



<b>Trạng thái</b>	Đã xong
<b>Bắt đầu vào lúc</b>	Thứ Tư, 27 tháng 11 2024, 1:20 PM
<b>Kết thúc lúc</b>	Thứ Tư, 27 tháng 11 2024, 1:50 PM
<b>Thời gian thực hiện</b>	29 phút 40 giây

## Câu hỏi 1

Hoàn thành

Đạt điểm 1,00

Which of the following provide the Spark Core's fast scheduling capability to perform streaming analytics.

- ☐ a. GraphX
- ☒ b. Spark Streaming
- ☐ c. Spark R
- ☐ d. RDD

## Câu hỏi 2

Hoàn thành

Đạt điểm 1,00

\_\_\_\_\_ is a shell utility which can be used to run Hive queries in either interactive or batch mode.

- ☐ a. \$HIVE/bin/hive
- ☒ b. \$HIVE\_HOME/bin/hive
- ☐ c. \$HIVE\_HOME/hive
- ☐ d. All of the mentioned

## Câu hỏi 3

Hoàn thành

Đạt điểm 1,00

Which of the following is not a component of Spark Ecosystem?

- ☐ a. MLlib
- ☐ b. GraphX
- ☒ c. Sqoop
- ☐ d. BlinkDB

## Câu hỏi 4

Hoàn thành

Đạt điểm 1,00

What are the parameters defined to specify window operation

- ☐ a. State size, sliding interval
- ☐ b. None of the above
- ☒ c. Window length, sliding interval
- ☐ d. State size, window length

## Câu hỏi 5

Hoàn thành

Đạt điểm 1,00

Apache Hive can be run in

- ☐ a. Interactive Mode
- ☐ b. Non Interactive Mode
- ☐ c. None
- ☒ d. Both Interactive and Non-Interactive Mode

## Câu hỏi 6

Hoàn thành

Đạt điểm 1,00

How many tasks does Spark run on each partition?

- ☐ a. Any number of task
- ☒ b. One
- ☐ c. More than one less than five

## Câu hỏi 7

Hoàn thành

Đạt điểm 1,00

Specific static Java functions can be used in Pig like UDFs. Take a look at the following Pig script: `define hex InvokeForString('java.lang.Integer.toHexString','int');nums = load 'numbers' as (n:int);inHex = foreach nums generate hex 🍌`; Apart from these three lines of code, what additional coding responsibilities do we have as developer here?

- ☒ a. There is nothing else to be done.
- ☐ b. We need to write the `toHexString()` functionality, extending `java.lang.Integer`.
- ☐ c. We need to write `InvokeForString()`.
- ☐ d. We need to register the jar containing `java.lang.Integer`.

## Câu hỏi 8

Hoàn thành

Đạt điểm 1,00

How many Spark Context can be active per JVM?

- ☐ a. More than one
- ☒ b. Only one
- ☐ c. Not specific
- ☐ d. None of the above

## Câu hỏi 9

Hoàn thành

Đạt điểm 1,00

Which guarantee that Hadoop provides does Pig break?

- ☒ a. All values associated with a single key are processed by the same Reducer.
- ☐ b. Calls to the Reducer's `reduce()` method only occur after the last Mapper has finished running.
- ☐ c. Task stragglers due to slow machines (not data skew) can be sped up through speculative execution.
- ☐ d. The Combiner (if defined) may run multiple times, on the Map-side as well as the Reduce-side.

## Câu hỏi 10

Hoàn thành

Đạt điểm 1,00

Dstream internally is

- ☐ a. Continuous Stream of DataFrame
- ☐ b. Continuous Stream of DataSet
- ☐ c. None of the above
- ☒ d. Continuous Stream of RDD

## Câu hỏi 11

Hoàn thành

Đạt điểm 1,00

Apache Spark has API's in

- ☒ a. All of the above
- ☐ b. Java
- ☐ c. Python
- ☐ d. Scala

## Câu hỏi 12

Hoàn thành

Đạt điểm 1,00

Which of the following is not a transformation?

