1 - Qual o objetivo do comando cache em Spark?

**Resposta: comando cache é para manter na memória para reutilização no tempo de execução**

2 - O mesmo código implementado em Spark é normalmente mais rápido que a implementação equivalente em MapReduce. Por quê?

**Respostas: devido a paralelizar a execução.**

3 - Qual é a função do SparkContext ?

**Resposta: ele é  o framework do spark, precisa dele para ter o RDD e toda a orquestração do Spark.**

4 - Explique com suas palavras o que é Resilient Distributed Datasets (RDD).

**Resposta: Responsável por paralelizar, quebrando o arquivo em blocos.**

5 - GroupByKey é menos eficiente que reduceByKey em grandes dataset. Por quê?

**Resposta: ReduceByKey agrupa as mesmas chave e contabiliza as chaves, o GroupByKey ele faz comparações linha a linha**

6 - Codigo: **PROGRAMA WORD COUNT**

 val textFile = sc.textFile("hdfs://...”)                        **# Le o arquivo.**

 val counts = textFile.flatMap(line => line.split(" "))  **# quebra a linha por palavras virando um Array bidimensional.**

           .map(word => (word, 1))           **# conta quantas palavras tem e soma 1.**

           .reduceByKey(\_ + \_)                 **# contabiliza o total.**

 counts.saveAsTextFile("hdfs://…”)    **# salva o resultado em um arquivo e coloca dentro o ambiente.**