



Data Science Academy

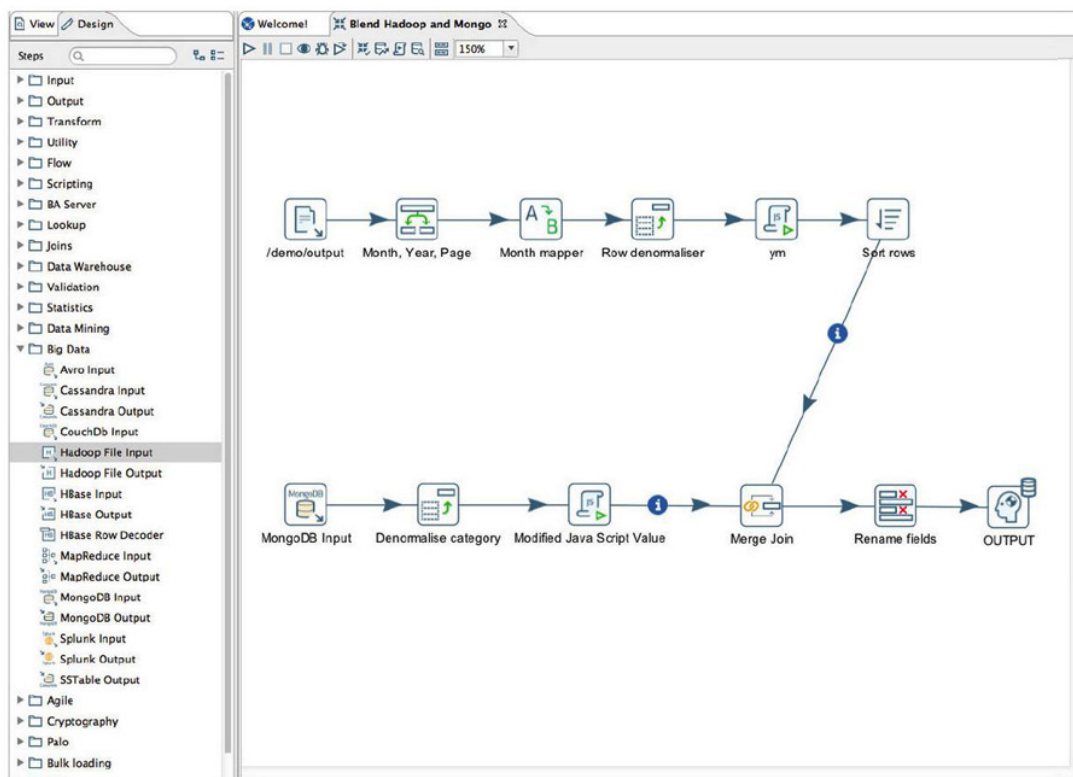
www.datascienceacademy.com.br

Data Lake – Design, Projeto e Integração

Desafio - Usando Pentaho Data Integration Para Carga de Dados no Hadoop

Você gosta e desafios? Nós adoramos! Então aqui vai um para você!

No primeiro curso da Formação Engenheiro de Dados, você teve algumas aulas sobre o Pentaho Data Integration como ferramenta ETL. Por que não experimentá-lo agora com seu Data Lake, movendo dados do banco relacional para o Hadoop?



Nós teremos um lab sobre isso mais a frente no curso, mas porque não começar a praticar desde já? Faça o download da versão trial do Pentaho, instale na sua máquina e tente configurar no Pentaho o mesmo que fizemos com o Sqoop. Fique atento ao uso das suas horas com as instâncias AWS (mantenha as instâncias paradas até que esteja pronto para testar o Pentaho). O download do PDI pode ser feito aqui:

<https://www.hitachivantara.com/en-us/products/big-data-integration-analytics/pentaho-data-integration.html>