



Data Science Academy

www.datascienceacademy.com.br

Formação Cientista de Dados

Projeto com Feedback 5

Implementando Slow Changing Dimensions em um DW Usando Hive e Spark



Um dos mais amplos usos do Hadoop atualmente é construir uma plataforma de Data Warehousing sobre um Data Lake, através do Apache Hive. E na construção de um Data Warehouse, as tradições deixadas por Kimball e Inmon ainda são mais válidas do que nunca (Kimball e Inmon são os “papas” do DW).

A dimensão de alteração lenta (Slowly Change Dimension) de um DW, é uma dimensão que raramente muda. No entanto, quando muda, deve haver uma abordagem sistemática para capturar essa mudança. Exemplos de SCDs são informações de clientes e produtos.

Neste projeto, você deve implementar as SCDs no Hive e no Spark.

Os detalhes sobre nossas recomendações de ferramenta e datasets e como enviar seu projeto para análise, você encontra quando concluir a avaliação final.