

APACHE-HADOOP-DEVELOPER^{Q&As}

Hadoop 2.0 Certification exam for Pig and Hive Developer

Pass Hortonworks APACHE-HADOOP-DEVELOPER Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

<https://www.leads4pass.com/apache-hadoop-developer.html>

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by
Hortonworks Official Exam Center

-  **Instant Download** After Purchase
-  **100% Money Back** Guarantee
-  **365 Days** Free Update
-  **800,000+** Satisfied Customers



QUESTION 1

Which Two of the following statements are true about hdfs? Choose 2 answers

- A. An HDFS file that is larger than dfs.block.size is split into blocks
- B. Blocks are replicated to multiple datanodes
- C. HDFS works best when storing a large number of relatively small files
- D. Block sizes for all files must be the same size

Correct Answer: AB

QUESTION 2

Which one of the following statements describes a Hive user-defined aggregate function?

- A. Operates on multiple input rows and creates a single row as output
- B. Operates on a single input row and produces a single row as output
- C. Operates on a single input row and produces a table as output
- D. Operates on multiple input rows and produces a table as output

Correct Answer: A

QUESTION 3

Which TWO of the following statements are true regarding Hive? Choose 2 answers A. Useful for data analysts familiar with SQL who need to do ad-hoc queries

- B. Offers real-time queries and row level updates
- C. Allows you to define a structure for your unstructured Big Data
- D. Is a relational database

Correct Answer: AC

QUESTION 4

You write MapReduce job to process 100 files in HDFS. Your MapReduce algorithm uses TextInputFormat: the mapper applies a regular expression over input values and emits key- values pairs with the key consisting of the matching text, and the value containing the filename and byte offset. Determine the difference between setting the number of reduces to one and settings the number of reducers to zero.

- A. There is no difference in output between the two settings.

B. With zero reducers, no reducer runs and the job throws an exception. With one reducer, instances of matching patterns are stored in a single file on HDFS.

C. With zero reducers, all instances of matching patterns are gathered together in one file on HDFS. With one reducer, instances of matching patterns are stored in multiple files on HDFS.

D. With zero reducers, instances of matching patterns are stored in multiple files on HDFS. With one reducer, all instances of matching patterns are gathered together in one file on HDFS.

Correct Answer: D

Explanation: * It is legal to set the number of reduce-tasks to zero if no reduction is desired.

In this case the outputs of the map-tasks go directly to the FileSystem, into the output path set by `setOutputPath(Path)`. The framework does not sort the map-outputs before writing them out to the FileSystem.

* Often, you may want to process input data using a map function only. To do this, simply set `mapreduce.job.reduces` to zero. The MapReduce framework will not create any reducer tasks. Rather, the outputs of the mapper tasks will be the final output of the job.

Note:

Reduce

In this phase the `reduce(WritableComparable, Iterator, OutputCollector, Reporter)` method is called for each pair in the grouped inputs.

The output of the reduce task is typically written to the FileSystem via `OutputCollector.collect(WritableComparable, Writable)`.

Applications can use the Reporter to report progress, set application-level status messages and update Counters, or just indicate that they are alive.

The output of the Reducer is not sorted.

QUESTION 5

Which of the following tool was designed to import data from a relational database into HDFS?

A. HCatalog B. Sqoop

C. Flume

D. Ambari

Correct Answer: B

QUESTION 6

You want to understand more about how users browse your public website, such as which pages they visit prior to

placing an order. You have a farm of 200 web servers hosting your website. How will you gather this data for your analysis?

- A. Ingest the server web logs into HDFS using Flume.
- B. Write a MapReduce job, with the web servers for mappers, and the Hadoop cluster nodes for reduces.
- C. Import all users\' clicks from your OLTP databases into Hadoop, using Sqoop.
- D. Channel these clickstreams into Hadoop using Hadoop Streaming.
- E. Sample the weblogs from the web servers, copying them into Hadoop using curl.

Correct Answer: A

QUESTION 7

Which one of the following statements regarding the components of YARN is FALSE?

- A. A Container executes a specific task as assigned by the ApplicationMaster
- B. The ResourceManager is responsible for scheduling and allocating resources
- C. A client application submits a YARN job to the ResourceManager
- D. The ResourceManager monitors and restarts any failed Containers

Correct Answer: D

QUESTION 8

You need to perform statistical analysis in your MapReduce job and would like to call methods in the Apache Commons Math library, which is distributed as a 1.3 megabyte Java archive (JAR) file. Which is the best way to make this library available to your MapReducer job at runtime?

