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: 6.0

Question: 1

You are having a CDP Data platform already setup by your BigData team which has following Hybrid Setup
- 2 Cluster on AWS Public Cloud
- 5 Cluster on Your Company Data Centre
Out of which 2 cluster which are in AWS are for Development and Testing purpose, now you want that. How much load is on the cluster which are on AWS public cloud, so that you can find the budget and further planning. So, which of the following you would use to find the Workload on your public cloud cluster?

A. Using AWS Cloud Watch
B. Using AWS Billing
C. Using CDP Cloudera Manager
D. Using CDP Replication Manager
E. Using CDP Workload XM
F. Using CDP Telemetry Publisher

Answer: E

Explanation:
AWS CloudWatch: This is a monitoring tool provided by AWS, which provides you with data and actionable insights to monitor your applications, respond to system-wide performance changes, and optimize resource utilization. CloudWatch collects monitoring and operational data in the form of logs, metrics, and events. You get a unified view of operational health and gain complete visibility of your AWS resources, applications, and services running on AWS and on-premises. You can use CloudWatch to detect anomalous behavior in your environments, set alarms, visualize logs and metrics side by side, take automated actions, troubleshoot issues, and discover insights to keep your applications running smoothly. However, this is not a good choice for checking workload on your CDP Cluster. Since, this would include all the resources which are being used for a particular AWS Account.
AWS Billing: Is just to check your billing amount for the resources you are using from AWS Cloud and not for the managing Workload on your CDP cluster. Cloudera Manager: Cloudera Manager is an application you use to manage, configure, and monitor CDP Private Cloud Base clusters and Cloudera Runtime services. The Cloudera Manager server runs on a host in your CDP Private Cloud Base deployment and manages one or more clusters using Cloudera Manager Agents that run on each host in the cluster.
The Cloudera Manager Admin Console is a web application administrators and others can use to manage CDP Private Cloud Base deployments. using the Cloudera Manager Admin Console, you can start and stop the cluster and individual services, configure and add new services, manage security, and upgrade the cluster. You can also use the Cloudera Manager API to programmatically perform management tasks. Since, Cloudera Manager is an Admin console, it gives you capability to monitor and manager your CDP cluster. But not a right tool for finding workload on your cluster. Hence, this is also not a right choice.

CDP Replication Manager: As name suggests, it is for replicating data to Public Cloud Clusters, which are managed using CDP. Replication Manager is a service in a CDP. That support copying data from HDFS, Hive and HBase to CDP Public Cloud clusters. However, there are different versions dependency what it supports or not. For that you need to look into Support Matrix for Replication Manager on CDP Public Cloud. Hence, this cannot be a correct option. Workload XM: How do you get to know; how much workload is being handled by your Cloudera CDP platform? For example, you want to know that
- Your workload is processed by which cluster.
- Which Service was involved in processing your workload?
- What data is processed as part of your workload.
These all things you can understand using Workload XM.
Hence, we can say that Workload XM is a correct choice for this question.

Question: 2

Which of the following things, you can check using Workload XM in your CDP cluster?

A. Troubleshooting failed jobs.
B. Optimizing slow jobs
C. Workload XM displays metrics about a Job's performance and compares the current Job execution with previous execution by creating baselines.
D. Sending an email alert for failed and long running jobs.

Answer: A,B,C

Explanation:
How do you get to know: how much workload is being handled by your Cloudera CDP platform? For example, you want to know that
• Your workload is processed by which cluster.
• Which Service was involved in processing your workload?
• What data is processed as part of your workload.
These all things you can understand using workload XM. Other things for your
• CDP, Workload XM can help is
• Troubleshooting failed jobs.
• Optimizing slow jobs
Workload XM displays metrics about a Job's performance and compares the current Job execution with previous execution by creating baselines.

Question: 3

Which of the following tool or component can be used for following requirement?
When any jobs get completed, then information about the job and the cluster that processed the job is sent to Workload XM using.

A. Cloudera Manager
B. Cloudera Manager Agent
C. Workload XM Agent

D. AWS Cloud logs
E. Telemetry Publisher

Answer: E

Explanation:
Telemetry Publisher: Telemetry Publisher is a role, in a Cloudera Manager Management Service. When any jobs get completed, then information about the job and the cluster that processed the job is sent to Workload XM with Telemetry Publisher.

Question: 4

For which of the following types of the Job, you can use Workload XM telemetry service?

