



Data Science Academy

www.datascienceacademy.com.br

Engenharia de Dados com Hadoop e Spark

Mini-Projeto 3
Design de um Job MapReduce para Gastos
Totais por Cliente

Para a solução desse projeto, são necessárias 3 etapas:

- 1- Instalação do Hadoop e Spark. Você pode utilizar o ambiente virtual criado ao longo do curso ou as máquinas virtuais Cloudera e Hortonworks.
- 2- Criar um código que faça a leitura do dataset e então para cada código de cliente, busque todos os gastos efetuados e faça a soma. Ao final, deve ser apresentado o resultado.
- 3- Executar o job via linha de comando usando o Spark-submit

Em anexo você encontra os scripts com a solução do projeto.