- A. Have your system administrator copy the JAR to all nodes in the cluster and set its location in the HADOOP_CLASSPATH environment variable before you submit your job.
- B. Have your system administrator place the JAR file on a Web server accessible to all cluster nodes and then set the HTTP_JAR_URL environment variable to its location.
- C. When submitting the job on the command line, specify the `-jars` option followed by the JAR file path.
- D. Package your code and the Apache Commons Math library into a zip file named JobJar.zip

Correct Answer: C

Explanation: The usage of the `jar` command is like this,

Usage: `hadoop jar [mainClass] args...`

If you want the commons-math3.jar to be available for all the tasks you can do any one of these

1. Copy the jar file in \$HADOOP_HOME/lib dir

2.

or

Use the generic option -libjars.

QUESTION 9

In a MapReduce job with 500 map tasks, how many map task attempts will there be?

- A. It depends on the number of reduces in the job.
- B. Between 500 and 1000.
- C. At most 500.
- D. At least 500.
- E. Exactly 500.

Correct Answer: D

From Cloudera Training Course: Task attempt is a particular instance of an attempt to execute a task ?There will be at least as many task attempts as there are tasks ?If a task attempt fails, another will be started by the JobTracker ?Speculative execution can also result in more task attempts than completed tasks

QUESTION 10

Examine the following Pig commands:

```
customers = LOAD 'data' USING PigStorage(',')
AS (name:chararray,age:int,zip:int,receipts:double);
s = SAMPLE customers 0.2;
DUMP s;
```

Which one of the following statements is true?

- A. The SAMPLE command generates an "unexpected symbol" error
- B. Each MapReduce task will terminate after executing for 0.2 minutes
- C. The reducers will only output the first 20% of the data passed from the mappers
- D. A random sample of approximately 20% of the data will be output

Correct Answer: D

QUESTION 11

What data does a Reducer reduce method process?

- A. All the data in a single input file.
- B. All data produced by a single mapper.
- C. All data for a given key, regardless of which mapper(s) produced it.
- D. All data for a given value, regardless of which mapper(s) produced it.

Correct Answer: C

Explanation: Reducing lets you aggregate values together. A reducer function receives an iterator of input values from an input list. It then combines these values together, returning a single output value.

All values with the same key are presented to a single reduce task.

Reference: Yahoo! Hadoop Tutorial, Module 4: MapReduce

QUESTION 12

Which one of the following Hive commands uses an HCatalog table named x?

- A. `SELECT * FROM x;`
- B. `SELECT x.-FROM org.apache.hcatalog.hive.HCatLoader(\\'x\\');`
- C. `SELECT * FROM org.apache.hcatalog.hive.HCatLoader(\\'x\\');`
- D. Hive commands cannot reference an HCatalog table

Correct Answer: C

QUESTION 13

Which one of the following statements is true regarding a MapReduce job?

- A. The job's Partitioner shuffles and sorts all (key.value) pairs and sends the output to all reducers
- B. The default Hash Partitioner sends key value pairs with the same key to the same Reducer
- C. The reduce method is invoked once for each unique value
- D. The Mapper must sort its output of (key.value) pairs in descending order based on value

Correct Answer: A

QUESTION 14

You want to perform analysis on a large collection of images. You want to store this data in HDFS and process it with MapReduce but you also want to give your data analysts and data scientists the ability to process the data directly from HDFS with an interpreted high- level programming language like Python. Which format should you use to store this data in HDFS?

- A. SequenceFiles
- B. Avro
- C. JSON
- D. HTML
- E. XML
- F. CSV

Correct Answer: B

Reference: Hadoop binary files processing introduced by image duplicates finder

QUESTION 15

What are the TWO main components of the YARN ResourceManager process? Choose 2 answers

- A. Job Tracker
- B. Task Tracker
- C. Scheduler
- D. Applications Manager

Correct Answer: CD

[Latest APACHE-HADOOP-DEVELOPER Dumps](#)

[APACHE-HADOOP-DEVELOPER Practice Test](#)

[APACHE-HADOOP-DEVELOPER Study Guide](#)