

Spark SQL

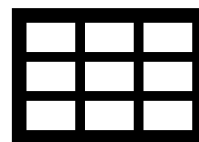
Spark

Utiliza o
Metastore do
Hive

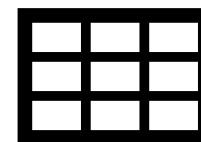
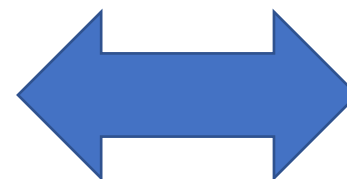
Não é preciso ter
o Hive instalado
para usar o Spark

Tabela

- Persistente
- Objeto Tabular que reside em um banco de dados
- Pode ser gerenciado e consultado utilizando SQL
- Totalmente interoperável com DataFrame
- Ex: Você pode transformar um DataFrame que importamos (Parquet, json, orc csv) em tabela

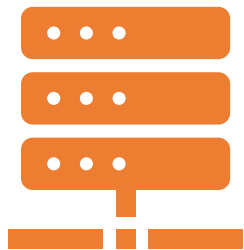


Tabela



DataFrame

Tabelas



Gerenciadas: Spark gerencia dados e metadados

Armazenadas no warehouse do spark

Se excluirmos, tudo é apagado (dados e metadados)



Não Gerenciadas (External): Spark apenas gerencia metadados

Informamos onde a tabela está (arquivo, por exemplo orc)

Se excluirmos, Spark só exclui os metadados, dados permanecem onde estavam

Views

- Mesmo conceito de banco de dados relacionais
- São um “alias” para uma tabela (por exemplo, vendas_rs pode mostrar vendas do estado já com filtro aplicado)
- Não contém dados

Views



GLOBALIS: VISÍVEIS EM
TODAS AS SESSÕES



SESSÃO: VISÍVEIS APENAS
NA PRÓPRIA SESSÃO