



# Cloudera Semantic Search: Deploying Cluster in K8s(ECS) for Private Cloud

Version 0.1

Document Prepared by: Abhradeep Kundu

Revision History			
Version	Author(s)	Description	Date
0.1	Abhradeep Kundu Vaibhav Joshi	Initial version	2024/11/12

## Important Notice

© 2010-2024 Cloudera, Inc. All rights reserved.

Cloudera, the Cloudera logo, and any other product or service names or slogans contained in this document, except as otherwise disclaimed, are trademarks of Cloudera and its suppliers or licensors, and may not be copied, imitated or used, in whole or in part, without the prior written permission of Cloudera or the applicable trademark holder.

Hadoop and the Hadoop elephant logo are trademarks of the Apache Software Foundation. All other trademarks, registered trademarks, product names and company names or logos mentioned in this document are the property of their respective owners to any products, services, processes or other information, by trade name, trademark, manufacturer, supplier or otherwise does not constitute or imply endorsement, sponsorship or recommendation thereof by us.

Complying with all applicable copyright Azure is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Cloudera.

Cloudera may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Cloudera, the furnishing of this document does not give you any license to these patents, trademarks copyrights, or other intellectual property.

The information in this document is subject to change without notice. Cloudera shall not be liable for any damages resulting from technical errors or omissions which may be present in this document, or from use of this document.

### **Cloudera, Inc.**

Santa Clara, California  
5470 Great America Pkwy  
Santa Clara, CA 95054

US: +1 888 789 1488

Outside the US: +1 650 362 0488

<http://www.cloudera.com>

## Disclaimer

This document is only for internal use purposes and should not be provided to customers.

# Introduction

This document contains instructions on how to onboard a CSS(Cloudera Semantic Search) cluster on a Private Cloud Experience Environment (K8s On ECS).

## Prerequisites

Helm (v 3.14.x)	<a href="#">Installing Helm</a>
Kubectl	<a href="#">Install Tools   Kubernetes</a>
curl	Linux curl commands
Hosts for ECS	<p>Min Nodes: 4 with following spec:</p> <ul style="list-style-type: none"><li>• Min CPU: 16 cores</li><li>• Memory: 64 GB</li><li>• Min Disk: 400GB at / mount and 400GB at /ecs mount</li></ul> <p>For more specific requirements you can check this article <a href="#">CDP Private Cloud Data Services Software Requirements</a>. For /ecs directory we need +100GB on top of the requirement specified in the doc. for our CSS basic setup to run smoothly</p>
ECS	<a href="#">Installing CDP Private Cloud Data Services using ECS</a>
CDP PvCDS parcel V1.5.4+	<a href="#">How do I get it the current PvCDS parcel</a>
Private Cloud Base Cluster	<a href="https://docs.cloudera.com/cdp-private-cloud-base/7.1.9/index.html">https://docs.cloudera.com/cdp-private-cloud-base/7.1.9/index.html</a>
Steps to create manual instances	<a href="#">CSS on PvC DS - Prod deployment</a>

## Document Verification

This document is verified with the below-mentioned cluster setup.

	CM Version	CDH Version	CDP-PvC Version	Status
Base Cluster + DataService Cluster	7.11.3.4	7.1.9.0	1.5.0	<ul style="list-style-type: none"><li>• Verified with manual steps to add the ECS cluster</li></ul>
Base Cluster + DataService Cluster	7.11.3.9	7.1.9.0	1.5.4	<ul style="list-style-type: none"><li>• Verified with jenkins job to add the ecs cluster</li></ul>

Base Cluster + DataService Cluster	7.13.1.0	7.3.1.0	1.5.4	<ul style="list-style-type: none"> <li>• Verified with manual steps to add the ecs cluster</li> <li>• Verified with jenkins job to add the ecs cluster</li> </ul>
------------------------------------	----------	---------	-------	---

