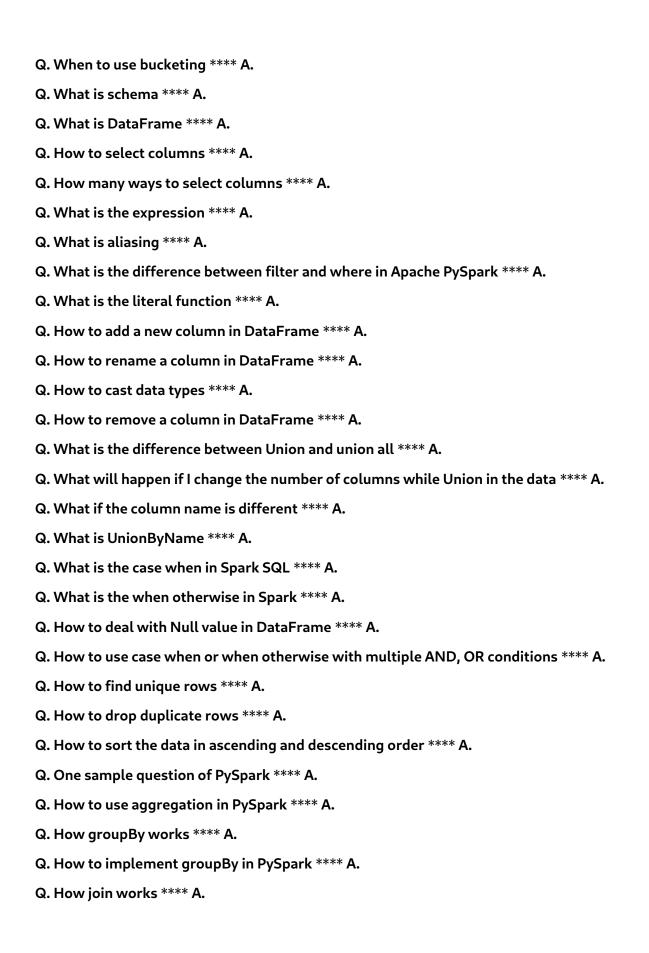
- Q. How to create Schema in PySpark \*\*\*\* A.
- Q. What are other ways to creating schema \*\*\*\* \*\*A.\*
- Q. What is the structType and structField in schema \*\*\*\* A.
- Q. What if I have a header in my DataFrame \*\*\*\* A.
- Q. Have you worked with corrupted records \*\*\*\* A.
- Q. When do you say that records are corrupted \*\*\*\* A.
- Q. What happens when we encounter corrupted records in different read modes \*\*\*\* A.
- Q. How can we print bad records \*\*\*\* A.
- Q. Where do you store corrupted records and how can we access them later \*\*\*\* A.
- Q. What is JSON data and how to read it in Apache PySpark \*\*\*\* A.
- Q. What if I have 3 keys in all lines and 1 key in one line in the JSON file \*\*\*\* A.
- Q. What is multi-line and line-delimited JSON \*\*\*\* A.
- Q. Which one works faster: multi-Line or Line-delimited in JSON in file format \*\*\*\* A.
- Q. How to convert nested JSON into PySpark DataFrame \*\*\*\* A.
- Q. What will happen if I have a corrupted JSON file \*\*\*\* A.
- Q. What is Parquet as a file format \*\*\*\* A.
- Q. Why do we need Parquet \*\*\*\* A.
- Q. How to read a Parquet file \*\*\*\* A.
- Q. What makes Parquet the default choice \*\*\*\* A.
- Q. What encoding is done on Parquet data \*\*\*\* A.
- Q. What comparison technique is used in the Parquet file format \*\*\*\* A.
- Q. How to optimize the Parquet file \*\*\*\* A.
- Q. What are the modes available in DataFrame write \*\*\*\* A.
- Q. What is Partition By and Bucket \*\*\*\* A.
- Q. How to write data into multiple partitions \*\*\*\* A.
- Q. What is a partition in Apache Spark \*\*\*\* A.
- Q. What is a bucket in Apache Spark \*\*\*\* A.
- Q. Why do we need these two: partitioning and bucketing \*\*\*\* A.
- Q. When to use partitioning \*\*\*\* A.



- Q. Why do we need join \*\*\*\* A.
- Q. What to do after joining two tables \*\*\*\* A.
- Q. What if tables have the same column name \*\*\*\* A.
- Q. How to join on two more columns \*\*\*\* A.
- Q. How many types of join \*\*\*\* A.
- Q. What is the window function \*\*\*\* A.
- Q. What is the row number rank dense rank in PySpark \*\*\*\* A.
- Q. How to calculate the top two salary holders from each department \*\*\*\* A.
- Q. What is LEAD and LAG in PySpark \*\*\*\* A.
- Q. What is nested JSON in PySpark \*\*\*\* A.
- Q. What is SCD2 \*\*\*\* A.