Cloudera

CDP-ADMIN-PVC

Cloudera Administrator on premises Certification Exam

Latest Version: 6.0

Question: 1

Which of the following components are essential for the functioning of the Cloudera Data Platform (CDP) on-premises Data Lake, focusing on data storage and metadata management?

A. HDFS (Hadoop Distributed File System) for data storage.
B. Hive Metastore for metadata management.
C. YARN (Yet Another Resource Negotiator) for resource management.
D. ZooKeeper for coordinating distributed processes.
E. Kafka for real-time data streaming.

Answer: A,B,C,D

Explanation:
HDFS provides distributed storage, Hive Metastore manages metadata, YARN manages resources and ZooKeeper coordinates distributed processes are all essential for a functioning Data Lake. Kafka is more related to streaming ingest rather than core Data Lake functionality, though it may be part of the overall architecture. The question specifically asks about Data Lake functioning.

Question: 2

You are configuring resource pools in YARN for different departments in your organization. The marketing department requires a guaranteed minimum of 40% of the cluster's resources, while the engineering department requires at least 50%. What is the most appropriate configuration for the 'yarn.scheduler.capacity.maximum-am-resource-percent' property in 'yarn-site.xmP , considering efficient resource utilization and potentially variable workloads?

A. Set 'yarn.scheduler.capacity.maximum-am-resource-percent' to 0.1 (10%).
B. Set 'yarn.scheduler.capacity.maximum-am-resource-percent' to 0.4 (40%).
C. Set 'yarn.scheduler.capacity.maximum-am-resource-percent' to 0.5 (50%).
D. Set 'yarn.scheduler.capacity.maximum-am-resource-percent' to 0.9 (90%).
E. Set 'yarn.scheduler.capacity.maximum-am-resource-percent' to 1.0 (100%).

Answer: D

Explanation:
'yarn.scheduler.capacity.maximum-am-resource-percent' defines the maximum percentage of resources in the cluster that can be used for ApplicationMasters (AMs). While marketing and engineering require 40% and 50% respectively of overall cluster resources, the AMs need resources too. The maximum usage percentage should allow enough space for the Application Master processes to start for a resource intensive job. Setting this to a lower value might lead to resource starvation for the application masters.

Question: 3

A critical Hive query that performs complex aggregations on a large dataset is consistently slow. You suspect the issue is related to suboptimal data partitioning and bucketing. How can you re-design the Hive table definition to improve query performance?

A. Remove partitioning and bucketing altogether to simplify the table structure.
B. Partition the table by a low-cardinality column and bucket it by a high-cardinality column frequently used in 'JOIN' or 'GROUP BY operations.
C. Partition the table by a high-cardinality column and bucket it by a low-cardinality column.
D. Use dynamic partitioning with a very large number of partitions.
E. Bucket the table only, with a very small number of buckets.

Answer: B

Explanation:
Partitioning by a low-cardinality column reduces the amount of data scanned for queries filtering on that column. Bucketing by a high- cardinality column frequently used in 'JOIN' or 'GROUP BY operations distributes the data evenly across buckets, allowing for parallel processing and improved join performance (bucket map joins). Using high cardinality column as partition would result in large number of small files and low cardinality column as bucket will result in data skewness. Dynamic partitioning is useful but can be inefficient if not managed properly.

Question: 4

Which of the following statements accurately describe the role of the Cloudera Manager Agent in a CDP on-premises cluster?

A. The Cloudera Manager Agent is responsible for distributing the Cloudera Manager Server software across the cluster.
B. The Cloudera Manager Agent collects metrics and status information from the services running on its host and reports them to the Cloudera Manager Server.
C. The Cloudera Manager Agent executes commands issued by the Cloudera Manager Server, such as starting, stopping, and reconfiguring services.
D. The Cloudera Manager Agent is only required on the Cloudera Manager Server host.
E. The Cloudera Manager Agent automatically optimizes query performance without direct intervention.

Answer: B,C

Explanation:
Cloudera Manager Agents monitor the services on each host, reporting back to the Cloudera Manager Server, and execute commands issued by the Server. They don't distribute the Server software, and are needed on all cluster hosts, not just the Server host. They don't directly optimize query performance without configuration.

Question: 5

You are configuring Kerberos authentication for your CDP on-premises cluster After enabling Kerberos, you notice that some of your existing Spark applications are failing with authentication errors. What are the likely causes and how do you resolve them?

