Cloudera

Exam Questions CCA-500

Cloudera Certified Administrator for Apache Hadoop (CCAH)

NEW QUESTION 1
You are migrating a cluster from MApReduce version 1 (MRv1) to MapReduce version 2 (MRv2) on YARN. You want to maintain your MRv1 TaskTracker slot capacities when you migrate. What should you do/

A. Configure yarn.applicationmaster.resource.memory-mb and yarn.applicationmaster.resource.cpu-vcores so that ApplicationMaster container allocations match the capacity you require.
B. You don't need to configure or balance these properties in YARN as YARN dynamically balances resource management capabilities on your cluster
C. Configure mapred.tasktracker.map.tasks.maximum and mapred.tasktracker.reduce.tasks.maximum ub yarn-site.xml to match your cluster's capacity set by the yarn-scheduler.minimum-allocation
D. Configure yarn.nodemanager.resource.memory-mb and yarn.nodemanager.resource.cpu-vcores to match the capacity you require under YARN for each NodeManager

Answer: D

NEW QUESTION 2
You have a Hadoop cluster HDFS, and a gateway machine external to the cluster from which clients submit jobs. What do you need to do in order to run Impala on the cluster and submit jobs from the command line of the gateway machine?

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

Answer: D

NEW QUESTION 3
You observed that the number of spilled records from Map tasks far exceeds the number of map output records. Your child heap size is 1GB and your io.sort.mb value is set to 1000MB. How would you tune your io.sort.mb value to achieve maximum memory to disk I/O ratio?

A. For a 1GB child heap size an io.sort.mb of 128 MB will always maximize memory to disk I/O
B. Increase the io.sort.mb to 1GB
C. Decrease the io.sort.mb value to 0
D. Tune the io.sort.mb value until you observe that the number of spilled records equals (or is as close to equals) the number of map output records.

Answer: D

NEW QUESTION 4
You are planning a Hadoop cluster and considering implementing 10 Gigabit Ethernet as the network fabric. Which workloads benefit the most from faster network fabric?

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

Answer: A

NEW QUESTION 5
A slave node in your cluster has 4 TB hard drives installed (4 x 2TB). The DataNode is configured to store HDFS blocks on all disks. You set the value of the dfs.datanode.du.reserved parameter to 100 GB. How does this alter HDFS block storage?

A. 25GB on each hard drive may not be used to store HDFS blocks
B. 100GB on each hard drive may not be used to store HDFS blocks
C. All hard drives may be used to store HDFS blocks as long as at least 100 GB in total is available on the node
D. A maximum if 100 GB on each hard drive may be used to store HDFS blocks

Answer: B

NEW QUESTION 6
You are running a Hadoop cluster with a NameNode on host mynamenode, a secondary NameNode on host mysecondarynamenode and several DataNodes. Which best describes how you determine when the last checkpoint happened?

A. Execute hdfs namenode –report on the command line and look at the Last Checkpoint information
B. Execute hdfs dfsadmin –saveNamespace on the command line which returns to you the last checkpoint value in fstime file
C. Connect to the web UI of the Secondary NameNode (http://mysecondary:50090/) and look at the "Last Checkpoint" information
D. Connect to the web UI of the NameNode (http://mynamenode:50070) and look at the "Last Checkpoint" information

Answer: C

Explanation: Reference:https://www.inkling.com/read/hadoop-definitive-guide-tom-white-3rd/chapter- 10/hdfs

NEW QUESTION 7

Your Hadoop cluster contains nodes in three racks. You have not configured the dfs.hosts property in the NameNode's configuration file. What results?

