![Pass2Lead logo](https://Pass2Lead.com)
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.pass2lead.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

On a cluster running MapReduce v1 (MRv1), a TaskTracker heartbeats into the JobTracker on your cluster, and alerts the JobTracker it has an open map task slot.

What determines how the JobTracker assigns each map task to a TaskTracker?

A. The amount of RAM installed on the TaskTracker node.

B. The amount of free disk space on the TaskTracker node.

C. The number and speed of CPU cores on the TaskTracker node.

D. The average system load on the TaskTracker node over the past fifteen (15) minutes.

E. The location of the InsputSplit to be processed in relation to the location of the node.

Correct Answer: E

Explanation: The TaskTrackers send out heartbeat messages to the JobTracker, usually every few minutes, to reassure the JobTracker that it is still alive. These message also inform the JobTracker of the number of available slots, so the JobTracker can stay up to date with where in the cluster work can be delegated. When the JobTracker tries to find somewhere to schedule a task within the MapReduce operations, it first looks for an empty slot on the same server that hosts the DataNode containing the data, and if not, it looks for an empty slot on a machine in the same rack.

Reference: 24 Interview Questions and Answers for Hadoop MapReduce developers, How JobTracker schedules a task?

QUESTION 2

You have the following key-value pairs as output from your Map task:

(the, 1)

(fox, 1)

(faster, 1)

(than, 1)

(the, 1) (dog, 1)

How many keys will be passed to the Reducer\\'s reduce method?

A. Six

B. Five

C. Four

D. Two

E. One

![Pass2Lead](https://Pass2Lead.com)
F. Three

Correct Answer: B

Explanation: Only one key value pair will be passed from the two (the, 1) key value pairs.

QUESTION 3

Given the following Pig command:

logevents = LOAD andapos;input/my.logandapos; AS (date:chararray, levehstring, code:int, message:string);

Which one of the following statements is true?

A. The logevents relation represents the data from the my.log file, using a comma as the parsing delimiter

B. The logevents relation represents the data from the my.log file, using a tab as the parsing delimiter

C. The first field of logevents must be a properly-formatted date string or table return an error

D. The statement is not a valid Pig command

Correct Answer: B

QUESTION 4

Which one of the following statements is FALSE regarding the communication between DataNodes and a federation of NameNodes in Hadoop 2.0?

A. Each DataNode receives commands from one designated master NameNode.

B. DataNodes send periodic heartbeats to all the NameNodes.

C. Each DataNode registers with all the NameNodes.

D. DataNodes send periodic block reports to all the NameNodes.

Correct Answer: A

QUESTION 5

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.

[APACHE-HADOOP-DEVELOPER PDF Dumps](#) | [APACHE-HADOOP-DEVELOPER VCE Dumps](#) | [APACHE-HADOOP-DEVELOPER Braindumps](#)

3 / 9

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 6

MapReduce v2 (MRv2/YARN) splits which major functions of the JobTracker into separate daemons? Select two.

A. Heath states checks (heartbeats)

B. Resource management

C. Job scheduling/monitoring

D. Job coordination between the ResourceManager and NodeManager

E. Launching tasks

F. Managing file system metadata

G. MapReduce metric reporting H. Managing tasks

Correct Answer: BC

Explanation: The fundamental idea of MRv2 is to split up the two major functionalities of the JobTracker,

resource management and job scheduling/monitoring, into separate daemons. The idea is to have a global

ResourceManager (RM) and per-application ApplicationMaster (AM). An application is either a single job in

the classical sense of Map- Reduce jobs or a DAG of jobs.

Note:

The central goal of YARN is to clearly separate two things that are unfortunately smushed together in

current Hadoop, specifically in (mainly) JobTracker:

/ Monitoring the status of the cluster with respect to which nodes have which resources available. Under

YARN, this will be global.

/ Managing the parallelization execution of any specific job. Under YARN, this will be done separately for

each job.

Reference: Apache Hadoop YARN ?Conceptsand; Applications

QUESTION 7

Examine the following Hive statements:

```
CREATE TABLE x (name STRING, age INT, zip INT, salary DOUBLE)
ROW FORMAT DELIMITED FIELDS TERMINATED BY
',' LOCATION '/user/joe/x;
LOAD DATA INPATH 'input/File1' OVERWRITE INTO TABLE x;
```

Assuming the statements above execute successfully, which one of the following statements is true?

A. Hive reformats File1 into a structure that Hive can access and moves into to/user/joe/x/

B. The file named File1 is moved to to/user/joe/x/

C. The contents of File1 are parsed as comma-delimited rows and loaded into /user/joe/x/

D. The contents of File1 are parsed as comma-delimited rows and stored in a database

Correct Answer: B

QUESTION 8

When can a reduce class also serve as a combiner without affecting the output of a MapReduce program?

A. When the types of the reduce operation\'s input key and input value match the types of the reducer\'s output key and output value and when the reduce operation is both communicative and associative.

B. When the signature of the reduce method matches the signature of the combine method.

C. Always. Code can be reused in Java since it is a polymorphic object-oriented programming language.

D. Always. The point of a combiner is to serve as a mini-reducer directly after the map phase to increase performance.

E. Never. Combiners and reducers must be implemented separately because they serve different purposes.

Correct Answer: A

Explanation: You can use your reducer code as a combiner if the operation performed is commutative and associative.