A. The Spark application is not configured to use Kerberos credentials. You need to configure the ‘spark.yarn.principar and ‘spark.yarn.keytab' properties in the Spark configuration.
B. The Kerberos ticket-granting ticket (TGT) has expired on the client machine. You need to run 'kinit' to
C. The necessary Java Cryptography Extension (JCE) Unlimited Strength Jurisdiction Policy Files are not installed on the Spark driver and executor nodes.
D. The clock skew between the Kerberos Key Distribution Center (KDC) and the Spark nodes is too large. You need to synchronize the clocks using NTP (Network Time Protocol).
E. Spark applications automatically inherit Kerberos authentication settings from the operating system; no specific configuration is required.

Answer: A,B,C,D

Explanation:
Spark applications need to be explicitly configured to use Kerberos with correct principal and keytab. Expired TGTs prevent authentication. JCE policy files are required for strong encryption used by Kerberos. Clock skew can cause authentication failures. So all options except E are correct.

Question: 6

Your Cloudera Data Platform (CDP) on-premises cluster is experiencing frequent HDFS NameNode failures, leading to data unavailability. You want to implement High Availability (HA) for the NameNode. What steps are essential to configure NameNode HA using Quorum Journal Manager (QJM)?

A. Install and configure a single ZooKeeper server.
B. Configure two NameNode hosts and share a single edit log directory between them.
C. Install and configure a quorum of JournalNodes (typically 3 or more) to store the edit log.
D. Configure the 'dfs.nameservices’ and 'dfs.ha.namenodes.[nameserviceId]' properties in 'hdfs-site.xmr.
E. Format the NameNode using the '-format' option on both NameNode hosts.

Answer: C,D

Explanation:
QJM requires a quorum of JournalNodes to store the edit log redundantly. You also need to configure 'dfs.nameserviceS and dfs.ha.namenodes.[nameserviceId]' to enable HA. A single ZooKeeper server is not enough for HA. NameNodes should not share a single edit log directory. Formatting with -format should be done only on the initial NameNode setup and not on both after configuring HA.

Question: 7

Assume you are facing issues related to Parcel distribution. The below Python script is intended to download parcel files. Find out which of the below code Snippet will resolve the problem if CM API call is failing due to invalid credentials, connection refused errors or invalid URL etc. Assume other code is correct. In other words provide code to handle exceptions

A.

```
try:
  response = requests.get(url, stream=True)
  response.raise_for_status()
  with open(parcel_file, 'wb') as f:
    for chunk in response.iter_content(chunk_size=8192):
      f.write(chunk)
except requests.exceptions.RequestException as e:
  print(f"Download failed: {e}")
  return False
```

B.

```
response = requests.get(url, stream=True)
with open(parcel_file, 'wb') as f:
  for chunk in response.iter_content(chunk_size=8192):
    f.write(chunk)
```

C.

```
try:
  response = requests.get(url, stream=True)
  with open(parcel_file, 'wb') as f:
    for chunk in response.iter_content(chunk_size=8192):
      f.write(chunk)
except Exception as e:
  print(f"Download failed: {e}")
  return False
```

D.

```
try:
  response = requests.get(url, stream=True)
  with open(parcel_file, 'wb') as f:
    for chunk in response.iter_content(chunk_size=8192):
      f.write(chunk)
f.close()
```

E.

```
try:
  response = requests.get(url, stream=True, auth=('username', 'password'), verify=False)
  response.raise_for_status()
  with open(parcel_file, 'wb') as f:
    for chunk in response.iter_content(chunk_size=8192):
      f.write(chunk)
except requests.exceptions.RequestException as e:
  print(f"Download failed: {e}")
  return False
```

Answer: A,E

Explanation:
Snippet A, E incorporates robust error handling using a 'try...except' block to catch 'requests.exceptions.RequestException' , which covers various potential issues, including connection errors, invalid URLs, and HTTP errors. Additionally, using 'response.raise_for_status()' is to make sure that the URL exists, if not its thrown exception.

Question: 8

You are deploying a CDP Private Cloud Base cluster with Kerberos enabled. After the initial deployment, you notice that some Hadoop services are failing to start, with errors related to keytab files. What are the MOST likely causes of this issue? (Select TWO)

A. The keytab files were not properly distributed to all nodes in the cluster.
B. The Kerberos principal used for the service does not exist in the KDC.
C. The clocks on the Hadoop nodes are not synchronized with the KDC server.
D. Incorrect permissions are set on the keytab files, preventing the service from reading them.
E. The Java Cryptography Extension (JCE) Unlimited Strength Jurisdiction Policy Files are not installed.

