



Hewlett Packard
Enterprise

HPE Private Cloud AI 1.5 Deployment Guide (Service Edition)

Part Number: 20-PCAI-DGSE
Published: September 5, 2025

HPE Private Cloud AI 1.5 Deployment Guide (Service Edition)

Abstract

This document describes how to deploy the on-premises components of the HPE Private Cloud AI engineered system at a customer site. It supports HPE Private Cloud AI version 1.5 and applies to the Small, Medium, and Large configurations. Do not use this document for the Developer System configuration. This document is updated frequently; the information contained is accurate as of the publication date listed. This document is CONFIDENTIAL and is intended only for the HPE personnel.

Part Number: 20-PCAI-DGSE
Published: September 5, 2025

© Copyright 2024–2025 Hewlett Packard Enterprise Development LP

Notices

The information provided here is subject to change without notice. Hewlett Packard Enterprise's products and services are covered only by the express warranty statements that come with them. This document does not constitute an additional warranty. Hewlett Packard Enterprise is not responsible for any technical or editorial errors or omissions in this document.

Confidential computer software. You must have a valid license from Hewlett Packard Enterprise to possess, use, or copy the software. In accordance with FAR 12.211 and 12.212, Commercial Computer Software, Computer Software Documentation, and Technical Data for Commercial Items are licensed to the U.S. Government under the vendor's standard commercial license.

Links to third-party websites will take you outside of the Hewlett Packard Enterprise website. Hewlett Packard Enterprise has no control over and is not responsible for the information outside the Hewlett Packard Enterprise website.

Acknowledgments

Microsoft® and Windows® are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.

Linux® is the registered trademark of Linus Torvalds in the U.S. and other countries.

NVIDIA® and NVIDIA logos are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries.

VMware®, VMware NSX®, VMware vCenter®, and VMware vSphere® are registered trademarks or trademarks of VMware, Inc. and its subsidiaries in the United States and other jurisdictions.

Red Hat® Enterprise Linux is a registered trademark of Red Hat, Inc. in the United States and other countries.

Revision history

Publication date	Solution version	Configuration type	Summary
September 2025	1.5	Small, Medium, and Large	Initial release

Table of contents

- Overview
 - Onsite deployment workflow
- Preparing
 - Reviewing safety and compliance information
 - Gathering required tools
 - Acclimatizing the storage system
 - Unpacking the components
 - Firewall rules
 - Networking requirements
- Powering
 - Powering on the rack
 - Powering on the HPE ProLiant servers
- Installing
 - Validating the rack and cable connections
 - Connect the Private Cloud AI system to the external network
 - Connecting the Private Cloud AI solution to the data center network via high speed uplink ports
 - Connecting the Private Cloud AI solution to the data center network via gateway servers
 - Validating the connectivity of the system
- Configuring components
 - Upgrading the existing systems to the latest HPE Private Cloud AI version
 - Applying the Red Hat license to each worker nodes
 - Validating NTP server
 - Validating DNS entries
 - Setting up HPE GreenLake for File Storage
 - Rebooting the CBoxes
 - Launching and activating a HPE GreenLake for File Storage cluster
 - Initial TUI configuration
 - Enabling HPE GreenLake cloud services
- Connecting to Private Cloud AI service from Data Services Cloud Console
 - Configuring the Data Services Connector using the Private Cloud AI On-Prem Setup wizard
 - Adding a system using the Private Cloud AI Setup wizard
 - Setup Wizard: Configuring infrastructure
 - Setup Wizard: Configuring Private Cloud AI
 - Protecting the Data Services Connector VM
 - Applying licenses
- Component identification
 - HPE ProLiant server component identification
 - HPE GreenLake for File Storage component identification
 - Aruba 6300 series switch component identification

- HPE M-series switch component identification
- Specifications
 - Environmental specifications
- Setting up OpsRamp gateway component of the HPE Private Cloud AI
- Getting started with HPE Private Cloud AI documentation
- Appendix A: Alternative virtualization - VMware-based procedures
 - Setting up HPE GreenLake for File Storage and Private Cloud AI service on VMware-based systems
 - Mounting the HPE GreenLake for File Storage NFS share as a datastore in ESXi
 - Using the Private Cloud AI Setup wizard on VMware based systems
- Websites
 - Documentation feedback

Overview

HPE Private Cloud AI is a purpose-built infrastructure optimized for AI workloads, particularly generative AI applications. It leverages extensive experience in high-performance computing, data management, and enterprise IT to create a unified environment where businesses can build, deploy and operate AI models with unprecedented ease and efficiency.

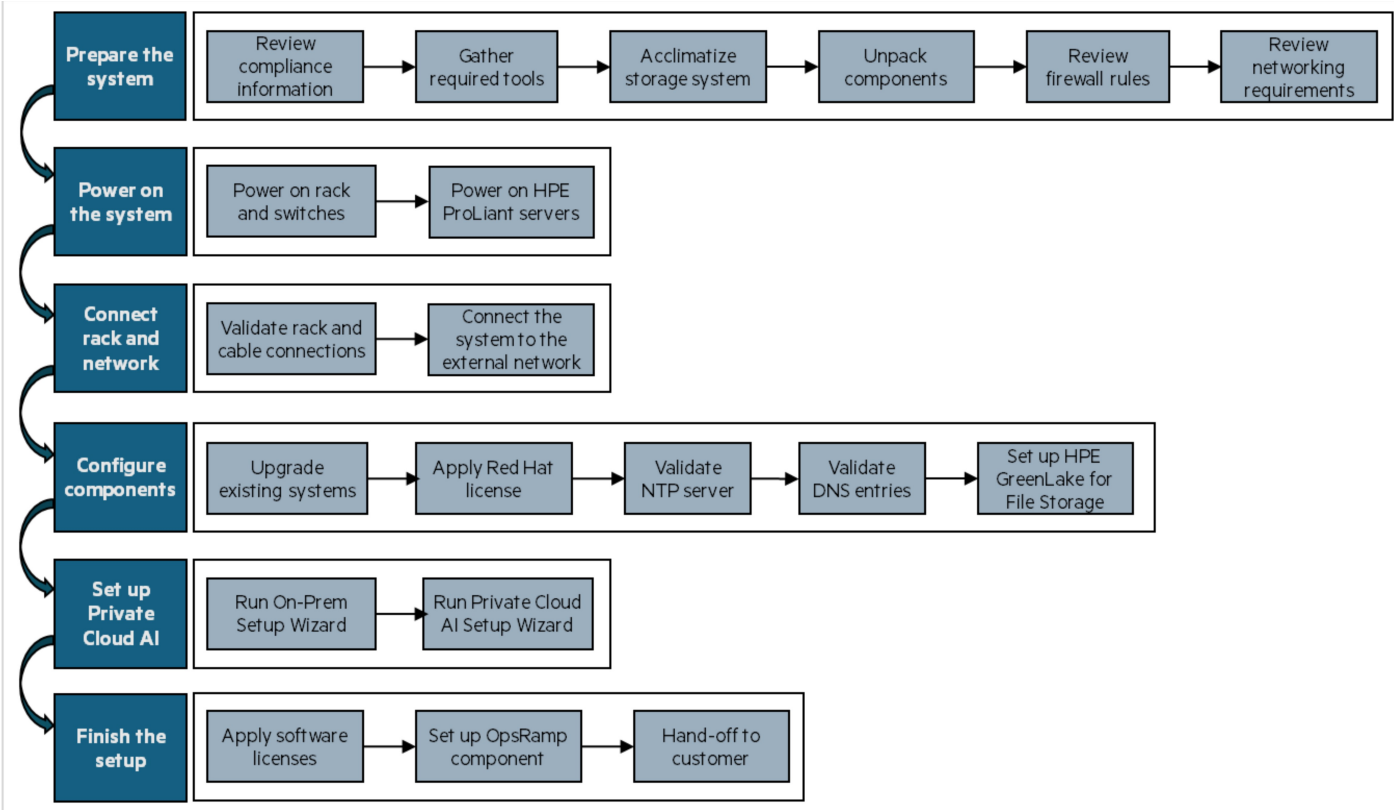
HPE AI Essentials, the core engine of HPE Private Cloud AI, offers a comprehensive turnkey solution for Generative AI applications. This platform integrates pre-configured AI and data foundation tools under a unified control plane, providing adaptable solutions tailored to specific business needs. It combines cutting-edge NVIDIA AI computing technology with enterprise-grade security and compliance measures, enabling swift deployment, management, and scaling of AI initiatives.

Subtopics

Onsite deployment workflow

Onsite deployment workflow

The following chart shows the high level workflow for the deployment of HPE Private Cloud AI engineered system at a customer site:



Preparing

Subtopics

- Reviewing safety and compliance information
- Gathering required tools
- Acclimatizing the storage system

Reviewing safety and compliance information

About this task

Review the [Safety and Compliance Information for Server, Storage, Power, Networking, and Rack Products](#) guide before you begin unpacking and installing the storage system.

Gathering required tools

Procedure

1. Gather the following tools before you begin the unpacking and installation process:

Purpose	Tools
Preventing ESD	- ESD mat - ESD grounding strap
Opening shipping boxes	- Scissors or snips - Box cutter
Configuring switches	Serial port cable

2. Verify that the following tools and software links can be accessed to complete deployment of the HPE Private Cloud AI solution:
- A connection to the onsite management network is required to initiate deployment of the solution. If the customer does not allow access to the management network with a HPE laptop, the customer must provide a workstation that has access to the management network with the following tools:
 - SSH client installed
 - Web browser installed
 - The HPE Private Cloud AI Installer must have access to the following links to successfully install or update HPE AI Essentials Software as part of the solution on the cluster nodes:
 - <https://solutionhub-metadata.s3.us-east-1.amazonaws.com/>
 - <https://marketplace.us1.greenlake-hpe.com>
 - <https://gl-vas-harbor-prod-images.s3.us-west-2.amazonaws.com/>
 - <https://github.com/HPEEzmeral>

Acclimatizing the storage system



About this task



CAUTION

To prevent damage to the storage system, do not power on the rack or storage components until they are fully acclimatized.

Procedure

1. Note the temperature of the location where the shipment package has been stored and the temperature of the data center.
2. If the difference between the two temperatures is greater than 36°F (20°C):
 - a. Move the package to the data center.
 - b. Leave the package in a dry area away from any direct source of moist air. Do not open the packaging right away.
3. Check for condensation on the shipment package. If it exists, leave the shipment package as is for at least 24 hours, and wait for the condensation to fully evaporate.
4. After the package has been fully acclimatized to the data center environment, unpack the package.

Unpacking the components

Prerequisites

Ensure that there is adequate space in the location where you unpack the contents of the shipment.

To protect the components, Hewlett Packard Enterprise recommends moving them into the computer room in their shipping containers and unpacking them near the installation location. If facility requirements mandate that packing material cannot be taken into the computer room, unpack the components in an ESD protected environment. Carefully transport them to the installation location using ESD protective measures.

About this task



CAUTION

- Observe local occupational safety requirements and guidelines for heavy equipment handling.
- Use proper grounding techniques to prevent Electrostatic Discharge (ESD) damage to the components.

The factory integrated system arrives on a pallet.

Procedure

1. Inspect the packaging for damage and report any issues to the Hewlett Packard Enterprise Support Center.
2. Verify that the shipment contents match the SKUs and contents listed on the packing slip.
3. Follow the instructions printed on the outer packaging before unpacking the hardware components. Pay attention to cautions on the boxes and read any Read This First document.
4. Save the shipping boxes and packing material for shipping the system without the rack if necessary.
5. Cut all the packing straps and remove the shrink wrap, if present.

Firewall rules

Firewall rules

This section provides information about the firewall rules that are required to gain access through the customer's firewall.

For firewall and port information, see the [Site Readiness Checklist](#).