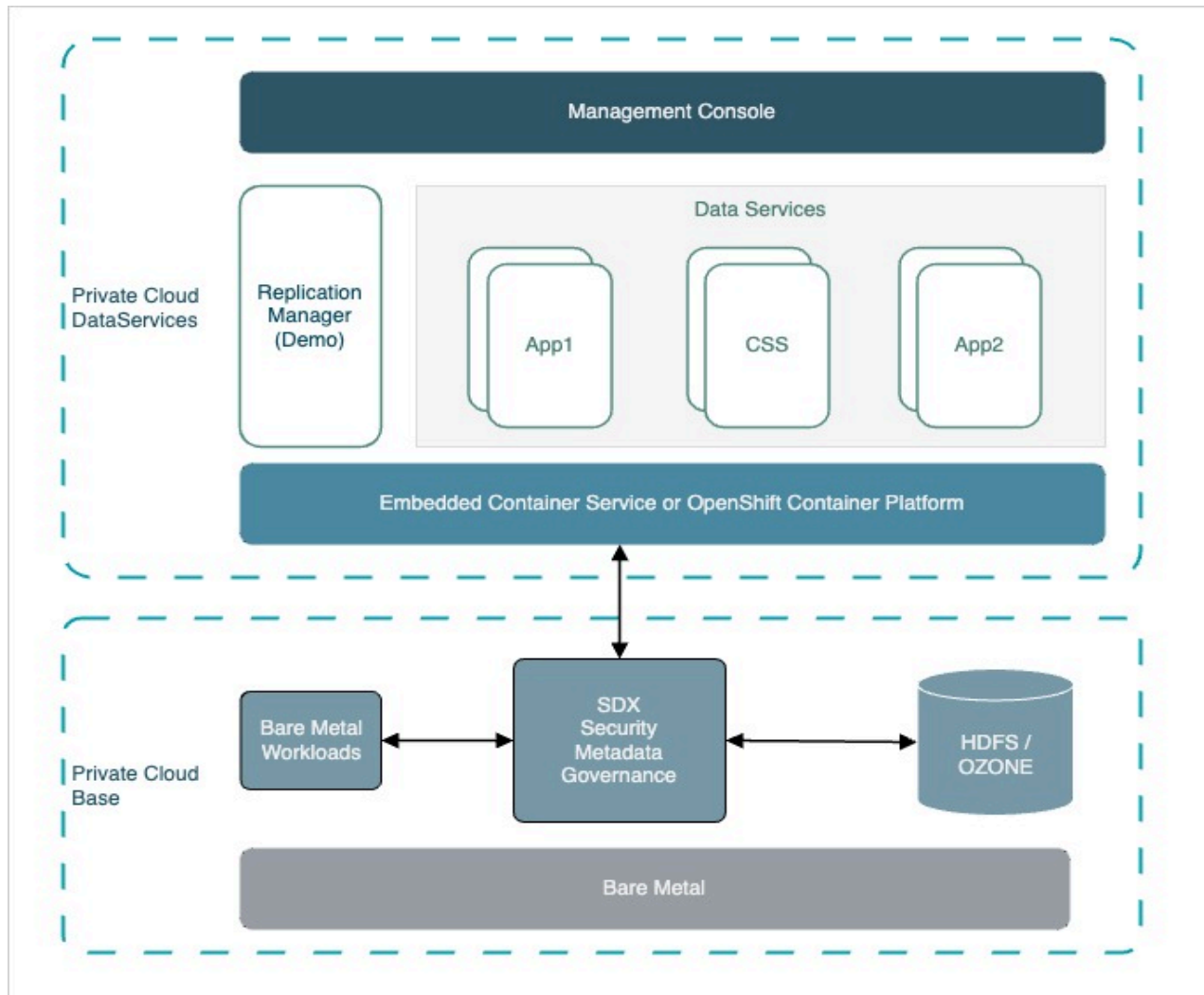
## Assumptions

- 1) You have access to Jenkins <https://master-01.jenkins.cloudera.com>
- 2) You have access to Cloudcat  
<https://cloudcat.infra.cloudera.com/provisionedInstanceGroup/ec2Create>
- 3) You have access to the ocker private registry <https://docker-private.infra.cloudera.com/>

## Overview of Private Cloud Experiences

CDP Private Cloud Experiences works on top of CDP Private Cloud Base and is the on-premise offering of CDP that brings many of the benefits of the public cloud deployments to the on-premise CDP deployments. Like the cloud-native applications, the CDP Private Cloud Experiences are designed to be easy to use and offer tenant-level isolation and self-service with auto-scale. All of this is made possible by the new Experiences Compute Service (ECS) which manages the compute infrastructure and ease of deployment for the Experiences.

The Private Cloud Experiences rely on and work with customers' existing data storage and governance clusters, which we refer to as the Private Cloud Base Cluster. The Experiences, once deployed, seamlessly and securely connect with the Private Cloud Base cluster. The following diagram is a typical example of a Private Cloud Experiences deployment:



Before you can install Private Cloud Experiences, you need a running instance of CDP Private Cloud Base. You need an isolated hardware environment with dedicated infrastructure and networking for Private Cloud Experiences.