A. The NameNode will update the dfs.hosts property to include machines running the DataNode daemon on the next NameNode reboot or with the command dfsadmin–refreshNodes
B. No new nodes can be added to the cluster until you specify them in the dfs.hosts file
C. Any machine running the DataNode daemon can immediately join the cluster
D. Presented with a blank dfs.hosts property, the NameNode will permit DataNodes specified in mapred.hosts to join the cluster

Answer: C

NEW QUESTION 8
Identify two features/issues that YARN is designated to address:(Choose two)

A. Standardize on a single MapReduce API
B. Single point of failure in the NameNode
C. Reduce complexity of the MapReduce APIs
D. Resource pressure on the JobTracker
E. Ability to run framework other than MapReduce, such as MPI
F. HDFS latency

Answer: DE

Explanation: Reference:http://www.revelytix.com/?q=content/hadoop-ecosystem(YARN, first para)

NEW QUESTION 9
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 cluster?

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

Answer: A

Explanation: http://wiki.apache.org/hadoop/FAQ#I_have_a_new_node_I_want_to_add_to_a_running_H adoop_cluster.3B_how_do_I_start_services_on_just_one_node.3F

NEW QUESTION 10
You are working on a project where you need to chain together MapReduce, Pig jobs. You also need the ability to use forks, decision points, and path joins. Which ecosystem project should you use to perform these actions?

A. Oozie
B. ZooKeeper
C. HBase
D. Sqoop
E. HUE

Answer: A

NEW QUESTION 10
You decide to create a cluster which runs HDFS in High Availability mode with automatic failover, using Quorum Storage. What is the purpose of ZooKeeper in such a configuration?

A. It only keeps track of which NameNode is Active at any given time
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 fil
D. Which is a log of changes to the HDFS filesystem
E. If only manages the Edits file, which is log of changes to the HDFS filesystem
F. Clients connect to ZooKeeper to determine which NameNode is Active

Answer: A

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

NEW QUESTION 11
Each node in your Hadoop cluster, running YARN, has 64GB memory and 24 cores. Your yarn.site.xml has the following configuration:
```
<property>
<name>yarn.nodemanager.resource.memory-mb</name>
<value>32768</value>
</property>
<property>
<name>yarn.nodemanager.resource.cpu-vcores</name>
<value>12</value>
```

</property>
You want YARN to launch no more than 16 containers per node. What should you do?

A. Modify yarn-site.xml with the following property:<name>yarn.scheduler.minimum-allocation-mb</name><value>2048</value>
B. Modify yarn-sites.xml with the following property:<name>yarn.scheduler.minimum-allocation-mb</name><value>4096</value>
C. Modify yarn-site.xml with the following property:<name>yarn.nodemanager.resource.cpu-vccores</name>
D. No action is needed: YARN's dynamic resource allocation automatically optimizes the node memory and cores

Answer: A

NEW QUESTION 15
Which scheduler would you deploy to ensure that your cluster allows short jobs to finish within a reasonable time without starting long-running jobs?

A. Complexity Fair Scheduler (CFS)
B. Capacity Scheduler
C. Fair Scheduler
D. FIFO Scheduler

Answer: C

Explanation: Reference:http://hadoop.apache.org/docs/r1.2.1/fair_scheduler.html

NEW QUESTION 16
What two processes must you do if you are running a Hadoop cluster with a single NameNode and six DataNodes, and you want to change a configuration parameter so that it affects all six DataNodes.(Choose two)

A. You must modify the configuration files on the NameNode onl
B. DataNodes read their configuration from the master nodes
C. You must modify the configuration files on each of the six SataNodes machines
D. You don't need to restart any daemon, as they will pick up changes automatically
E. You must restart the NameNode daemon to apply the changes to the cluster
F. You must restart all six DatNode daemon to apply the changes to the cluster

Answer: BD

NEW QUESTION 21
Your company stores user profile records in an OLTP databases. You want to join these records with web server logs you have already ingested into the Hadoop file system. What is the best way to obtain and ingest these user records?

A. Ingest with Hadoop streaming
B. Ingest using Hive's IQAD DATA command
C. Ingest with sqoop import
D. Ingest with Pig's LOAD command
E. Ingest using the HDFS put command

Answer: C

NEW QUESTION 22
Your cluster's mapred-start.xml includes the following parameters
<name>mapreduce.map.memory.mb</name>
<value>4096</value>
<name>mapreduce.reduce.memory.mb</name>
<value>8192</value>
And any cluster's yarn-site.xml includes the following parameters
<name>yarn.nodemanager.vmen-pmen-ration</name>
<value>2.1</value>
What is the maximum amount of virtual memory allocated for each map task before YARN will kill its Container?