Networking requirements

The HPE Private Cloud AI solution must be connected to an external network to be fully functional:

- Administrators are able to connect to the management plane for managing and monitoring the system
- Developers can access HPE AI Essentials for implementing AI workloads
- Users can access the HPE Private Cloud AI from their personal devices to use AI services implemented on top of the solution
- Some components in the system are able to connect to the Internet

Uplink ports on switches

Each HPE Private Cloud AI size defines some dedicated switch ports on the data network, to connect to an external infrastructure.

Solution configuration size	Switches used	Uplink ports on each switch	Number of ports required (min/max)	Max throughput
Small	SN4600 switches	51, 52	2/4	400Gb/s
Medium	SN4700 switches	25	2/2	800Gb/s
Large	SN4700 (storage rack)	27, 28	2/4	1600Gb/s



NOTE

If additional uplinks are required, consult HPE Support to know how to configure them.

Network bandwidth and software updates

HPE Private Cloud AI requires a sustained internet connection of at least 100 Mbps to ensure that software updates finish in hours and not days. For example, at 100 Mbps, the pre-staging step for HPE AI Essentials can take up to eight hours if 100% of the software bits need to be updated. In practice, it generally takes about three or four hours as only a subset of the components usually need to be updated.

Powering

Subtopics

- [Powering on the rack](#)
- [Powering on the HPE ProLiant servers](#)

Powering on the rack

Prerequisites

All power cables have been connected to the components in the rack.

Procedure



1. Connect the power cables to the PDUs in the rack.
2. Flip the breaker on the PDUs to power on the switches and HPE GreenLake for File Storage components.

**NOTE**

The network switches and HPE GreenLake for File Storage components are now powered on. The HPE ProLiant servers are powered on individually later.

3. Verify that the System Status and System Power LEDs on the switches are green.

Powering on the HPE ProLiant servers

Procedure

1. Power on the HPE ProLiant control plane servers.
 - a. Remove the bezel.
 - b. Press the Power On/Stand By button to power on the server.
 - c. Replace the bezel.
 - d. Verify that the Power and Health LEDs are green indicating that the server is powered on.
 - e. Repeat the previous steps for all HPE ProLiant control plane servers.
2. Power on the HPE ProLiant worker nodes.
 - a. Press the Power On/Standby button to power on the server.
 - b. Verify that the LED changes from amber to green indicating that the server is powered on.
 - c. Repeat the previous steps for all HPE ProLiant worker nodes.

Installing

Subtopics

[Validating the rack and cable connections](#)

[Connect the Private Cloud AI system to the external network](#)

Validating the rack and cable connections

Prerequisites

The rack must be installed in the customer data center.

About this task

This section describes the procedure to verify that no cables were disconnected during shipping or installation of the rack.

Procedure

1. Verify that the power cables are connected to all components in the rack.

2. Verify that all network cables are connected.

**NOTE**

Ensure that all network cables and rack-to-rack internal cables are connected according to the customer-specific cabling guidance for the solution, also known as the point-to-point cabling guide.

- a. Verify that all network cables are connected from the switches to the HPE ProLiant servers.
- b. Verify that all network cables are connected from the switches to the HPE GreenLake for File Storage components.

Connect the Private Cloud AI system to the external network

About this task

This section outlines the procedure to connect the Private Cloud AI system to the external network.

Procedure

1. Connect the Private Cloud AI system to the external network using one of the following steps:
 - a. [Connect the Private Cloud AI solution to the data center network through high speed uplink ports](#).
 - b. [Connect the HPE Private Cloud AI solution to the data center network through one or two gateway servers](#).
2. [Validate the connectivity of the system after the uplinks are connected](#).

Subtopics

[Connecting the Private Cloud AI solution to the data center network via high speed uplink ports](#)
[Connecting the Private Cloud AI solution to the data center network via gateway servers](#)
[Validating the connectivity of the system](#)

Connecting the Private Cloud AI solution to the data center network via high speed uplink ports

Prerequisites

- **Physical connectivity:** Two or four high speedlinks must be used to connect to the external data center.

These high speed links must be distributed to at least two switches and must be bonded using LACP. The list of compatible transceiver for the NVIDIA switches is available in the [NVIDIA Hardware Compatibility List](#). In case of DAC cable or AOC, ensure that the cable is also compatible with the remote switch.

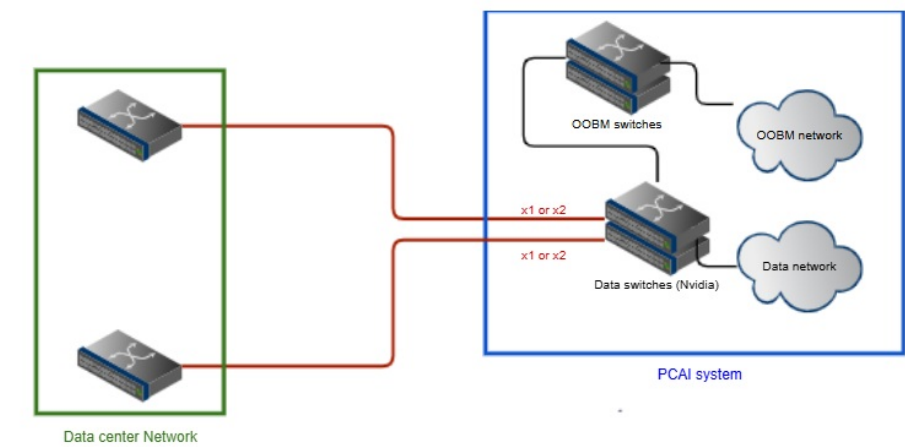
**NOTE**

The actual capacity of these links will affect the network bandwidth to the network. Hewlett Packard Enterprise recommends that the capacity of these links must be at least 100Gb total bandwidth.

- **Network layout:** Ensure that both the Management network and Production network VLANs are extended through the network uplinks.
 - Both the networks must be configured on the data center network with the corresponding VLAN ID.
 - Both the networks provide access to the external network wherein a gateway exists on these networks that can easily route to the external network.

About this task

This section provides describes the procedure to connect the HPE Private Cloud AI solution to the data center network through the high speed uplink ports.



Procedure

Data Center configuration:

- 1. Verify the ingress data center network ports are configured.
 - a. Verify that all the ports are a member of the LACP bond.
 - b. Verify the bond is configured as a trunk port where both the Management and Production VLANs are allowed.

HPE Private Cloud AI configuration:

- 2. Verify the required configuration based on the type of cable used to connect to the data center.
 - a. Verify the port can work at the correct speed if straight cables are used.
 - b. Split the corresponding ports if splitter cables are used.

Connecting the Private Cloud AI solution to the data center network via gateway servers

Prerequisites

- **Physical connectivity:** Two uplinks must be connected to one or two servers, which play the role of gateways. These servers must have several high performance Ethernet ports.

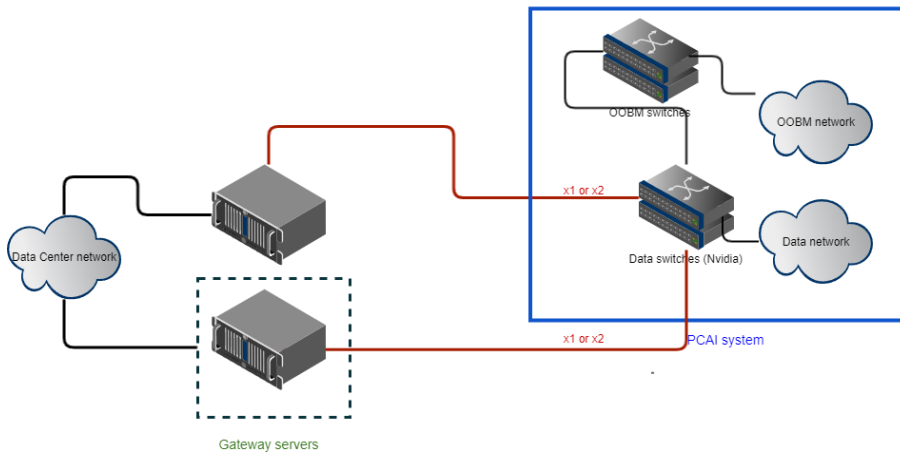
The following table includes a list of (non-exhaustive) adapters with the corresponding uplinks that can be used to connect to the NVIDIA switch:

Adapters	Connectors
2p 100Gb Cx6 IB/EN adapters	2x QSFP28 to QSFP28 Cables (R9F76A)
2p 200Gb Cx7 IB/EN adapters	2x QSFP56 to QSFP56 Cable

- **Network layout:** No action is required at the data center network side. The following features must be enabled at the gateway server side:
 - Routing must be enabled for egress traffic to the data center network
 - NAT must be configured for any packet coming from the Private Cloud AI setup

About this task

This section describes the procedure to connect the HPE Private Cloud AI solution to the data center network through one or two gateway servers. This procedure is useful in an environment where creating VLANs could be complex on a corporate network. Usually, this procedure requires a user or the admin to connect to a gateway before having access to the system.



Procedure

Data Center configuration:

1. Check the configuration of each gateway server.
 - a. Verify that one LACP bond on the NIC port is connected to the Private Cloud AI solution.
 - b. Verify that one virtual interface is configured for the management network.
 - c. Verify that one virtual interface is configured for the production network.
 - d. Verify that all the egress traffic from the gateway to the data center network must be configured as NAT.

HPE Private Cloud AI configuration:

2. Verify the required configuration to connect to the data center.
 - a. Verify that the default configuration is removed for the uplink port.
 - b. Verify that the LACP bond is created for every port group, which is connected to the same gateway server.
 - c. Verify that both the Management and Production network are allowed through the bond.
 - d. Verify the port can work at the correct speed if straight cables are used.
 - e. Split the corresponding ports if splitter cables are used.

Validating the connectivity of the system

About this task

This section provides the procedure to perform the post configuration check to validate the connectivity of the system to the Private Cloud AI solution.

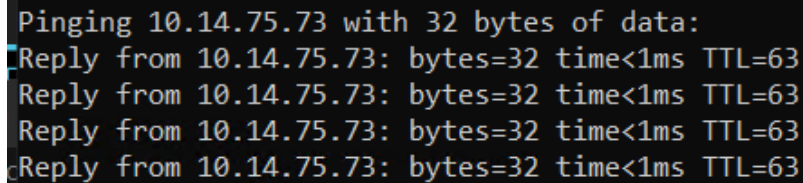
Procedure

1. Retrieve the IP address and user credentials assigned to the worker nodes from the `base-configuration.json` and `infra-layout.json` SCID files.
 - In the `base-configuration.json` file, find the `"ipAddress"` field for all the servers starting from the `"componentId": "server-004"` under `"infrastructure" > "racks" > "servers" > "networkConnections"` to retrieve the IP addresses for the worker nodes.

- In the `infra-layout.json` file, find the `"userName"` and `"password"` fields for `"type": "DL380a"` under `"racks" > "servers" > "accessCredentials"` to retrieve the user credentials.
2. Check the list of iLO addresses assigned to the worker nodes.
 3. Validate the system has access to the management network.
 4. On a Linux or Windows system terminal, run the following command to validate that the iLO IP address is reachable:

```
ping  
<iilo-address>
```

The output displays a response over the network indicating that the connection is established as illustrated in the following figure:



```
Pinging 10.14.75.73 with 32 bytes of data:  
Reply from 10.14.75.73: bytes=32 time<1ms TTL=63  
Reply from 10.14.75.73: bytes=32 time<1ms TTL=63  
Reply from 10.14.75.73: bytes=32 time<1ms TTL=63  
Reply from 10.14.75.73: bytes=32 time<1ms TTL=63
```

5. Connect to `https://<iilo-address>` from the web browser to verify that you can log in to the iLO web interface.
6. Repeat steps 4 and 5 for all worker node iLO IP addresses.