## Deployment Steps

If you already have the required hosts (4 Hosts minimum, here is the [configuration](#)) you can refer to the below [section](#) otherwise you need to follow either of the following paths:

- Follow [these steps to manually create the nodes](#), for this case you need to continue from [here](#).
- Follow [these steps to create and setup the ECS cluster](#) and add it to the Base cluster through Jenkins job. If you choose this path then you can continue from [here](#).

## Create and Add Data Service Cluster in CM of Base Cluster

To manually create and add a Data Service cluster in the Base Cluster we need to follow the steps mentioned in the following document:

<https://docs.cloudera.com/cdp-private-cloud-data-services/latest/installation-ecs/topics/cdppvc-installation-ecs-steps.html>

*Note: While performing the steps from the above doc if you find the parcel is older than 1.5.4 then replace the URL with the latest repo URL. [How to find the latest repo url?](#)*

*In case of any queries please reach out to the PvC DS team.*

## Deploying CSS Cluster on ECS

To deploy the CSS cluster we need to deploy the cert manager and a self-sign certificate first. After this, we can deploy the helm charts specific to CSS. In this CSS deployment, we will have 3 master nodes and 3 data nodes by default. And the coordinator node will get deployed along with the data node.

### Step 1: Download the Kube config

To Download the kube config we need to sftp `/etc/rancher/rke2/rke2.yaml` file from the node with the ECS master role. Also, we need to update the IP address 127.0.0.1 to the actual ECS server IP address in the same file after doing sftp.

```
Unset
sftp root@<ecs master host>
passwd: <root password>
sftp> get /etc/rancher/rke2/rke2.yaml
mv rke2.yaml ~/.kube/ops-cluster-configs
```

### Step 2: Export the Kube config to set the context

```
Unset
export KUBECONFIG=~/.kube/ops-cluster-configs
```

### Step 3: Test using Kubernetes command

Unset

```
kubectl get pods -A
```

It should show all the infra pods created by the ECS cluster.

## Step 4: Deploy Helm charts

### A. Pull the charts

**We need to get this 0.1.0-b20 version while pulling the charts**

Unset

```
VERSION=0.1.0-b28
```

```
mkdir css-helm-charts
```

```
cd css-helm-charts
```

```
helm pull oci://docker-private.infra.cloudera.com/cloudera-helm/solr/opensearch  
--version $VERSION
```

```
helm pull
```

```
oci://docker-private.infra.cloudera.com/cloudera-helm/solr/opensearch-dashboards  
--version $VERSION
```

### B. Unpack helm charts bundle

Unset

```
for file in *.tgz; do tar -vxf "$file"; done
```

After this, your css-helm-charts will have 2 more directories

Unset

```
opensearch
```

```
opensearch-dashboards
```

### C. Execute Helm commands for CSS charts

*Note: You may need to execute [these steps](#) in prod based on the registry you are using*  
Before executing the master and coordinator helm you need to modify the ingress host *ecs-ingress.yaml* in the folder *opensearch*.

- 1) You need to replace the value of `example.vpc.cloudera.com` with the exact value of the host where the ECS server role is running. To get the host DNS follow the instructions [here](#).

Below is an example of how it is done in our cluster

Unset

```
hosts:  
  - opensearch-cluster.example.vpc.cloudera.com
```

## 2) Deploy master nodes

Unset

```
helm install opensearch-master opensearch -f opensearch/values.yaml -f  
opensearch/master-pvcds.yaml --set adminPassword=Cloudera@Test4321 --namespace  
css --create-namespace --set k8sProvider=ecs
```

## 3) Deploy Data nodes along with the coordinator node

Unset

```
helm install opensearch-data opensearch -f opensearch/values.yaml -f  
opensearch/data-pvcds.yaml -f opensearch/ecs-ingress.yaml --set  
adminPassword=Cloudera@Test4321 --set coordinatorService=data --namespace css  
--create-namespace --set k8sProvider=ecs
```

## 4) Deploy opensearch-dashboards helm chart

If you would like to access opensearch from a browser interface, follow the dashboard setup [document](#).

# Validation Steps For CSS

Please check [this section](#) on how to validate the CSS cluster deployments.

# Validation Steps For Dashboard

Please check [this section](#) on how to validate the CSS Dashboard.

# ML Node and Ingest Node steps

Please follow [this section](#) for ML Node and Ingest Node



# Uninstall steps

Refer to the following steps to delete a Cloudera Semantic Search cluster from your private cloud environment.

## Step 1: Uninstall the helm charts

Unset

```
helm uninstall opensearch-master -n css  
helm uninstall opensearch-data -n css
```

## Limitations

- 1) We need to change a few values of the helm chart manually to work properly.