A. 4 GB
B. 17.2 GB
C. 8.9 GB
D. 8.2 GB
E. 24.6 GB

Answer: D

NEW QUESTION 25
You have recently converted your Hadoop cluster from a MapReduce 1 (MRv1) architecture to MapReduce 2 (MRv2) on YARN architecture. Your developers are accustomed to specifying map and reduce tasks (resource allocation) tasks when they run jobs: A developer wants to know how specify to reduce tasks when a specific job runs. Which method should you tell that developers to implement?

A. MapReduce version 2 (MRv2) on YARN abstracts resource allocation away from the idea of "tasks" into memory and virtual cores, thus eliminating the need for a developer to specify the number of reduce tasks, and indeed preventing the developer from specifying the number of reduce tasks.
B. In YARN, resource allocations is a function of megabytes of memory in multiples of 1024m
C. Thus, they should specify the amount of memory resource they need by executing –D mapreduce-reduces.memory-mb-2048
D. In YARN, the ApplicationMaster is responsible for requesting the resource required for a specific launc
E. Thus, executing –D yarn.applicationmaster.reduce.tasks=2 will specify that the ApplicationMaster launch two task contains on the worker nodes.

F. Developers specify reduce tasks in the exact same way for both MapReduce version 1 (MRv1) and MapReduce version 2 (MRv2) on YAR

G. Thus, executing –D mapreduce.job.reduces-2 will specify reduce tasks.

H. In YARN, resource allocation is function of virtual cores specified by the ApplicationManager making requests to the NodeManager where a reduce task is handeled by a single container (and thus a single virtual core). Thus, the developer needs to specify the number of virtual cores to the NodeManager by executing –p yarn.nodemanager.cpu-vcores=2

Answer: D

NEW QUESTION 27
Your cluster has the following characteristics:
? A rack aware topology is configured and on
? Replication is set to 3
? Cluster block size is set to 64MB
Which describes the file read process when a client application connects into the cluster and requests a 50MB file?

A. The client queries the NameNode for the locations of the block, and reads all three copie

B. The first copy to complete transfer to the client is the one the client reads as part of hadoop's speculative execution framework.

C. The client queries the NameNode for the locations of the block, and reads from the first location in the list it receives.

D. The client queries the NameNode for the locations of the block, and reads from a random location in the list it receives to eliminate network I/O loads by balancing which nodes it retrieves data from any given time.

E. The client queries the NameNode which retrieves the block from the nearest DataNode to the client then passes that block back to the client.

Answer: B

NEW QUESTION 32
You have a cluster running with a FIFO scheduler enabled. You submit a large job A to the cluster, which you expect to run for one hour. Then, you submit job B to the cluster, which you expect to run a couple of minutes only.
You submit both jobs with the same priority.
Which two best describes how FIFO Scheduler arbitrates the cluster resources for job and its tasks?(Choose two)

A. Because there is a more than a single job on the cluster, the FIFO Scheduler will enforce a limit on the percentage of resources allocated to a particular job at any given time

B. Tasks are scheduled on the order of their job submission

C. The order of execution of job may vary

D. Given job A and submitted in that order, all tasks from job A are guaranteed to finish before all tasks from job B

E. The FIFO Scheduler will give, on average, and equal share of the cluster resources over the job lifecycle

F. The FIFO Scheduler will pass an exception back to the client when Job B is submitted, since all slots on the cluster are use

Answer: AD

NEW QUESTION 35
Table schemas in Hive are:

A. Stored as metadata on the NameNode

B. Stored along with the data in HDFS

C. Stored in the Metadata

D. Stored in ZooKeeper

Answer: B

NEW QUESTION 38
What does CDH packaging do on install to facilitate Kerberos security setup?

