

[View on Web](#)

How Data Lakes Can Empower Enterprises

22nd Apr, 2022

What spurred Google to go for a data lake solution when its Big Query data warehouse has had a successful run for 11 years? The need to break data silos and put in place seamless data management. Recently, Google unveiled a preview of Big Lake- a solution that lets enterprises unify data warehouses and data lakes without fussing about underlying storage systems. Google isn't the first to embrace Data lakehouses. Databricks pioneered the data lakehouse concept and has been upping the game to get more [data warehouse](#) -like performance for its Spark data lakes. Amazon Redshift, Microsoft Azure Synapse Analytics and others have been using data lakes too.



A data lake architecture has massive scalability in handling data. Plus, it's flexible enough to support new data analyses on polyglot data.

Why do you need Data Lake?

Data lakes can enable organizations to perform analytics like SQL queries, big data analytics, full-text search, real-time analytics, and Machine Learning (ML) to uncover insights. The

findings from Aberdeen Research show that the average company sees its data swell by 50 per cent every year. Besides the volume, the companies are managing data pulled from 33 unique sources. Unless they implement data lake technologies, they will find it challenging to navigate the volume and variety of data. *A study by Markets and Markets says the global data lake software and services market is expected to grow from \$7.9 billion in 2019 to \$20.1 billion in 2024. Enterprises that implemented a data lake outperformed similar companies by nine per cent in organic growth.*



The value that can be unlocked

Data lakes can harness more data from a multitude of sources. They empower users to collaborate and analyze data in different ways, leading to better, faster decision making. Here are the examples where data lakes can create and multiply value:-

Improved customer interactions:

In a Data Lake, customer data from a CRM can be combined with social media analytics, a marketing platform that includes buying history, and incident tickets so that the business can better understand the most profitable customer cohort, the cause of churn, or the promotions that will increase customer loyalty.

Improve R&D innovation choices:

Your R&D teams can use a data lake to test hypotheses, refine assumptions, and assess results. For example, choosing the right materials in your product design can result in faster performance, and genomic research can lead to more effective medication or understanding

customers' willingness to pay for different attributes.

Increase operational efficiencies

Through the Internet of Things (IoT), manufacturers have access to real-time data on processes like manufacturing. Machine-generated IoT data is easily stored and analyzed with a data lake, and ways to reduce operational costs and improve quality are discovered.

Challenges in managing data lakes

Data swamps:

The biggest challenge is keeping a data lake from becoming a data swamp. Unless properly designed and managed, the data lake can become a messy dumping ground for data.

Technology overload:

Deploying data lakes can also be complicated by the wide range of technologies available. To meet their particular data management and analytics requirements, organizations must choose the right technologies.

Unexpected costs:

Even if upfront technology costs are not high, this can change if organizations do not carefully manage data lake environments. A company may receive surprise bills for cloud-based data lakes if they use them more than expected.

Data governance:

One of the reasons for setting up a data lake is to store raw data for various analytical uses. But in the absence of data governance, organizations may be hit with data quality, consistency and reliability issues.

The future of Big Data

As big data gets bigger, it can overwhelm the best of data scientists. To reach a data-driven decision, organizations consult at least five data sources. **The worrisome fact is that 80 per cent of the data that bombards enterprises today is unstructured and hence, incapable of being handled by a data warehouse. The solution is in a [data lake](#).** To evolve with emerging technologies and deliver transformative business outcomes.



AUTHOR:

Jayajit Dash

Senior Manager- Corporate Communications (Marketing)