- ☐ a. Filter
- ☒ b. Reduce
- ☐ c. Map
- ☐ d. Flatmap

## Câu hỏi 13

Hoàn thành

Đạt điểm 1,00

Which of the following will prefix the query string with parameters?

- ☐ a. SET hive.exec.compress.output=true
- ☒ b. SET hive.exec.compress.output=false
- ☐ c. All of the mentioned
- ☐ d. SET hive.compress.output=false

## Câu hỏi 14

Hoàn thành

Đạt điểm 1,00

Which of the following operator executes a shell command from the Hive shell?

- ☐ a. |
- ☐ b. +
- ☒ c. !
- ☐ d. ^

## Câu hỏi 15

Hoàn thành

Đạt điểm 1,00

In addition to stream processing jobs, what all functionality do Spark provides?

- ☐ a. Machine learning
- ☐ b. Graph processing
- ☒ c. All of the above
- ☐ d. Batch processing

## Câu hỏi 16

Hoàn thành

Đạt điểm 1,00

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

- ☐ a. Flume
- ☐ b. Kafka
- ☒ c. All of the above
- ☐ d. Kinesis

## Câu hỏi 17

Hoàn thành

Đạt điểm 1,00

By default, the location of the warehouse is

- ☐ a. file:///user/hive/warehouse
- ☐ b. file://user/hive/warehouse
- ☐ c. metastore.db
- ☒ d. metastore\_db

## Câu hỏi 18

Hoàn thành

Đạt điểm 1,00

Map in Apache Hive is a collection of

- ☐ a. Tables
- ☒ b. Key Value Pair
- ☐ c. homogeneous elements
- ☐ d. Geographic Locations

## Câu hỏi 19

Hoàn thành

Đạt điểm 1,00

What do you trigger by setting up a 'spark.cleaner.ttl' parameter?

- ☒ a. Automatic cleanup
- ☐ b. Automatic recycling
- ☐ c. Automatic recovery
- ☐ d. Automatic delete

## Câu hỏi 20

Hoàn thành

Đạt điểm 1,00

When you want Hive to completely manage the lifecycle of the table and data, then use \_\_\_\_\_

- ☐ a. Permanent Table
- ☒ b. Internal Table
- ☐ c. External Table
- ☐ d. View

## Câu hỏi 21

Hoàn thành

Đạt điểm 1,00

Which of the following is not a function of Spark Context in Apache Spark?

- ☐ a. To set the configuration
- ☒ b. Entry point to Spark SQL
- ☐ c. To Access various services
- ☐ d. To get the current status of Spark Application

## Câu hỏi 22

Hoàn thành

Đạt điểm 1,00

Major components of the Hive architecture includes the following except

- ☒ a. Interpreter
- ☐ b. Compiler
- ☐ c. Drivers
- ☐ d. Metastore

## Câu hỏi 23

Hoàn thành

Đạt điểm 1,00

The database stores it's files in a directory called

- ☐ a. Metastore
- ☐ b. Metastoredb
- ☐ c. Metastore.db
- ☒ d. Metastore\_db

## Câu hỏi 24

Hoàn thành

Đạt điểm 1,00

Which Cluster Manager do Spark Support?

- ☒ a. All of the above
- ☐ b. MESOS
- ☐ c. YARN
- ☐ d. Standalone Cluster Manager

## Câu hỏi 25

Hoàn thành

Đạt điểm 1,00

Views in Apache Hive are

- ☒ a. Virtual Tables
- ☐ b. Physical Files
- ☐ c. Slightly different with SQL Views
- ☐ d. not meant for DML operations

**Câu hỏi 26**

Hoàn thành

Đạt điểm 1,00

Hive converts queries to all except

- ☐ a. Apache tez
- ☐ b. Spark jobs
- ☒ c. Spark ten
- ☐ d. Map reduce

**Câu hỏi 27**

Hoàn thành

Đạt điểm 1,00

Which of the following algorithm is not present in MLlib?

- ☒ a. Tanimoto distance
- ☐ b. None of the above
- ☐ c. Streaming Linear Regression
- ☐ d. Streaming KMeans

**Câu hỏi 28**

Hoàn thành

Đạt điểm 1,00

Which is the abstraction of Apache Spark?