A. Automatically configures permissions for log files at & MAPRED_LOG_DIR/userlogs

B. Creates users for hdfs and mapreduce to facilitate role assignment

C. Creates directories for temp, hdfs, and mapreduce with the correct permissions

D. Creates a set of pre-configured Kerberos keytab files and their permissions

E. Creates and configures your kdc with default cluster values

Answer: B

NEW QUESTION 39
You have a cluster running with the fair Scheduler enabled. There are currently no jobs running on the cluster, and you submit a job A, so that only job A is running on the cluster. A while later, you submit Job B. now Job A and Job B are running on the cluster at the same time. How will the Fair Scheduler handle these two jobs?(Choose two)

A. When Job B gets submitted, it will get assigned tasks, while job A continues to run with fewer tasks.

B. When Job B gets submitted, Job A has to finish first, before job B can gets scheduled.

C. When Job A gets submitted, it doesn't consumes all the task slots.

D. When Job A gets submitted, it consumes all the task slots.

Answer: B

NEW QUESTION 40
Your cluster is running MapReduce version 2 (MRv2) on YARN. Your ResourceManager is configured to use the FairScheduler. Now you want to configure your scheduler such that a new user on the cluster can submit jobs into their own queue application submission. Which configuration should you set?

A. You can specify new queue name when user submits a job and new queue can be created dynamically if the property yarn.scheduler.fair.allow-undecleared-pools = true
B. Yarn.scheduler.fair.user.fair-as-default-queue = false and yarn.scheduler.fair.allow- undecleared-pools = true
C. You can specify new queue name when user submits a job and new queue can be created dynamically if yarn .schedule.fair.user-as-default-queue = false
D. You can specify new queue name per application in allocations.xml file and have new jobs automatically assigned to the application queue

Answer: A

NEW QUESTION 45
During the execution of a MapReduce v2 (MRv2) job on YARN, where does the Mapper place the intermediate data of each Map Task?

A. The Mapper stores the intermediate data on the node running the Job's ApplicationMaster so that it is available to YARN 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)/apache/application_&(appid) directory for the user who ran the job
C. The Mapper transfers the intermediate data immediately to the reducers as it is generated by the Map Task
D. YARN holds the intermediate data in the NodeManager's memory (a container) until it is transferred to the Reducer
E. The Mapper stores the intermediate data on the underlying filesystem of the local disk in the directories yarn.nodemanager.locak-DIFS

Answer: E

NEW QUESTION 48
Your cluster implements HDFS High Availability (HA). Your two NameNodes are named nn01 and nn02. What occurs when you execute the command: hdfs haadmin –failover nn01 nn02?

A. nn02 is fenced, and nn01 becomes the active NameNode
B. nn01 is fenced, and nn02 becomes the active NameNode
C. nn01 becomes the standby NameNode and nn02 becomes the active NameNode
D. nn02 becomes the standby NameNode and nn01 becomes the active NameNode

Answer: B

Explanation: failover – initiate a failover between two NameNodes
This subcommand causes a failover from the first provided NameNode to the second. If the first
NameNode is in the Standby state, this command simply transitions the second to the Active statewithout error. If the first NameNode is in the Active state, an attempt will be made to gracefullytransition it to the Standby state. If this fails, the fencing methods (as configured bydfs.ha.fencing.methods) will be attempted in order until one of the methods succeeds. Only afterthis process will the second NameNode be transitioned to the Active state. If no fencing methodsucceeds, the second NameNode will not be transitioned to the Active state, and an error will bereturned.

NEW QUESTION 53
You want to node to only swap Hadoop daemon data from RAM to disk when absolutely necessary. What should you do?