A. Impala
B. Oozie
C. Hive
D. YARN
E. Spark

Answer: A,B,C,D,E

Explanation:
As soon as you enable the Workload XM, Cloudera Management Service will starts the Telemetry Publisher Role as well. For the following jobs
- Impala
- Oozie
- Hive
- YARN
- Spark

Question: 5

Which of the following data would be collected and transmitted by Telemetry Service in Workload XM?

A. Metrics data
B. Configuration Data
C. Log files
D. Data processed by Job
E. Host information on which Task was executed

Answer: A,B,C

Explanation:

As soon as you enable the Workload XM, Cloudera Management Service will starts the Telemetry Publisher Role as well. For the following jobs
- Impala
- Oozie
- YARN
- Spark
Telemetry Publisher collects and transmits
- Metrics data
- Configuration Data
- Log files
For all of the above-mentioned services to Workload XM. You can have multiple servers configured in your CDP and Telemetry Publisher Role collects metrics for all the clusters that use Workload-XM environments. For a single cluster how that is performed is shown below.
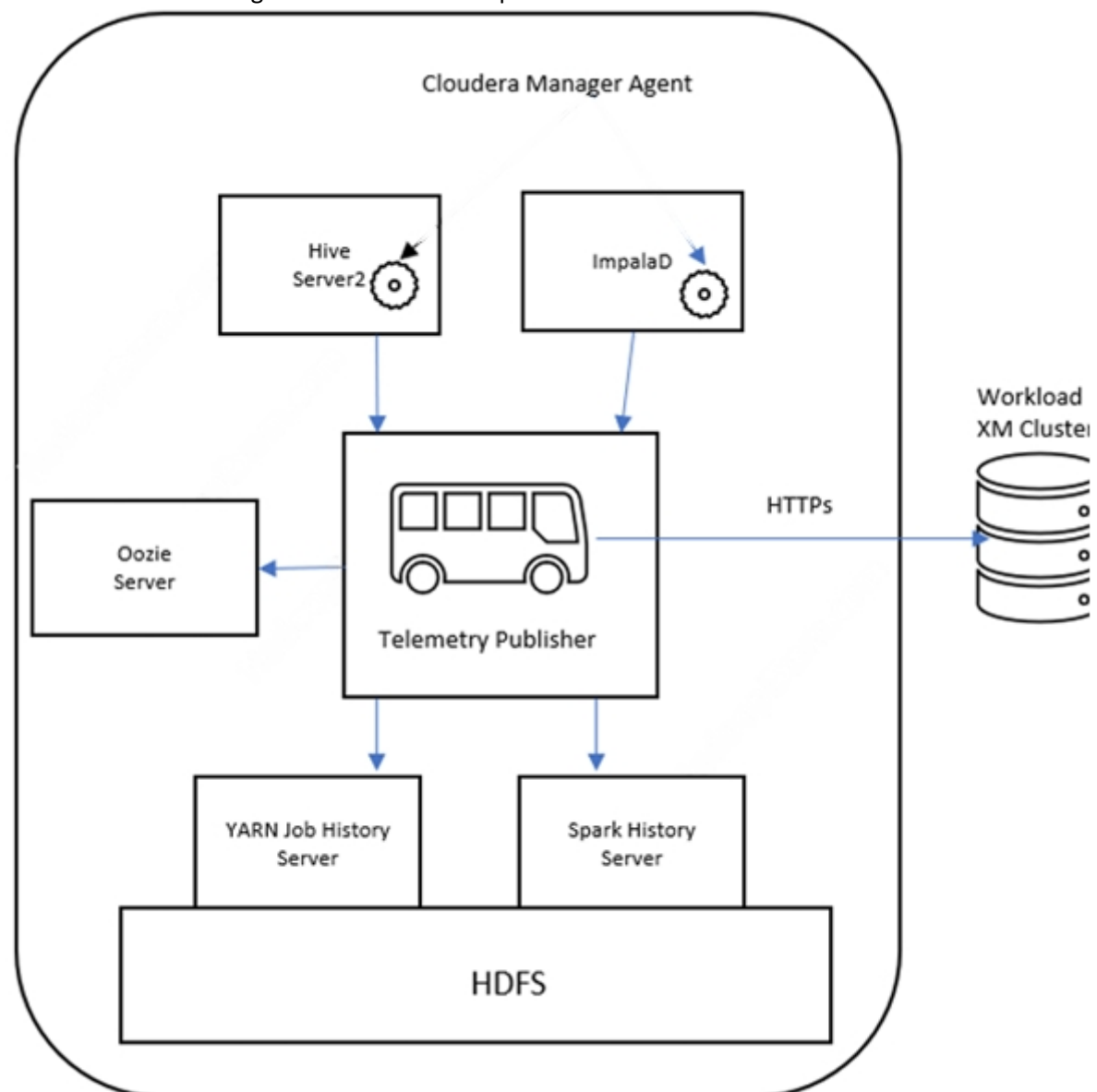
Question: 6

Which of the following tool you can use to replicate data among cluster while using CDPcluster in public cloud?

