HADOOP-PR000007 Q&As

Hortonworks Certified Apache Hadoop 2.0 Developer (Pig and Hive Developer)

Pass Hortonworks HADOOP-PR000007 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.pass4itsure.com/hadoop-pr000007.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Hortonworks Official Exam Center

QUESTION 1

Your cluster\\'s HDFS block size in 64MB. You have directory containing 100 plain text files, each of which is 100MB in size. The InputFormat for your job is TextInputFormat.

Determine how many Mappers will run?

A. 64

B. 100

C. 200

D. 640

Correct Answer: C

Explanation: Each file would be split into two as the block size (64 MB) is less than the file size (100 MB), so 200 mappers would be running.

Note:

If you\\'re not compressing the files then hadoop will process your large files (say 10G), with a number of mappers related to the block size of the file.

Say your block size is 64M, then you will have ~160 mappers processing this 10G file (160*64 ~= 10G).

Depending on how CPU intensive your mapper logic is, this might be an acceptable blocks size, but if you find that your mappers are executing in sub minute times, then you might want to increase the work done by each mapper (by increasing the block size to 128, 256, 512m - the actual size depends on how you intend to process the data). Reference: http://stackoverflow.com/

QUESTION 2

Which one of the following statements is false about HCatalog?

A. Provides a shared schema mechanism

B. Designed to be used by other programs such as Pig, Hive and MapReduce

C. Stores HDFS data in a database for performing SQL-like ad-hoc queries

D. Exists as a subproject of Hive

Correct Answer: C

QUESTION 3

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 4

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 5

MapReduce v2 (MRv2/YARN) is designed to address which two issues?

A. Single point of failure in the NameNode.

B. Resource pressure on the JobTracker.

C. HDFS latency.

D. Ability to run frameworks other than MapReduce, such as MPI.

E. Reduce complexity of the MapReduce APIs.

F. Standardize on a single MapReduce API.

Correct Answer: AB

Reference: Apache Hadoop YARN ?Conceptsand; Applications