

4 - Data Analysis - Hive



#Objetivo

Criar tabelas hive para consulta de dados por linguagem SQL.

#Link da apresentação

<https://drive.google.com/open?id=0B9LAB06v0XyRWXFpV3RrTGxmaFk>

#Referências

<https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-CreateTable>

<https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-ExternalTables>

<https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Loadingfilesintotables>

<https://cwiki.apache.org/confluence/display/Hive/CompressedStorage>

#Exercícios

1.1 - Criar uma MANAGED TABLE "salaries" para que contenha os dados de salários dos jogadores. Fazer carga dos dados a partir do arquivo salaries.csv no sistema de arquivos Linux.

1.2 - Criar uma MANAGED TABLE chamada "master" que contenha os dados dos jogadores. Fazer carga dos dados a partir de arquivos do hdfs.

1.3 - Criar uma tabela chamada "teams" que contenha os dados dos times. Esta tabela deve ser uma EXTERNAL TABLE.

1.4 - Criar uma tabela chamada "teamsfranchises" para armazenar os dados das franquias. Esta tabela deve ser uma EXTERNAL TABLE e utilizar o formato de compressão BZ2.

1.5 - Dropar as tabelas dos exercícios 1.1 e 1.3 e ver o que acontece com os arquivos de dados.

OBS:

1 - Salvar os scripts de criação de tabelas.

2 - Remover o cabeçalho do arquivo utilizando uma propriedade da hive table.