

Cloudera

CDP-0011

CDP Generalist Exam



For More Information – Visit link below:

<https://www.examsboost.com/>

Product Version

- ✓ Up to Date products, reliable and verified.
- ✓ Questions and Answers in PDF Format.

Latest Version: 6.3

Question: 1

A CDP Private Cloud Base cluster utilizes HDFS High Availability (HA). If the Active NameNode fails, the Standby NameNode must quickly become Active. Which component ensures the Standby NameNode has the absolute most recent transaction log entries (edits) necessary for a seamless failover?

- A. The DataNode, by sending block reports and edit logs directly to the Standby NameNode.
- B. The JournalNodes, which maintain a quorum-based log of namespace modifications accessible by both NameNodes.
- C. The Secondary NameNode, which periodically merges the active NameNode's edits into a new fsimage.
- D. The Failover Controller, which manages ZooKeeper znodes used for leadership election and stores the latest checkpoint.
- E. The Resource Manager, which provides application resource allocation and ensures HDFS services remain running.

Answer: B

Explanation:

In an HDFS HA configuration, the JournalNodes (JNs) are critical. HA is implemented using quorum-based storage, where a set of JournalNodes maintains a local edits directory that logs all modifications to the namespace metadata. The Active NameNode writes these changes to a majority of JNs, and the Standby NameNode reads these changes, ensuring it is up-to-date for a fast failover.

Question: 2

An administrator successfully registers a CDP environment and subsequently provisions a Data Hub cluster dedicated to running Streaming Analytics. Which core CDP component is responsible for providing the unified security policies (via Ranger) and governance (via Atlas) foundation for this new Data Hub cluster?

- A. The Cloudera Management Console (CMC), as the centralized control plane.
- B. The Cloudera Data Hub (CDH), as it hosts the workload services.
- C. The Data Lake, which creates a protective ring of security and governance around the data.
- D. The FreeIPA server, which manages Kerberos principals and synchronization.
- E. The Operational Database with SQL cluster deployed in the same environment.

Answer: C

Explanation:

A Data Lake is automatically provisioned when an environment is created. It creates a protective ring of security and governance around data and hosts shared security and governance services, such as

Apache Ranger and Apache Atlas. All Data Hub clusters created in that environment are attached to the Data Lake and rely on these shared services.

Question: 3

A critical DataNode in a healthy HDFS cluster (default replication factor 3) stops sending heartbeats to the NameNode. If the NameNode determines the DataNode is 'dead,' what action does the NameNode immediately prioritize to ensure data durability and consistency?

- A. It performs an immediate rolling restart of the HDFS service to force the dead DataNode to rejoin.
- B. It automatically deletes the records associated with the failed DataNode from the fsimage and edit logs.
- C. It recalculates the overall cluster disk usage using the HDFS Balancer utility.
- D. It instructs other DataNodes holding replicas of the affected blocks to copy them to new DataNodes to restore the configured replication factor.
- E. It initiates the Checkpointing process to merge the current edits log into a new fsimage before attempting recovery.

Answer: D

Explanation:

When a DataNode is considered dead, the NameNode detects which blocks were on the failed DataNode. To maintain the configured replication factor (3x by default), the NameNode instructs other DataNodes possessing replicas of those blocks to copy them to new DataNodes, thus ensuring recovery from the loss of data copies.

Question: 4

In a Virtual Private Cluster (VPC) architecture, a Compute Cluster running Spark workloads needs to access data stored in HDFS and retrieve authorization policies from Ranger, both residing on the Base Cluster. Which architectural construct facilitates this shared access to data, metadata, and security services?

- A. The Cloudbreak microservice, which manages cluster provisioning.
- B. The Shared Data Context, acting as a connector to the Base cluster's data, metadata, and security services.
- C. The Failover Controller, ensuring continuous connectivity to the NameNode.
- D. The CDP Control Plane, which handles all cross-cluster data transfer.
- E. The embedded Kubernetes Cluster, managing the Spark containerized execution.

Answer: B

Explanation:

A Compute cluster accesses data and related governance/security services by connecting to a Shared Data Context (Data Context) associated with the Base cluster. This context defines the data, metadata, and security services deployed in the Base cluster that are required for access.

Question: 5

A data engineer is optimizing an existing HDFS cluster (default replication factor 3x, block size 128 MB). Which two scenarios require modifying either the replication factor or the block size from their default values?

- A. You are performing a benchmark test using TeraGen to measure the raw disk I/O performance baseline, avoiding network bandwidth constraints. `-Ddfs.replication=1`
- B. The cluster runs exclusively CPU-heavy data processing that requires minimal network transfer latency.
- C. A small subset of files is frequently accessed by many users, leading to observed resource bottlenecks on the DataNodes hosting those blocks.
- D. You are enabling HDFS Transparent Data Encryption (TDE) for existing Hive tables.
- E. You need to perform a NameNode rollback after a failed major cluster upgrade.

