Q.1 In multiprocessing package of python _____class is used to implement parallel execution.

1. Pool class

2. Multi class

3. Thread class

4. All of the above

Q.2 A _____ execution is one the processes are completed in the same order in which it was started.

1. **Synchronous**

2. Asynchronous

3. Multi-threading

4. None

Q.3 The multiprocessing.Pool() class spawns a set of processes called _____

1. Salves

2. **Workers**

3. Clients

4. Stations

Q.4 Select the python libraries and frameworks for multiprocessing:

i) Ray
ii) Dask
iii) Sklearn

1. **i and ii**
2. ii and iii
3. i, ii and iii
4. only i

Q.5 _____ function of pool class is used to perform asynchronous parallel processing.

1. **Pool.apply_async()**

2. Pool.apply()

3. Pool.starmap()

4. None

Q.6 We can use Spark over Hadoop in-

1.Standalone

2.YARN

3.SIMR (Spark in MapReduce)

4. All of the above

Q.7 Which of the following is module for Structured data processing?

1. GraphX

2. MLlib

3. Spark SQL

4. Spark R

Q.8 Which of the following is fundamental data structure of Spark?

1.RDD

2. DataFrame

3. Dataset

4. None of the above

Q.9 DataFrame in Apache Spark prevails over RDD and doesnot contain any feature of RDD.

1. TRUE

2. FALSE

Q.10 Which of the following are common feature of RDD and DataFrame?

1. immutability

2. in-memory

3. resilient

4. All of the above

Q.11 Which of the following is not true for DataFrame?

1. DataFrame in Apache Spark is behind RDD

2. We can build DataFrame from different data sources. structured data file, tables in Hive

3. The Application Programming Interface (APIs) of DataFrame is available in various languages.

4. Both in Scala and Java, we represent DataFrame as Dataset of rows.

Q.12 We can create DataFrame using-

1.Tables in Hive

2. Structured data files

3. External databases

4. All of the above

Q.13 Which of the following is not the feature of spark?

1. Supports in memory computation

2. Fault tolerant

3. It is cost efficient

4. None of the above

Q.14 Which of the following is not true for Apache Spark Execution?

1.To simplify working with structured data it provides DataFrame abstraction in Python, Java, and Scala.

2. The data can be read and written in a variety of structured formats. For example, JSON, Hive Tables, and Parquet.

3. Using SQL we can query data,only from inside a Spark program and not from external tools.

4. The best way to use Spark SQL is inside a Spark application. This empowers us to load data and query it with SQL.

Q.15 If we launch the application through Spark submit, application jar doesnot gets distributed to all worker nodes.

1. TRUE

2. FALSE

Q.16 Apache Spark has APIs in:

1.Java

2. Scala

3. Python

4. All of the above

Q.17 Which of the following are uses of Apache Spark SQL?

1. It executes SQL queries.

2. We can read data from existing Hive installation using SparkSQL.

3. When we run SQL within another programming language we will get the result as Dataset/DataFrame.

4. All of the above

Q.18 Spark is developed in which language:

1. Scala

2. R

3. Java

4. None of the above

Q.19 Which of the following is true for Spark SQL?

1.Hive transactions are not supported by Spark SQL.

2. No support for time-stamp in Avro table.

3. Even if the inserted value exceeds the size limit, no error will occur.

4. All of the above

Q.20 Which of the followimg is true for Catalyst optimizer?

1.The optimizer helps us to run queries much faster than their counter RDD part.

2.The optimizer helps us to run queries little faster than their counter RDD part.

3.The optimizer helps us to run queries in the same speed as their counter RDD part.

4. None of the above

Q.21 The Dataset API is accessible in-

1. Java and Scala

2. Java, Scala and python

3. Scala and Python

4. Scala and R

Q.22 Does Dataset API support Python and R.

1. yes

2. no

Q.23 which of the following provide object-oriented programming interface?

1.RDD

2.DataFrame

3.Dataset

4.All of the above

Q.24 Which of the following the data is organized into named column?

1.RDD

2.DataFrame

3.Dataset

4.a and b

Q.25 Spark SQL can connect through JDBC or ODBC.

1.TRUE

2.FALSE

Q.26 Which type of processing Apache Spark can handle?

1. Batch Processing

2. Interactive Processing

3. Stream Processing

4. Graph Processing

5. All of the above

Q.27 Spark is developed in

1. Scala

2. Java

Q.28 Which of the following statements are correct?

1. Spark can run on the top of Hadoop

2. Spark can process data stored in HDFS

3. Spark can use Yarn as resource management layer

4. All of the above

Q.29 Spark's core is a batch engine

1. TRUE

2. FALSE

Q.30 Spark is 100x faster than Map Reduce due to

1. In-memory computing

2. Development in Scala