

Spark - Exercícios de DataFrame

1. Enviar o diretório local “/input/exercises-data/juros_selic” para o HDFS em “/user/aluno/<nome>/data”
2. Criar o DataFrame jurosDF para ler o arquivo no HDFS “/user/aluno/<nome>/data/juros_selic/juros_selic.json”
3. Visualizar o Schema do jurosDF
4. Mostrar os 5 primeiros registros do jurosDF
5. Contar a quantidade de registros do jurosDF
6. Criar o DataFrame jurosDF10 para filtrar apenas os registros com o campo “valor” maior que 10
7. Salvar o DataFrame jurosDF10 como tabela Hive “<nome>.tab_juros_selic”
8. Criar o DataFrame jurosHiveDF para ler a tabela “<nome>.tab_juros_selic”
9. Visualizar o Schema do jurosHiveDF
10. Mostrar os 5 primeiros registros do jurosHiveDF
11. Salvar o DataFrame jurosHiveDF no HDFS no diretório “/user/aluno/nome/data/save_juros” no formato parquet
12. Visualizar o save_juros no HDFS
13. Criar o DataFrame jurosHDFS para ler o diretório do “save_juros” da questão 8
14. Visualizar o Schema do jurosHDFS
15. Mostrar os 5 primeiros registros do jurosHDFS