



Big Data Fundamentos 3.0

Big Data Fundamentos – Versão 3.0

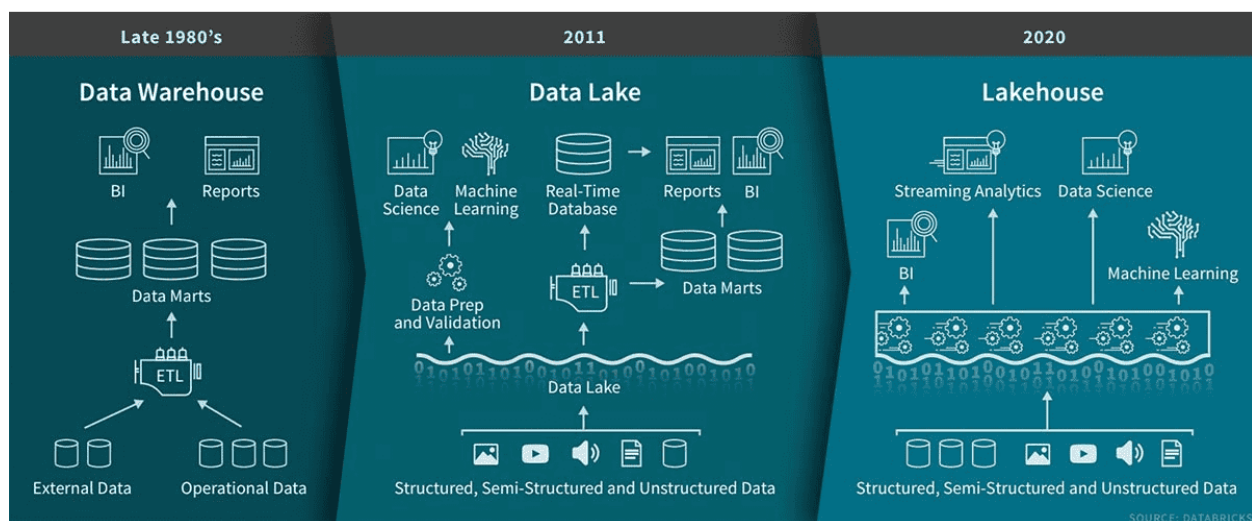
O Que é Data Lakehouse?



Um Data Lakehouse é uma nova arquitetura de gerenciamento de dados que combina a flexibilidade, economia e escala de Data Lakes com o gerenciamento de dados e transações ACID (Atomicidade, Consistência, Isolamento, Durabilidade) de Data Warehouses, permitindo Business Intelligence (BI) e Machine Learning (ML) em todos dados armazenados em um único repositório.

Data Lakehouses são habilitados por um novo design de sistema aberto: implementação de estruturas de dados e recursos de gerenciamento de dados semelhantes aos de um Data Warehouse, diretamente no tipo de armazenamento de baixo custo usado para Data Lakes. Mesclá-los em um único sistema significa que as equipes de dados podem se mover mais rapidamente, pois podem usar os dados sem a necessidade de acessar vários sistemas. Os Data Lakehouses também garantem que as equipes tenham os dados mais completos e atualizados disponíveis para projetos de ciência de dados, aprendizado de máquina e análise de negócios.

O diagrama abaixo ilustra a ideia por trás do Data Lakehouse.



O termo Data Lakehouse foi definido pela Databricks (plataforma que fornece infraestrutura para processamento de grandes volumes de dados). Abaixo o paper de pesquisa que originou o termo:

Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics

https://databricks.com/wp-content/uploads/2020/12/cidr_lakehouse.pdf