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

Review the following data and Pig code.

M,38,95111

F,29,95060

F,45,95192

M,62,95102

F,56,95102

A = LOAD andapos;dataandapos; USING PigStorage(andapos;.andapos;) as (gender:Chararray, age:int,

zlp:chararray);

B = FOREACH A GENERATE age;

Which one of the following commands would save the results of B to a folder in hdfs named myoutput?

A. STORE A INTO andapos;myoutputandapos; USING PigStorage(andapos;,andapos;);

B. DUMP B using PigStorage(andapos;myoutputandapos;);

C. STORE B INTO andapos;myoutputandapos;;

D. DUMP B INTO andapos;myoutputandapos;;

Correct Answer: C

QUESTION 2

Which Two of the following statements are true about hdfs? Choose 2 answers

A. An HDFS file that is larger than dfs.block.size is split into blocks

B. Blocks are replicated to multiple datanodes

C. HDFS works best when storing a large number of relatively small files

D. Block sizes for all files must be the same size

Correct Answer: AB

QUESTION 3

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 4

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

QUESTION 5

On a cluster running MapReduce v1 (MRv1), a TaskTracker heartbeats into the JobTracker on your cluster, and alerts the JobTracker it has an open map task slot.

What determines how the JobTracker assigns each map task to a TaskTracker?

A. The amount of RAM installed on the TaskTracker node.

B. The amount of free disk space on the TaskTracker node.

C. The number and speed of CPU cores on the TaskTracker node.

D. The average system load on the TaskTracker node over the past fifteen (15) minutes.

E. The location of the InsputSplit to be processed in relation to the location of the node.

Correct Answer: E

Explanation: The TaskTrackers send out heartbeat messages to the JobTracker, usually every few minutes, to reassure the JobTracker that it is still alive. These message also inform the JobTracker of the number of available slots, so the JobTracker can stay up to date with where in the cluster work can be delegated. When the JobTracker tries to find somewhere to schedule a task within the MapReduce operations, it first looks for an empty slot on the same server that hosts the DataNode containing the data, and if not, it looks for an empty slot on a machine in the same rack.

Reference: 24 Interview Questions and Answers for Hadoop MapReduce developers, How JobTracker schedules a task?

QUESTION 6

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 7

You need to perform statistical analysis in your MapReduce job and would like to call methods in the Apache Commons Math library, which is distributed as a 1.3 megabyte Java archive (JAR) file. Which is the best way to make this library available to your MapReducer job at runtime?

A. Have your system administrator copy the JAR to all nodes in the cluster and set its location in the HADOOP_CLASSPATH environment variable before you submit your job.

B. Have your system administrator place the JAR file on a Web server accessible to all cluster nodes and then set the HTTP_JAR_URL environment variable to its location.

C. When submitting the job on the command line, specify the ?ibjars option followed by the JAR file path.

D. Package your code and the Apache Commands Math library into a zip file named JobJar.zip

Correct Answer: C

Explanation: The usage of the jar command is like this,

Usage: hadoop jar [mainClass] args...

If you want the commons-math3.jar to be available for all the tasks you can do any one of these

1. Copy the jar file in $HADOOP_HOME/lib dir

2.

or

Use the generic option -libjars.

