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

Given the following Hive commands:

```
CREATE TABLE mytable (name chararray, age int) ROW FORMAT
DELIMITED FIELDS TERMINATED BY ',' STORED AS
TEXTFILE;
LOAD DATA INPATH '/home/user/mydata.txt' INTO TABLE mytable;
```

Which one of the following statements Is true?

A. The file mydata.txt is copied to a subfolder of /apps/hive/warehouse

B. The file mydata.txt is moved to a subfolder of /apps/hive/warehouse

C. The file mydata.txt is copied into Hive\\'s underlying relational database 0.

D. The file mydata.txt does not move from Its current location in HDFS

Correct Answer: A

QUESTION 2

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 3

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 4

Can you use MapReduce to perform a relational join on two large tables sharing a key? Assume that the two tables are formatted as comma-separated files in HDFS.

A. Yes.

B. Yes, but only if one of the tables fits into memory

C. Yes, so long as both tables fit into memory.

D. No, MapReduce cannot perform relational operations.

E. No, but it can be done with either Pig or Hive.

Correct Answer: A

Explanation: Note:

*

Join Algorithms in MapReduce A) Reduce-side join B) Map-side join C) In-memory join / Striped Striped variant variant / Memcached variant

*

Which join to use? / In-memory join > map-side join > reduce-side join / Limitations of each? In-memory join: memory Map-side join: sort order and partitioning Reduce-side join: general purpose

QUESTION 5

Examine the following Hive statements:

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

![Pass2Lead](https://Pass2Lead.com)
Assuming the statements above execute successfully, which one of the following statements is true?

A. Each reducer generates a file sorted by age

B. The SORT BY command causes only one reducer to be used

C. The output of each reducer is only the age column

D. The output is guaranteed to be a single file with all the data sorted by age

Correct Answer: A

QUESTION 6

You need to move a file titled "weblogs" into HDFS. When you try to copy the file, you can\\\'t. You know you have ample space on your DataNodes. Which action should you take to relieve this situation and store more files in HDFS?

A. Increase the block size on all current files in HDFS.

B. Increase the block size on your remaining files.

C. Decrease the block size on your remaining files.

D. Increase the amount of memory for the NameNode.

E. Increase the number of disks (or size) for the NameNode.

F. Decrease the block size on all current files in HDFS.

Correct Answer: C

QUESTION 7

Which one of the following statements describes the relationship between the NodeManager and the ApplicationMaster?

A. The ApplicationMaster starts the NodeManager in a Container

B. The NodeManager requests resources from the ApplicationMaster

C. The ApplicationMaster starts the NodeManager outside of a Container

D. The NodeManager creates an instance of the ApplicationMaster

Correct Answer: D

QUESTION 8

Identify the tool best suited to import a portion of a relational database every day as files into HDFS, and generate Java classes to interact with that imported data?

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

B. Flume

C. Pig

D. Hue

E. Hive

F. Sqoop

G. fuse-dfs

Correct Answer: F

Sqoop ("SQL-to-Hadoop") is a straightforward command-line tool with the following capabilities:

Imports individual tables or entire databases to files in HDFS Generates Java classes to allow you to

interact with your imported data Provides the ability to import from SQL databases straight into your Hive

data warehouse

Note:

Data Movement Between Hadoop and Relational Databases Data can be moved between Hadoop and a

relational database as a bulk data transfer, or relational tables can be accessed from within a MapReduce

map function.

Note:

* Cloudera\\'s Distribution for Hadoop provides a bulk data transfer tool (i.e., Sqoop) that imports individual tables or entire databases into HDFS files. The tool also generates Java classes that support interaction with the imported data. Sqoop supports all relational databases over JDBC, and Quest Software provides a connector (i.e., OraOop) that has been optimized for access to data residing in Oracle databases.

Reference: http://log.medcl.net/item/2011/08/hadoop-and-mapreduce-big-data-analytics- gartner/ (Data Movement between hadoop and relational databases, second paragraph)

QUESTION 9

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 inot Hadoop using Hadoop Streaming.

E. Sample the weblogs from the web servers, copying them into Hadoop using curl.

Correct Answer: A

QUESTION 10

Workflows expressed in Oozie can contain:

A. Sequences of MapReduce and Pig. These sequences can be combined with other actions including forks, decision points, and path joins.

B. Sequences of MapReduce job only; on Pig on Hive tasks or jobs. These MapReduce sequences can be combined with forks and path joins.

C. Sequences of MapReduce and Pig jobs. These are limited to linear sequences of actions with exception handlers but no forks.

D. Iterntive repetition of MapReduce jobs until a desired answer or state is reached.

Correct Answer: A

Explanation: Oozie workflow is a collection of actions (i.e. Hadoop Map/Reduce jobs, Pig jobs) arranged in a control dependency DAG (Direct Acyclic Graph), specifying a sequence of actions execution. This graph is specified in hPDL (a XML Process Definition Language).

hPDL is a fairly compact language, using a limited amount of flow control and action nodes. Control nodes define the flow of execution and include beginning and end of a workflow (start, end and fail nodes) and mechanisms to control the workflow execution path (decision, fork and join nodes).

Workflow definitions Currently running workflow instances, including instance states and variables

Reference: Introduction to Oozie

Note: Oozie is a Java Web-Application that runs in a Java servlet-container - Tomcat and uses a database to store:

QUESTION 11

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 12

Which describes how a client reads a file from HDFS?

A. The client queries the NameNode for the block location(s). The NameNode returns the block location(s) to the client. The client reads the data directory off the DataNode(s).

B. The client queries all DataNodes in parallel. The DataNode that contains the requested data responds directly to the client. The client reads the data directly off the DataNode.

C. The client contacts the NameNode for the block location(s). The NameNode then queries the DataNodes for block locations. The DataNodes respond to the NameNode, and the NameNode redirects the client to the DataNode that holds the requested data block(s). The client then reads the data directly off the DataNode.

D. The client contacts the NameNode for the block location(s). The NameNode contacts the DataNode that holds the requested data block. Data is transferred from the DataNode to the NameNode, and then from the NameNode to the client.

Correct Answer: A

Reference: 24 Interview Questions and Answers for Hadoop MapReduce developers, How the Client communicates with HDFS?

QUESTION 13

Given the following Hive command:

INSERT OVERWRITE TABLE mytable SELECT * FROM myothertable;

Which one of the following statements is true?

A. The contents of myothertable are appended to mytable

B. Any existing data in mytable will be overwritten

C. A new table named mytable is created, and the contents of myothertable are copied into mytable

D. The statement is not a valid Hive command

Correct Answer: B

QUESTION 14

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 YARW job to the ResourceManager

D. The ResourceManager monitors and restarts any failed Containers

Correct Answer: D

QUESTION 15

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

[HDPCD Practice Test](#)   [HDPCD Exam Questions](#)   [HDPCD Braindumps](#)