A. Workload XM
B. HDFS
C. 400zie
D. ImpalaD
E. Replication Manager

Answer: E

Explanation:
About Replication Manager
As name suggests, it is for replicating data to Public Cloud Clusters, which are managed using CDP.
Replication Manager is a service in a CDP. That support copying data from HDFS, Hive and HBase to CDP Public Cloud clusters.
However, there are different versions dependency what it supports or not. For that you need to look into Support Matrix for Replication Manager on CDP Public Cloud.
Replication Manager can be used to replicate data from CDH Clusters and CDP Private Cloud Base Cluster to CDP Public Cloud Clusters.

Question: 7

Match the following items with their descriptions:

Policy		This requires minimal services for metadata and governance, such as Hive MetaStore, Ranger and Atlas.
Job		This represent an instance of rules that can be in running state or already completed.
Data Lake		These are set of rules, that would be applied during replication.
Cloud storage		You can have data in public cloud like AWS, Azure

Answer:

Policy		These are set of rules, that would be applied during replication.
Job		This represent an instance of rules that can be in running state or already completed.
Data Lake		This requires minimal services for metadata and governance, such as Hive MetaStore, Ranger and Atlas.
Cloud storage		You can have data in public cloud like AWS, Azure

Explanation:

• Policy: These are set of rules, that would be applied during replication. For example, which cluster is a source and destination, the type of data to replicate, schedule for replication etc.

• Job: This represent an instance of a policy that can be in running state or already completed.

• Data Lake: You can have a CDP cluster on Cloud, using virtual machines, and keeping data retained on Cloud storage. A Cloud data lake requires minimal services for metadata and governance, such as Hive Metastore, Ranger and Atlas.

• Cloud storage: You can have data in public cloud like AWS, Azure etc. So, if your storage is in cloud account like AWS S3 of AWS or Azure then that is called cloud storage.

Question: 8

Please select the wrong statement regarding the Version delete in HBase

A. Deleting all the versions older than a certain timestamp can be performed on a row, column family or column

B. Delete version at a specific timestamp and can only be performed on a column

C. For deletion if not timestamp is specified currentTimeMiIlis is used

D. For deletion if not timestamp is specified all the versions will be deleted

<div style="border: 1px solid black; padding: 10px; text-align: center;">

Answer: D

</div>

Explanation:

For Deleting Versions

Delete all versions older than a certain timestamp

Can be performed on a row, column family or a column

Delete version at a specific timestamp

Can only be performed on a column

If no timestamp is specified, currrentTimeMillis is used

Question: 9

There is a feature provided in QuickTechie.com website that any Software Professional can create an article as well as can update and delete the article. You decided to use HBase rather than HDFS to store this article. What would be the reason, you preferred the HBase over HDFS.

A. Fault tolerance

B. Batch processing

C. Random writes

D. Even Distribution of Data.

<div style="border: 1px solid black; padding: 10px; text-align: center;">

Answer: C

</div>

Explanation:

Apache HBase provides random, realtime read/write access to your data. HDFS does not allow random writes. HDFS is built for scalability, fault tolerance, and batch processing. HDFS is a distributed file system that is well suited for the storage of large files. It's documentation states that it is not, however, a general purpose file system, and does not provide fast individual record lookups in files. HBase, on the other hand, is built on top of HDFS and provides fast record lookups (and updates) for large tables. This can sometimes be a point of conceptual confusion. HBase internally puts your data in indexed "StoreFiles" that exist on HDFS for high-speed lookups. Features of HBase Strongly consistent reads/writes: HBase is not an "eventually consistent" Datastore. This makes it very suitable for tasks such as high-speed counter aggregation. Automatic sharding: HBase tables are distributed on the cluster via regions, and regions are automatically split and re-distributed as your data grows. Automatic RegionServer failover Hadoop/HDFS Integration: HBase supports HDFS out of the box as its distributed file system. MapReduce: HBase supports massively parallelized processing via MapReduce for using HBase as both source and sink. Java Client API: HBase supports an easy to use Java API for programmatic access.

Thrift/REST API: HBase also supports Thrift and REST for non-Java front-ends. Block Cache and Bloom Filters: HBase supports a Block Cache and Bloom Filters for high volume query optimization. Operational Management: HBase provides build-in web-pages for operational insight as well as JMX metrics.

