

Exam Questions CCA-500

Cloudera Certified Administrator for Apache Hadoop (CCA-H)

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



NEW QUESTION 1

You are migrating a cluster from MAppReduce 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` in `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 of 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

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 6

Which two features does Kerberos security add to a Hadoop cluster?(Choose two)

- A. User authentication on all remote procedure calls (RPCs)
- B. Encryption for data during transfer between the Mappers and Reducers
- C. Encryption for data on disk ("at rest")
- D. Authentication for user access to the cluster against a central server
- E. Root access to the cluster for users `hdfs` and `mapred` but non-root access for clients

Answer: AD

NEW QUESTION 7

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 8

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 MapReeeduc 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 9

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 10

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

NEW QUESTION 10

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 the third command?

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

Answer: CE

NEW QUESTION 12

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 17

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

NEW QUESTION 20

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-vcores</name>
- D. No action is needed: YARN's dynamic resource allocation automatically optimizes the node memory and cores

Answer: A

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 27

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 29

For each YARN job, the Hadoop framework generates task log file. Where are Hadoop task log files stored?

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

Answer: D

NEW QUESTION 33

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 38

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.vmem-pmem-ratio</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 43

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 44

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 46

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 51

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 54

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 57

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 58

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 60

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 64

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 69

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.local-DIFS

Answer: E

NEW QUESTION 74

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 78

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 79

You are running Hadoop cluster with all monitoring facilities properly configured. Which scenario will go undeselected?

- A. HDFS is almost full
- B. The NameNode goes down
- C. A DataNode is disconnected from the cluster
- D. Map or reduce tasks that are stuck in an infinite loop
- E. MapReduce jobs are causing excessive memory swaps

Answer: B

NEW QUESTION 84

You need to analyze 60,000,000 images stored in JPEG format, each of which is approximately 25 KB. Because your Hadoop cluster isn't optimized for storing and processing many small files, you decide to do the following actions:

1. Group the individual images into a set of larger files
2. Use the set of larger files as input for a MapReduce job that processes them directly with python using Hadoop streaming.

Which data serialization system gives the flexibility to do this?

- A. CSV
- B. XML
- C. HTML
- D. Avro
- E. SequenceFiles
- F. JSON

Answer: E

Explanation: Sequence files are block-compressed and provide direct serialization and deserialization of several arbitrary data types (not just text). Sequence files can be generated as the output of other MapReduce tasks and are an efficient intermediate representation for data that is passing from one MapReduce job to another.

NEW QUESTION 87

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 90

.....

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