

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

A NameNode in Hadoop 2.2 manages _____.

- A. Two namespaces: an active namespace and a backup namespace
- B. A single namespace
- C. An arbitrary number of namespaces
- D. No namespaces

Correct Answer: B

QUESTION 2

What is the disadvantage of using multiple reducers with the default HashPartitioner and distributing your workload across your cluster?

- A. You will not be able to compress the intermediate data.
- B. You will no longer be able to take advantage of a Combiner.
- C. By using multiple reducers with the default HashPartitioner, output files may not be in globally sorted order.
- D. There are no concerns with this approach. It is always advisable to use multiple reducers.

Correct Answer: C

Explanation: Multiple reducers and total ordering If your sort job runs with multiple reducers (either because `mapreduce.job.reduces` in `mapred-site.xml` has been set to a number larger than 1, or because you've used the `-r` option to specify the number of reducers on the command-line), then by default Hadoop will use the HashPartitioner to distribute records across the reducers. Use of the HashPartitioner means that you can't concatenate your output files to create a single sorted output file. To do this you'll need total ordering,

Reference: [Sorting text files with MapReduce](#)

QUESTION 3

Which project gives you a distributed, Scalable, data store that allows you random, realtime read/write access to hundreds of terabytes of data?

- A. HBase
- B. Hue
- C. Pig
- D. Hive
- E. Oozie

F. Flume

G. Sqoop

Correct Answer: A

Explanation: Use Apache HBase when you need random, realtime read/write access to your Big Data.

Note: This project's goal is the hosting of very large tables -- billions of rows X millions of columns -- atop clusters of commodity hardware. Apache HBase is an open-source, distributed, versioned, column-oriented store modeled after Google's Bigtable: A Distributed Storage System for Structured Data by Chang et al. Just as Bigtable leverages the distributed data storage provided by the Google File System, Apache HBase provides Bigtable-like capabilities on top of Hadoop and HDFS.

Features

Linear and modular scalability.

Strictly consistent reads and writes.

Automatic and configurable sharding of tables

Automatic failover support between RegionServers.

Convenient base classes for backing Hadoop MapReduce jobs with Apache HBase tables.

Easy to use Java API for client access.

Block cache and Bloom Filters for real-time queries.

Query predicate push down via server side Filters

Thrift gateway and a REST-ful Web service that supports XML, Protobuf, and binary data encoding options

Extensible jruby-based (JIRB) shell

Support for exporting metrics via the Hadoop metrics subsystem to files or Ganglia; or via JMX

Reference: <http://hbase.apache.org/> (when would I use HBase? First sentence)

QUESTION 4

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 5

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

[Latest HDPCD Dumps](#)

[HDPCD Practice Test](#)

[HDPCD Exam Questions](#)