Spark Questions:

1. Explain the Spark Architecture in detail
2. How the fault tolerance work in Spark
3. What is the difference between “Cores” & “Executors” in Spark....how to increase the number of cores in spark
4. What is the difference between “map” & “flatMap” in Spark
5. Where does “shuffling” comes into picture in Spark?
6. Difference between “reduceByKey” & “groupByKey” transformations? Which one is suggested w.r.to performance and why?
7. What are the different stages in Spark Execution...explain a use case where different stages will be skipped off?
8. How Spark is a “DAG”?
9. Difference between “Coalesce” & “Repartition”? which is a costly operation
10. What is the difference between “map” & “mapPartition” transformations
11. Why “Collect” action usage is NOT recommended in Spark? What are the alternatives for the same?
12. Explain the steps for integrating “HIVE” & “SPARK SQL” in detail
13. What is the difference between “DF” & “DS” with examples
14. What is the concept of UDFs in SPARK SQL? how to achieve the same
15. Explain the partitioning concept of Spark RDD?