Configuring components

Subtopics

- [Upgrading the existing systems to the latest HPE Private Cloud AI version](#)
- [Applying the Red Hat license to each worker nodes](#)
- [Validating NTP server](#)
- [Validating DNS entries](#)
- [Setting up HPE GreenLake for File Storage](#)

Upgrading the existing systems to the latest HPE Private Cloud AI version

This section provides a list of issues and workarounds to upgrade the existing systems to the latest HPE Private Cloud AI version with the corresponding recipe version at a customer site.

For a list of issues that are yet to be fixed or are fixed in a specific version, see [HPE Private Cloud AI 1.4.1 Known Issues](#).

Applying the Red Hat license to each worker nodes

Prerequisites

Confirm that the customer has received an email receipt from Hewlett Packard Enterprise containing the order details and instructions to register the products. This email is sent to the contact on the Purchase Order.

Procedure

1. Customer activates the purchased products in the My HPE Software Center.

HPE sends an email confirmation of successful registration within the My HPE Software Center.

2. Customer completes the subscription assignment online using the following steps:

- a. Log in to the customer account at <https://redhat.com>.
- b. Navigate to Manage Accounts > Subscriptions > Subscription Inventory > Subscription Activation.
- c. Enter the Subscription Activation number that was provided during HPE registration.

3. Service personnel registers the Red Hat subscription on the worker node.

- a. Log in to the worker node.
- b. Register the worker node using the following command:

```
subscription-manager register
```



NOTE

Alternatively, you can run the following command to register the worker node:

```
subscription-manager register --org=  
<RedHat org number>  
--activationkey="  
<activation key name>
```

- c. When prompted, the customer enters their Red Hat login credentials.
- d. Attach the appropriate subscription to your system using the following command:

```
subscription-manager attach --auto
```



NOTE

Alternatively, you can run the following commands to manually specify a pool to attach any available subscription that matches the registered system:

```
sudo subscription-manager list --available  
sudo subscription-manager attach --pool=  
<pool-id>
```

If the Red Hat account of the customer is configured for Simple Content Access (SCA), this command displays Ignoring the request to auto-attach. Attaching subscriptions is disabled for organization "#####" because Simple Content Access (SCA) is enabled. Ignore this message.

Validating NTP server

Procedure

1. Open an SSH session to the worker node.
2. Test NTP server reachability using the following command:

```
chronyc sources
```

Example:

chronyc sources

```
MS Name/IP address         Stratum Poll Reach LastRx Last sample
=====
====
^* time.example.com        2   10   377   195   +25us[ +34us]
+/-      17ms
```

In the output, look for `*` that displays the currently selected source of time.

3. Check the synchronization status.

```
chronyc tracking | grep "Leap status"
```

If the output of this command shows `Leap status` as `Normal`, it signifies that the system is synchronized.

Example:

```
# chronyc tracking | grep "Leap status"
Leap status      : Normal
```

Validating DNS entries

Procedure

1. Open an SSH session to the worker node.
2. Verify that the DNS entries are resolved using the `nslookup` command.
3. Validate the FQDN for all A records and the PTR records for each IP address.

Setting up HPE GreenLake for File Storage

Installing and setting up the HPE GreenLake for File Storage is divided into two phases. Phase 1 is to be executed at the integration center. This document covers phase 2, install and setup of HPE GreenLake for File Storage at the customer site.

Subtopics

[Rebooting the CBoxes](#)

[Launching and activating a HPE GreenLake for File Storage cluster](#)

[Initial TUI configuration](#)

[Enabling HPE GreenLake cloud services](#)

Rebooting the CBoxes

Prerequisites

- The racking and cabling of the system must be completed.
- The network is configured as per the customer requirements at the Integration Center.

If the customer network is not configured at the Integration Center, update the following network configurations with the required

customer site IP addresses as mentioned in the [HPE Private Cloud AI Onsite Customer Network Update Guide](#):

- Change the IP addresses on the PDUs
- Change the IP addresses on the NVIDIA switches
- Change the IP addresses on the HPE Aruba 6300 series switches
- Change the IP addresses on the worker nodes
- Change the IP addresses on the HPE Alletra Storage MP network switches
- The required drive firmware must be updated on the DNodes.

If the firmware is not updated, perform the steps mentioned in the [Drive firmware is not updated on the DNodes for DBoxes](#) section of the HPE Private Cloud AI 1.4.1 Known Issues.

About this task

The HPE GreenLake for File Storage components are powered on when the power is applied to the PDU. Reboots of the CBoxes are required after powering on to ensure correct operation of the cluster.

Procedure

1. Open an SSH connection to cnode1 using the management IP.

2. Log in with the credentials

```
root/rootvastdata
```

.

3. Reboot the system using the following command:

```
sudo -u vastdata clush -g cnodes 'sudo shutdown -r'
```



NOTE

This command has a one minute delay to ensure that all nodes in the cluster receive the command before the node running the `clush` command reboots.

Wait 10 minutes to ensure all the CBoxes have rebooted.

Launching and activating a HPE GreenLake for File Storage cluster

About this task



NOTE

Use MGMT CNode IP to remove the files (VMS IP comes up after deleting the files).

Procedure

1. Log in to CNode management IP as

```
root/rootvastdata
```

.

2. Change the user to

```
vastdata
```

using the following command:

```
su vastdata
```

3. Remove the disable leader election file from all CNodes using the following command:

```
clush -g cnodes 'sudo rm -f /vast/data/DONT_RUN_LEADER'
```

4. Verify that the file is removed from all CNodes using the following command:

```
clush -g cnodes 'ls /vast/data/DONT_RUN_LEADER'
```

Example:

```
Rack-CB1-U-bottom ~:$ su vastdata
Rack-CB1-U-bottom ~:$ clush -g cnodes 'rm -f
/vast/data/DONT_RUN_LEADER'
Rack-CB1-U-bottom ~:$ clush -g cnodes 'ls /vast/data/DONT_RUN_LEADER'

172.16.128.1: ls: cannot access '/vast/data/DONT_RUN_LEADER': No such
file or directory

clush: 172.16.128.1: exited with exit code 2

172.16.128.2: ls: cannot access '/vast/data/DONT_RUN_LEADER': No such
file or directory

clush: 172.16.128.2: exited with exit code 2
```



NOTE

If the file is removed correctly, an error message is displayed that no such file or directory exists. This is expected behavior.

Wait for about two to three minutes until the VMS starts running.

5. Check if the output returns an IP address indicating that the VMS is running:

```
find-vms
```

6. Ensure that the leader is up and running using the following command:

```
find-leader
```

```
Rack-CB1-U-bottom ~:$Rack-CB1-U-bottom ~:$ find-leader

172.16.128.1
Rack-CB1-U-bottom ~:$
```

Use the following commands to monitor or troubleshoot the process:

```
Rack-CB1-U-bottom ~:$ clush -g cnodes "ls -l /vast/data/vms-
bringup.log"

172.16.128.2: -rw-r--r--. 1 vastdata vastdata 1680273 Jun 19 12:32
/vast/data/vms-bringup.log

172.16.128.1: -rw-r--r--. 1 vastdata vastdata 7329877 Jun 19 12:32
/vast/data/vms-bringup.log
Rack-CB1-U-bottom ~:$

Rack-CB1-U-bottom ~:$ docker ps

CONTAINER ID    IMAGE
COMMAND        CREATED        STATUS        PORTS        NAMES
```

```

8017d8f35403  vastdata.registry.local:5000/dev/orion:release-5.1.0-
1340687-mc-vms  "/bin/sh -c 'envsubs..."  33 seconds ago  Up 32
seconds          mcvms

6fbadb6f6dd9c  vastdata.registry.local:5000/dev/orion:release-5.1.0-
1340687-vms     "/usr/bin/dumb-init ..."  2 minutes ago   Up 2 minutes
vast_vms

98c5b900a679   vastdata.registry.local:5000/dev/orion:release-5.1.0-
1340687-x86_64  "dumb-init /vast/bin..."  46 hours ago    Up 2 hours
vast_platform

7322f27f916f   local_registry:latest
"/entrypoint.sh /etc..."  7 days ago      Up 2 hours
registry

```

7. Open a web browser and use an https session to connect to the management IP for the HPE GreenLake for File Storage UI.



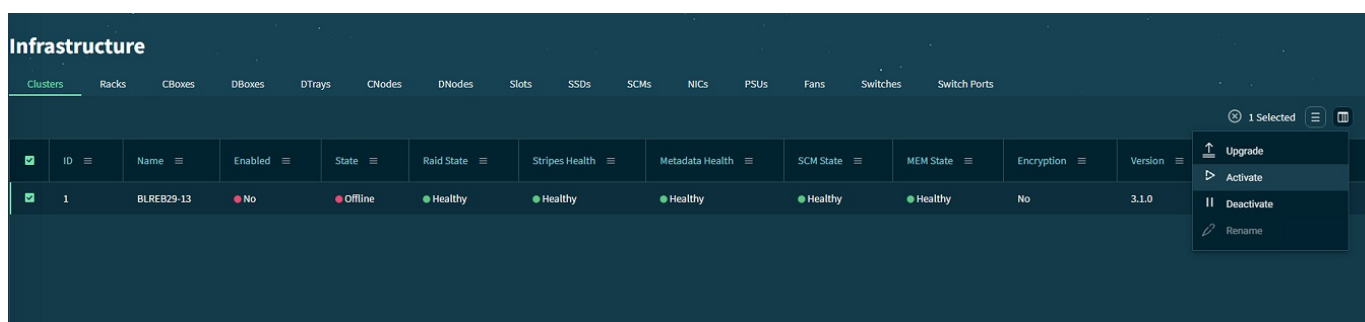
NOTE

The management IP is labeled as *sstor001* in the HPE Smart CID file.

8. Log in to the HPE GreenLake for File Storage UI using the user name and password specified under the **storageArrays > accessCredentials** section in the infra-layout HPE Smart CID file.

Wait for a few minutes until all nodes are activated.

9. Navigate to  Infrastructure from the left navigation menu to activate all the nodes and clusters.



10. In the Clusters tab, activate the cluster if it is not enabled using the following steps:
 - a. Select the cluster if the Enabled column shows No.
 - b. Select Activate from the Actions drop-down list displayed as a hamburger (three lines) icon on the top-right side of the page.
 - c. Click Yes to confirm.
11. In the CNodes tab, activate all the nodes that are not enabled or where the status is inactive using the following steps:
 - a. Select all the nodes where the Enabled column shows No or the State column shows Inactive.
 - b. Select Activate from the Actions drop-down list displayed as a hamburger (three lines) icon on the top-right side of the page.
 - c. Click Yes to confirm.
12. In the DNodes tab, verify that the State column shows Active for both the nodes.
13. Navigate to  Activities from the left navigation menu and check whether all the CNodes and clusters have been successfully activated.

14. Disable the collapse option on any CNode to resolve the VM Essentials Manager corruption issue.

For more information, see the [HPE GreenLake for File Storage becomes inconsistent while running VM Essentials Manager](#) section in the HPE Private Cloud AI 1.4.1 Known Issues

Initial TUI configuration

Procedure

1. Launch the HPE File Storage Setup Wizard through a TUI.
 - a. Open an SSH session to the management node using the management virtual IP address (VMS IP) for the storage system.
 - b. Log in to the TUI using the user name as

hpeadmin

and password as

hpedata

.

The default VMS IP is 192.168.11.10, but be sure to confirm the VMS IP with the network information.



