

Data Lakes

Prof. Álisson R. Arantes

Data Lakes

- Grande demanda por dados de diversas naturezas
- Tomada de decisão: altamente depende de dados em quantidade e variedade
- Utilização de dados pelas organizações:
 - diversos conceitos colocados em prática
 - Data Warehouse, Data Mart, Data Lake, ...

Data Lakes

- Data Lake:
 - repositório de dados disponíveis para consumo na organização
 - fontes independem de formato, estrutura e origem
 - fontes internas (mantidas pela própria organização) ou externas

Data Lakes

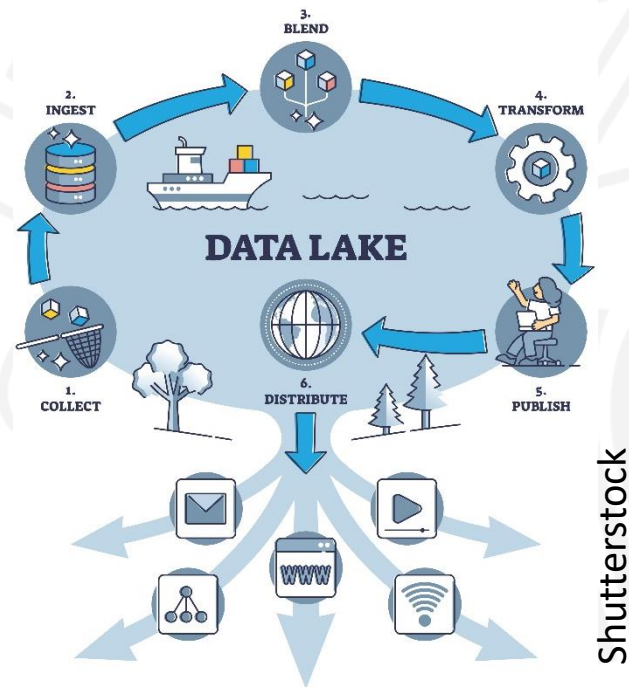
- Qualquer fonte de dados em potencial pode estar em um Data Lake de uma organização:
 - acessível aos usuários
 - agregue valor à organização
 - contribua para decisões estratégicas
- Importância: potencial que os dados representam

Data Lakes

- Fontes de dados em potencial:
 - bancos de dados transacionais
 - arquivos texto
 - planilhas eletrônicas
 - dados abertos de terceiros
 - sensores de dispositivos IoT
 - ...

Data Lakes

- Foco na coleta (ingestão), armazenamento, tratamento e consumo dos dados
- Tratamento: cientistas de dados, engenheiros de dados, especialistas do conhecimento, ...



Data Lakes

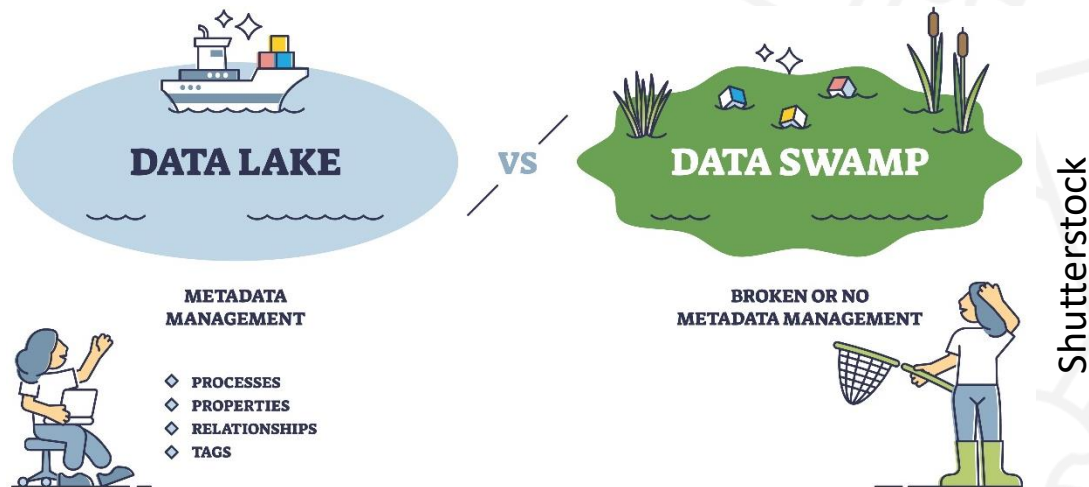
- Dados quando ingeridos podem não estar prontos para consumo

Data Lakes

- Desafio: atingir nível de maturidade suficiente na gestão e uso do Data Lake
- Possível deterioração ao longo do tempo, armazenando dados que:
 - não possam ser consumidos
 - não agreguem valor

Data Lakes

- Fundamental a gestão adequada do Data Lake:
 - evitar que o Data Lake se torne um Data Swamp





PUC Minas
Virtual