

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.leads4pass.com/hdpcd.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

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

- A. The default partitioner assigns key-values pairs to reducers 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 even 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 value 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 partitions is then equal to the number of reduce tasks for the job.

Reference: Getting Started With (Customized) Partitioning

QUESTION 2

You use the `hadoop fs -put` command to write a 300 MB file using an 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 file?

- A. They would see Hadoop throw a `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 is written and closed.

Correct Answer: C

QUESTION 3

Assuming the following Hive query executes successfully:

```
from inputdata select context_ngrams(sentences(lines),  
array("you", "are", null), 80);
```

Which one of the following statements describes the result set?

- A. A bigram of the top 80 sentences that contain the substring "you are" in the lines column of the input data A1 table.
- B. An 80-value ngram of sentences that contain the words "you" or "are" in the lines column of the inputdata table.
- C. A trigram of the top 80 sentences that contain "you are" followed by a null space in the lines column of the inputdata table.
- D. A frequency distribution of the top 80 words that follow the subsequence "you are" in the lines column of the inputdata table.

Correct Answer: D

QUESTION 4

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 5

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 6

For each intermediate key, each reducer task can emit:

- A. As many final key-value pairs as desired. There are no restrictions on the types of those key-value pairs (i.e., they can be heterogeneous).
- B. As many final key-value pairs as desired, but they must have the same type as the intermediate key-value pairs.
- C. As many final key-value pairs as desired, as long as all the keys have the same type and all the values have the same type.
- D. One final key-value pair per value associated with the key; no restrictions on the type.
- E. One final key-value pair per key; no restrictions on the type.

Correct Answer: C

Reference: Hadoop Map-Reduce Tutorial; Yahoo! Hadoop Tutorial, Module 4: MapReduce

QUESTION 7

Determine which best describes when the reduce method is first called in a MapReduce job?

- A. Reducers start copying intermediate key-value pairs from each Mapper as soon as it has completed. The programmer can configure in the job what percentage of the intermediate data should arrive before the reduce method begins.
- B. Reducers start copying intermediate key-value pairs from each Mapper as soon as it has completed. The reduce method is called only after all intermediate data has been copied and sorted.
- C. Reduce methods and map methods all start at the beginning of a job, in order to provide optimal performance for map-only or reduce-only jobs.
- D. Reducers start copying intermediate key-value pairs from each Mapper as soon as it has completed. The reduce method is called as soon as the intermediate key-value pairs start to arrive.

Correct Answer: B

Reference: 24 Interview Questions and Answers for Hadoop MapReduce developers , When is the reducers are started in a MapReduce job?

QUESTION 8

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 9

To process input key-value pairs, your mapper needs to load a 512 MB data file in memory. What is the best way to accomplish this?

- A. Serialize the data file, insert in it the JobConf object, and read the data into memory in the configure method of the mapper.
- B. Place the data file in the DistributedCache and read the data into memory in the map method of the mapper.
- C. Place the data file in the DataCache and read the data into memory in the configure method of the mapper.
- D. Place the data file in the DistributedCache and read the data into memory in the configure method of the mapper.

Correct Answer: C

QUESTION 10

Which HDFS command copies an HDFS file named foo to the local filesystem as localFoo?

- A. `hadoop fs -get foo LocalFoo`
- B. `hadoop -cp foo LocalFoo`
- C. `hadoop fs -ls foo`
- D. `hadoop fs -put foo LocalFoo`

Correct Answer: A

QUESTION 11

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 12

Examine the following Pig commands: Which one of the following statements is true?

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

- 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 13

You are developing a MapReduce job for sales reporting. The mapper will process input keys representing the year (IntWritable) and input values representing product identifiers (Text).

Identify what determines the data types used by the Mapper for a given job.

- A. The key and value types specified in the JobConf.setMapInputKeyClass and JobConf.setMapInputValuesClass methods
- B. The data types specified in HADOOP_MAP_DATATYPES environment variable
- C. The mapper-specification.xml file submitted with the job determine the mapper's input key and value types.
- D. The InputFormat used by the job determines the mapper's input key and value types.

Correct Answer: D

Explanation: The input types fed to the mapper are controlled by the InputFormat used. The default input format, "TextInputFormat," will load data in as (LongWritable, Text) pairs. The long value is the byte offset of the line in the file. The Text object holds the string contents of the line of the file.

Note: The data types emitted by the reducer are identified by setOutputKeyClass()

and setOutputValueClass(). The data types emitted by the reducer are identified by setOutputKeyClass()

and `setOutputValueClass()`.

By default, it is assumed that these are the output types of the mapper as well. If this is not the case, the methods `setMapOutputKeyClass()` and `setMapOutputValueClass()` methods of the `JobConf` class will override these.

Reference: Yahoo! Hadoop Tutorial, THE DRIVER METHOD

QUESTION 14

For each input key-value pair, mappers can emit:

- A. As many intermediate key-value pairs as designed. There are no restrictions on the types of those key-value pairs (i.e., they can be heterogeneous).
- B. As many intermediate key-value pairs as designed, but they cannot be of the same type as the input key-value pair.
- C. One intermediate key-value pair, of a different type.
- D. One intermediate key-value pair, but of the same type.
- E. As many intermediate key-value pairs as designed, as long as all the keys have the same types and all the values have the same type.

Correct Answer: E

Explanation: Mapper maps input key/value pairs to a set of intermediate key/value pairs.

Maps are the individual tasks that transform input records into intermediate records. The transformed intermediate records do not need to be of the same type as the input records. A given input pair may map to zero or many output pairs.

Reference: Hadoop Map-Reduce Tutorial

QUESTION 15

In a MapReduce job, you want each of your input files processed by a single map task. How do you configure a MapReduce job so that a single map task processes each input file regardless of how many blocks the input file occupies?

- A. Increase the parameter that controls minimum split size in the job configuration.
- B. Write a custom `MapRunner` that iterates over all key-value pairs in the entire file.
- C. Set the number of mappers equal to the number of input files you want to process.
- D. Write a custom `FileInputFormat` and override the method `isSplittable` to always return false.

Correct Answer: D

Explanation: `FileInputFormat` is the base class for all file-based `InputFormats`. This provides a generic implementation of

getSplits(JobContext). Subclasses of FileInputFormat can also override the isSplittable (JobContext, Path) method to ensure input-files are not split-up and are processed as a whole by Mappers.

Reference: org.apache.hadoop.mapreduce.lib.input, Class FileInputFormat

[Latest HDPCD Dumps](#)

[HDPCD Practice Test](#)

[HDPCD Study Guide](#)