NOTE

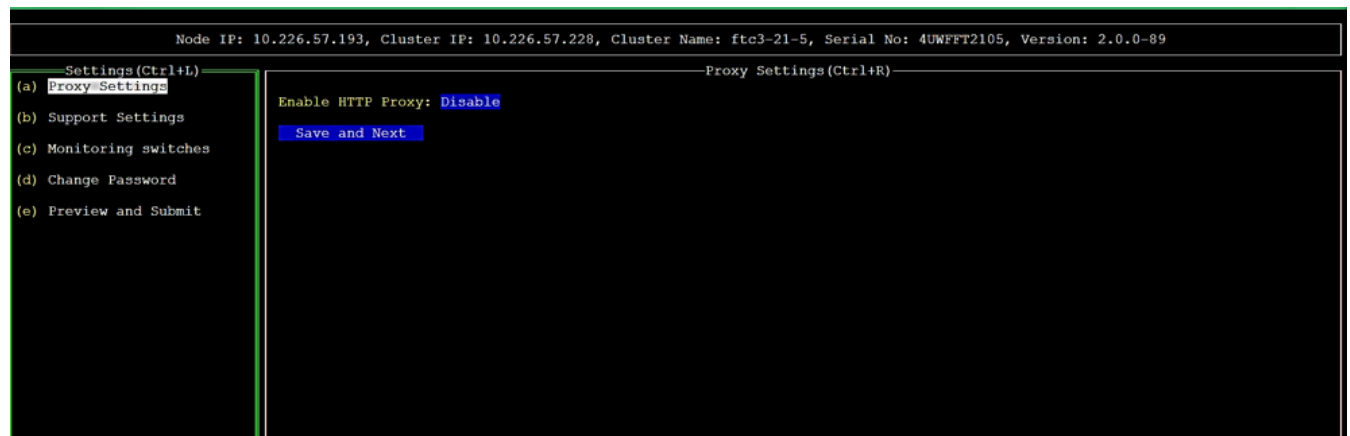
If the TUI does not start, use the SSH that is open from the previous session and run the following commands:

```
clush -g cnodes "sudo rm -rf /opt/sharedFolder/hpe/fileservice/"  
clush -g cnodes "sudo rm -f /vast/hpe/tuiconfig/setup_status.json"
```

2. Configure the HPE GreenLake Cloud Services through the Setup Wizard.

- a. On the Welcome screen, select Next to proceed to the Settings menu.

The Proxy Settings screen from the Settings menu is displayed.



- b. On the Proxy Settings screen, enter the appropriate customer proxy information from the base-configuration.json HPE Smart CID file if it is used and select Save and Next to proceed to the Support Settings screen:

Node IP: 10.226.57.193, Cluster IP: 10.226.57.228, Cluster Name: ftc3-21-5, Serial No: 4UWFFT2105, Version: 2.0.0-89

Settings (Ctrl+L)
(a) Proxy Settings
(b) Support Settings
(c) Monitoring switches
(d) Change Password
(e) Preview and Submit

Proxy Settings (Ctrl+R)
Enable HTTP Proxy: **Enable**
*Proxy Port: **443**
*Proxy Server: **proxy.am.hpecore.net**
Authentication Required: **Disable**
Save and Next

- c. On the Support Settings screen, verify that the Send Call Home is disabled and select Save and Next to proceed to the Monitoring switches screen.
- d. On the Monitoring switches screen under Switch Details, select Add Switch Entry to add switch information.

Node IP: 10.226.57.193, Cluster IP: 10.226.57.228, Cluster Name: ftc3-21-5, Serial No: 4UWFFT2105, Version: 2.0.0-89

Settings (Ctrl+L)
(a) Proxy Settings
(b) Support Settings
(c) Monitoring switches
(d) Change Password
(e) Preview and Submit

Switch Details (Ctrl+R) Scroll Up(↑), Scroll Down(↓)

Sequence No.	Switch IP	Switch User

Switch Setting (Ctrl+D)
Back **Save and Next** **Add Switch Entry** **Modify Switch Entry** **Delete Switch Entry**

- e. On the Switch Settings screen, add the appropriate switch information and select Add Switch.

Node IP: 10.226.57.193, Cluster IP: 10.226.57.228, Cluster Name: ftc3-21-5, Serial No: 4UWFFT2105, Version: 2.0.0-89

Switch Settings
*Switch IP: **10.226.57.191**
*Switch User: **admin**
*Switch Password: *********
Back **Add Switch**



NOTE

Repeat this step for the second switch.

The information for both the switches is displayed on the Switch Details screen.

Node IP: 10.226.57.193, Cluster IP: 10.226.57.228, Cluster Name: ftc3-21-5, Serial No: 4UWFFT2105, Version: 2.0.0-89

Settings(Ctrl+L)

- (a) Proxy Settings
- (b) Support Settings
- (c) Monitoring switches
- (d) Change Password
- (e) Preview and Submit

Switch Details (Ctrl+R) Scroll Up(↑), Scroll Down(↓)

Sequence No.	Switch IP	Switch User
1	10.226.57.190	admin
2	10.226.57.191	admin

Switch Setting (Ctrl+D)

Back Save and Next Add Switch Entry Modify Switch Entry Delete Switch Entry

When the information for both the switches is displayed, select **Save and Next** to proceed to the **Change Password** screen.

- f. On the **Change Password** screen, create a new password for all the accounts using the [Password guidelines](#) and select **Save and Next**.



NOTE

Ensure that you record the created passwords and share them with the customer after the deployment is completed.

Node IP: 10.226.57.193, Cluster IP: 10.226.57.228, Cluster Name: ftc3-21-5, Serial No: 4UWFFT2105, Version: 2.0.0-89

Settings(Ctrl+L)

- (a) Proxy Settings
- (b) Support Settings
- (c) Monitoring switches
- (d) Change Password
- (e) Preview and Submit

Change Password (Ctrl+R)

User: hpeadmin(*)
New Password: ***** Confirm Password: *****

User: hpecliadmin(*)
New Password: ***** Confirm Password: *****

User: hpelogs(*)
New Password: ***** Confirm Password: *****

User: admin (ipmi user) (*)
New Password: ***** Confirm Password: *****

User: admin (vms admin User) (*)
New Password: ***** Confirm Password: *****

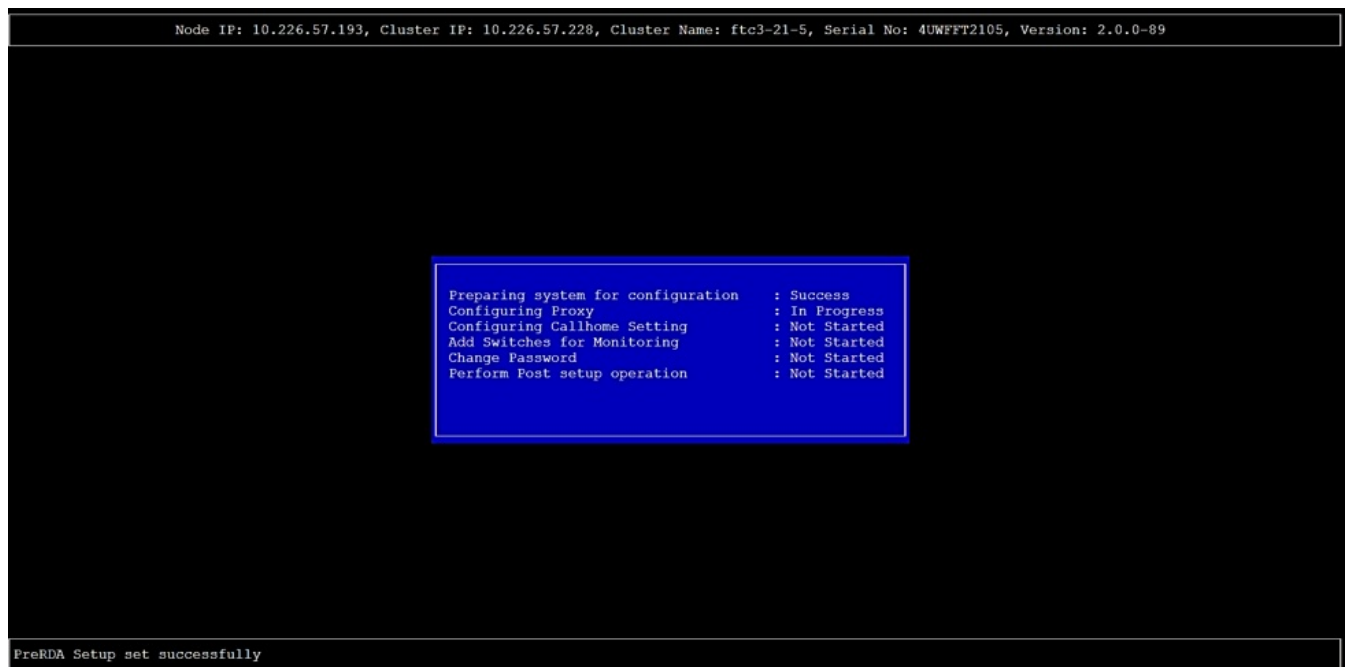
User: support (vms support User) (*)
New Password: ***** Confirm Password: *****

Save and Next Back View Password Rule

The **Preview and Submit** screen is displayed.

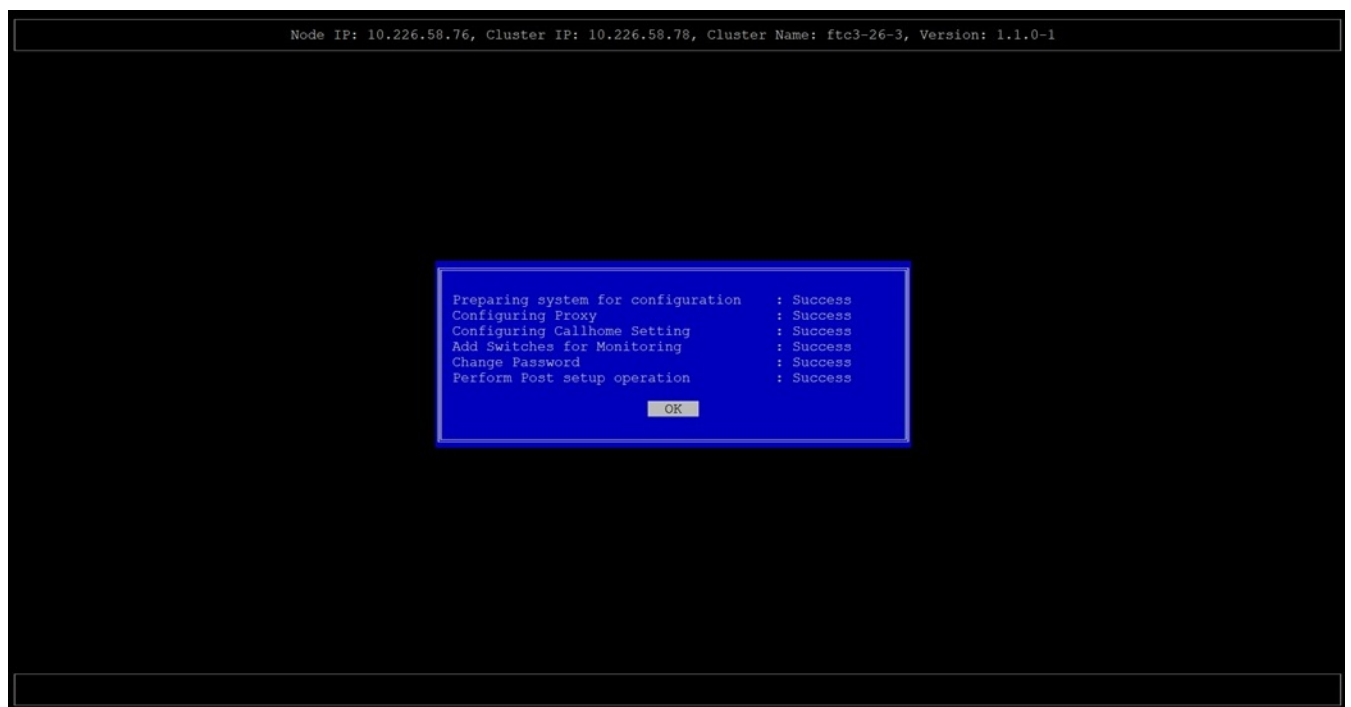
- g. On the **Preview and Submit** screen, review the configuration information and select **Submit**.