## Troubleshooting

- 1) If your pod is stuck in a pending state, you need to check the log of the pod as well as the events in the describe pod command

Sample command to get the log

Unset

```
kubectl logs pod/hello
```

Sample command to describe a pod

Unset

```
kubectl describe pods/hello
```

## Additional Instructions

*Note: For any queries related to the host creation or ECS cluster creation, please reach out to the PvC DS team.*

## Additional Steps to use prod registry

- A. Modify the helm charts to use Cloudera's public repository.

**Change the values.yaml for opensearch:**

- For image field, Replace docker-private.infra.cloudera.com with container.repository.cloudera.com
- For imagePullSecrets field, Replace [] with [{"name": "jfrog-dev"}]

Unset

```
vi opensearch/values.yaml

# replace docker-private.infra.cloudera.com with
container.repository.cloudera.com
image:
  repository: "container.repository.cloudera.com/cloudera/opensearch"
  tag: "" ## Use Release version from Chart
  pullPolicy: "IfNotPresent"

# also replace [] with [{"name": "jfrog-dev"}]
imagePullSecrets: [{"name": "jfrog-dev"}]
```

### Change the values.yaml for opensearch-dashboards:

- For image field, Replace docker-private.infra.cloudera.com with container.repository.cloudera.com
- For imagePullSecrets field, Replace [] with [{"name": "jfrog-dev"}]

Unset

```
vi opensearch-dashboards/values.yaml

# replace docker-private.infra.cloudera.com with
container.repository.cloudera.com
image:
  repository:
"container.repository.cloudera.com/cloudera/opensearch-dashboards"
  tag: "" ## Use Release version from Chart
  pullPolicy: "IfNotPresent"

# also replace [] with [{"name": "jfrog-dev"}]
imagePullSecrets: [{"name": "jfrog-dev"}]
```

## How to get Host details

Step 1) From CM you need to go to the Data Service Cluster which you added earlier. Click on the Hosts link.

✓ Experience Cluster 1

Actions ▾

Status

Health Issues

Configuration ▾

Status

ECS 1.5.4 (Parcels)

✓ 4 Hosts

✓ DOCKER-1

✓ ECS-1

Tags

Edit Tags

\_cldr\_cm\_ek8s\_control\_plane=75865fa2-17f8-414b-a44f-7bcec425703a

Step 2) Click on the roles to find where the “ECS server” is installed

Experience Cluster 2

CDEP Deployment from 2024-Oct-20 23:58

Hosts

Configuration

Add Hosts

Review Upgrade Status

Inspect Hosts in Cluster

Inspect Cluster Network Performance

Q Search

Filters

Last Updated: Oct 22, 12:54:02 AM PDT

Columns: 11 Selected ▾

Filters

STATUS

Good Health 4

CLUSTERS

CORES

COMMISSION STATE

LAST HEARTBEAT

Actions for Selected ▾

<input type="checkbox"/>	Status	Name	IP	Roles	Tags	Commission State	Last Heartbeat	Load Average	Disk Use
<input checked="" type="checkbox"/>	✓	DOCKER Docker Server		2 Roles		Commissioned	2.15s	1.14 1.16 1.05	56.5 GiB
<input checked="" type="checkbox"/>	✓	ECS Ecs Server		2 Roles		Commissioned	2.81s	0.02 0.18 0.40	57 GiB
<input type="checkbox"/>	✓	akundu3-3.vpc.cloudera.com	10.65.156.244	2 Roles		Commissioned	2.58s	0.02 0.17 0.60	58.4 GiB
<input type="checkbox"/>	✓	akundu3-4.vpc.cloudera.com	10.65.147.218	2 Roles		Commissioned	2.55s	0.10 0.23 0.33	55.4 GiB

