1 - entrar no servidor do spark

- > docker.exe exec -it spark bash  
  
2 - criar uma pasta do HDFS dados\_processamento/dados  
  
-> hdfs dfs -mkdir -p dados\_processamento/dados  
  
3 - enviar os arquivos para esta pasta ( linux -> hdfs)

-> Copiar arquivos :

hdfs dfs -copyFromLocal /unixpath /hdfspath  
  
4 - processar os dados no Spark efetuando as limpezas, tratando os campos vazios  
  
5 - efetuar as responder as perguntas:  
  
- quantos pedidos foram realizados  
- quantos clientes tem em nossa base  
- quantos clientes temos por Região  
- quantidade de vendas em 2018  
  
OBS. podem fazer a importação dos arquivos direto do hdfs não sendo necessario criar uma tabela no hive, ok?  
  
Ex. dataframe\_alunos\_csv =  spark.read.option("header","true").option("inferSchema" , "true")\  
        .csv("/user/aluno/caiua/dados\_aluno/alunos.csv")  
  
**se atentem apenas com o separador no meu exemplo estou usando ',' e o do arquivo em anexo eh ';', vocês precisaram mudar isso na criação do dataframe.**  
  
**me enviem o script criado no final com as respostas**