A. Delete the /dev/vmswap file on the node
B. Delete the /etc/swap file on the node
C. Set the ram.swap parameter to 0 in core-site.xml
D. Set vm.swapfile file on the node
E. Delete the /swapfile file on the node

Answer: D

NEW QUESTION 57
You're upgrading a Hadoop cluster from HDFS and MapReduce version 1 (MRv1) to one running HDFS and MapReduce version 2 (MRv2) on YARN. You want to set and enforce version 1 (MRv1) to one running HDFS and MapReduce version 2 (MRv2) on YARN. You want to set and enforce a block size of 128MB for all new files written to the cluster after upgrade. What should you do?

A. You cannot enforce this, since client code can always override this value
B. Set dfs.block.size to 128M on all the worker nodes, on all client machines, and on the NameNode, and set the parameter to final
C. Set dfs.block.size to 128 M on all the worker nodes and client machines, and set the parameter to fina
D. You do not need to set this value on the NameNode
E. Set dfs.block.size to 134217728 on all the worker nodes, on all client machines, and on the NameNode, and set the parameter to final
F. Set dfs.block.size to 134217728 on all the worker nodes and client machines, and set the parameter to fina
G. You do not need to set this value on the NameNode

Answer: C

NEW QUESTION 60
A user comes to you, complaining that when she attempts to submit a Hadoop job, it fails. There is a Directory in HDFS named /data/input. The Jar is named j.jar, and the driver class is named DriverClass.
She runs the command:
Hadoop jar j.jar DriverClass /data/input/data/output The error message returned includes the line:
PrivilegedActionException as:training (auth:SIMPLE) cause:org.apache.hadoop.mapreduce.lib.input.invalidInputException:
Input path does not exist: file:/data/input What is the cause of the error?

A. The user is not authorized to run the job on the cluster
B. The output directory already exists
C. The name of the driver has been spelled incorrectly on the command line
D. The directory name is misspelled in HDFS

E. The Hadoop configuration files on the client do not point to the cluster

Answer: A

NEW QUESTION 61
Which two are features of Hadoop's rack topology?(Choose two)

A. Configuration of rack awareness is accomplished using a configuration fil
B. You cannot use a rack topology script.
C. Hadoop gives preference to intra-rack data transfer in order to conserve bandwidth
D. Rack location is considered in the HDFS block placement policy
E. HDFS is rack aware but MapReduce daemon are not
F. Even for small clusters on a single rack, configuring rack awareness will improve performance

Answer: BC

NEW QUESTION 65
Which YARN daemon or service negotiations map and reduce Containers from the Scheduler, tracking their status and monitoring progress?

A. NodeManager
B. ApplicationMaster
C. ApplicationManager
D. ResourceManager

Answer: B

Explanation: Reference:http://www.devx.com/opensource/intro-to-apache-mapreduce-2-yarn.html(See resource manager)

NEW QUESTION 70
You have installed a cluster HDFS and MapReduce version 2 (MRv2) on YARN. You have no dfs.hosts entry(ies) in your hdfs-site.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 sorting HDFS blocks?

A. Without creating a dfs.hosts file or making any entries, run the commands hadoop.dfsadmin-refreshModes on the NameNode
B. Restart the NameNode
C. Creating a dfs.hosts file on the NameNode, add the worker Node's name to it, then issue the command hadoop dfsadmin –refresh Nodes = on the Namenode
D. Nothing; the worker node will automatically join the cluster when NameNode daemon is started

Answer: A

NEW QUESTION 73
Which process instantiates user code, and executes map and reduce tasks on a cluster running MapReduce v2 (MRv2) on YARN?

A. NodeManager
B. ApplicationMaster
C. TaskTracker
D. JobTracker
E. NameNode
F. DataNode
G. ResourceManager

Answer: A

NEW QUESTION 75
......

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questons and Answers in PDF Format

CCA-500 Practice Exam Features:

* CCA-500 Questions and Answers Updated Frequently

* CCA-500 Practice Questions Verified by Expert Senior Certified Staff

* CCA-500 Most Realistic Questions that Guarantee you a Pass on Your FirstTry

* CCA-500 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

100% Actual & Verified — Instant Download, Please Click
Order The CCA-500 Practice Test Here