

Looking for Real Exam Questions for IT Certification Exams!

We guarantee you can pass any IT certification exam at your first attempt with just 10-12 hours study of our guides.

Our study guides contain actual exam questions; accurate answers with detailed explanation verified by experts and all graphics and drag-n-drop exhibits shown just as on the real test.

To test the quality of our guides, you can download the one-fourth portion of any guide from <http://www.certificationking.com> absolutely free. You can also download the guides for retired exams that you might have taken in the past.

For pricing and placing order, please visit <http://certificationking.com/order.html>

We accept all major credit cards through www.paypal.com

For other payment options and any further query, feel free to mail us at info@certificationking.com

Question No : 1

You have installed a cluster running HDFS and MapReduce version 2 (MRv2) on YARN. You have no `dfs.hosts` entry()ies in your `hdfs-alte.xml` configuration file. You configure a new worker node by setting `fs.default.name` in its configuration files to point to the NameNode on your cluster, and you start the DataNode daemon on that worker node.

What do you have to do on the cluster to allow the worker node to join, and start storing HDFS blocks?

- A. Nothing; the worker node will automatically join the cluster when the DataNode daemon is started.
- B. Without creating a `dfs.hosts` file or making any entries, run the command `hadoop dfsadmin -refreshHadoop` on the NameNode
- C. Create a `dfs.hosts` file on the NameNode, add the worker node's name to it, then issue the command `hadoop dfsadmin -refreshNodes` on the NameNode
- D. Restart the NameNode

Answer: B

Question No : 2

Given:

```
[user1@host1 ~] $ yarn application -list
Total Applications: 3
```

Application-Id	Application-Name	Application-Type	User	Queue	State	Final State	Process	Tracking-URL
Application_1374638600275_0109	Sleep job	MAPREDUCE	user1	default	KILLED	KILLED	100%	host1:54059
Application_1374638600275_0121	sleep job	MAPREDUCE	user1	default	SUCCEEDED	SUCCEEDED	100%	host1:19888/jobhistory/job/job_1374638600275_0121
Application_1374638600275_0020	sleep job	MAPREDUCE	user1	default	SUCCEEDED	SUCCEEDED	100%	host1:19888/jobhistory/job/job_1374638600275_0020

You want to clean up this list by removing jobs where the state is KILLED. What command you enter?

- A. `Yarn application -kill application_1374638600275_0109`
- B. `Yarn rmdadmin -refreshQueue`
- C. `Yarn application -refreshJobHistory`
- D. `Yarn rmdadmin -kill application_1374638600275_0109`

Answer: A

Reference: http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.1-latest/bk_using-apache-hadoop/content/common_mrv2_commands.html

Question No : 3

Assuming a cluster running HDFS, MapReduce version 2 (MRv2) on YARN with all settings at their default, what do you need to do when adding a new slave node to a cluster?

- A. Nothing, other than ensuring that DNS (or /etc/hosts files on all machines) contains an entry for the new node.
- B. Restart the NameNode and ResourceManager daemons and resubmit any running jobs
- C. Increase the value of dfs.number.of.replicas in hdfs-site.xml
- D. Add a new entry to /etc/nodes on the NameNode host.
- E. Restart the NameNode daemon.

Answer: B

Question No : 4

You have a 20 node Hadoop cluster, with 18 slave nodes and 2 master nodes running HDFS High Availability (HA). You want to minimize the chance of data loss in your cluster. What should you do?

- A. Add another master node to increase the number of nodes running the JournalNode which increases the number of machines available to HA to create a quorum
- B. Configure the cluster's disk drives with an appropriate fault tolerant RAID level
- C. Run the ResourceManager on a different master from the NameNode in the order to load share HDFS metadata processing
- D. Run a Secondary NameNode on a different master from the NameNode in order to load provide automatic recovery from a NameNode failure
- E. Set an HDFS replication factor that provides data redundancy, protecting against failure

Answer: C

Question No : 5

You decide to create a cluster which runs HDFS in High Availability mode with automatic failover, using Quorum-based Storage. What is the purpose of ZooKeeper in such a configuration?

- A. It manages the Edits file, which is a log changes to the HDFS filesystem.
- B. It monitors an NFS mount point and reports if the mount point disappears
- C. It both keeps track of which NameNode is Active at any given time, and manages the Edits file, which is a log of changes to the HDFS filesystem
- D. It only keeps track of which NameNode is Active at any given time
- E. Clients connect to ZoneKeeper to determine which NameNode is Active

Answer: D

Reference: <http://www.cloudera.com/content/cloudera-content/cloudera-docs/CDH4/latest/PDF/CDH4-High-Availability-Guide.pdf> (page 15)

Question No : 6

During the execution of a MapReduce v2 (MRv2) job on YARN, where does the Mapper place the intermediate data each Map task?