1 - 4 of 4

Step 3) Copy the hostname

Experience Cluster 2

CDEP Deployment from 2024-Oct-20 23:58

Hosts

Configuration

Add Hosts

Review Upgrade Status

Inspect Hosts in Cluster

Inspect Cluster Network Performance

Q Search

Filters

Last Updated: Oct 22, 12:54:02 AM PDT

Columns: 11 Selected ▾

Filters

STATUS

Good Health 4

CLUSTERS

CORES

COMMISSION STATE

Actions for Selected ▾

<input type="checkbox"/>	Status	Name	IP	Roles	Tags	Commission State	Last Heartbeat	Load Average	Disk Use
<input type="checkbox"/>	✓	akundu3-1.vpc.cloudera.com	10.65.144.218	2 Roles		Commissioned	2.15s	1.14 1.16 1.05	56.5 GiB
<input type="checkbox"/>	✓	akundu3-2.vpc.cloudera.com	10.65.144.178	2 Roles		Commissioned	2.81s	0.02 0.18 0.40	57 GiB
<input type="checkbox"/>	✓	akundu3-3.vpc.cloudera.com	10.65.156.244	2 Roles		Commissioned	2.58s	0.02 0.17 0.60	58.4 GiB
<input type="checkbox"/>	✓	akundu3-4.vpc.cloudera.com	10.65.147.218	2 Roles		Commissioned	2.55s	0.10 0.23 0.33	55.4 GiB

# How do I get the current PvCDS parcel

Step 1) Log in to canary build

<https://release.infra.cloudera.com/stackpage?stack=CDP-PVC&filter=SHIPPED>

Step 2) Find the latest release version of the build and click on it

Home / CDP-PVC

Builds Hotfixes Shipped Bulletin Info

Public Private

Search

Release Version	JIRA	ReleasedBuild	GBN	Base Release	Released Date	Internal Release Name	Repo url
1.5.4	Not Available	1.5.4-b99	53854554	Not Available	May 29, 2024	CDP-PVC-1.5.4	generic <a href="#">↓</a>
1.5.3	RELENG-24511	1.5.3-b297	50802651	Not Available	Not Available	CDP-PVC-1.5.3	generic <a href="#">↓</a>
1.5.2	Not Available	1.5.2-b886	46792599	Not Available	November 2, 2023	CDP-PVC-1.5.2	generic <a href="#">↓</a>
1.5.1	Not Available	1.5.1-b626	42068229	Not Available	June 13, 2023	CDP-PVC-1.5.1	generic <a href="#">↓</a>
1.5.0	Not Available	1.5.0-b448	36846926	Not Available	January 20, 2023	CDP-PVC-1.5.0	generic <a href="#">↓</a>

Step 3) Find the latest GBN where the platform status is green and click on the GBN number

CDP-PVC : 1.5.4 Branch: NA

BUILD NUMBER BUILD NUMBER

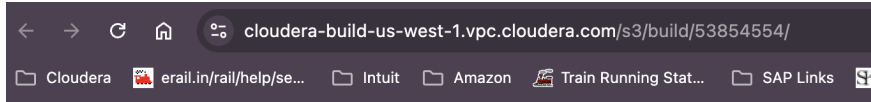
Search

Build Dashboard Stats Dependency

20 SEARCH

Build Number	GBN	Platform Status	Changes	Triggered	Expiry	Links
1.5.4-b99	53854554	<span>R<sub>0</sub></span>	1	5 months ago	3 months ago	<a href="#">📄</a> <a href="#">📁</a> <a href="#">🔗</a> <a href="#">🔧</a> <a href="#">🔍</a>
1.5.4-b98	53717659	<span>R<sub>0</sub></span>	2	5 months ago	3 months ago	<a href="#">📄</a> <a href="#">📁</a> <a href="#">🔗</a> <a href="#">🔧</a> <a href="#">🔍</a>
1.5.4-b97	53703916	<span>R<sub>0</sub></span>	1	5 months ago	3 months ago	<a href="#">📄</a> <a href="#">📁</a> <a href="#">🔗</a> <a href="#">🔧</a> <a href="#">🔍</a>
1.5.4-b96	53682355	<span>R<sub>0</sub></span>	1	5 months ago	4 months ago	<a href="#">📄</a> <a href="#">📁</a> <a href="#">🔗</a> <a href="#">🔧</a> <a href="#">🔍</a>

Step 4) Click on the cdp-pvc link



## Index for /s3/build/53854554/

Name	Last Modified	Size
<a href="#">Parent Directory</a>	-	-
<a href="#">AQUASEC CI REPORTS/</a>	-	-
<a href="#">LOGS/</a>	-	-
<a href="#">PATCH FILES/</a>	-	-
<a href="#">cdp-pvc/</a>	-	-
<a href="#">build-mutable.json</a>	2024-08-05 16:03	614.00 B
<a href="#">build.json</a>	2024-05-28 13:20	5.41 KB

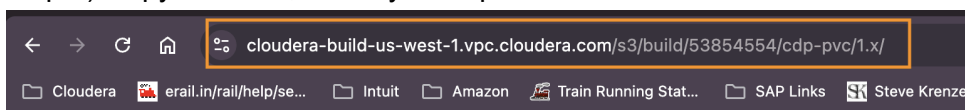
Step 5) Click on the 1.x link

## Index for /s3/build/53854554/cdp-pvc/

Name	Last Modified	Size
<a href="#">Parent Directory</a>	-	-
<a href="#">1.x/</a>	-	-

Unless otherwise specified herein, downloads of software from this site and its use are governed by the [Standard License](#). If you do not wish to be bound by these terms, then do not

Step 6) Copy the URL, this is your repo-url.