Wait until the processing is complete.



NOTE

You can also perform post setup updates to the configuration from TUI.



- h. Select OK to complete the configuration of the file services and return to the main TUI menu.

```

File Services
(a) Proxy settings
    Setup Node configurations -Proxy
(b) DNS settings
    Setup DNS settings
(c) Date & Time Settings
    Setup Date & Time Settings
(d) Support/Callhome settings
    Setup Support/Callhome settings
(e) Collect Support Data
    Ondemand initiate support data collection
(f) Password Management
    Password Management
(g) Manage Certificates
    Manage Device Certificates
(h) Manage File Services
    Restart callhome and cloud enablement services
(i) Set Logger Level
    Change logging level for file services
(q) Quit
    Press to exit

```

- i. Record the serial number for the system from the top of the menu.
3. Record the ciphertext for the root and hpesupport users from the TUI menu.
 - a. On the TUI main menu, select Password Management.
 - b. Select Secure Password Management.
 - c. Select View/Update Encryption or Time Based Password .
 - d. Select

root

 from the User drop-down list and click View Ciphertext.
 - e. Copy the blob including the first and last lines that use dashes.
 - f. Repeat the previous steps for the

hpesupport

 user.
4. Generate a strong password for the root and hpesupport users on the Navigator with the following steps:
 - a. Log in to <https://navigator.service.suptools.hpecorp.net/> using your HPE domain account.
 - b. Click the Systems menu at the top of the screen.
 - c. Enter the serial number recorded previously and press Enter to look it up.
 - d. Click the Request Password link at the top-right of the screen.
 - e. Select

user=root

 .
 - f. Paste the blob copied previously into the ciphertext box, including the starting and ending lines.
 - g. Click Submit.
 - h. Repeat the previous steps for the

hpesupport

 user.
5. Record the strong password generated for the root and hpesupport users to be used later during the deployment.

Enabling HPE GreenLake cloud services

Prerequisites

- The network on the CNodes and DNodes of the HPE GreenLake for File Storage switches is configured as per the customer requirements at the Integration Center.

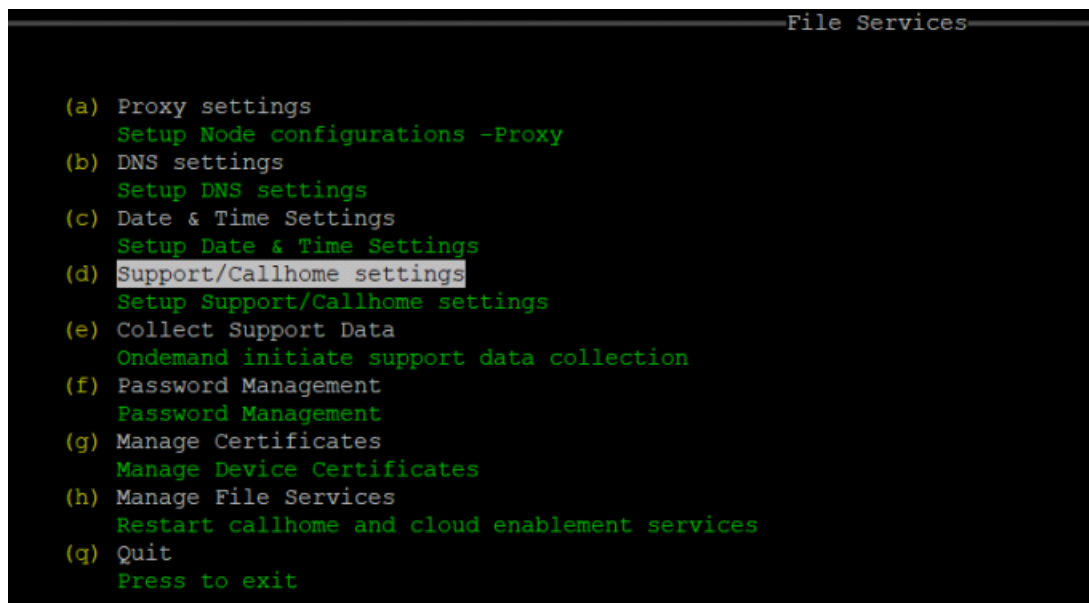
If the customer network is not configured at the Integration Center, update the CNodes and DNodes network configurations with the required customer site IP addresses as mentioned in the HPE GreenLake for File Changes section of the [HPE Private Cloud AI Onsite Customer Network Update Guide](#).

- The required virtual IP pool, view policies, views, user, and user group are created in the HPE GreenLake for File Storage cluster at the Integration Center.

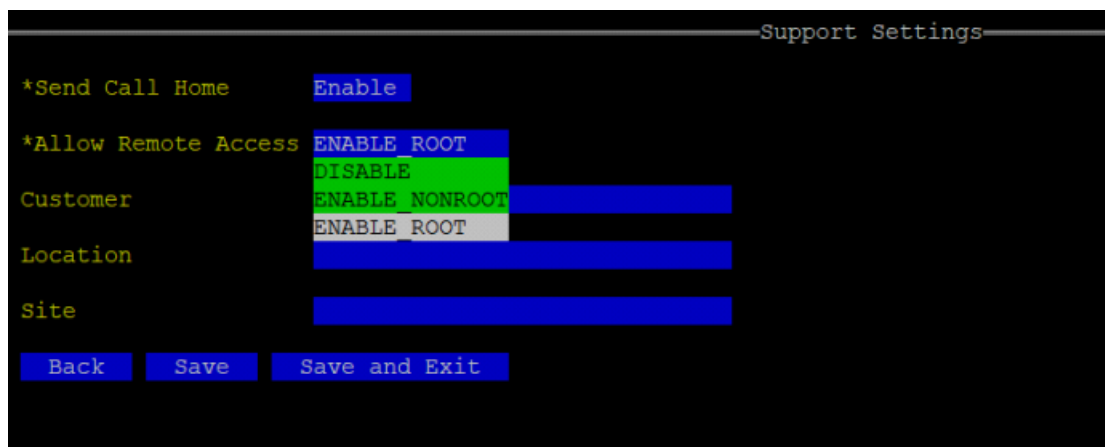
Procedure

1. Enable HPE Call Home from within the TUI.

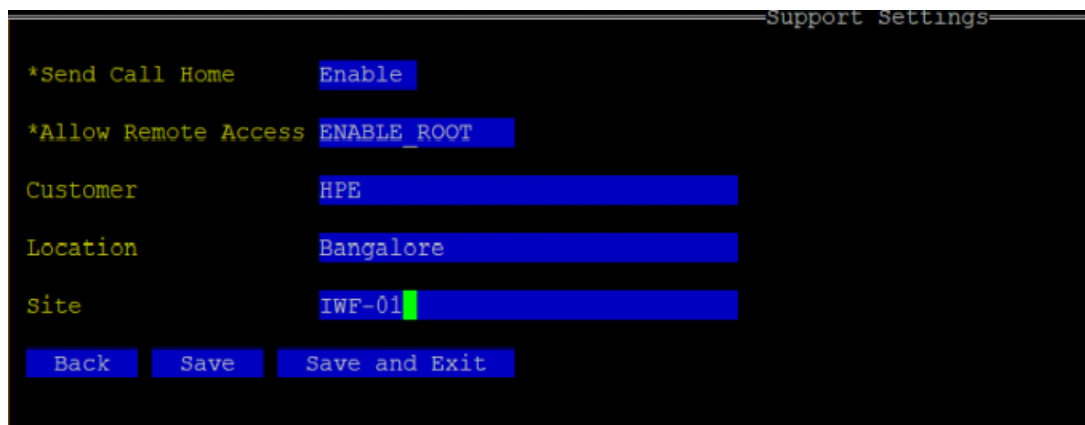
- a. Choose (d) Support/Callhome settings.



- b. Select Enable for the Send Call Home option.
- c. Select one of the values from DISABLE, ENABLE_NONROOT, and ENABLE_ROOT in the Allow Remote Access drop-down list as per the security requirements for the customer.



- d. Enter customer information and click Save.



2. Enable HPE GLCP connectivity.

- a. Open an SSH session to the management node using the management virtual IP address (VMS IP) for the storage system.

- b. Use the

hpesupport

user and the strong password generated previously.

- c. Switch to the

root

user using the `SU root` command and the strong password generated previously to get root access.

- d. Remove the Dark Site exception touch file from all CNodes.

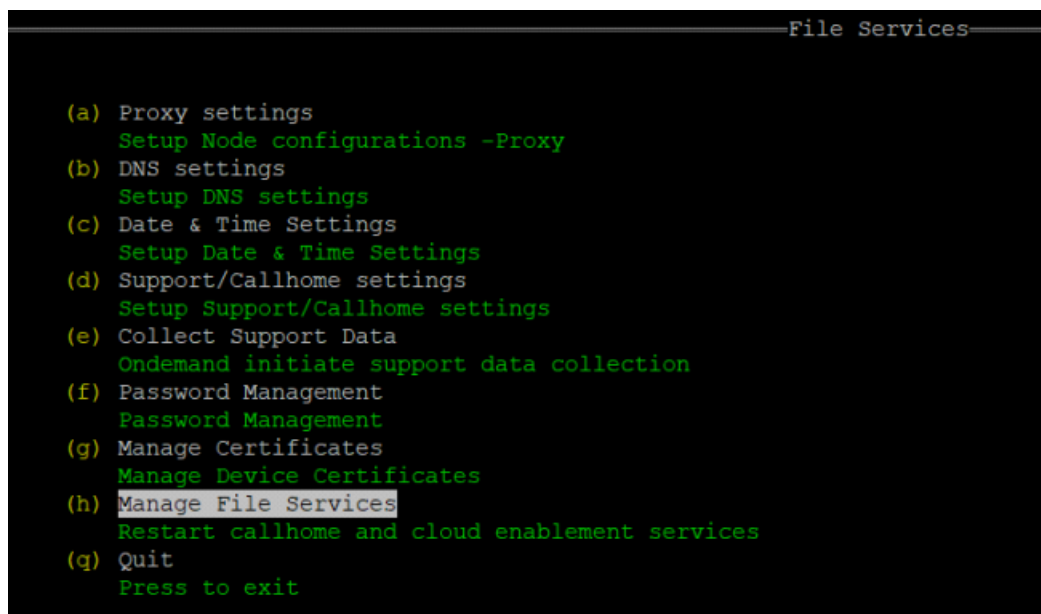
```
$ sudo -u vastdata clush -g cnodes sudo rm -f /opt/hpe/dark_site_exception_ccd
```

- e. Remove the Dark Site exception touch file from the shared folder.

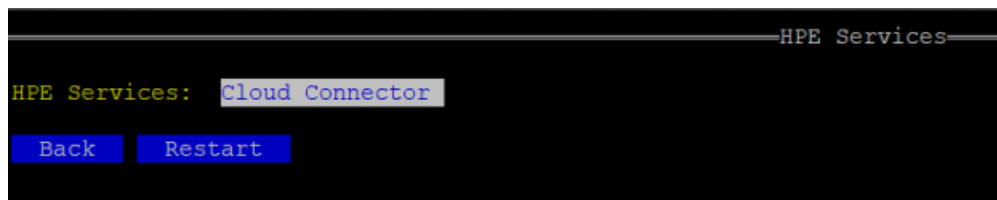
```
$ sudo rm -f /opt/sharedFolder/hpe/dark_site_exception_ccd
```

3. Restart the Cloud Connector service.

- a. In the TUI, select (h) Manage File Services.



