4, The export and import of data between sqoop and relational system happens through

which of the following programs?

B. Mapreduce job submitted by the sqoop command

4. Which of the following is NOT a supported data format for import/export in Sqoop?

D) JSON

10. Which of these format does Sqoop use for importing the

data from SQL to Hadoop?

***C. Text File Format***

11. Which technology is best suited for batch data processing?

***A. Hive***

1. What is Sqoop?

***A) A tool to transfer data between Hadoop and structured data stores***

2. Which of the following commands is used to list the available databases in a database server using Sqoop?

***C) sqoop list-databases***

7. Mention the best features of Apache Sqoop.

A Parallel import/export

B. Connectors for all major RDBMS Databases

C. Import results of SQL query

***D.All of these***

8. Which of the following commands is used to export data from Hadoop to a database using Sqoop?

***B) sqoop export***

3. What is the default delimiter used by Sqoop?

***A) Tab (\t)***

2. Which of the following commands is used to import data from a database into Hadoop using Sqoop?

***C) sqoop import***

7. What is a Hive partition?

***A. A subset of data in a Hive table (\*)***

7. Mention the best features of Apache Sqoop.

A Parallel import/export

B. Connectors for all major RDBMS Databases

C. Import results of SQL query

**D.All of these**

8. Hive is a type of software

**C. Data warehouse**

1. What is Hive?

B. A database management system ?? chọn a?? – > câu này đáp án a data processing tool (tài long ) -> dạ ck

10. What is Hive metastore?

**C. A component that stores metadata for Hive tables and partitions**

*(What is Hive Metastore? Metastore is the central repository of Apache Hive metadata. It stores metadata for Hive tables (like their schema and location) and partitions in a relational database. It provides client access to this information by using metastore service API.)*

11. Major components of the Hive architecture includes the following except

A. Metastore

B. Drivers

C. Compiler

**D. Interpreter**

12. Apache Hive is data warehouse software project built on top

A. Apache groove

***B. Apache Hadoop***

C. Apache net

D. Apache loof

13. Hive has how many execution engines

***A.2***

B.3 -> đây mới đúng (AN)

C.4

D.5

6. What is the default file format used by Hive to store data in HDFS?

A. CSV

B. Avro

C. Parquet

**D. ORC**

7. Which of the following is the reason for Spark being Speedy than MapReduce?

A. RDDs are immutable and fault-tolerant

B. Support for different language APIs like Scala, Java, Python and R

**C. DAG execution engine and in-memory computation**

D. None of the above

8. Which of the following is not true for DataFrame?

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

Hive IN SPARK

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

languages

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

***D. DataFrame in Apache Spark is behind RDD***

B)

Cc)

D)

What is action in Spark RDD?

A Creates one or many new RDDs

**B The ways to send result from executors to the driver**

C Takes RDD as input and produces one or more RDD as output.

D All of the above

is a component on top of Spark Core.

A RDDs

**B Spark SQL**

C Spark Streaming

D None of the above

What does Spark Engine do?

A Scheduling

B Monitoring data across a cluster

C Distributing data across a cluster

**D All of the above**

when processed in memory than MapReduce can?

10 times faster

**100 times faster**

200 times faster

300 times faster

. In how many ways Spark uses Hadoop?

a. 2 -> 2 mới đúng (AN)

**b. 3**

c. 4

d. 5

How can you use Spark to access and analyze dota stored

Cassandra databases? .

‘A By sing Spark Speci Keys

B. Byusing Sota

C. By using Sparse Vector

**D By using Spark Cassandra Connector**

How can you describe RDDS?

A Mutable

B Imeruabie

C. Positive

D.Negative

How many cluster managers are in Spark?

C 3

8)

°

Dd)

SparkSQL translates commands into codes. These codes are processed by.

Executor Nodes

Dd)

In Spark Streaming the data can be from what all sources?

Flume

Kafka

Kinesis

All of the above

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

RDD

Dataset

DataFrame

None of the above

We can create DataFrame using.

A Tables in Hive in spark

B External databases

C Structured data files

**D All of the above**

Which of the following is not a Spark cluster manager?

A. YARN

B. Standalone deployment

***C. Groovy***

D. Apache Mesos

B)

c)

D)

Spark is is developed in which language

Scala

11. Which of the following is true about narrow transformation\_\_\_

The data required to compute resides on the single partition.

—----------------------

1. In Spark Streaming the data can be from what all sources?

D ) all of bove

9 )

7 Which of the following is true about narrow transformation\_\_

b )the data required to compute resides on the single partition.

10 Spark is developed in which language

Scala

6 Which of the following Features of Apache Spark?

All of the above

5. Which of the following is a module for Structured data processing?

D) Spark SQL