Answer: A,C

Explanation:

Option A: When obtaining a performance baseline using tools like TeraGen, the replication factor should be overridden and set to 1 (`Ddfs.replication=1`) to prevent data replication over the network from obscuring raw disk performance. Option C: If bottlenecks occur on a small number of nodes due to heavy access to specific files, increasing the replication factor of those files allows blocks to be spread over more nodes, thus alleviating the load.

Question: 6

A Cloudera DataFlow data stream uses the FetchHDFS processor to read data from Amazon S3 via the HDFS compatibility layer, leveraging Cloudera's IDBroker authentication method. Which special parameter must be set as the value for the 'Hadoop configuration files' property within the processor to automatically retrieve necessary configuration (e.g., `core-site.xml`) during deployment?

- A. `#{AWSIDBrokerCloudCredentialsProvider}`
- B. `#{SchemaRegistryHost}`
- C. `#{IDBrokerConfig}`
- D. `#{CDPEnvironment}`
- E. `#{ClusterTemplateCrn}`

Answer: D

Explanation:

When using HDFS processors (like FetchHDFS) in Cloudera DataFlow to access cloud object stores (like S3), they require Hadoop configuration files (such as `core-site.xml`). By using the special parameter for the configuration file property, Cloudera DataFlow `#{CDPEnvironment}` automatically obtains the required files during flow deployment, enabling the use of IDBroker authentication to access S3/ABFS.

Question: 7

A cluster administrator is performing a backup of HDFS NameNode metadata in a non-HA cluster. The NameNode process reads this metadata upon startup to load the namespace into memory. Which two file components found within the NameNode data directory (dfs. namenode. name. dir) represent the persistent record of the HDFS file system state and all subsequent modifications?

- ☐ The **VERSION** file, containing the NameNode's **storageType** and **clusterID**.
- ☐ The **fsimage_*** file, which contains the complete state of the file system at a specific point in time.
- ☐ The **datanode_blockpool_mapping.xml**.
- ☐ The **edits_start-end** file(s), which log every file system change since the most recent **fsimage**.
- ☐ The **seen_txid** file, which confirms the transaction ID of the last checkpoint or edit log roll.

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

Answer: B,D

Explanation:

HDFS metadata persistence relies on two main categories of files: the file, which contains the complete state of the file fsimage system at a point in time up to a specific transaction ID; and the (or segments), which contain a log of every file system change edits file (creation, deletion, modification) that occurred after the most recent fsimage. Both must be backed up to ensure the complete state is restorable.

Question: 8

A data engineer is configuring a Kafka Connect job to sink data directly into a specific volume and bucket within Ozone. Based on the configuration requirements to replace generic HDFS paths, which path scheme must be used for the 'hdfs.uri' and 'hdfs.output' properties to ensure valid Ozone access?

- ☐ **hdfs://<ozone_server>/<vol>/<buck>/<path>**
- ☐ **s3a://<ozone_server>/<vol>/<buck>/<path>**
- ☐ **ofs://<ozone_server>/<vol>/<buck>/<path>**
- ☐ **o3fs://<ozone_server>/<vol>/<buck>/<path>**
- ☐ **webhdfs://<ozone_server>/<vol>/<buck>/<path>**

- A. Option A

- B. Option B
- C. Option C
- D. Option D
- E. Option E

Answer: C

Explanation:

Configurations for Kafka Connect targeting Ozone require replacing generic paths with valid Ozone paths. The structure shown in the source material for Ozone paths uses the Open File System (OFS) scheme, typically formatted as `ofs://ozone/voll/buckl/` followed by the key path.

Question: 9

A data engineering team is developing custom SQL functions to extend Phoenix's capabilities. Before these UDFs can be used in queries, the system must be explicitly enabled to execute user-defined code. Which configuration property must be set to `*true` to permit the use of custom UDFs in Phoenix?

- ☐ `hbase.regionserver.allow.userdefined.scripts`
- ☐ `phoenix.sql.execute.udf.enabled`
- ☐ `phoenix.functions.allowUserDefinedFunctions`
- ☐ `hive.udf.security.enabled`
- ☐ `hbase.master.function.udf.whitelist`

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

Answer: C

Explanation:

To allow the use of User Defined Functions (UDFs) with Phoenix, the property

`phoenix.functions.allowUserDefinedFunctions` must be set to `'true'`. The administrator must also restart the role and service after saving changes.

Question: 10

An application relies on Phoenix Query Server (PQS) for its read/write operations. When configuring this application to connect to the PQS service in a secured CDP Data Hub, the client must obtain a non-JDBC

driver to interface with the operational database. Which programming language is explicitly supported for accessing data stored in the operational database via this method?

- A. Rust
- B. R
- C. Haskell
- D. PHP
- E. Ruby

Answer: D

Explanation: The list of supported programming languages to access data stored in your operational database includes Apache Groovy, C, C++, Go, Java, Jython, Python, PHP, and Scala. REST API support is also available. PHP is one of the explicitly supported languages.

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**



Visit us at: <https://www.examsboost.com/test/cdp-0011>