Interview question

1.Difference between union and union all in spark?

2.What is the difference between hive and Impala?

3.What is an API with an example?

4.What is dag?

5.What is swap memory??

6.what is rdd??why do we need it ??

//what is the difference between dataset and dataframe in spark

7.how do you inject the data ??

8.what is spark architecture

9. what are the window function in sql

10.how to run a spark job

11.what are the written data type in python

12. What is the difference between Client mode and Cluster mode in Spark?

13.how many data types in python and what are they?

14.which version spark are you using

15. Is the Namenode machine the same as a DataNode machine as in terms of hardware?

16.What is a block?

17.How to copy a file into HDFS with a different block size to that of existing block size configuration?

// Can you change the block size of HDFS files?

18.How To Show Hidden Files on Linux

19. How is a Class different from an Object?

20. How Many Mapreduce jobs and Tasks will be submitted for Sqoop copying into HDFS?

21. What are the main components of MapReduce Job?

22.How to set mappers and reducers for Hadoop jobs?

23.1) I have 20000 records in a table. I want to copy them to two separate files( records equally distributed) into HDFS (using Sqoop).

How do we achieve this, if the table does not have a primary key or unique key?

24.how many ways you can create a hive table ??What is the difference between external table and managed table?what's gonna happen if you delete both tables ??