

Spark - Exercícios de Esquema e Join

1. Criar o DataFrame **alunosDF** para ler o arquivo no hdfs “/user/aluno/<nome>/data/escola/alunos.csv” sem usar as “option”
2. Visualizar o esquema do alunosDF
3. Criar o DataFrame **alunosDF** para ler o arquivo “/user/aluno/<nome>/data/escola/alunos.csv” com a opção de Incluir o cabeçalho
4. Visualizar o esquema do alunosDF
5. Criar o DataFrame **alunosDF** para ler o arquivo “/user/aluno/<nome>/data/escola/alunos.csv” com a opção de Incluir o cabeçalho e inferir o esquema
6. Visualizar o esquema do alunosDF
7. Salvar o DataFrame alunosDF como tabela Hive “tab_alunos” no banco de dados <nome>
8. Criar o DataFrame **cursosDF** para ler o arquivo “/user/aluno/<nome>/data/escola/cursos.csv” com a opção de Incluir o cabeçalho e inferir o esquema
9. Criar o DataFrame **alunos_cursosDF** com o inner join do alunosDF e cursosDF quando o id_curso dos 2 forem o mesmo
10. Visualizar os dados, o esquema e a quantidade de registros do alunos_cursosDF