# HDPCD Q&As

Hortonworks Data Platform Certified Developer

## Pass Hortonworks HDPCD Exam with 100% Guarantee

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

**https://www.pass4itsure.com/hdpcd.html**

100% Passing Guarantee
100% Money Back Assurance

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

**QUESTION 1**

Which two of the following statements are true about Pig\\'s approach toward data? Choose 2 answers

A. Accepts only data that has a key/value pair structure

B. Accepts data whether it has metadata or not

C. Accepts only data that is defined by metadata tables stored in a database

D. Accepts tab-delimited text data only

E. Accepts any data: structured or unstructured

Correct Answer: BE

**QUESTION 2**

You have user profile records in your OLPT database, that you want to join with web logs you have already ingested into the Hadoop file system. How will you obtain these user records?

A. HDFS command

B. Pig LOAD command

C. Sqoop import

D. Hive LOAD DATA command

E. Ingest with Flume agents

F. Ingest with Hadoop Streaming

Correct Answer: C

Reference: Hadoop and Pig for Large-Scale Web Log Analysis

**QUESTION 3**

In the reducer, the MapReduce API provides you with an iterator over Writable values. What does calling the next () method return?

A. It returns a reference to a different Writable object time.

B. It returns a reference to a Writable object from an object pool.

C. It returns a reference to the same Writable object each time, but populated with different data.

D. It returns a reference to a Writable object. The API leaves unspecified whether this is a reused object or a new object.

E. It returns a reference to the same Writable object if the next value is the same as the previous value, or a new Writable object otherwise.

Correct Answer: C

Explanation: Calling Iterator.next() will always return the SAME EXACT instance of IntWritable, with the contents of that instance replaced with the next value.

Reference: manupulating iterator in mapreduce

**QUESTION 4**

Which one of the following statements is FALSE regarding the communication between DataNodes and a federation of NameNodes in Hadoop 2.2?

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, the reducer receives all values associated with same key. Which statement best describes the ordering of these values?

A. The values are in sorted order.

B. The values are arbitrarily ordered, and the ordering may vary from run to run of the same MapReduce job.

C. The values are arbitrary ordered, but multiple runs of the same MapReduce job will always have the same ordering.

D. Since the values come from mapper outputs, the reducers will receive contiguous sections of sorted values.

Correct Answer: B

Note:

*

Input to the Reducer is the sorted output of the mappers.

*

The framework calls the application\\'s Reduce function once for each unique key in the sorted order.

*

Example:

For the given sample input the first map emits:

The second map emits:

**QUESTION 6**

In a large MapReduce job with m mappers and n reducers, how many distinct copy operations will there be in the sort/shuffle phase?

A. mXn (i.e., m multiplied by n)

B. n

C. m

D. m+n (i.e., m plus n) E. mn (i.e., m to the power of n)

Correct Answer: A

Explanation: A MapReduce job with m mappers and r reducers involves up to m * r distinct copy operations, since each mapper may have intermediate output going to every reducer.

**QUESTION 7**

You want to populate an associative array in order to perform a map-side join. You\\'ve decided to put this information in a text file, place that file into the DistributedCache and read it in your Mapper before any records are processed.

Indentify which method in the Mapper you should use to implement code for reading the file and populating the associative array?

A. combine

B. map

C. init

D. configure

Correct Answer: D

Reference: org.apache.hadoop.filecache , Class DistributedCache

**QUESTION 8**

Review the following data and Pig code.

M,38,95111

F,29,95060

F,45,95192

M,62,95102

F,56,95102

A = LOAD andapos;dataandapos; USING PigStorage(andapos;.andapos;) as (gender:Chararray, age:int,

zlp:chararray);

B = FOREACH A GENERATE age;

Which one of the following commands would save the results of B to a folder in hdfs named myoutput?

A. STORE A INTO andapos;myoutputandapos; USING PigStorage(andapos;,andapos;);

B. DUMP B using PigStorage(andapos;myoutputandapos;);

C. STORE B INTO andapos;myoutputandapos;;

D. DUMP B INTO andapos;myoutputandapos;;

Correct Answer: C

**QUESTION 9**

Analyze each scenario below and indentify which best describes the behavior of the default partitioner?

A. The default partitioner assigns key-values pairs to reduces based on an internal random number generator.

B. The default partitioner implements a round-robin strategy, shuffling the key-value pairs to each reducer in turn. This ensures an event partition of the key space.

C. The default partitioner computes the hash of the key. Hash values between specific ranges are associated with different buckets, and each bucket is assigned to a specific reducer.

D. The default partitioner computes the hash of the key and divides that valule modulo the number of reducers. The result determines the reducer assigned to process the key-value pair.

E. The default partitioner computes the hash of the value and takes the mod of that value with the number of reducers. The result determines the reducer assigned to process the key-value pair.

Correct Answer: D

Explanation: The default partitioner computes a hash value for the key and assigns the partition based on this result.

The default Partitioner implementation is called HashPartitioner. It uses the hashCode() method of the key objects modulo the number of partitions total to determine which partition to send a given (key, value) pair to.

In Hadoop, the default partitioner is HashPartitioner, which hashes a record\\'s key to determine which partition (and thus which reducer) the record belongs in.The number of partition is then equal to the number of reduce tasks for the job.

Reference: Getting Started With (Customized) Partitioning

**QUESTION 10**

All keys used for intermediate output from mappers must:

A. Implement a splittable compression algorithm.

B. Be a subclass of FileInputFormat.

C. Implement WritableComparable.

D. Override isSplitable.

E. Implement a comparator for speedy sorting.

Correct Answer: C

Explanation: The MapReduce framework operates exclusively on pairs, that is, the framework views the input to the job as a set of pairs and produces a set of pairs as the output of the job, conceivably of different types.

The key and value classes have to be serializable by the framework and hence need to implement the Writable interface. Additionally, the key classes have to implement the WritableComparable interface to facilitate sorting by the framework.

Reference: MapReduce Tutorial

**QUESTION 11**

You need to create a job that does frequency analysis on input data. You will do this by writing a Mapper that uses TextInputFormat and splits each value (a line of text from an input file) into individual characters. For each one of these characters, you will emit the character as a key and an InputWritable as the value. As this will produce proportionally more intermediate data than input data, which two resources should you expect to be bottlenecks?

A. Processor and network I/O

B. Disk I/O and network I/O

C. Processor and RAM

D. Processor and disk I/O

Correct Answer: B

**QUESTION 12**

You are developing a combiner that takes as input Text keys, IntWritable values, and emits Text keys, IntWritable values. Which interface should your class implement?

A. Combiner

B. Mapper

C. Reducer

D. Reducer

E. Combiner

Correct Answer: D

**QUESTION 13**

What is a SequenceFile?

A. A SequenceFile contains a binary encoding of an arbitrary number of homogeneous writable objects.

B. A SequenceFile contains a binary encoding of an arbitrary number of heterogeneous writable objects.

C. A SequenceFile contains a binary encoding of an arbitrary number of WritableComparable objects, in sorted order.

D. A SequenceFile contains a binary encoding of an arbitrary number key-value pairs. Each key must be the same type. Each value must be same type.

Correct Answer: D

Explanation: SequenceFile is a flat file consisting of binary key/value pairs.

There are 3 different SequenceFile formats:

Uncompressed key/value records.

Record compressed key/value records - only \\'values\\' are compressed here. Block compressed key/value

records - both keys and values are collected in \\'blocks\\' separately and compressed. The size of the \\'block\\'

is configurable.

Reference: http://wiki.apache.org/hadoop/SequenceFile

**QUESTION 14**

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 15**

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

F. Three

Correct Answer: B

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