

1.Explain Architecture of Spark

The spark is master slave architecture and it consists of driver,cluster manger ,worker nodes and executor.It consists of 3 APIs which are RDD,datframe and dataset.

2. Difference between Hadoop n spark

Hadoop is slower than spark.

Both are distributive.

Hadoop was written in java whereas spark in scala.

All tasks are done in terminal in terminal and for spark it is not.

All process are done in disk whereas all are done in memory.

Both are Fault tolerant.

3. Difference between RDD, Dataframe, Dataset

RDD:Oops style API,On heap JVM object,Serialization cant be avoided,Strongly type.

Dataset:SQL style API,Off heap JVM object,Serialization can be avoided,Less type safely.

Dataframe:SQL style API,Off heap JVM object,Serialization can be avoided,strongly type.

4. Explain the similarities in all API of Spark .

5. What is Transformation? Explain in detail

Transformation is like a function in which a dataframe change from one form to another and it is based on lazy evaluation of DAG..examples of transformation are filter(),groupby() etc.

6.What is Actions in spark? Explain in detail.

Actions is function which is executed in the dataframe lastly to get the required result.when the action statement is executed the transformations wil start to executed oneby one in the the reverse order in which the are used in the statement.

7. What is the Wide Transformation ?, explain with example.

Narrow transformation is a transformation which shuffles the datas across different nodes.example:groupby

8. What is Narrow Transformation? Explain with example .

Narrow transformation is a transformation which dont shuffle the datas across different nodes.example:filter.

9. Write down the query of wide n narrow transformation with example?

Narrow transformation-example:filter()

Wide tranformation-example:group(),join()