## Index for /s3/build/53854554/cdp-pvc/1.x/

Name	Last Modified	Size
<a href="#">Parent Directory</a>	-	-
<a href="#">maven-repository/</a>	-	-
<a href="#">parcels/</a>	-	-
<a href="#">redhat8/</a>	-	-
<a href="#">cdp-private-1.5.4-b99.tgz</a>	2024-05-28 13:20	113.73 MB
<a href="#">manifest.json</a>	2024-05-28 13:20	204.36 KB

Unless otherwise specified herein, downloads of software from this site and its use are governed by the [Standard License](#). If you do not wish to be bound by these terms, then do not download or use the soft

# How to create a Data Service cluster using Jenkins job?

Step 1) Log in to <https://master-01.jenkins.cloudera.com/job/Cluster-Setup-Add-EC2-ECS>

Step 2) Click on the “Build with Parameters” links from the left panel.

Step 3) Fill in all the parameters as shown below:

Parameters

CLUSTER\_SHORTNAME

ak731ecs-{1..4}

REQUIRED: Hosts where we will add the ECS cluster.

CM\_VERSION

bquery://product=cm:version=7.13.1.0:official:L0\_PASSED

Version of Cloudera Manager. Accepted formats:

- Global Build Numbers (e.g.: 'gbrn://123456')

CDH

bquery://product=cdh:version=7.3.1.0:official:L0\_PASSED

Version of CDH. This is usually meaningless in an ECS add\_cluster context and you can just use the default. It's required though to satisfy CDEP cross-checking of parameter values. Modify only if the default is incompatible with the CM version.

DOMAIN

vpc.cloudera.com

The domain where the hosts are/will be. Manowar for airgap.

CLOUDCAT\_OS

redhat86

The CDEP os name on Amazon EC2.

CLOUDCAT\_MASTER\_SIZE

16core32gb

(Provision only) Size minimal criteria of the master node (CM Server) instance (Recommended: 2xlarge)

**Note:** Among the above parameters *CM\_Version* and *CDH version* should match exactly with your Base Cluster

**CLOUDCAT\_SLAVE\_SIZE**

(Provision only) Size minimal criteria of the slave node instances

**CLOUDCAT\_BUDGET**

(Provision only, optional) Allows you to charge a specified project rather than your personal allocation. When empty, Cloudcat uses your set default budget.

**CLOUDCAT\_SHORTNAME**

(Optional) All hosts in the cloudcat provisioning group. This parameter is typically used when one needs extra hosts alongside the private cloud hosts, e.g. a load balancer host. This is the group definition for Cloudcat (total hosts to provision). If left empty, value defaults to 'CLUSTER\_SHORTNAME'.

**Example usage:** ECS HA setup needs 6 hosts: 3 ECS servers, 2 ECS agents and 1 (outside the cluster) for load balancer.

CLUSTER\_SHORTNAME = my-ecs-hosts-{1..5}

CLOUDCAT\_SHORTNAME = my-ecs-hosts-{1..6}

Don't forget to set OPTIONAL\_ARGS --private-cloud-haproxy-host=host6\_fullname and --ha-service-types=ECS for this use case.

☐ AIRGAP

**CLOUDCAT\_EXPIRATION\_DAYS**

(Provision only) Days/hours to retain host groups, maximum 21 days. Pick the minimum for your needs, saving cost!

**CLOUDCAT\_USERNAME**

(Provision only) Specify your cloudcat username. BUILD\_USER\_ID is your username in Jenkins.

**MASTER\_NODE**

(REQUIRED) ^^^^^ Specify the host running the CM server ^^^^^ ^^^^^

**Note:** Among the above parameters master node should match exactly with your Base Cluster CM host, this has to be filled

#### EMAIL\_ADDRESS

(Optional) In addition to the user who starts the job, these recipients will get an email on completion. Accepted special values: none, nonelfBuildSuccess, nonelfBuildFailure

#### PROVISION\_ACTIONS

The make\_group\_available action will re-use hosts if they already exist, but will not update the expiration time.

When used with extend\_group the expiration time is updated.

The destroy\_and\_create\_group combination destroys the group if it exists, and then creates fresh guaranteed clean hosts.

