APACHE-HADOOP-DEVELOPER$^{Q\&As}$

Hadoop 2.0 Certification exam for Pig and Hive Developer

Pass Hortonworks APACHE-HADOOP-DEVELOPER Exam with 100% Guarantee

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

https://www.passapply.com/apache-hadoop-developer.html

100% Passing Guarantee
100% Money Back Assurance

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

QUESTION 1

Which process describes the lifecycle of a Mapper?

A. The JobTracker calls the TaskTracker\\'s configure () method, then its map () method and finally its close () method.

B. The TaskTracker spawns a new Mapper to process all records in a single input split.

C. The TaskTracker spawns a new Mapper to process each key-value pair.

D. The JobTracker spawns a new Mapper to process all records in a single file.

Correct Answer: B

Explanation: For each map instance that runs, the TaskTracker creates a new instance of your mapper.

Note:

*

 The Mapper is responsible for processing Key/Value pairs obtained from the InputFormat. The mapper

may perform a number of Extraction and Transformation functions on the Key/Value pair before ultimately

outputting none, one or many Key/Value pairs of the same, or different Key/Value type.

*

 With the new Hadoop API, mappers extend the org.apache.hadoop.mapreduce.Mapper class. This class

defines an \\'Identity\\' map function by default - every input Key/Value pair obtained from the InputFormat is

written out.

Examining the run() method, we can see the lifecycle of the mapper:

/**

*

 Expert users can override this method for more complete control over the

*

 execution of the Mapper.

*

 @param context

*

 @throws IOException

*/

public void run(Context context) throws IOException, InterruptedException { setup(context);

while (context.nextKeyValue()) {

map(context.getCurrentKey(), context.getCurrentValue(), context); }

cleanup(context);

}

setup(Context) - Perform any setup for the mapper. The default implementation is a no-op method. map(Key, Value, Context) - Perform a map operation in the given Key / Value pair. The default implementation calls Context.write(Key, Value) cleanup(Context) - Perform any cleanup for the mapper. The default implementation is a no-op method.

Reference: Hadoop/MapReduce/Mapper

QUESTION 2

Workflows expressed in Oozie can contain:

A. Sequences of MapReduce and Pig. These sequences can be combined with other actions including forks, decision points, and path joins.

B. Sequences of MapReduce job only; on Pig on Hive tasks or jobs. These MapReduce sequences can be combined with forks and path joins.

C. Sequences of MapReduce and Pig jobs. These are limited to linear sequences of actions with exception handlers but no forks.

D. Iterntive repetition of MapReduce jobs until a desired answer or state is reached.

Correct Answer: A

Explanation: Oozie workflow is a collection of actions (i.e. Hadoop Map/Reduce jobs, Pig jobs) arranged in a control dependency DAG (Direct Acyclic Graph), specifying a sequence of actions execution. This graph is specified in hPDL (a XML Process Definition Language).

hPDL is a fairly compact language, using a limited amount of flow control and action nodes. Control nodes define the flow of execution and include beginning and end of a workflow (start, end and fail nodes) and mechanisms to control the workflow execution path (decision, fork and join nodes).

Workflow definitions Currently running workflow instances, including instance states and variables

Reference: Introduction to Oozie

Note: Oozie is a Java Web-Application that runs in a Java servlet-container - Tomcat and uses a database to store:

QUESTION 3

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 4

Given the following Pig commands: Which one of the following statements is true?

```
logevents = LOAD 'input/my.log';
severe = FILTER logevents BY ($1 =='severe' AND $2 >= 500);
grouped = GROUP severe BY $2;
DUMP grouped;
```

A. The $1 variable represents the first column of data in \\'my.log\\'

B. The $1 variable represents the second column of data in \\'my.log\\'

C. The severe relation is not valid

D. The grouped relation is not valid

Correct Answer: B

QUESTION 5

You need to create a job that does frequency analysis on input data. You will do this by writing a Mapper that uses TextInputFormat and splits each value (a line of text from an input file) into individual characters. For each one of these characters, you will emit the character as a key and an InputWritable as the value. As this will produce proportionally more intermediate data than input data, which two resources should you expect to be bottlenecks?

A. Processor and network I/O

B. Disk I/O and network I/O

C. Processor and RAM

D. Processor and disk I/O

Correct Answer: B

Latest APACHE-HADOOP-DEVELOPER Dumps

APACHE-HADOOP-DEVELOPER Exam Questions

APACHE-HADOOP-DEVELOPER Braindumps