- b. Choose Cloud Connector from the HPE Services drop down.



- c. Click Restart.
4. Click Quit to close the TUI and end the SSH session.

Connecting to Private Cloud AI service from Data Services Cloud Console

This section guides you through the steps to set up your Private Cloud AI system for the first time in the data center.

Subtopics

- [Configuring the Data Services Connector using the Private Cloud AI On-Prem Setup wizard](#)
- [Adding a system using the Private Cloud AI Setup wizard](#)
- [Protecting the Data Services Connector VM](#)
- [Applying licenses](#)

Configuring the Data Services Connector using the Private Cloud AI On-Prem Setup wizard

Prerequisites

- Completed the prerequisite worksheets with site-specific network and device details
- Completed the solution hardware deployment, including HPE GreenLake for File Storage configuration

About this task

This section describes the procedure to complete the Private Cloud AI On-Prem Setup and generate the Data Services Connector activation code.

Procedure

1. Connect to the Data Services Connector VM from your workstation.

If your workstation is using the Windows operating system, perform the following steps to add the link-local IPv4 address configuration to the management network of your workstation:

- a. Navigate to the Network settings on your workstation.
- b. Right-click the network adapter that is connected to the management network and select Properties.
- c. Select Change adapter options.
- d. Double-click Internet Protocol Version 4 (TCP/IPv4).
- e. Click Advanced.
- f. In the IP addresses section, click Add.
- g. Enter the IP address as **169.254.1.2**
for the management network of your workstation and the subnet mask as **255.255.0.0**

and click Add.



NOTE

If the IP address 169.254.1.2 is not accepted, use another one from the IP address range **169.254.x.y**

h. Click OK and close the network settings.

2. On the configured workstation, log in to the Data Services Connector VM using the <https://initial-setup.local/> with a link-local IPv4 address from a web browser.

Follow through the Private Cloud AI On-Prem Setup wizard:

3. On the Domain Time and Proxy screen, perform the following steps:

- a. Enter the IPv4 address for a DNS server.

If you want to add multiple DNS servers, click the + icon and enter the IP address for each DNS server.

- b. Enter the search domains.

If you want to add multiple search domains, click the + icon and enter the search domains, separated by commas.



NOTE

You can find the search domains under the *dnsSearchDomains* in the **base_configuration.json** Smart CID file.

- c. Enter the IPv4 address for an NTP server.

If you want to add multiple NTP servers, click the + icon and add the IPv4 address for each NTP server.

- d. Select the Include option if you have a proxy server and enter the following details:

Address

Enter the proxy server address.

Port

Enter the proxy server port.

Username

Enter the proxy server user name.

Password

Enter the proxy server password.

- e. Click Next.

4. On the Storage screen, perform the following steps:

- a. In the Is your Private Cloud AI system utilizing a converged storage system? dialog box, select either Yes, my system is utilizing a converged storage system or No, my system is using the built in storage from the nodes .

- If you select Yes, my system is utilizing a converged storage system , perform the next step for entering the serial number for the storage system.
- If you select No, my system is using the built in storage from the nodes , click Next to proceed to step 5.

- b. Enter the serial number for the storage system and click Submit.

The storage system with the serial number is successfully submitted.

- c. Click Next.

5. On the Data Services Connector Network screen, perform the following steps:

- a. Provide the following management network details for the Data Services Connector:

NIC Name

The name of the network interface is set automatically in systems using the VM Essentials.

Address Type

Select Static radio button for the Address Type.

Host Name

Enter the host name for the management network.

IP Address

Enter the IP address for the management network.

Subnet Mask

Enter the subnet mask for the management network.

Default Gateway

Enter the default gateway for the management network.

Default Docker Bridge Network

The default IP address for the Docker bridge network is set automatically.

Data Services Connector Bridge Network

The IP address for the Data Services Connector bridge network is set automatically.

- b. Click Next.

6. On the Review screen, verify the information and click Submit.

The system changes the web address on the browser from the first run wizard address <https://initial-setup.local/> to <https://<IP address of the DSC VM>>.

7. On the Activation screen, click Copy to Clipboard to make a note of the Data Services Connector activation code.

This code is required to manage this Data Services Connector in the Data Services Cloud Console.

Adding a system using the Private Cloud AI Setup wizard

Prerequisites

- Completed the HPE Smart CID with site-specific network and device details.
- Verify that you have the activation code that was generated after completing the Private Cloud AI On-Prem Setup.
- HPE GreenLake account credentials have been configured, including user name and password.
- HPE GreenLake account is added to the customer workspace.
- Do not assign a temporary IPv4 static IP address to the HVM host to ensure that the server discovery and infrastructure setup functions correctly. The server discovery and deployment process use only the link-local IPv4 or IPv6 addresses assigned to the HVM host in the HPE factory.

About this task

This section describes the procedure to activate the Data Services Connector and set up the Private Cloud AI service from the Data Services Cloud Console on the systems using the VM Essentials.

Procedure

1. Log in to HPE GreenLake and launch the Data Services Cloud Console.

2. Click Launch for the Private Cloud AI service.
3. On the Welcome to Private Cloud AI screen, enter the activation code obtained after completing the Private Cloud AI On-Prem Setup wizard and click Activate. If you are not automatically prompted to enter the activation code, perform the following steps:
 - a. Navigate to Systems from the main menu and click the + icon.
 - b. Enter the activation code generated by the Private Cloud AI On-Prem Setup wizard.
 - c. Click Activate.

Wait for the Data Services Connector registration process to complete.

After the Data Services Connector is registered, the Private Cloud AI Setup wizard appears.

4. Follow through the Private Cloud AI Setup wizard to complete the following modules:
 - a. [Configure the Infrastructure module.](#)
 - b. [Configure the Private Cloud AI module.](#)

Subtopics

[Setup Wizard: Configuring infrastructure](#)

[Setup Wizard: Configuring Private Cloud AI](#)

Setup Wizard: Configuring infrastructure

About this task

In the Infrastructure module, perform the following steps:

Procedure

1. On the General screen, perform the following steps:
 - a. In the System Details section, enter a unique and descriptive name for the Private Cloud AI system in the System Name field.

This name is used throughout the UI to identify and reference the system.



NOTE

Ensure that the name starts with a letter or number. It allows underscores, hyphens, and periods.

- b. In the Storage Details section, enter the following details:



NOTE

Find the start and end IP addresses, subnet mask, and gateway address from the HPE Smart CID files.

Virtual Pools Start and End IP Address

Enter the start and end IP addresses for the virtual IP pools of the storage system in the Virtual Pools Start IP Address and the Virtual Pools End IP Address fields.

Data Subnet Mask

Enter the data subnet mask for the storage network in the Data Subnet Mask field that can be obtained from the **Storage Network** tab in the HPE Smart CID file.

Data Gateway

Enter the data gateway address for the storage network in the Data Gateway field that can be obtained from the **Storage Network** tab in the HPE Smart CID file.

c. Click Next.

2. On the Control Plane screen, perform the following steps:

- a. On the Servers screen, select three control plane servers from the list and click Discover Servers to activate the nodes. After the three servers are discovered, click Next.



NOTE

If the "DSC VM is an unsupported server" error is displayed, the system will automatically upgrade to the required version. Ignore this error and continue with the setup process.

- b. On the Network: Server Management screen, enter the following details and click Next:

Start IP Address

Enter the start IP address for the management network in the control plane servers and click Apply.



NOTE

This field is optional. The Apply button is enabled only when the start IP address is provided.

Control Plane Server Management IP Address

Enter a unique management IP address for each control plane server.



NOTE

Ensure that all management IP addresses are unique.

Subnet Mask

Enter the subnet mask for the management network in the control plane servers.

Gateway

Enter the gateway address for the management network in the control plane servers.

- c. On the Network: iLO Management screen, enter the following details and click Next:

Use the same or different Server Management Network

- If the iLO and server management networks are same, select Use the same network as the Server Management Network radio button and enter the IP addresses for the control plane servers.
- If both the networks are different, select Use a different network than the Server Management Network radio button and enter the IP addresses, gateway, and subnet mask for the iLO management network.

Start IP Address

Enter the start IP address for the iLO network in the control plane servers if the same network as the Server Management Network is used and click Apply.



NOTE

This field is optional. The Apply button is enabled only when the start IP address is provided.

Control Plane Server Management IP Address

Enter a unique management IP address for each control plane server.

**NOTE**

Ensure that all management IP addresses are unique.

Subnet Mask

Enter the subnet mask for the iLO network in the control plane servers if a different network than the Server Management Network is used. This field is enabled only when the Use a different network than the Server Management Network radio button is selected.

Gateway

Enter the gateway address for the iLO network in the control plane servers if a different network than the Server Management Network is used. This field is enabled only when the Use a different network than the Server Management Network radio button is selected.

- d. On the Network: Data Subnet screen, enter the following details and click **Next**:

Data Network IP Address

Enter the data network IP address for the control plane servers that can be obtained from the **Storage Network** tab in the HPE Smart CID file.

Data Subnet Mask

This field is disabled and is prefilled with the value from the General screen of the Infrastructure module.

Data Gateway

This field is disabled and is prefilled with the value from the General screen of the Infrastructure module.

- e. On the Credentials screen, enter the following details and click **Next**:

HPE VME Hypervisor Root Credentials

Select the Hypervisor manager root secret for the HPE VM Essentials user.

iLO Admin Credentials

Select the iLO admin secret for the ilo_admin user.

3. On the HPE VM Essentials Software screen, perform the following steps to create and deploy a new VM Essentials manager and to assign or create credentials:

- a. In the HPE VME Manager Details section, enter the following details:

HPE VME Manager Name

Enter the FQDN for the VM Essentials manager.

HPE VME Manager IP Address

Enter the IP address for the VM Essentials manager.

HPE VME Manager Gateway

Enter the gateway address for the VM Essentials manager.

HPE VME Manager Subnet Mask

Enter the subnet mask for the VM Essentials manager.

- b. In the HPE VME Hypervisor & Linux System Credentials section, enter the following details:

HPE VME Hypervisor Credentials

Select the SSH credentials for accessing the VM Essentials manager. The credentials are stored in the Secrets service as BASIC_AUTH types. If a new secret must be created, click **+Create New** and follow the prompts that appear on the screen.

HPE VME Admin User Email

Enter the email address of the main contact for the solution.

Linux System Admin Credentials

Select the SSH credentials for accessing the Linux system hosting the VM Essentials manager. The credentials are stored in the Secrets service as BASIC_AUTH types. If a new secret must be created, click +Create New and follow the prompts that appear on the screen.

c. Click Next.

4. On the Summary screen, verify the Infrastructure information and click Submit.

The Infrastructure Setup dialog box is displayed.

a. On the Infrastructure Setup dialog box, select I acknowledge that these inputs cannot be changed after moving onto the next step and click Yes, Setup to complete this module.

Wait for some time after the Infrastructure Setup is complete before proceeding to the Private Cloud AI Setup module.



NOTE

The setup can take about 10 minutes to complete.

b. Click Next to proceed to the Private Cloud AI Setup module.

Setup Wizard: Configuring Private Cloud AI

About this task

In the Private Cloud AI module, perform the following steps:

Procedure

1. On the Control Plane screen, perform the following steps to define the network details for the three control plane VMs and the single Orchestrator VM:

a. Select the SSH key pair for the pcadmin user in the SSH Key for VM Access field.



NOTE

Ensure that you retrieve the SSH key pair using the `ssh-keygen` command on any Linux node, accept default parameters, and copy the contents of the public key and private key files.

b. Enter the ai-cluster IP address in the Control Plane Cluster IP Address field.



NOTE

The ai-cluster IP address for the management network can be obtained from the HPE Smart CID file.

c. In the Management Network section, enter the following details:

Start IP Address

Enter the IP address for the management network in the Start IP Address field and click Apply.