Answer: A,D

Explanation:
Keytab files are essential for Kerberos authentication. If they are not distributed correctly, or the service doesn't have the correct permissions to read them, services will fail to start. Clock skew is also a common Kerberos issue, but is not directly related to keytab distribution. While JCE can be a cause of other Kerberos issues, its is not the first cause to investigate in this case. Ensuring that Kerberos principals exist in the KDC is part of pre-deployment setup.

Question: 9

During the installation of CDP Private Cloud Base, you encounter an error in Cloudera Manager related to network connectivity between the Cloudera Manager Server and the hosts. Specifically, you receive a message indicating that the 'Ping Test' failed. What is the MOST likely root cause of this issue?

A. The Cloudera Manager Agent is not installed on the target hosts.
B. Firewall rules are blocking ICMP traffic between the Cloudera Manager Server and the hosts.
C. The /etc/hosts file is not correctly configured on either the Cloudera Manager Server or the hosts.
D. There is a DNS resolution problem, preventing the Cloudera Manager Server from resolving the hostnames of the target hosts.
E. Insufficient memory is allocated to the Cloudera Manager Server.

Answer: B

Explanation:
The 'Ping Test' in Cloudera Manager specifically checks for basic network connectivity using ICMP. A firewall blocking ICMP is the most direct and common reason for this test to fail. While DNS resolution issues and incorrect /etc/hosts entries can cause connectivity problems, the 'Ping Test' is primarily reliant on ICMP, and firewall settings are more likely cause a Ping test failure directly.

Question: 10

You're tasked with automating the creation of a new Cloudera Manager user with specific roles and permissions using the REST API. You need to create a user 'data_analyst' with 'Read-Only' access to the 'MyCluster' cluster and 'Operator' access to the 'Hive' service. Which of the following API call(s) and body structures would accomplish this? Note that a single user creation might require multiple API calls to set all the desired permissions.

- ○ 1. POST to `/api/v32/users` with body: `{"name": "data_analyst", "password": "password123", "roles": [{"name": "ReadOnly", "scope": "CLUSTER", "entityName": "MyCluster"}, {"name": "Operator", "scope": "SERVICE", "entityName": "Hive"}]}`
- ○ 1. POST to `/api/v32/users` with body: `{"name": "data_analyst", "password": "password123"}` 2. POST to `/api/v32/users/data_analyst/roles` with body: `{"name": "ReadOnly", "scope": "CLUSTER", "entityName": "MyCluster"}` 3. POST to `/api/v32/users/data_analyst/roles` with body: `{"name": "Operator", "scope": "SERVICE", "entityName": "Hive"}`
- ○ 1. POST to `/api/v32/users` with body: `{"name": "data_analyst", "password": "password123"}`. 2. PUT to `/api/v32/clusters/MyCluster/users/data_analyst/roles` with body: `{"roleName": "ReadOnly"}` 3. PUT to `/api/v32/services/Hive/users/data_analyst/roles` with body: `{"roleName": "Operator"}`
- ○ 1. POST to `/api/v32/users` with body: `{"name": "data_analyst", "password": "password123"}`. 2. POST to `/api/v32/clusters/MyCluster/users/data_analyst/roles` with body: `{"roleName": "ReadOnly"}` 3. POST to `/api/v32/services/Hive/users/data_analyst/roles` with body: `{"roleName": "Operator"}`
- ○ User creation and Role Assignment are only possible through UI.

A. Option A
B. Option B
C. Option C
D. Option D
E. Option E

Answer: B

Explanation:
The correct approach involves first creating the user using `POST` to `/api/v32/users`. Then, you need to make separate `POST` requests to `/api/v32/users/{username}/roles` to assign each role individually. The request body should specify the role name, scope (CLUSTER or SERVICE), and the entity name (cluster or service name). Role assignments require separate calls per role.

Thank You for Trying Our Product

For More Information – Visit link below:

https://www.examsboost.com/

15 USD Discount Coupon Code:

G74JA8UF

FEATURES

- ✓ 90 Days Free Updates
- ✓ Money Back Pass Guarantee
- ✓ Instant Download or Email Attachment
- ✓ 24/7 Live Chat Support
- ✓ PDF file could be used at any Platform
- ✓ 50,000 Happy Customer