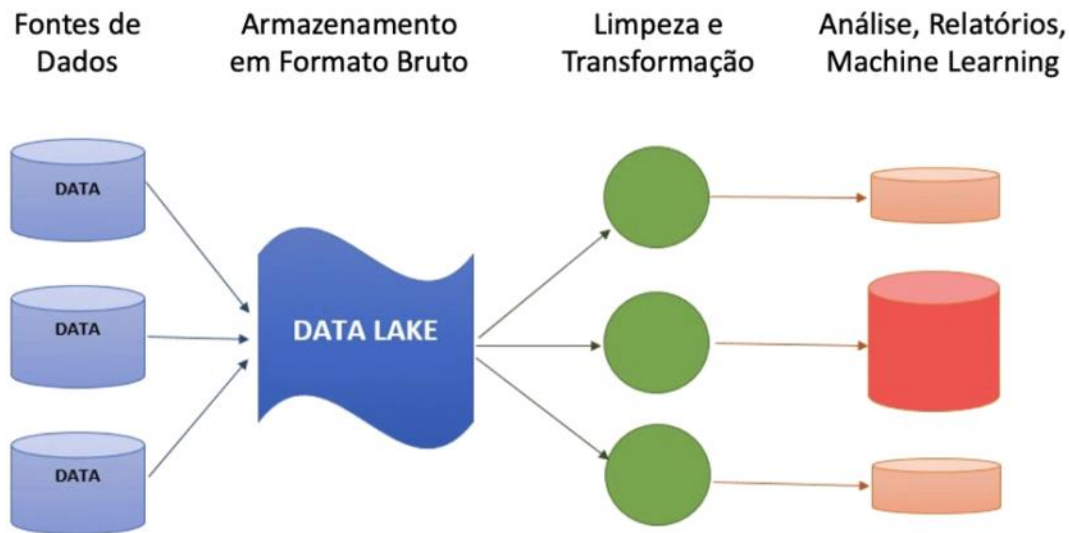


04.Data Lakes

terça-feira, 18 de julho de 2023

16:14

Os dados são armazenados brutos, podendo ser redes sociais, fotos, vídeos, e diversos outros tipos



No Data Lake diferente do DW você pega todos os dados da forma original sem a limpeza para não perder dados, o Data Lake serve para isso mesmo, podendo pegar a porção que precisa dos dados para utilizar.

É um repositório centralizado usado para dados estruturados e não estruturados para executar diferentes tipos de análises de dados, podendo até armazenar dados estruturados também no Data Lake, porém o DW é mais prático para usar em outra ferramenta, porém no Data Lake precisa fazer toda a estruturação para usar em outra ferramenta.

Apache Hadoop serve para armazenar os dados no Data Lake criando uma camada com o apache hive para usar a linguagem SQL nos Data Lakes.

O desafio da arquitetura de Data Lake é que os dados são armazenados sem supervisão do conteúdo, para ter dados utilizáveis precisa de uma proteção de dados. Sem esses elementos os dados não atendem as necessidades.

Tomar cuidado com governança dos dados, gestão de metadados, consistência semântica e controles de acesso.

Data Lake é um conceito e pode ser construído com diferentes tecnologias como apache hadoop ou bancos de dados NoSQL.

Podemos pegar dados do DW para o Data Lake e vice-versa.

Para o DW e Data Lake usamos ETL - Data Lakes e DWs podem fazer parte de

uma grande estrutura central de armazenamento chamada Data Hub.

Benefícios do Data Lake

- **Armazenamento em formato bruto**
- **Importação de qualquer quantidade de dados em tempo real**
- **Repositório Central para todos os dados da empresa**