Question: 10

Type	It represents the configuration of Cloudera Manager and all the clusters it manages.
Service	This can be for one of the services.
Role	This is an instance of a service running on the Cloudera Manager.
Deployment	This is a one of the managed functionalities in Cloudera Manager, which may or may not be distributed.

Answer:

Type	This is an instance of a service running on the Cloudera Manager.
Service	This is a one of the managed functionalities in Cloudera Manager, which may or may not be distributed.
Role	This can be for one of the services.
Deployment	It represents the configuration of Cloudera Manager and all the clusters it manages.

Explanation:
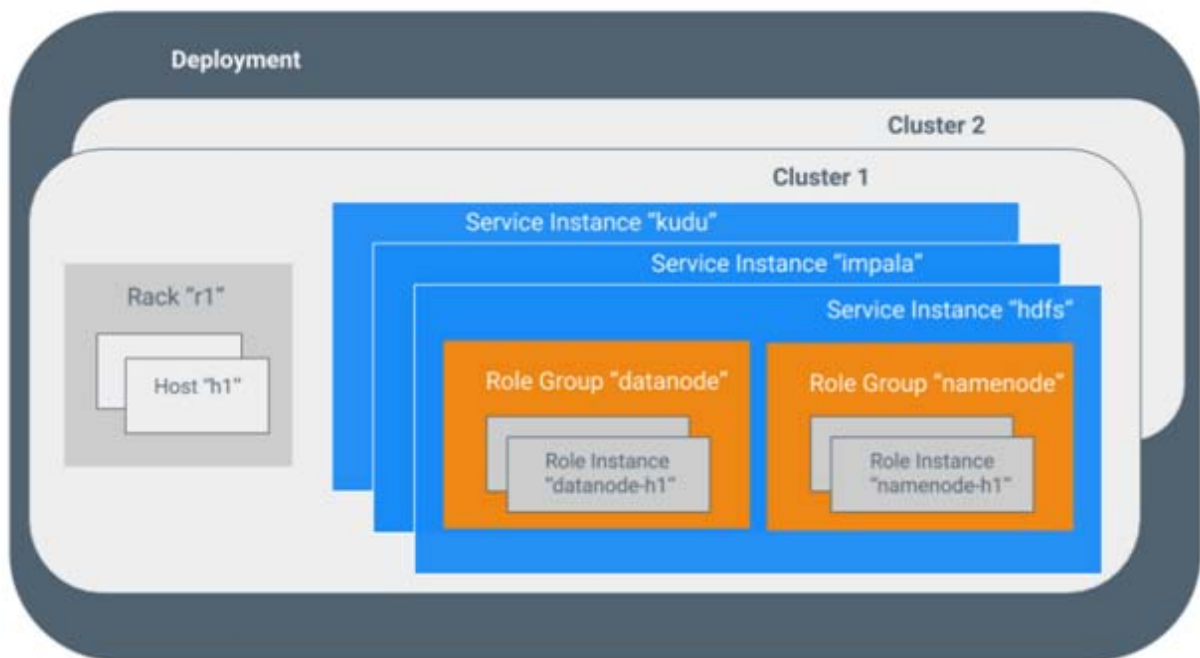
Type & instance If you understand Object Oriented concepts a type is a Template for example in Java a java.lang.String is a Class of a template and below all are instances of Type String. "hadoopexam", "Amit", "Rakesh" , "QuickTechie" etc.

Services & Instance Concepts

Similarly, in case of Cloudera Manager there are various services and they can have instances. For example, Spark is one Service in Cloudera Manager. However, it one of the instances is running and you can check its status using "Home>Status" or "Clusters>ClusterName", which list all the services. Hence, please keep an eye where we use instance and type while discussion.

Service

To understand you can imagine HDFS, MapReduce, Spark etc. are considered as service.

If you understand more in a Linux or Unix environment a command whichruns a "System init script in /etc/init.d/.

This is a one of the managed functionality in Cloudera Manager, which may or may not be distributed.

• Hive
• HBase
• HDFS
• YARN

Service Instance

This is an instance of the above service running on the Cloudera Manager.

And you can even give name to those service instance as per your choice for example.

HE-HDFS Service instance represent a running HDFS service.

HE_YARN Service instance represent a running YARN service instance.

A single Service instance can have multiple roles as well.

You can have one of the instances running from this service.