NOTE

This field is optional. The Apply button is enabled only when the start IP address is provided. The first IP address for the management network can be obtained from the **Management Network** tab in the `networks-excel` file where the **HostName** is `ezmaster01`.

Control Plane VM IP Address

Enter a unique management IP address for each control plane server.

**NOTE**

The first IP address for the management network can be obtained from the **Management Network** tab in the **networks-excel** file where the **HostName** is `ezmaster01` . For the next set of IP addresses to be assigned for additional VMs, use the IP addresses where the **HostName** values are `ezmaster02` , `ezmaster03` , and `ezfab` .

Subnet Mask

Enter the subnet mask for the management network in the Subnet Mask field.

Gateway

Enter the gateway address for the management network in the Gateway field.

d. Click Next.

2. On the Worker Nodes screen, perform the following steps:

- a. On the Nodes screen, select the required worker nodes from the list of preselected servers and click **Discover Worker Nodes**. Ensure that at least one worker node is selected. After the worker nodes are discovered, click Next.

**NOTE**

If you see the "Unsupported servers found." error causing the details of those nodes to show that there are no GPUs, reboot and rediscover the worker node.

- b. On the Network: Management screen, enter the following details and click **Next**:

Start IP Address

Enter the start IP address for the management network in the worker nodes and click **Apply**.

**NOTE**

This field is optional. The **Apply** button is enabled only when the start IP address is provided.

Worker Node Management Network IP Address

Enter a unique management IP address for all worker nodes.

**NOTE**

Ensure that all management IP addresses are unique.

Subnet Mask

Enter the subnet mask for the management network in the worker nodes.

Gateway

Enter the gateway address for the management network in the worker nodes.

- c. On the Network: iLO screen, enter the following details and click **Next**:

Start IP Address

Enter the start IP address for the iLO network in the worker nodes and click **Apply**.

**NOTE**

This field is optional. The **Apply** button is enabled only when the start IP address is provided.

Worker Node iLO IP Address

Enter a unique iLO network IP address for each worker node.



NOTE

Ensure that all IP addresses for iLO network are unique.

Subnet Mask

Enter the subnet mask for the iLO network in the worker nodes.

Gateway

Enter the gateway address for the iLO network in the worker nodes.

d. On the Network: Storage screen, enter the following details and click **Next**:

Section	Steps
Storage 1 Network	Enter the following details: Storage 1 Start IP Address Enter the start IP address for the storage network in the worker nodes. Subnet Mask Enter the subnet mask for the storage network in the worker nodes. Gateway Enter the gateway address for the storage network in the worker nodes.
Storage 2 Network	Enter the following details: Storage 2 Start IP Address Enter the start IP address for the storage network in the worker nodes. This IP address must be N more than the start IP of Storage 1 Network, where N is the maximum number of worker nodes in the configuration. Subnet Mask Enter the same subnet mask as the one used for the Storage 1 Network. Gateway Enter the gateway address for the storage network in the worker nodes.

e. On the Credentials screen, enter the following details and click **Next**:

Node iLO Credentials

Select the node iLO admin credentials.

OS Credentials

Select the same SSH key pair for the worker_os as provided in the SSH Key for VM Access field on the **Control Plane** screen. If a new SSH key pair must be created, run the `ssh-keygen` command on any Linux node, accept default parameters, and copy the contents of the public key and private key files.

3. On the Platform screen, perform the following steps:

a. On the HPE AI Essentials screen, enter the following details and click **Next**:

Application IP address



Enter the IP address used to connect to the HPE AI Essentials application.



NOTE

The application IP address can be obtained from the **Production Network** tab in the **networks-excel** file where the **HostName** is `ai-application`.

Application Domain Name

Enter the domain name for the HPE AI Essentials application.

Storage Virtual IP Pools Domain Name

Enter the domain name for the storage virtual IP pools.

Pod Network CIDR

Enter the IP addresses to individual pods in the Kubernetes cluster. This CIDR block provides a range of IP addresses that ensures each pod has a unique address.



NOTE

You can accept the defaults if they do not interfere with your existing network settings.

Default: 10.224.0.0/19

Service Network CIDR

Enter the IP addresses to HPE Kubernetes Service defined by a separate CIDR block.



NOTE

You can accept the defaults if they do not interfere with your existing network settings.

Default: 10.96.0.0/22

End User License Agreement

Review and accept the license agreements to continue with the system setup.

- b. On the HPE Data Lakehouse Gateway screen, enter the following details and click **Next**:

Start IP Address

Enter the start IP address of the management network for the first Data Lakehouse Gateway VM. The IP addresses for the next two Data Lakehouse Gateway VMs and six Virtual IP pool entries are assigned sequentially.



NOTE

The start IP address can be obtained from the **Management Network** tab in the **networks-excel** file where the **HostName** is `dl1`.

Subnet Mask

This field is set automatically depending on the management IP address used for the Data Lakehouse Gateway VM.

Gateway

This field is set automatically depending on the management IP address used for the Data Lakehouse Gateway VM.

4. On the Summary screen, verify the Private Cloud AI information and click **Submit**.

Wait for the Private Cloud AI Setup process to complete.

5. Change the production network configuration on each worker node.

For more information, see the HPE AI Essentials UI is not accessible due to the load balancer routing issue on the worker nodes section in the HPE Private Cloud AI 1.5 Known Issues.

Protecting the Data Services Connector VM

After the Private Cloud AI deployment is complete, HPE recommends implementing robust security controls within your environment to safeguard the Data Services Connector VM. The recommendations include:

- Assigning appropriate access controls so that only users with demonstrated business needs are allowed to access the Data Services Connector VM.
- Minimizing privileges so that each user is allowed to perform only the tasks required by their job role.
- Using controls to limit cloning or copying of the Data Services Connector VM, and recording any operations that copy or clone the Data Services Connector VM in an audit trail.

Carefully assigning access controls prevents authorized but unprivileged users from accessing the Data Services Connector VM, and stops unauthorized access to the Data Services Connector VM by malicious users.

Your proactive steps to secure this image help maintain the integrity and confidentiality of your information.

Applying licenses

The following table describes how you receive and apply license information for the software components that make up the HPE Private Cloud AI engineered system:

Product	How to receive the license information	How to apply the license information
VM Essentials	HPE sends an email receipt with the order details.	No action is required.
Red Hat Enterprise Linux (RHEL)	HPE sends an email receipt with the order details and instructions to register the products. The email is sent to the contact on the Purchase Order.	You must have already applied for the Red Hat license on the worker node. If you haven't applied for this license, see Apply the Red Hat license to each worker node section.

Product	How to receive the license information	How to apply the license information
HPE AI Essentials	HPE sends an email receipt with the order details and instructions to register the products. The email is sent to the contact on the Purchase Order.	After completing the "Private Cloud AI Setup wizard" procedure for initial deployment: 1. HPE TC launches the HPE AI Essentials application. 2. The first time the application is launched, the platform ID is displayed to the HPE TC. 3. The HPE TC provides the customer with the platform ID either directly or via the IPM. 4. The customer goes to the HPE Software Center and entitles both the Ezmeral vCPU and Ezmeral GPU licenses. 5. HPE sends an email confirmation to the customer of successful registration within the My HPE Software Center. 6. The customer provides HPE the license key files. 7. The installer drags and drops the file for the Ezmeral vCPU into the application. Then they open the HPE AI Essentials application. 8. Once in the application, they apply the Ezmeral GPU license by clicking Administration > Settings > Activation Key > Upload Activation Key. Select the file for the Ezmeral GPU license, and upload it.
HPE GreenLake for File with Object Storage Enabled	HPE sends an email receipt with the order details and instructions to register the products. The email is sent to the contact on the Purchase Order.	1. The customer activates the purchased products in the My HPE Software Center. 2. HPE sends the customer an email confirmation of successful registration within the My HPE Software Center. 3. The customer must add the subscription license to their HPE GreenLake workspace by using the Adding a customer-owned device to HPE GreenLake instructions. 4. They need to use Product Number S1F24A.

Product	How to receive the license information	How to apply the license information
NVIDIA AI Enterprise Small or Medium Configuration (NVIDIA L40S)	HPE sends an email receipt with the order details and instructions to register the products. The email is sent to the contact on the Purchase Order.	1. The customer activates the purchased products in the My HPE Software Center. 2. HPE sends an email confirmation to the customer of successful registration within the My HPE Software Center. 3. NVIDIA also sends the customer an email with instructions for how to register. 4. The customer must register on the NVIDIA website to receive NVIDIA support. 5. No additional on-prem licensing is required.
NVIDIA AI Enterprise Large Configuration (NVIDIA H100 NVL)	The customer does not receive any notification.	1. Log in to the NGC account at ngc.nvidia.com/signin. 2. Navigate to Organization. 3. Select Activate Subscription. 4. Complete the company information form. 5. Enter the GPU serial numbers. 6. Click Activate Subscription. 7. Review information in the pop-up window. 8. Click Request Subscription. 9. Wait for approval (up to 48 hours). 10. Access the Enterprise Catalog from the left menu. 11. Download the NVIDIA AI Enterprise software. 12. Watch for the NVIDIA email (within 24 hours). 13. Create your Enterprise Support account using the email link. 14. Access the licensing portal through the email link.
HPE Private Cloud AI Platform	The subscription is provided through HPE GreenLake Cloud.	No action is required.
OpsRamp Enterprise	The subscription is provided through HPE GreenLake Cloud.	No action is required.

Component identification

Subtopics



[HPE ProLiant server component identification](#)
[HPE GreenLake for File Storage component identification](#)
[Aruba 6300 series switch component identification](#)
[HPE M-series switch component identification](#)

HPE ProLiant server component identification

For more information on HPE ProLiant server components, go to the Hewlett Packard Enterprise Support Center to consult the User Guide for your HPE ProLiant server model.



NOTE

HPE supports the use of iLO dedicated and shared ports. For more information, see HPE Integrated Lights Out (iLO) – Configuring the NIC Settings, available on the HPESC (<https://www.hpe.com/support/hpesc>).

HPE GreenLake for File Storage component identification

For more information on HPE GreenLake for File Storage server components, go to the Hewlett Packard Enterprise Support Center (<https://www.hpe.com/support/hpesc>) to consult the Maintenance and Service guide for your HPE GreenLake for File Storage server model.

Aruba 6300 series switch component identification

For more information on Aruba 6300 series switch components, go to the Hewlett Packard Enterprise Support Center (<https://www.hpe.com/support/hpesc>) to consult the documentation for your Aruba 6300 series switch model.

HPE M-series switch component identification

For more information on HPE M-series switch components, go to the Hewlett Packard Enterprise Support Center (<https://www.hpe.com/support/hpesc>) to consult the QuickSpecs for your HPE M-series switch model.

Specifications

Subtopics

[Environmental specifications](#)

Environmental specifications



Specification	Value
Temperature range ¹	—
Operating	10°C to 35°C (50°F to 95°F)
Nonoperating	-30°C to 60°C (-22°F to 140°F)
Relative humidity (noncondensing)	—
Operating	8% to 90%. 28°C (82.4°F), maximum wet bulb temperature
Nonoperating	5% to 95% 38.7°C (101.7°F), maximum wet bulb temperature
Altitude	—
Operating	3050 m (10,000 ft). This value may be limited by the type and number of options installed. Maximum allowable altitude change rate is 457 m/min (1500 ft/min).
Non-operating	9144 m (30,000 ft). Maximum allowable altitude change is 457 m/min (1500 ft/min).

¹ All temperature ratings shown are for sea level. An altitude derating of 1.0°C per 305 m (1.8°F per 1,000 ft) to 3,050 m (10,000 ft) is applicable. No direct sunlight allowed. Maximum rate of change is 20°C per hour (36°F per hour). The upper limit and rate of change might be limited by the type and number of options installed.

