SPARK Interview questions:

* Resource Allocation in Spark
* How to limit the number of output files in hive and spark
* Question regarding file formats used
* Broadcast joins in Spark
* What is coalesce and repartitioning in Spark
* How to implement incremental update in hive
* What windowing and analytical functions have you used in hive. Rank, Row Number, Dense Rank.
* What is partitioning and bucketing in Hive and types of partitions.
* What are the updates in Spark2.0
* If used, questions on EC2, S3 and EMR.
* Spark Architecture and explain?
* How do you set no of cores and executors on what basis you set them.
* How do you do Broad cast join and explain the process of broad cast join
* How do you handle updates and deletes in Spark and explain
* Groupby vs reduce by in spark?
* Client mode and cluster mode in spark architecture
* Hive partitioning
* Mapside join in hive
* Spark sql vs hive
* How do you add a schema to a schema less file in spark
* Spark context
* Broad cast in spark
* How do you differentiate the code with driver and executor?
* What is cache and broadcasting and need in spark and how do you do?
* Deduplication function by using rank function in hive
* Whats going on you project with AWS cloud migration
* How do you process JSON and XML files in Spark
* Memory level optimization done in Spark
* How spark memory optimizations will work explain them.
* Optimizations in Spark
* Coalesce and Repartition explain which is better and why
* How EMR is being used in you project
* What is the use case of Spark and AWS in your project explain.
* 1.Performance tuning in spark:
* 2. Name two technologies in Hadoop why did you use them.
* 3. Inner join and outer join diff
* 4. Issues you have face and how did you solve them.