- ☐ a. Shared Variable
- ☐ b. Both the above
- ☒ c. RDD

**Câu hỏi 29**

Hoàn thành

Đạt điểm 1,00

Apache Spark was made open-source in which year?

- ☐ a. 2011
- ☐ b. 2009
- ☒ c. 2010
- ☐ d. 2008



## Câu hỏi 30

Hoàn thành

Đạt điểm 1,00

Which of the following is not true for Hadoop and Spark?

- ☐ a. Both are cluster computing environments
- ☒ b. Both have their own file system
- ☐ c. Both are data processing platforms
- ☐ d. Both use open source APIs to link between different tools

## Câu hỏi 31

Hoàn thành

Đạt điểm 1,00

Lets take the following file dataset.txt:Frank,19,44,1st\_year,12John,23,,2nd\_year,-1Tom,21,,,0and the following Pig Latin script:A = load 'dataset.txt' using PigStorage(';');B = filter A by \$1>20;C = group B by \$2;dump C;

- ☒ a. 1
- ☐ b. 2
- ☐ c. 0
- ☐ d. 3

## Câu hỏi 32

Hoàn thành

Đạt điểm 1,00

By default, Hive stores metadata in an embedded

- ☐ a. Apache hadoop
- ☐ b. Apache tez
- ☐ c. Apache hood
- ☒ d. Apache derby

## Câu hỏi 33

Hoàn thành

Đạt điểm 1,00

What is the representation of dependencies in-between RDDs called?

- ☐ a. Quadratic graph
- ☒ b. Lineage graph
- ☐ c. [oval\_new.svg?v=1] Graph
- ☐ d. Directed graph

**Câu hỏi 34**

Hoàn thành

Đạt điểm 1,00

Which of the following statements is correct?

- ☒ a. Pig is an execution engine that utilizes the MapReduce core in Hadoop.
- ☐ b. Pig is an execution engine that compiles Pig Latin scripts into HDFS.
- ☐ c. Pig is an execution engine that compiles Pig Latin scripts into database queries.
- ☐ d. Pig is an execution engine that replaces the MapReduce core in Hadoop.

**Câu hỏi 35**

Hoàn thành

Đạt điểm 1,00

Point out the correct statement.

- ☐ a. Streams are compressed using a codec, which is specified as a table property for all streams in that table
- ☒ b. The ODC file dump utility analyzes ORC files
- ☐ c. The Avro file dump utility analyzes ORC files
- ☐ d. All of the mentioned

**Câu hỏi 36**

Hoàn thành

Đạt điểm 1,00

Is Spark included in every major distribution of Hadoop?

- ☒ a. No
- ☐ b. Yes

**Câu hỏi 37**

Hoàn thành

Đạt điểm 1,00

In order to display column names in the result-set, which property you'll use?

- ☐ a. SET hive.columns.names.display=true;
- ☒ b. SET hive.cli.print.header=true;
- ☐ c. SET hive.cli.display.column.names=true;
- ☐ d. SET hive.cli.exec.print.header=true;

**Câu hỏi 38**

Hoàn thành

Đạt điểm 1,00

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

- ☐ a. None of the above
- ☐ b. Support for different language APIs like Scala, Java, Python and R
- ☒ c. DAG execution engine and in-memory computation
- ☐ d. RDDs are immutable and fault-tolerant

**Câu hỏi 39**

Hoàn thành

Đạt điểm 1,00

When \$HIVE\_HOME/bin/hive is run without either the -e or -f option, it enters \_\_\_\_\_ mode.

- ☒ a. Interactive shell
- ☐ b. Multiple
- ☐ c. Batch
- ☐ d. None of the mentioned

**Câu hỏi 40**

Hoàn thành

Đạt điểm 1,00

Let's consider the file above once more. You are tasked with writing a Pig Latin script that outputs the unique names (first column) occurring in this file. Which Pig Latin operators do you use (choose the minimum number)?

- ☐ a. filter, distinct
- ☒ b. foreach, distinct
- ☐ c. filter
- ☐ d. foreach