

Questão 1 - Trabalhando com o Apache Hive

1. Crie um banco de dados no Hive chamado “jogos”

Para criarmos uma base usamos o comando CREATE DATABASE jogos

2. Crie uma tabela chamada “video_game”.

hive> use jogos

hive> create table clientes (rank inteiro, name string, platform string, year inteiro, genre string, publisher string, na_sales float, eu_sales float, jp_sales float, other_sales float, global_sales float) partitioned by (period string) row format delimited fields terminated by ',';

3. Vamos trabalhar com informações sobre vendas de vídeo games fornecidas pelo Kaggle.

Para tanto, baixe o csv com os dados do site:

<https://gist.github.com/zhonglism/f146a9423e2c975de8d03c26451f841e>

4. Carregue os dados chamado vgsales na tabela video_game do Hive na partição “1980_2016”.

hive> load data local inpath 'arquivoshive/vgsales.csv' into table video_game partition (period='1980_2016');

5. Mostre as 10 primeiras linhas da base

hive> select top 10 * from video_game;

6. Mostre todas as linhas ocorridas após o ano de 2010.

select * from video_game where Year >= '2010';

7. Mostre todas as linhas como que possuem a plataforma Wii na partição “1980_2016”.

select * from video_game where Platform = 'Wii' and PARTITION = ('1980_2016');