Nisha Kumari -2211566

Q1) In distributed computing paradigm, which of the following is/are compiled to Java bytecodes that can run on a JVM (Java Virtual Machine)?

1. Scala
2. Pyspark
3. Both Scala and Pyspark
4. Neither Scala nor Pyspark

Your Answer: C)

Q2) Which component is the biggest bottleneck when working with big data?

1. Reading and Writing data from SSD
2. CPU Operations
3. Transferring data across a network
4. Reading and Writing data from RAM

Your Answer: C)

Q3) Which DataFrame method accepts aggregate functions as arguments and can be used to apply different aggregate functions on different columns?

1. avg ( )
2. agg ( )
3. avn ( )
4. ast ( )

Your Answer: B)

Q4) Spark's Machine Learning Library known as?

1. MLlib
2. MLL
3. RDD
4. S3

Your Answer: A)

Q5) Which of the following represents the connection to a Spark cluster, and can be used to create RDDs on that cluster?

1. SparkConf
2. AppName
3. SparkContext
4. ApacheLog

Your Answer:C)

Q6) The scheduling layer of the Apache Spark architecture that implements stage-oriented scheduling is called?

1. MapReduce
2. Directed Acyclic Graph
3. Spark driver
4. Cluster Manager

Your Answer: B)

Q7) What type of functions accept existing RDDs as input and gives new RDDs as output?

1. Actions
2. DAGs
3. graphContexts
4. Transformations

Your Answer: D)

Q8) Which of the following is not an Action is Spark?

1. take ( )
2. first ( )
3. map ( )
4. reduce ( )

Your Answer: C)

Q9) Which of the following is a Transformation in Spark?

1. foreach ( )
2. collect ( )
3. count ( )
4. filter ( )

Your Answer: D)

Q10) Which of the following is True?

1. Only one SparkContext can be active per JVM Boosting
2. Spark’s cluster manager cannot be used with YARN
3. RDDs are mutable datasets
4. Directed Acyclic Graph (DAG) can have a maximum of two stages

Your Answer: A)