Reference: 24 Interview Questions and Answers for Hadoop MapReduce developers, What are combiners? When should I use a combiner in my MapReduce Job?

QUESTION 9

You have written a Mapper which invokes the following five calls to the OutputColletor.collect method: output.collect (new Text ("Apple"), new Text ("Red")) ;

output.collect (new Text ("Banana"), new Text ("Yellow")) ;

output.collect (new Text ("Apple"), new Text ("Yellow")) ;

output.collect (new Text ("Cherry"), new Text ("Red")) ;

output.collect (new Text ("Apple"), new Text ("Green")) ;

How many times will the Reducer\\'s reduce method be invoked?

A. 6

B. 3

C. 1

D. 0

E. 5

Correct Answer: B

Explanation: reduce() gets called once for each [key, (list of values)] pair. To explain, let\\'s say you called:

out.collect(new Text("Car"),new Text("Subaru");

out.collect(new Text("Car"),new Text("Honda");

out.collect(new Text("Car"),new Text("Ford");

out.collect(new Text("Truck"),new Text("Dodge");

out.collect(new Text("Truck"),new Text("Chevy");

Then reduce() would be called twice with the pairs reduce(Car,)

reduce(Truck,)

Reference: Mapper output.collect()?

QUESTION 10

What does the following WebHDFS command do?

Curl -1 -L "http://host:port/webhdfs/v1/foo/bar?op=OPEN"

A. Make a directory /foo/bar

B. Read a file /foo/bar

C. List a directory /foo

D. Delete a directory /foo/bar

Correct Answer: B

QUESTION 11

You use the hadoop fs –put command to write a 300 MB file using and HDFS block size of 64 MB. Just after this

command has finished writing 200 MB of this file, what would another user see when trying to access this life?

A. They would see Hadoop throw an ConcurrentFileAccessException when they try to access this file.

B. They would see the current state of the file, up to the last bit written by the command.

C. They would see the current of the file through the last completed block.

D. They would see no content until the whole file written and closed.

Correct Answer: C

QUESTION 12

Indentify the utility that allows you to create and run MapReduce jobs with any executable or script as the mapper and/or the reducer?

A. Oozie

B. Sqoop

C. Flume

D. Hadoop Streaming

E. mapred

Correct Answer: D

Explanation: Hadoop streaming is a utility that comes with the Hadoop distribution. The utility allows you to create and run Map/Reduce jobs with any executable or script as the mapper and/or the reducer.

Reference: http://hadoop.apache.org/common/docs/r0.20.1/streaming.html (Hadoop Streaming, second sentence)

QUESTION 13

You want to count the number of occurrences for each unique word in the supplied input data. You\'ve decided to implement this by having your mapper tokenize each word and emit a literal value 1, and then have your reducer increment a counter for each literal 1 it receives. After successful implementing this, it occurs to you that you could optimize this by specifying a combiner. Will you be able to reuse your existing Reduces as your combiner in this case and why or why not?

A. Yes, because the sum operation is both associative and commutative and the input and output types to the reduce method match.

B. No, because the sum operation in the reducer is incompatible with the operation of a Combiner.

C. No, because the Reducer and Combiner are separate interfaces.

D. No, because the Combiner is incompatible with a mapper which doesn\'t use the same data type for both the key and value.

E. Yes, because Java is a polymorphic object-oriented language and thus reducer code can be reused as a combiner.

Correct Answer: A

Explanation: Combiners are used to increase the efficiency of a MapReduce program. They are used to aggregate intermediate map output locally on individual mapper outputs. Combiners can help you reduce the amount of data that needs to be transferred across to the reducers. You can use your reducer code as a combiner if the operation performed is commutative and associative. The execution of combiner is not guaranteed, Hadoop may or may not execute a combiner. Also, if required it may execute it more then 1 times. Therefore your MapReduce jobs should not depend on the combiners execution.

Reference: 24 Interview Questions and Answers for Hadoop MapReduce developers, What are combiners? When should I use a combiner in my MapReduce Job?

QUESTION 14

You have a directory named jobdata in HDFS that contains four files: _first.txt, second.txt, .third.txt and #data.txt. How many files will be processed by the FileInputFormat.setInputPaths () command when it\\'s given a path object representing this directory?

A. Four, all files will be processed

B. Three, the pound sign is an invalid character for HDFS file names

C. Two, file names with a leading period or underscore are ignored

D. None, the directory cannot be named jobdata

E. One, no special characters can prefix the name of an input file

Correct Answer: C

Explanation: Files starting with \\'_\\' are considered \\'hidden\\' like unix files starting with \\'.\\'. # characters are allowed in HDFS file names.

QUESTION 15

MapReduce v2 (MRv2/YARN) is designed to address which two issues?

A. Single point of failure in the NameNode.

B. Resource pressure on the JobTracker.

C. HDFS latency.

D. Ability to run frameworks other than MapReduce, such as MPI.

E. Reduce complexity of the MapReduce APIs.

F. Standardize on a single MapReduce API.

Correct Answer: AB

![Pass2Lead](https://Pass2Lead.com)
Reference: Apache Hadoop YARN ?Conceptsand; Applications

APACHE-HADOOP-DEVELOPER PDF Dumps

APACHE-HADOOP-DEVELOPER VCE Dumps

APACHE-HADOOP-DEVELOPER Braindumps