouse

|  |  |
| --- | --- |
| Data warehouses and Business Intelligence tools support reporting and analytics on historical data | data lakes support newer use cases that leverage data for machine learning, predictions, and real-time analysis |
| Data warehouses require sequential ETL to ingest and transform the data before its usage for analytics | data lakes are ideal for machine learning use cases. |
| Data warehouses support sequential ETL operations, where data flows in a waterfall model from the raw data format to a fully transformed set, optimized for fast performance | Data lakes are exceptionally strong for use cases that require continuous data engineering. In data lakes, the waterfall approach of ETL is replaced by iterative and continuous data engineering. The raw data that lands in a data lake can be accessed and transformed iteratively via SQL and programmatic interfaces to meet the changing needs of the use case. |



Big Data

Big data is data that contains greater variety, arriving in increasing volumes and with more velocity.Big data is larger, more complex data sets, especially from new data sources.

Database

A database is an organized collection of structured information, or data, typically stored electronically in a computer system. A database is usually controlled by a [database management system (DBMS)](https://www.oracle.com/pk/database/what-is-database/" \l "WhatIsDBMS).