- A. The Mapper stores the intermediate data on the node running the job's ApplicationMaster so that is available to YARN's ShuffleService before the data is presented to the Reducer
- B. The Mapper stores the intermediate data in HDFS on the node where the MAP tasks ran in the HDFS /usercache/&[user]sppcache/application_&(appid) directory for the user who ran the job
- C. YARN holds the intermediate data in the NodeManager's memory (a container) until it is transferred to the Reducers
- D. The Mapper stores the intermediate data on the underlying filesystem of the local disk in the directories yarn.nodemanager.local-dirs
- E. The Mapper transfers the intermediate data immediately to the Reducers as it generated by the Map task

Answer: D

Question No : 7

Which Yarn daemon or service monitors a Container's per-application resource usage (e.g, memory, CPU)?

- A. NodeManager
- B. ApplicationMaster

- C. ApplicationManagerService
- D. ResourceManager

Answer: A

Reference: http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.0.0.2/bk_using-apache-hadoop/content/ch_using-apache-hadoop-4.html (4th para)

Question No : 8

You are planning a Hadoop cluster and considering implementing 10 Gigabit Ethernet as the network fabric. Which workloads benefit the most from a faster network fabric?

- A. When your workload generates a large amount of output data, significantly larger than amount of intermediate data
- B. When your workload generates a large amount of intermediate data, on the order of the input data itself
- C. When workload consumers a large amount of input data, relative to the entire capacity of HDFS
- D. When your workload consists of processor-intensive tasks

Answer: B

Question No : 9

For each YARN Job, the Hadoop framework generates task log files. Where are Hadoop's files stored?

- A. In HDFS, In the directory of the user who generates the job
- B. On the local disk of the slave node running the task
- C. Cached In the YARN container running the task, then copied into HDFS on job completion
- D. Cached by the NodeManager managing the job containers, then written to a log directory on the NameNode

Answer: B

Reference: <http://hortonworks.com/blog/simplifying-user-logs-management-and-access-in-yarn/>

Question No : 10

You are the hadoop fs –put command to add a file “sales.txt” to HDFS. This file is small enough that it fits into a single block, which is replicated to three nodes in your cluster (with a replication factor of 3). One of the nodes holding this file (a single block) fails. How will the cluster handle the replication of this file in this situation/

- A.** The cluster will re-replicate the file the next time the system administrator reboots the NameNode daemon (as long as the file's replication doesn't fall two)
- B.** This file will be immediately re-replicated and all other HDFS operations on the cluster will halt until the cluster's replication values are restored
- C.** The file will remain under-replicated until the administrator brings that nodes back online
- D.** The file will be re-replicated automatically after the NameNode determines it is under replicated based on the block reports it receives from the DataNodes

Answer: B

Question No : 11

You are configuring your cluster to run HDFS and MapReduce v2 (MRv2) on YARN. Which daemons need to be installed on your clusters master nodes? (Choose Two)

- A.** ResourceManager
- B.** DataNode
- C.** NameNode
- D.** JobTracker
- E.** TaskTracker
- F.** HMaster

Answer: A,C

Question No : 12

Assume you have a file named foo.txt in your local directory. You issue the following three commands:

Hadoop fs -mkdir input

Hadoop fs -put foo.txt input/foo.txt

Hadoop fs -put foo.txt input

What happens when you issue that third command?

- A. The write succeeds, overwriting foo.txt in HDFS with no warning
- B. The write silently fails
- C. The file is uploaded and stored as a plain named input
- D. You get an error message telling you that input is not a directory
- E. You get a error message telling you that foo.txt already exists. The file is not written to HDFS
- F. You get an error message telling you that foo.txt already exists, and asking you if you would like to overwrite
- G. You get a warning that foo.txt is being overwritten

Answer: E

Question No : 13

You have a Hadoop cluster running HDFS, and a gateway machine external to the cluster from which clients submit jobs. What do you need to do in order to run on the cluster and submit jobs from the command line of the gateway machine?

- A. Install the impslad daemon, statestored daemon, and catalogd daemon on each machine in the cluster and on the gateway node
- B. Install the impalad daemon on each machine in the cluster, the statestored daemon and catalogd daemon on one machine in the cluster, and the impala shell on your gateway machine
- C. Install the impalad daemon and the impala shell on your gateway machine, and the statestored daemon and catalog daemon on one of the nodes in the cluster
- D. Install the impalad daemon, the statestored daemon, the catalogd daemon, and the impala shell on your gateway machine
- E. Install the impalad daemon, statestored daemon, and catalogd daemon on each machine in the cluster, and the impala shell on your gateway machine

Answer: B

Question No : 14