#### EXTRA\_PROVISIONING\_ARGS

#### DB

The database to use for the cluster.

☒ AUTO\_TLS

#### KERBEROS

Enable Kerberos authentication

#### JAVA\_VERSION

Select the Java version to use for the cluster

☒ AUTO\_TLS

#### KERBEROS

Enable Kerberos authentication

#### JAVA\_VERSION

Select the Java version to use for the cluster

#### ACTIONS

Select the deployment actions to perform on the cluster.

print\_hosts is only for running/confirming a host provision worked.

remove\_cluster\_if\_exists removes the CM cluster containing the agents you specify and then removes the agent hosts from CM.

replace\_cluster removes existing cluster if it exists then re-uses agent hostnames.

#### CDEP\_HASH

Specify the QE/deploy Git repo hash/branch for CDEP

#### OPTIONAL\_ARGS

The log4j vulnerability is no longer removed by infrastructure scripts. Our product builds now remove the vulnerability. If you need to run the infrastructure script, you can add --remove-jndi-lookup-from-jars-value=TRUE.



*Note: We need to mention what version of the private service cluster we want to install. In this case we are using 1.5.4*

Step 4) Click on the Build button at the bottom.

## How to create a Base Cluster using Jenkins job?

Step 1) Log in to <https://master-01.jenkins.cloudera.com/job/Cluster-Setup-CDPD>

Step 2) Click on the “Build with Parameters” links from the left panel.

Step 3) Fill in all the parameters as shown below:

**Parameters**

**CLUSTER\_SHORTNAME**

**REQUIRED:** Hosts in the cluster to deploy CM/CDH on.  
The CLUSTER\_SHORTNAME format varies according to cloud type.  
**When using ycloud (root.comops.site), the prefix should be ccycloud (Provisioning will be done via Cloudcat).**  
**When not using ycloud, the prefix should be completely removed.**  
Sample GCE value: jsmith-hive-{1..3}  
Sample YCLOUD value: ccycloud-{1..3}.jsmith-hive  
Sample Openstack value: jsmith-openstack-{1..3}  
Change CLUSTER\_SHORTNAME and DOMAIN based on where the Ycloud Queue resides.

Note: you can destroy ycloud hosts [here](#), adjust expiration [here](#), and find clusters (with search) [here](#)  
Specify just the base hostname, e.g. jsmith-hive, and expiration time is expressed in hours.

**CM\_VERSION**

Version of Cloudera Manager. Accepted formats:

- Labels (e.g.: specific version such as 'cm7.2.0')
- Global Build Numbers (e.g.: 'gbn://123456')
- BQueries (e.g.: 'bquery://product=cm:version=7.2.0:official')

**Latest CM build labels:**

- **C7:** cm7.12.1, cm7.12.0, cm7.11.3, cm7.11.2, cm7.11.1, cm7.11.0, cm7.10.1, cm7.10.0, cm7.9.7, cm7.9.6

**CDH**

Version of CDPD (merged CDH+HDP). Note the usage of versions with 4 fields.  
Labels, GBNs and BQueries are supported, as with CM\_VERSION.  
The BQuery :L0\_PASSED tag indicates you want a build that has passed 'L0' testing. Remove it to use latest and greatest, untested.

**DOMAIN**

The domain where the hosts are/will be. root.comops.site is internal YCloud, vpc.cloudera.com is Amazon EC2.

*Note: We need to mention what version of the CM and CDH we want to install. In this case, we are using 7.13.1.0 for CM and 7.3.1 for CDH. For more information about the compatibility you can follow this page <https://cloudera.atlassian.net/wiki/spaces/ENG/pages/1040351461/Cloudera+Manager+versions>*

**CLOUDCAT\_OS**

**(Provision only)** OS for provisioned hosts.

**CLOUDCAT\_MASTER\_SIZE**

**(Provision only)** Size minimal criteria of the master node (CM Server) instance (Recommended: 2xlarge)

**CLOUDCAT\_SLAVE\_SIZE**

**(Provision only)** Size minimal criteria of the slave node instances

**CLOUDCAT\_EXPIRATION\_DAYS**

**(Provision only)** Days/hours to retain cloudcat host groups, maximum 21 days. Pick the mininum for your needs, saving cost!

