**Difference between Data Warehouses and Data Lakes**

According to [Google](http://www.google.com/trends/explore#q=big%20data%2C%20data%20lake&date=1%2F2010%2061m&cmpt=q&tz=), the interest in “Big Data” has been trending up for several years and has really gained steam in the last couple. The purpose of this article is to help understand the differences between data lakes and data warehouses to help you make an informed decision on how to manage your data.

**Data Warehouses**

A data warehouse is a storage repository for large volumes of data collected from multiple sources. It usually contains both historical and present data in a structured format. The data stored in a data warehouse is used by businesses to create annual and quarterly reports to measure business performance. It is highnly transformed and structured.

Popular companies that offer data warehouses include:

* [Snowflake](https://www.bmc.com/blogs/import-data-s3-snowflake/)
* Yellowbrick
* Teradata

**Data Lakes**

A data lake is a pool of raw data that flows like streams from data sources into the lake. Data lakes accept all data types, irrespective of whether or not it is structured or unstructured. First, the data is stored at the leaf level in an untransformed state, after which it is transformed, and schema is applied to fulfill the needs of analysis. All data is loaded form Source Systems. No data is turned away.

Popular data lake companies are:

* [Hadoop](https://www.bmc.com/blogs/hadoop-introduction/)
* Azure
* Amazon S3

**How are they different from each other?**

|  |  |  |
| --- | --- | --- |
|  | Data Warehouses | Data Lakes |
| Data Format | Structured | Raw and Unstructured |
| Cost | More | Less |
| Security | More | Less |
| Data Structure | Store organized and processed data | Store raw and unprocessed data |
| Accessibility | Easy to interpret and understand | Easily access and modify the data |
| Adaptability | They are not flexible or adaptable | Anyone, Anytime, any way you desire, without any restrictions. |

**Conclusion:**

Data lakes and data warehouses serve different purposes altogether. A data lake’s primary goal is to gather Big Data from disparate sources, whereas data warehouses are best for data analytics. While a data lake may work best for one organization, a data warehouse might be the best fit for another company, whereas some companies may require both.