View on Web

Data Lake: Unlocking the Power of Data for Business Transformation

27th Sep,2021

Think of the scale and volume of data permeating our landscape today. Be it businesses or governments, a virtual tidal wave of data is sweeping processes and systems. Big data is getting bigger and the enormous volume of data can overwhelm even the best data scientists unless there is an agile data management strategy. Tackling this data deluge can boggle any business- whether a pharmaceutical company looking for the right vaccine candidate, a security firm keen to conduct analytics with speed and precision or even governments that have to juggle with data from numerous sources to make decisions. *Around 80 per cent of this data is unstructured, which can't be handled by a data warehouse. The solution is in the adoption of an agile, scalable and configurable 'data lake' that pools all data- structured or unstructured into a single repository.* To put it plainly, it's a lake where any data can be fished in 'as-is' without pre-configuration and can be fished out for analysis as and when needed.



Why Do You Need Data Lake?

Global data is estimated to reach 175 million petabytes (one petabyte is one million GB). And by 2025, 60 per cent of the existing data would be created and managed by enterprise organizations compared to 30 per cent in 2015. On average, enterprises need to consult at least five data sources to reach a data-driven decision. What's worrying, though, is that 99.5 per cent of collected data remains unused, primarily due to a lack of infrastructure, resources, and management (Source Grow.com). Most organizations are amassing data from all sources without a clear strategy on how to tap this data. Much of this data turns into 'dark data', consuming space and wasting dollars since it's never used.

Having a data lake allows organizations to import data of all formats while saving time on defining data structures**.**

A data lake can harness more data from more sources in less time, thus empowering users to collaborate and analyze data in different ways leads to better, faster decision making. While data lakes are typically used in conjunction with traditional enterprise data warehouses (EDWs), they cost less to operate than EDWs.

Is There a Challenge in The Data Lake Setup?

The main challenge with a data lake architecture is that raw data is fed into it with no oversight of the contents. Making data more usable needs predefined mechanisms to catalogue and secure data. Without these elements, data cannot be found, or trusted resulting in a ‘data swamp’. Meeting the needs of wider audiences require data lakes to have governance, consistency, and access controls.

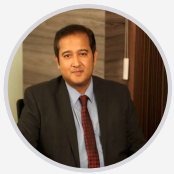
..But Benefits Outweigh Challenges

Companies are switching over to the data lake architecture for two primary reasons. First, they are keen to leverage its advanced and sophisticated analytical techniques. Second, for injecting more efficiency into traditional activities like data access and speed of retrieval.

The Future- Towards a Hybrid Environment Called ‘Lakehouse’

Over the past decade, enterprise data analytics attention has shifted away from the data warehouse architecture to the data lake architecture. Now, the question is, how can data lakes perform the functions of data warehouses to the optimum benefit of enterprises? The answer is in creating a hybrid environment titled ‘Lakehouse’. This provides a structured transactional layer to a data lake, allowing many of the use cases that traditionally required legacy data warehouses to be completed. Built-in integration with emerging technologies like Artificial Intelligence (AI) and Machine Learning (ML) can enable data lakes to process larger and more complex datasets.

This article was originally published on Priyadarshi Nanu Pany's **Medium** Account.



AUTHOR:

Priyadarshi Pany

Managing Director & CEO

© Copyright CSM Technologies, Level-6, E/56, Infocity, Chandrasekharpur, Bhubaneswar, Odisha - 751024