Q. What is the Apache Spark ** A.**

Q. Why Apache Spark, what problem does it solve ** A.**

Q. What is unified ** A.**

Q. What is cOOMputing engine ** A.**

Q. What is libraries ** A.**

Q. What is parallel data processing? ** A.**

Q. Why we need Apache Spark ** A.**

Q. Hadoop vs Apache Spark ** A.**

Q. What is the Apache Spark ecosystem? ** A.**

Q. what is Apache Spark architecture? ** A.**

Q. What is transformation and how many types of transformation do we have ** A.**

Q. What happens when we use group By or join transformation ** A.**

Q. When jobs are created in Apache Spark ** A.**

Q. What is the DAG in Apache Spark ** A.**

Q. What is the lazy evolution? ** A.**

Q. What are the job, task, and stages in Apache Spark? ** A.**

Q. How many types of transformation ** A.**

Q. What is the narrow and wide transformation in Apache Spark? ** A.**

Q. What is the catalyst optimizer or Spark SQL engine? ** A.**

Q. Why do we get AnalysisException error ** A.**

Q. What is the catalog? ** A.**

Q. What is the physical planning or Spark plan ** A.**

Q. Is Spark SQL engine a cOOMpiler or not ** A.**

Q. How many phases are involved in Spark SQL engine to convert a code into java bite code ** A.**

Q. What is the RDD ** A.**

Q. When do you we need an RDD ** A.**

Q. What is the Features of an RDD ** A.**

Q. What is the dataFrame and data set ** A.**

Q. Why we should not use an RDD ** A.**

Q. What is the difference between SparkSession and SparkContext ** A.**

Q. What is a job, stage, and task in Apache Spark ** A.**

Q. How many jobs will be created in the given question ** A.**

Q. How many stages will be created in the given question ** A.**

Q. How many tasks will be created in the given question ** A.**

Q. What is the repartitioning in Apache Spark ** A.**

Q. What is the coalesce in Spark ** A.**

Q. What is the difference between repartitioning and coalesce in Apache Spark ** A.**

Q. Which one will you choice and why (repartition or coalesce) ** A.**

Q. repartitioning vs coalesce ** A.**

Q. What are the join strategies in Apache Spark ** A.**

Q. Why join is expensive or wide dependency transformation ** A.**

Q. Difference between shuffle-hash join and shuffle-short-marge join ** A.**

Q. When do we need broadcast join ** A.**

Q. What is the accumulator in Apache Spark ** A.**

Q. Why do we need broadcast-hash join ** A.**

Q. How does broadcast join works ** A.**

Q. Difference between broadcast hash join and shuffle-hash join ** A.**

Q. How can we change broadcast size of table ** A.**

Q. When broadcast table is not good or it will fail ** A.**

Q. What is the OOM in Apache Spark ** A.**

Q. Why do we get driver OOM ** A.**

Q. What is the driver overhead memory ** A.**

Q. Coomon reason to get a driver OOM ** A.**

Q. How to handle OOM ** A.**

Q. Why do we get OOM when data can be spill to the disc ** A.**

Q. How Spark manage storage inside executor intelligently ** A.**

Q. How task is spill in executor ** A.**

Q. Why do we need overhead memory ** A.**

Q. When do we get executor OOM ** A.**

Q. What is the Spark-submit ** A.**

Q. How do you run your job on Spark cluster ** A.**

Q. Where is your Spark cluster ** A.**

Q. What is the deploy mode in Apache Spark submit ** A.**

Q. What is Master in Spark submit ** A.**

Q. What all deployment mode on their in Apache Spark ** A.**

Q. What is the edge node ** A.**

Q. Why do we need client and cluster modes ** A.**

Q. What will happen if I close my adge node ** A.**

Q. What is the adaptive query execution (AQE) in Apache Spark ** A.**

Q. Why do we need adaptive query execution (AQE) ** A.**

Q. What is the caching ** A.**

Q. Why do we need caching or persistance ** A.**

Q. When should we avoid cashing ** A.**

Q. How to uncache the data ** A.**

Q. Difference between cache and persist ** A.**

Q. What is the difference storage level in Spark ** A.**

Q. Which storage level to choice ** A.**

Q. What is the dynamic resource allocation in Spark ** A.**

Q. How resource manager provides the resources if dynamic resource allocation ** A.**

Q. What are the resource allocation techniques we have in Apache Spark ** A.**

Q. What are the resource allocation technique we had in Spark ** A.**

Q. What are the challenges involved with dynamic resources allocation ** A.**

Q. What is the dynamic partition pruning ** A.**

Q. Why do we need dynamic partition pruning (DPP) ** A.**

Q. When dynamic partition pruning will not work ** A.**

Q. What is the data skewness problem ** A.**

Q. What are the ways to remove skewness ** A.**

Q. What is salting ** A.**

Q. How can we implement salting ** A.**

Q. Which cluster manager have you used in project ** A.**

Q. What is the size of the cluster? ** A.**