



Data Science Academy

www.datascienceacademy.com.br

Data Lake – Design, Projeto e Integração

Lab 3 - Criando e Automatizando Jobs de Importação de Dados Para o Data Lake



O framework Sqoop também suporta o conceito de Jobs. Um job Sqoop pode ser definido e salvo para reutilização de comandos Sqoop. Um job Sqoop normalmente consiste em uma origem e um destino definidos com alguns outros parâmetros opcionais, sendo salvo em metadados Sqoop para reutilização e acesso remoto.

Um job é armazenado como uma entidade de metadados no metastore Sqoop. O objetivo principal da criação de uma tarefa é definir uma configuração única entre os sistemas de origem e de destino e reutilizá-la várias vezes para executar a exportação ou a importação. Como qualquer outra entidade, até mesmo o job suporta métodos de ciclo de vida, como criar, obter, atualizar e excluir com algumas outras funções. Para lista os jobs criados, usamos:

sqoop job --list

E para executar um job, usamos:

sqoop job --exec < nome_do_job >

Seu trabalho neste lab é criar um job Sqoop para importação de apenas alguns registros. Em anexo você encontra dois scripts para criação de uma tabela e carga de dados no banco de dados relacional. Os scripts foram criados para o PostgreSQL, seguindo o que foi estudado neste capítulo, mas fique à vontade para usar outro banco de dados se preferir. Bom trabalho!