

Atividade de Aprofundamento

Você deverá escolher uma das questões abaixo e produzir um tutorial com o passo a passo da sua solução. Caso você não consiga executar a máquina Cloudera, não se preocupe, pois você conseguirá responder as questões com base nos tutoriais apresentados na trilha.

Questão 1 - Trabalhando com o Apache Hive

Você deve criar um documento no formato PDF com os códigos usados em cada item da questão. Lembre-se que você deve usar o HQL.

1. Crie um banco de dados no Hive chamado "jogos".
2. Crie uma tabela chamada "video_game". Essa tabela deve ter os campos abaixo e deve ser particionada com um campo chamado "period" que deve ser um string.
 - 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

Lembre-se de definir o formato dos campos.

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".
5. Mostre as 10 primeiras linhas da base.
6. Mostre todas as linhas ocorridas após o ano de 2010.
7. Mostre todas as linhas como que possuem a plataforma Wii na partição "1980_2016".

Questão 2 - Trabalhando com o Apache Pig

Você deve criar um documento no formato PDF com os códigos usados em cada item da questão.

1. 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>
2. Carregue os dados chamado vgsales em uma relação do pig chamada games_vendas. Lembre-se de definir cada um dos campos usando o AS.
3. Apresente todas as linhas apenas das colunas Platform e Genre.
4. Apresente todas as linhas ocorridas no ano de 2010.
5. Apresente todas as linhas como o 'Platform' PS4.