Role or Role Type

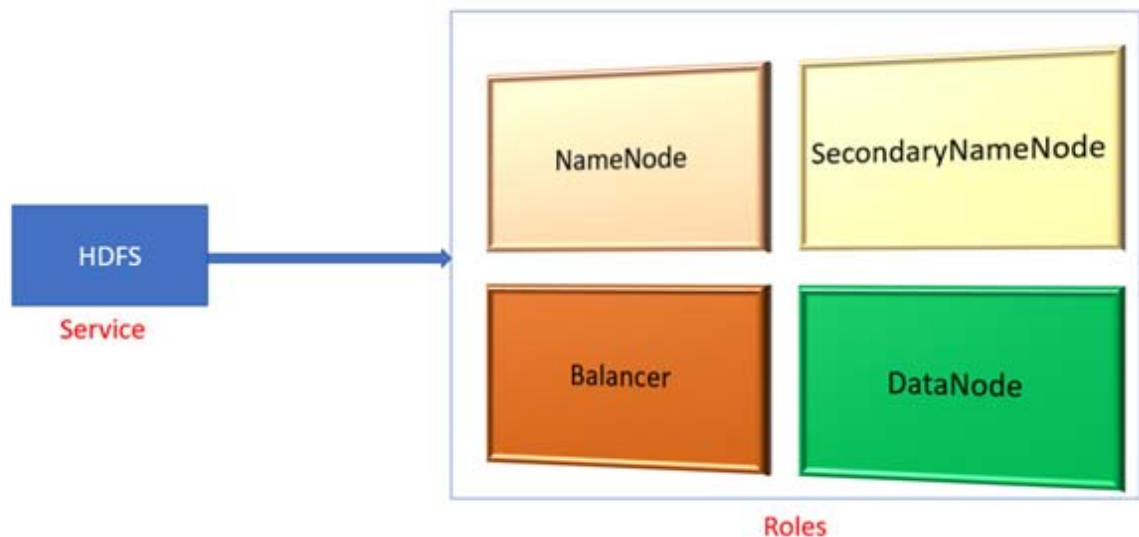There are various services and each service can have different roles.

For example, HDFS is a service and it can have following roles

• NameNode

• DataNode
• SecondaryNameNode
• Balancer
Remember sometime Role is referred as Role Type as well.
In Cloudera Manager, a category of functionality is considered as service.

Role Instance
On a host a role would be running of a service.
Role instance you can assume as a Windows or Unix process. If you are running three different Java Process on a single host. It means, it has 3 role instances running.
For NameNode Role you can have Role Instance name like NameNode-
HEI & similarly, for DataNode you can have DataNode-HE1
Role Group
This is a set of configuration properties for a set of Role instances.
Host
This represents a physical or virtual machine, that run role instances i.e. in case of HDFS, it could be either NameNode or DataNode Role instance.
One host can belong to only one cluster.
Host Template
A set of role groups in Cloudera Manager. When a template is applied to a host, a role instance from each role group is created and assigned to that host.
Rack
Rack represents a physical entity which contains a set of physical hosts and typically served by the same switch.
A Deployment
For Cloudera Manager, a deployment represents the configuration of Cloudera Manager and all the clusters it manages.
Dynamic Resource Pool
In Cloudera Manager, it is a named configuration of resources and a policy for scheduling the resources among YARN applications or Impala queries running in the pool.
Cluster

This represents a set of computers in a rack or rack of computers that contains an HDFS filesystem and runs MapReduce and other processes on that data.

Association: Please keep in mind that each individual cluster can only be associated with a single Cloudera Manager Server.

Logical Entity: In Cloudera Manager, a logical entity that contains a set of hosts, a single version of Cloudera Runtime installed on the hosts, and the service and role instances running on the hosts is a cluster.

Host: Please keep in mind that a host can belong to only one cluster.

Gateway or gateway node or edge node

This is a type of role that provides access to a client for a specific service from cluster services.

Hive, HDFS, Kafka, MapReduce, Solr and Spark each have gateway roles to provide access for their clients to their respective services.

Names: It is not mandatory that Gateway Role would have gateway name in their names and also, they may not be exclusively for client access.

Example: Hue Kerberos Ticket Renewer is a gateway role that proxies tickets from Kerberos.

The node supporting one or more gateway roles is sometimes referred to as the gateway node or edge node, with the notion of "edge" common in network or cloud environments.

In terms of the Cloudera cluster, the gateway nodes in the cluster receive the appropriate client configuration files when Deploy Client

Configuration is selected from the Actions menu in Cloudera Manager

Admin Console.

Parcel

This is a binary distribution format that contains compiled code and meta- information such as package description, version and dependencies.

Static Service Pool

In Cloudera Manager, a static partitioning of total cluster resources—CPU, memory, and 1/0 weight—across a set of services.

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**