QUESTION 8

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)
```

What is the output of the following Pig commands?

X = GROUP A BY S1;

DUMP X;

A. (group,{(apple,peach,cherry,orange)})

B. {apple,peach,cherry,orange}

C. {1,4,2,3}

D.
```
(apple,{(1,apple)})
(peach,{(4,peach)})
(cherry,{(2,cherry)})
(orange,{(3,orange)})
```

A. Option A

B. Option B

C. Option C

D. Option D

Correct Answer: D

QUESTION 9

Indentify the utility that allows you to create and run MapReduce jobs with any executable or script as the mapper and/or the reducer?

A. Oozie

B. Sqoop

C. Flume

D. Hadoop Streaming

E. mapred

Correct Answer: D

Explanation: Hadoop streaming is a utility that comes with the Hadoop distribution. The utility allows you to create and run Map/Reduce jobs with any executable or script as the mapper and/or the reducer.

Reference: http://hadoop.apache.org/common/docs/r0.20.1/streaming.html (Hadoop Streaming, second sentence)

QUESTION 10

How are keys and values presented and passed to the reducers during a standard sort and shuffle phase of MapReduce?

A. Keys are presented to reducer in sorted order; values for a given key are not sorted.

B. Keys are presented to reducer in sorted order; values for a given key are sorted in ascending order.

C. Keys are presented to a reducer in random order; values for a given key are not sorted.

D. Keys are presented to a reducer in random order; values for a given key are sorted in ascending order.

Correct Answer: A

Explanation: Reducer has 3 primary phases:

1.

Shuffle

The Reducer copies the sorted output from each Mapper using HTTP across the network.

2.

Sort

The framework merge sorts Reducer inputs by keys (since different Mappers may have output the same key).

The shuffle and sort phases occur simultaneously i.e. while outputs are being fetched they are merged.

SecondarySort

To achieve a secondary sort on the values returned by the value iterator, the application should extend the key with the secondary key and define a grouping comparator. The keys will be sorted using the entire key, but will be grouped using the grouping comparator to decide which keys and values are sent in the same call to reduce.

3. Reduce In this phase the reduce(Object, Iterable, Context) method is called for each in the sorted inputs.

The output of the reduce task is typically written to a RecordWriter via TaskInputOutputContext.write (Object, Object).

The output of the Reducer is not re-sorted.

Reference: org.apache.hadoop.mapreduce, Class Reducer

QUESTION 11

Given the following Pig command:

logevents = LOAD andapos;input/my.logandapos; AS (date:chararray, levehstring, code:int, message:string);

Which one of the following statements is true?

A. The logevents relation represents the data from the my.log file, using a comma as the parsing delimiter

B. The logevents relation represents the data from the my.log file, using a tab as the parsing delimiter

C. The first field of logevents must be a properly-formatted date string or table return an error

D. The statement is not a valid Pig command

Correct Answer: B

QUESTION 12

Examine the following Hive statements:

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

Assuming the statements above execute successfully, which one of the following statements is true?

A. Hive reformats File1 into a structure that Hive can access and moves into to/user/joe/x/

B. The file named File1 is moved to to/user/joe/x/

C. The contents of File1 are parsed as comma-delimited rows and loaded into /user/joe/x/

D. The contents of File1 are parsed as comma-delimited rows and stored in a database

Correct Answer: B

QUESTION 13

For each input key-value pair, mappers can emit:

A. As many intermediate key-value pairs as designed. There are no restrictions on the types of those key-value pairs (i.e., they can be heterogeneous).

B. As many intermediate key-value pairs as designed, but they cannot be of the same type as the input key-value pair.

C. One intermediate key-value pair, of a different type.

D. One intermediate key-value pair, but of the same type.

E. As many intermediate key-value pairs as designed, as long as all the keys have the same types and all the values have the same type.

Correct Answer: E

Explanation: Mapper maps input key/value pairs to a set of intermediate key/value pairs.

Maps are the individual tasks that transform input records into intermediate records. The transformed intermediate records do not need to be of the same type as the input records. A given input pair may map to zero or many output pairs.

Reference: Hadoop Map-Reduce Tutorial

QUESTION 14

You are developing a combiner that takes as input Text keys, IntWritable values, and emits Text keys, IntWritable values. Which interface should your class implement?

A. Combiner

B. Mapper

C. Reducer

D. Reducer

E. Combiner

Correct Answer: D

QUESTION 15

Which of the following tool was designed to import data from a relational database into HDFS?

A. HCatalog B. Sqoop

C. Flume

D. Ambari

Correct Answer: B