

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

Table metadata in Hive is:

- A. Stored as metadata on the NameNode.
- B. Stored along with the data in HDFS.
- C. Stored in the Metastore.
- D. Stored in ZooKeeper.

Correct Answer: C

Explanation: By default, hive use an embedded Derby database to store metadata information. The metastore is the "glue" between Hive and HDFS. It tells Hive where your data files live in HDFS, what type of data they contain, what tables they belong to, etc.

The Metastore is an application that runs on an RDBMS and uses an open source ORM layer called DataNucleus, to convert object representations into a relational schema and vice versa. They chose this approach as opposed to storing this information in hdfs as they need the Metastore to be very low latency. The DataNucleus layer allows them to plugin many different RDBMS technologies.

Note:

*

By default, Hive stores metadata in an embedded Apache Derby database, and other client/server databases like MySQL can optionally be used.

*

features of Hive include:

Metadata storage in an RDBMS, significantly reducing the time to perform semantic checks during query execution.

Reference: Store Hive Metadata into RDBMS

QUESTION 2

Given a directory of files with the following structure: line number, tab character, string: Example:

1abialkjjkaoasdfjksdlkjhweroij 2kadfjhuwqounahagtnbvaswslmnbfgy 3kjfteiomndscxeqalkzhtopedkfsikj You want to send each line as one record to your Mapper. Which InputFormat should you use to complete

the line: `conf.setInputFormat (____.class) ; ?`

- A. SequenceFileAsTextInputFormat
- B. SequenceFileInputFormat

C. KeyValueFileInputFormat

D. BDBInputFormat

Correct Answer: C

<http://stackoverflow.com/questions/9721754/how-to-parse-customwritable-from-text-in-hadoop>

QUESTION 3

A client application creates an HDFS file named foo.txt with a replication factor of 3. Identify which best describes the file access rules in HDFS if the file has a single block that is stored on data nodes A, B and C?

- A. The file will be marked as corrupted if data node B fails during the creation of the file.
- B. Each data node locks the local file to prohibit concurrent readers and writers of the file.
- C. Each data node stores a copy of the file in the local file system with the same name as the HDFS file.
- D. The file can be accessed if at least one of the data nodes storing the file is available.

Correct Answer: D

Explanation: HDFS keeps three copies of a block on three different datanodes to protect against true data corruption. HDFS also tries to distribute these three replicas on more than one rack to protect against data availability issues. The fact that HDFS actively monitors any failed datanode(s) and upon failure detection immediately schedules re-replication of blocks (if needed) implies that three copies of data on three different nodes is sufficient to avoid corrupted files. Note: HDFS is designed to reliably store very large files across machines in a large cluster. It stores each file as a sequence of blocks; all blocks in a file except the last block are the same size. The blocks of a file are replicated for fault tolerance. The block size and replication factor are configurable per file. An application can specify the number of replicas of a file. The replication factor can be specified at file creation time and can be changed later. Files in HDFS are write-once and have strictly one writer at any time. The NameNode makes all decisions regarding replication of blocks. HDFS uses rack-aware replica placement policy. In default configuration there are total 3 copies of a datablock on HDFS, 2 copies are stored on datanodes on same rack and 3rd copy on a different rack. Reference: 24 Interview Questions and Answers for Hadoop MapReduce developers , How the HDFS Blocks are replicated?

QUESTION 4

Which TWO of the following statements are true regarding Hive? Choose 2 answers

- A. Useful for data analysts familiar with SQL who need to do ad-hoc queries
- B. Offers real-time queries and row level updates
- C. Allows you to define a structure for your unstructured Big Data
- D. Is a relational database

Correct Answer: AC

QUESTION 5

Which one of the following files is required in every Oozie Workflow application?

- A. job.properties
- B. Config-default.xml
- C. Workflow.xml
- D. Oozie.xml

Correct Answer: C

[HDPCD Practice Test](#)

[HDPCD Exam Questions](#)

[HDPCD Braindumps](#)