Spark Interview Questions:

1. Describe the CAP Theorem
2. Describe some advantages of vertical vs. horizontal computer scaling
3. Describe some disadvantages of vertical vs. horizontal computer scaling
4. What is Hadoop – HDFS architecture and MapReduce programming paradigm
5. What is Spark – application in Hadoop ecosystem or programming interface to access Hadoop resource
6. What was the motivation for spark – IO operations are done using hard drive which is slow while spark does IO operations are done using RAM.
7. What is an executor process
8. What is a driver process
9. What is YARN – manages resources RAM, CPU, hard disks
10. What is a lineage graph – performance can be optimized with lineage graphs of RDDs
11. What is lazy execution and what benefit does it provide to spark

Map Reduce cannot be used if there is a dependence on order of the sequence.