

Arquivo

Objeto no DBFS



Tabelas

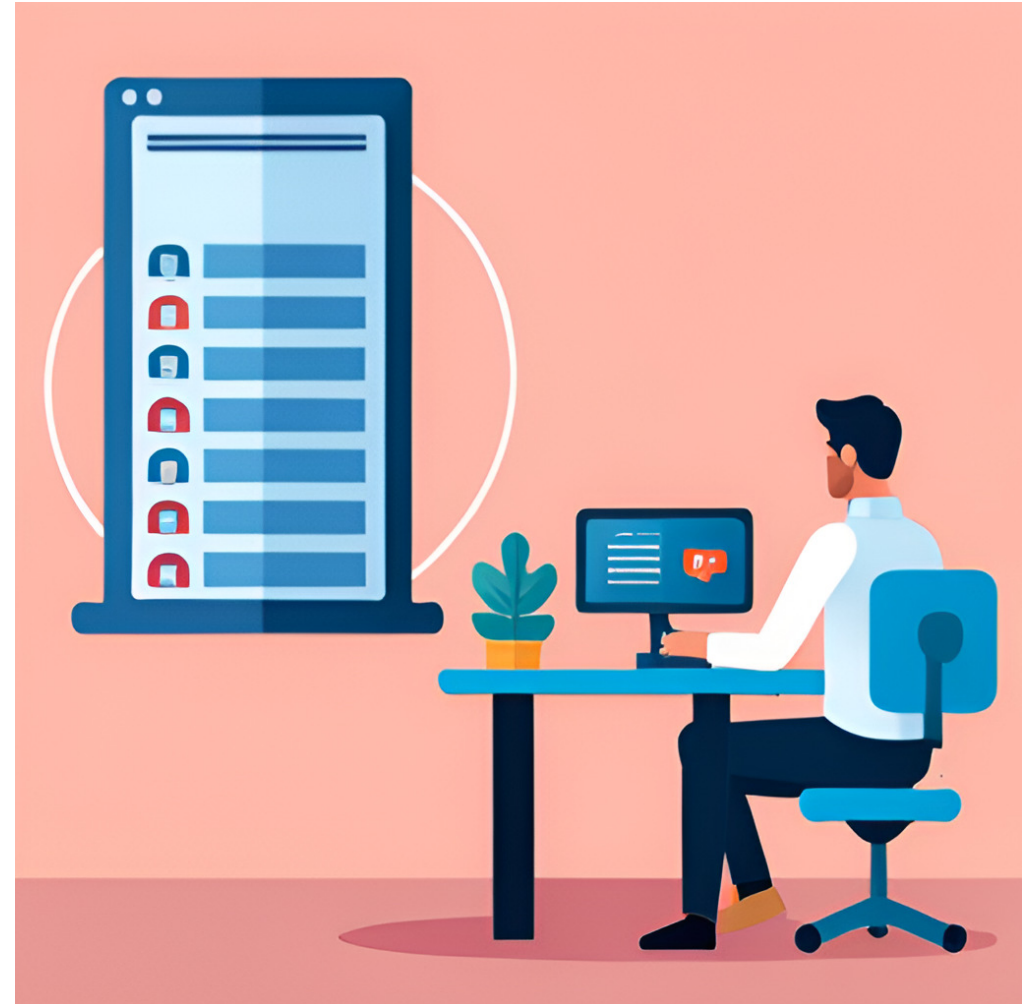
Tabela: dados em um formato específico com metadados no metastore.

Pode ser consultado usando SQL.

Criada a partir de um arquivo ou dataframe

Tabelas Externas: é criado um alias, sem cópia. Se o arquivo deixar de existir, a tabela não funciona mais.

Tabela Gerenciada: é criada uma cópia. Se você apagar o arquivo, a tabela continua existindo.



Data Frames

Formato ideal para processamento de dados

Dados efêmeros

São perdidos caso o cluster seja desligado

Para persistência devem ser salvos fisicamente (parquet)





Arquivo Delta (Delta Lake)

- Camada de armazenamento de código aberto que traz recursos ACID, controle de versão e controle de esquema
- Manipulados através de dataframes
- Podem ser consultados por SQL
- Não efemero