**CLOUDCAT\_USERNAME**

**(Provision only)** Specify your cloudcat username. BUILD\_USER\_ID is your username in Jenkins.

☒ CLOUDCAT\_PRESERVE\_PAUSED\_EXPIRED

**AUTO\_PAUSE**

**(Provision only)** Enables automatic pausing of your VM group to reduce cost. The group will be paused when there is minimal CPU load and there are no active pseudo terminals. You can constrain the hours to nighttime (8pm to 8am) for the selectable timezones, constrain to Saturday/Sunday (Pacific) only, or allow auto-pausing at any time. Specifying auto-pause as None suppresses auto-pausing. Applicable to vpc.cloudera.com and gce.cloudera.com only.

#### CLOUDCAT\_BUDGET

(Provision only, optional) Allows you to charge a specified project rather than your personal allocation. When empty, Cloudcat uses your set default budget. See [The Cloudcat Budgeting Primer](#)

☐ SPOT

#### SPOT\_DEFINED\_DURATION

**CAUTION: (Provision only)** This option can only be used on EC2. Use this option if you need the cluster for less than 6 hours. Using this option will lead to 30-50% saving on the cluster cost. Applicable to vpc.cloudera.com only.

#### CLOUDCAT\_SHORTNAME

(Optional) All hosts in the cloudcat provisioning group. This parameter is typically used when one needs extra hosts alongside the CDH cluster (e.g.: KTS etc.). This is the group definition for Cloudcat (total hosts to provision). If left empty, value defaults to 'CLUSTER\_SHORTNAME'.

**Example usage:** CM+CDH cluster needs 5 hosts: 3 for CDH nodes and 2 for KTS nodes.

**NOTE:** If your CLOUDCAT\_SHORTNAME does not equal CLUSTER\_SHORTNAME, and it is on ycloud, you should specify a suffix carefully.

CLUSTER\_SHORTNAME = example-hosts-{1..3}

CLOUDCAT\_SHORTNAME = example-hosts-{1..5}

#### PROVISION\_ACTIONS

The make\_group\_available action will re-use hosts if they already exist, but will not update the expiration time.

The destroy\_and\_create\_group combination destroys the group if it exists, and then creates fresh guaranteed clean hosts. The destroy\_group action is a convenience so you can destroy hosts here instead of using a cloud UI.

#### MASTER\_NODE

(Optional) Specify the host running the CM server. Defaults to the 1st host in the cluster

#### EMAIL\_ADDRESS

(Optional) In addition to the user who starts the job, these recipients will get an email on completion. Accepted special values: none, noneIfBuildSuccess, noneIfBuildFailure

#### LICENSE

CM license to install

#### EXTRA\_PROVISIONING\_ARGS

(Provision only, optional) Specify Cloudcat arguments not covered above here. See [here](#) for available options. In case ycloud is used a new parameter is available: --ycloud-queue, and you can consult list of valid ycloud queues here: <http://rm.yprod.comops.io/ws/v1/cluster/scheduler>

#### DB

Select the database to use for the cluster

☒ AUTO\_TLS

#### KERBEROS

Enable Kerberos authentication

#### JAVA\_VERSION

Select the Java version to use for the cluster

**Note:** Among the above parameters Auto TLS and Kerberos enabling is very important here

#### JAVA\_VERSION

DETERMINE\_BY\_CDH

Select the Java version to use for the cluster

#### ACTIONS

clean setup

Select the deployment actions to perform on the cluster

#### CDEP\_HASH

origin/master\_vetted

Specify the QE/deploy Git repo hash/branch for CDEP

#### OPTIONAL\_ARGS

--install-python-version=3.8 --include-service-types=ZOOKEEPER,HDFS,HBASE,OZONE

**(Optional)** Specify custom sysadmin arguments not covered above. Use './target/env/bin/sysadmin --help' from the CLI to get available options. Eg: --ha-service-types=ALL

A snapshot of the available options is listed [here](#).

**January 20 2022: the log4j vulnerability is no longer removed by infrastructure scripts. Our product builds now remove the vulnerability. If you need to run the infrastructure script, you can add --remove-jndi-lookup-from-jars-value=TRUE.**

The include-service-types above is recommended for CDPD parcel setups.

RANGER, ATLAS and KNOX are all suppressed unless KERBEROS is enabled. See QAINFRA-7946 for explanation.

If you do not need all the services listed, you can get a faster setup by eliminating undesired entries.

Other service types: RANGER\_KMS, RANGER\_KMS\_KTS, DAS

Newer CM with py3 requires the following arg: --install-python-version=3.8, for older builds, please remove

☐ SAVE\_TEMPLATE

☐ USE\_TEMPLATE\_IF\_AVAILABLE

**Note:** In --include-service-types you will specify the services you want to install as part of the base cluster.  
Step 4) Click on the Build button at the bottom.