

Avaliação

1. Cite e descreva 3 características (V's) de Big Data.
2. Explique o princípio da localidade de dados utilizado pelo Hadoop.
3. Quais são os principais componentes do Hadoop para armazenamento, escalonamento e processamento?
4. Apresente os componentes que formam o HDFS e explique a principal função de cada um deles.
5. Apresente os componentes que formam o YARN e explique a principal função de cada um deles.
6. Explique como o HDFS armazena arquivos e o objetivo pelo uso desta estratégia.
7. Apresente os comandos para criar a estrutura de diretórios e arquivos abaixo no HDFS. Considere que os arquivos a serem armazenados estão localizados no diretório local e o nome do usuário é "myuser".

```
/books/GeorgeOrwell_1984.txt  
/books/JRRTolkien_TheLordOfTheRings.txt  
/books/JaneAusten_PrideAndPrejudice.txt  
/user/myuser/datasets/dataset1.csv  
/user/myuser/datasets/dataset2.csv  
/user/myuser/datasets/dataset3.csv  
/user/myuser/provaGInfraBigData.txt
```

8. Qual a função do software Flume? Descreva como é a arquitetura do agente Flume.
9. Qual a função do software Sqoop? Apresente um cenário de utilização.
10. Descreva como os dados de entrada, saída e intermediários são representados no MapReduce.
11. Cite e descreva as fases do modelo MapReduce.
12. Qual a função do Combiner no MapReduce?
13. Qual a função do software Pig? Apresente um cenário de utilização.
14. Qual a função do software Hive? Apresente um cenário de utilização.
15. Qual a função do software Spark? Apresente a sua principal diferença em relação ao Hadoop MapReduce.