2passeasy

Exam Questions CCA-500

Cloudera Certified Administrator for Apache Hadoop (CCAH)

https://www.2passeasy.com/dumps/CCA-500/

NEW QUESTION 1
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 2
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 3
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 4
You have just run a MapReduce job to filter user messages to only those of a selected geographical region. The output for this job is in a directory named westUsers, located just below your home directory in HDFS. Which command gathers these into a single file on your local file system?

A. Hadoop fs –getmerge –R westUsers.txt
B. Hadoop fs –getemerge westUsers westUsers.txt
C. Hadoop fs –cp westUsers/* westUsers.txt
D. Hadoop fs –get westUsers westUsers.txt

Answer: B

NEW QUESTION 5
Which three basic configuration parameters must you set to migrate your cluster from MapReduce 1 (MRv1) to MapReduce V2 (MRv2)?(Choose three)

A. Configure the NodeManager to enable MapReduce services on YARN by setting the following property in yarn-
site.xml:<name>yarn.nodemanager.hostname</name><value>your_nodeManager_shuffle</value>
B. Configure the NodeManager hostname and enable node services on YARN by setting the following property in yarn-
site.xml:<name>yarn.nodemanager.hostname</name><value>your_nodeManager_hostname</value>
C. Configure a default scheduler to run on YARN by setting the following property in mapred-
site.xml:<name>mapreduce.jobtracker.taskScheduler</name><Value>org.apache.hadoop.mapred.JobQueueTaskScheduler</value>
D. Configure the number of map tasks per jon YARN by setting the following property in mapred:<name>mapreduce.job.maps</name><value>2</value>
E. Configure the ResourceManager hostname and enable node services on YARN by setting the following property in yarn-
site.xml:<name>yarn.resourcemanager.hostname</name><value>your_resourceManager_hostname</value>
F. Configure MapReduce as a Framework running on YARN by setting the following property in mapred-
site.xml:<name>mapreduce.framework.name</name><value>yarn</value>

Answer: AEF

NEW QUESTION 6
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 7

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 8
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. Set an HDFS replication factor that provides data redundancy, protecting against node failure
C. Run a Secondary NameNode on a different master from the NameNode in order to provide automatic recovery from a NameNode failure.
D. Run the ResourceManager on a different master from the NameNode in order to load- share HDFS metadata processing
E. Configure the cluster's disk drives with an appropriate fault tolerant RAID level

Answer: D

NEW QUESTION 9
On a cluster running MapReduce v2 (MRv2) on YARN, a MapReduce job is given a directory of 10 plain text files as its input directory. Each file is made up of 3 HDFS blocks. How many Mappers will run?

A. We cannot say; the number of Mappers is determined by the ResourceManager
B. We cannot say; the number of Mappers is determined by the developer
C. 30
D. 3
E. 10
F. We cannot say; the number of mappers is determined by the ApplicationMaster

Answer: E

NEW QUESTION 10
You are configuring your cluster to run HDFS and MapReducer v2 (MRv2) on YARN. Which two daemons needs to be installed on your cluster's master nodes?(Choose two)

A. HMaster
B. ResourceManager
C. TaskManager
D. JobTracker
E. NameNode
F. DataNode

Answer: BE

NEW QUESTION 10
Which YARN daemon or service monitors a Controller's per-application resource using (e.g., memory CPU)?

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

Answer: A

NEW QUESTION 15
You want to understand more about how users browse your public website. For example, you want to know which pages they visit prior to placing an order. You have a server farm of 200 web servers hosting your website. Which is the most efficient process to gather these web server across logs into your Hadoop cluster analysis?

A. Sample the web server logs web servers and copy them into HDFS using curl
B. Ingest the server web logs into HDFS using Flume
C. Channel these clickstreams into Hadoop using Hadoop Streaming
D. Import all user clicks from your OLTP databases into Hadoop using Sqoop
E. Write a MapReeeduce job with the web servers for mappers and the Hadoop cluster nodes for reducers

Answer: B

Explanation: Apache Flume is a service for streaming logs into Hadoop.
Apache Flume is a distributed, reliable, and available service for efficiently collecting, aggregating, and moving large amounts of streaming data into the Hadoop Distributed File System (HDFS). It has a simple and flexible architecture based on streaming data flows; and is robust and fault tolerant with tunable reliability

mechanisms for failover and recovery.

NEW QUESTION 20
Cluster Summary:
45 files and directories, 12 blocks = 57 total. Heap size is 15.31 MB/193.38MB(7%)

Configured capacity	:	17.33GB
DFS Used	:	144KB
Non DFS Used	:	5.49GB
DFS Remaining	:	11.84GB
DFS Used %	:	0%
DFS Remaining %	:	68.32GB
Live Nodes	:	6
Dead Nodes	:	1
Decommissioning Nodes	:	0
Number of Under-Replicated Blocks	:	6

Refer to the above screenshot.
You configure a Hadoop cluster with seven DataNodes and on of your monitoring UIs displays the details shown in the exhibit.
What does the this tell you?

A. The DataNode JVM on one host is not active
B. Because your under-replicated blocks count matches the Live Nodes, one node is dead, and your DFS Used % equals 0%, you can't be certain that your cluster has all the data you've written it.
C. Your cluster has lost all HDFS data which had bocks stored on the dead DatNode
D. The HDFS cluster is in safe mode

Answer: A

NEW QUESTION 21
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 22
On a cluster running CDH 5.0 or above, you use the hadoop fs –put command to write a 300MB file into a previously empty directory using an HDFS block size of 64 MB. Just after this command has finished writing 200 MB of this file, what would another use see when they look in directory?

