



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.geekcert.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

Identify which best defines 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 the 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 2

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 3

Consider the following two relations, A and B.

```
A = LOAD 'data1' AS (a1:int,a2:chararray);
DUMP A;
(1,apple)
(3,orange)
(4,peach)
(2,cherry)

B = LOAD 'data2' AS (b1:chararray,b2:int);
DUMP B;
(Jim,2)
(Brian,4)
(Kim,0)
(Terry,3)
(Chris,2)
```

A Pig JOIN statement that combined relations A by its first field and B by its second field would produce what output?

- A. 2 Jim Chris 2 3 Terry 3 4 Brian 4
- B. 2 cherry 2 cherry 3 orange 4 peach
- C. 2 cherry Jim, Chris 3 orange Terry 4 peach Brian
- D. 2 cherry Jim 2 2 cherry Chris 2 3 orange Terry 3 4 peach Brian 4

Correct Answer: D



QUESTION 4

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 5

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?

- 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)