



Formação Engenheiro de Dados

Data Lakes

Data Lakes

- ◊ Armazém corporativo de dados
- ◊ “Simples”, disponível, acessível, seguro
- ◊ Em conformidade com Governança de Dados
- ◊ Versão “Big Data” do Data Warehouse clássico
- ◊ Normalmente em Hadoop/HDFS
- ◊ Dados estruturados/não estruturados/semi-estruturados

Estruturas



Landing Layer: dados brutos



Analytic Layer: dados tratados para consumo



Arquivamento: dados históricos

Camadas

- ◊ Aquisição: extração (ingestão)
- ◊ Processamento
- ◊ Consumo

Ingestão de Dados

- ◇ Capturar dados para o Data Lake
 - ◇ Batch
 - ◇ Real-time: CDC: change data capture