A. The directory will appear to be empty until the entire file write is completed on the cluster
B. They will see the file with a ._COPYING_ extension on its nam
C. If they view the file, they will see contents of the file up to the last completed block (as each 64MB block is written, that block becomes available)
D. They will see the file with a ._COPYING_ extension on its nam
E. If they attempt to view the file, they will get a ConcurrentFileAccessException until the entire file write is completed on the cluster
F. They will see the file with its original nam
G. If they attempt to view the file, they will get a ConcurrentFileAccessException until the entire file write is completed on the cluster

Answer: B

NEW QUESTION 25
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 26
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 28
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 31
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 33
Assuming you're not running HDFS Federation, what is the maximum number of NameNode daemons you should run on your cluster in order to avoid a "split-brain" scenario with your NameNode when running HDFS High Availability (HA) using Quorum- based storage?

A. Two active NameNodes and two Standby NameNodes
B. One active NameNode and one Standby NameNode
C. Two active NameNodes and on Standby NameNode
D. Unlimite
E. HDFS High Availability (HA) is designed to overcome limitations on the number of NameNodes you can deploy

Answer: B

NEW QUESTION 35
In CDH4 and later, which file contains a serialized form of all the directory and files inodes in the filesystem, giving the NameNode a persistent checkpoint of the filesystem metadata?

A. fstime
B. VERSION

C. Fsimage_N (where N reflects transactions up to transaction ID N)
D. Edits_N-M (where N-M transactions between transaction ID N and transaction ID N)

Answer: C

Explanation: Reference:http://mikepluta.com/tag/namenode/

NEW QUESTION 37
Which command does Hadoop offer to discover missing or corrupt HDFS data?

A. Hdfs fs –du
B. Hdfs fsck
C. Dskchk
D. The map-only checksum
E. Hadoop does not provide any tools to discover missing or corrupt data; there is not need because three replicas are kept for each data block

Answer: B

Explanation: Reference:https://twiki.grid.iu.edu/bin/view/Storage/HadoopRecovery

NEW QUESTION 40
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 41
You are running a Hadoop cluster with a NameNode on host mynamenode. What are two ways to determine available HDFS space in your cluster?

A. Run hdfs fs –du / and locate the DFS Remaining value
B. Run hdfs dfsadmin –report and locate the DFS Remaining value
C. Run hdfs dfs / and subtract NDFS Used from configured Capacity
D. Connect to http://mynamenode:50070/dfshealth.jsp and locate the DFS remaining value

Answer: B

NEW QUESTION 43
Which YARN process run as "container 0" of a submitted job and is responsible for resource qrequests?

A. ApplicationManager
B. JobTracker
C. ApplicationMaster
D. JobHistoryServer
E. ResoureManager
F. NodeManager

Answer: C

NEW QUESTION 46
You use 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 replicationfactor of 3). One of the nodes holding this file (a single block) fails. How will the cluster handle the replication of file in this situation?

A. The file will remain under-replicated until the administrator brings that node back online
B. The cluster will re-replicate the file the next time the system administrator reboots the NameNode daemon (as long as the file's replication factor doesn't fall below)
C. This will be immediately re-replicated and all other HDFS operations on the cluster will halt until the cluster's replication values are resorted
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 NameNodes

Answer: D

NEW QUESTION 50
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 55
Choose three reasons why should you run the HDFS balancer periodically?(Choose three)

A. To ensure that there is capacity in HDFS for additional data
B. To ensure that all blocks in the cluster are 128MB in size
C. To help HDFS deliver consistent performance under heavy loads
D. To ensure that there is consistent disk utilization across the DataNodes
E. To improve data locality MapReduce

Answer: CDE

Explanation: http://www.quora.com/Apache-Hadoop/It-is-recommended-that-you-run-the-HDFS-balancer-periodically-Why-Choose-3

NEW QUESTION 60
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 63
Your Hadoop cluster is configuring with HDFS and MapReduce version 2 (MRv2) on YARN. Can you configure a worker node to run a NodeManager daemon but not a DataNode daemon and still have a functional cluster?

A. Ye
B. The daemon will receive data from the NameNode to run Map tasks
C. Ye
D. The daemon will get data from another (non-local) DataNode to run Map tasks
E. Ye
F. The daemon will receive Map tasks only
G. Ye
H. The daemon will receive Reducer tasks only

Answer: B

NEW QUESTION 65
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 66
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 70

You suspect that your NameNode is incorrectly configured, and is swapping memory to disk. Which Linux commands help you to identify whether swapping is occurring?(Select all that apply)

A. free
B. df
C. memcat
D. top
E. jps
F. vmstat
G. swapinfo

Answer: ADF

Explanation: Reference:http://www.cyberciti.biz/faq/linux-check-swap-usage-command/

NEW QUESTION 71

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 73
......

THANKS FOR TRYING THE DEMO OF OUR PRODUCT

Visit Our Site to Purchase the Full Set of Actual CCA-500 Exam Questions With Answers.

We Also Provide Practice Exam Software That Simulates Real Exam Environment And Has Many Self-Assessment Features. Order the CCA-500 Product From:

https://www.2passeasy.com/dumps/CCA-500/

Money Back Guarantee

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