Standard operating support

10° to 35°C (50° to 95°F) at sea level with an altitude derating of 1.0°C per every 305 m (1.8°F per every 1000 ft) above sea level up to a maximum of 3050 m (10,000 ft), no direct sustained sunlight. Maximum rate of change is 20°C/hr (36°F/hr). The upper limit and rate of change may be limited by the type and number of options installed.

System performance during standard operating support may be reduced if operating with a fan fault or above 30°C (86°F).

For certain approved hardware configurations, the supported system inlet temperature range is extended:

- 5°C to 10°C (41°F to 50°F) and 35°C to 40°C (95°F to 104°F) at sea level with an altitude derating of 1.0°C per every 175 m (1.8°F per every 574 ft) above 900 m (2,953 ft) to a maximum of 3,050 m (10,000 ft).
- 40°C to 45°C (104°F to 113°F) at sea level with an altitude derating of 1.0°C per every 125 m (1.8°F per every 410 ft) above 900 m (2,953 ft) to a maximum of 3,050 m (10,000 ft). The approved hardware configurations for this operating range require the High Performance Fan Kit.

The approved hardware configurations for this system are listed on the Hewlett Packard Enterprise website (<http://www.hpe.com/servers/ASHRAE>).

Setting up OpsRamp gateway component of the HPE Private Cloud AI



NOTE

The OpsRamp component is deployed after the initial Private Cloud AI deployment.

For information on setting up the OpsRamp components, see the [OpsRamp Deployment Guide](#).

Getting started with HPE Private Cloud AI documentation

As part of the deployment process, HPE personnel are encouraged to share the required documentation with Private Cloud AI customers.

Users who are new to Private Cloud AI can get started on learning how to use and manage the cluster by reviewing the following documentation first:

Reference	See these sections
HPE Private Cloud AI Release Notes	"New features"
HPE Private Cloud AI Administration Guide	"Overview of the HPE Private Cloud AI solution" "Cloud Administrator: Maintaining the solution hardware and software"
HPE AI Essentials Software Documentation	"Get Started" "Tutorials" "Data Engineering" "Data Analytics" "Data Science"
HPE Private Cloud AI Firmware and Software Compatibility Matrix	All
OpsRamp Documentation	All

Users can provide feedback to help us improve the documentation as mentioned in the [Documentation feedback](#) section.

Appendix A: Alternative virtualization - VMware-based procedures

This section describes the procedures to be performed on VMware-based systems only.

Mounting the HPE GreenLake for File Storage NFS share as a datastore in ESXi

Prerequisites

Ensure that the virtual IP pool, view policies, views, user, and user group are created properly in the HPE GreenLake for File Storage cluster at the Integration Center.

About this task

This section describes the procedure to mount the HPE GreenLake for File Storage NFS share as a datastore in ESXi after the copied OVA can be accessed from HPE GreenLake for File Storage.

Procedure

1. Configure Advanced Settings to mount the HPE GreenLake for File Storage NFS share as a datastore in ESXi.
 - a. Open an SSH session on each VMware host and log in using the root credentials with the factory default password.
 - b. Set the MaxConnectionsPerDatastore in ESXi using the following command:

```
esxcli system settings advanced set -o /NFS/MaxConnectionsPerDatastore -i 8
```

2. Continue the remaining steps to set up the HPE GreenLake for File Storage.



Using the Private Cloud AI Setup wizard on VMware based systems

Prerequisites

- Completed the HPE Smart CID with site-specific network and device details.
- Ensure that all the components are configured.
- Verify that you have the activation code that was generated after completing the Private Cloud AI On-Prem Setup.
- HPE GreenLake account credentials have been configured, including user name and password.
- HPE GreenLake account is added to the customer workspace.

About this task

This section describes the procedure to activate the Data Services Connector and set up the Private Cloud AI service from the Data Services Cloud Console on the systems using VMware.

Procedure

1. Log in to HPE GreenLake and launch the Data Services Cloud Console.
2. Launch the Private Cloud AI service.
3. On the Welcome screen, enter the activation code obtained after completing the Private Cloud AI On-Prem Setup wizard and click Activate. If you are not automatically prompted to enter the activation code, perform the following steps:
 - a. Navigate to Systems from the main menu and click the + icon.
 - b. Enter the activation code generated by the Private Cloud AI On-Prem Setup wizard.
 - c. Click Activate and wait for the registration process to complete.

If the device is connected to the Data Services Cloud Console, an Initializing status appears.

Follow through the Private Cloud AI Setup wizard:

In the Infrastructure Setup, perform the following steps:

4. On the General screen, perform the following steps:
 - a. In the System Details section, enter the appropriate name for the Private Cloud AI system in the System Name field.



NOTE

Ensure that the name starts with a letter or number and includes letters, numbers, underscore, hyphen or a period.

- b. In the Storage Details section, provide the following details:



NOTE

Find the start and end IP addresses, subnet mask, and gateway address from the HPE Smart CID files.

Virtual Pools Start and End IP Addresses

Enter the start and end IP addresses for the virtual IP pools in the Start IP and the End IP fields.

Data Subnet Mask

Enter the data subnet mask for the storage network in the Subnet Mask field.

Data Gateway

Enter the data subnet mask for the storage network in the Gateway field.

c. Click Next.

5. On the Control Plane screen, perform the following steps:

a. On the Servers screen, select the three control plane servers and click Discover. After the three servers are discovered, click Next.



NOTE

If the "DSC VM is an unsupported server" error is displayed, the system will automatically upgrade to the required version. Ignore this error and continue with the setup process.

b. On the Network screen, perform the following steps for each section and click Next:

Section	Steps
Server Management Network	Enter the following details: Start IP Enter the start IP address for the management network in the control plane servers. Gateway Enter the gateway address for the management network in the control plane servers. Subnet Mask Enter the subnet mask for the management network in the control plane servers.



iLO Management Network

Enter the following details:

Start IP

Enter the start IP address for the iLO network in the control plane servers if the same network as the Server Management Network is used.



NOTE

- If the iLO and server management networks are the same, select the Use the same network as the Server Management Network radio button and enter the start IP address for the control plane servers.
- If the networks are different, select the Use a different network than the Server Management Network radio button and enter the start IP address, gateway, and subnet mask for the iLO management network.

Subnet Mask

Enter the subnet mask for the iLO network in the control plane servers if a different network than the Server Management Network is used.

Gateway

Enter the gateway address for the iLO network in the control plane servers if a different network than the Server Management Network is used.

Data Subnet Configuration

Enter the following details:

Data Network IP Address

Enter the data network IP address for the control plane servers that can be obtained from the **Storage Network** tab in the HPE Smart CID file.

Data Subnet Mask

Enter the data subnet mask for the control plane servers that can be obtained from the **Storage Network** tab in the HPE Smart CID file.

Data Gateway

Enter the gateway address for the control plane servers that can be obtained from the **Storage Network** tab in the HPE Smart CID file.

- c. On the Credentials screen, enter the following details and click Next:

Hypervisor Root Credentials

Select the Hypervisor manager root credentials for the esxi_root user. These credentials are stored as secrets in the credentials vault and are used for logging into the Private Cloud AI system.

iLO Admin Credentials

Select the iLO admin credentials for the ilo_admin user.

6. On the Virtualization screen, perform the following steps:

- a. Enter the host name in the Hypervisor Manager Host Name field. This name must be resolved in the customer's DNS.
- b. Enter the IP address in the Hypervisor Manager IP Address field.
- c. Enter the SSO admin credentials for the vcenter_sso user in the Hypervisor Manager SSO Admin Credentials field. The user name for these credentials is typically **administrator@vsphere.local** and must be in the format of an email address.
- d. Enter the root credentials for the vcenter_root user in the Hypervisor Manager Root Credentials field.
- e. Click the link to open the Broadcom license and service terms and select the I accept the User License Agreements box to continue setting up the system.
- f. Click Next.

7. On the Summary screen, verify the Infrastructure information and click Next.

Wait for the Infrastructure Setup process to complete.



NOTE

Wait for some time after the Infrastructure Setup is complete before proceeding to the Private Cloud AI Setup process.

8. Click Next to proceed to the Private Cloud AI Setup process.

In the Private Cloud AI Setup, perform the following steps:

9. On the Control Plane screen, perform the following steps:

- a. Select the SSH key pair for the pcadmin user in the VM Access credentials field. To add the SSH key pair, connect to a Linux server and use the `ssh-keygen` command to generate a public and private key pair and use default parameters for all prompts. The private key is usually saved in `/<username>/.ssh/id_rsa` and the public key is usually in `/<username>/.ssh/id_rsa.pub`.



NOTE

Ensure that you save the SSH key pair data to a location so that it can be provided to the customer later.

- b. Enter the k8-mgmt or k8-cluster IP address in the Control Plane Cluster IP Address field.



NOTE

The k8-mgmt or k8-cluster IP address for the management network can be obtained from the HPE Smart CID file.

- c. In the Management Network section, enter the following details:

Start IP Address

Enter the IP address for the management network in the Control Plane VM IP Address field.



NOTE

The first IP address for the management network can be obtained from the **ezmaster01** in the HPE Smart CID file. The next set of IP addresses assigned for additional VMs can be obtained from the **ezmaster02** , **ezmaster03** , and **ezfab** in the HPE Smart CID file.

Subnet Mask

Enter the subnet mask for the management network in the Subnet Mask field.

Gateway

Enter the gateway address for the management network in the Gateway field.

d. Click Next.

10. On the Worker Nodes screen, perform the following steps:

a. On the Nodes screen, select the worker nodes and click Discover. After the worker nodes are discovered, click Next.



NOTE

If you see the "Unsupported servers found." error causing the details of those nodes to show that there are no GPUs, reboot and rediscover the worker node.

b. On the Network screen, perform the following steps and click Next:

Section	Steps
Server Management Network	Enter the following details: Start IP Enter the start IP address for the management network in the worker nodes. Gateway Enter the gateway address for the management network in the worker nodes. Subnet Mask Enter the subnet mask for the management network in the worker nodes.
iLO Management Network	Enter the following details: Start IP Enter the start IP address for the iLO network in the worker nodes. Gateway Enter the gateway address for the iLO network in the worker nodes. Subnet Mask Enter the subnet mask for the iLO network in the worker nodes.
Storage 1 Network	Enter the following details: Start IP Enter the start IP address for the storage network in the worker nodes. Subnet Mask Enter the subnet mask for the storage network in the worker nodes.
Storage 2 Network	Start IP Enter the start IP address for the storage network in the worker nodes. This IP address must be N more than the start IP of Storage 1 Network, where N is the maximum number of worker nodes in the configuration. Subnet Mask Enter the same subnet mask as the one used for the Storage 1 Network.

- c. On the Credentials screen, enter the following details and click **Next**:

Node iLO Admin Credentials

Enter the node iLO admin credentials.

OS Credentials

Enter the same SSH key pair for the worker_os as provided in the VM Access credentials field. If a new SSH key pair must be created, run the `ssh-keygen` command on any Linux node, accept default parameters, and copy the contents of the public

key and private key files.

11. On the Platform screen, enter the following details:

a. In the AI Essentials section, enter the following details:

Application IP address

Enter the IP address used to connect to the HPE AI Essentials application.



NOTE

This IP address can be obtained from the **Production Network** tab in the HPE Smart CID file.

Application Domain Name

Enter the HPE AI Essentials application domain name.

Virtual IP Pool Domain Name

Enter the storage virtual pools domain name.

b. Click Next.

12. On the Summary screen, verify the Private Cloud AI information and click Submit.

Wait for the Private Cloud AI Setup process to complete.

Websites

HPE Private Cloud AI websites

Log in to the Hewlett Packard Enterprise Support Center as an HPE Employee to access the internal documents.

HPE Private Cloud AI Documentation | Product Search

<https://www.hpe.com/support/Private-Cloud-AI-Docs>

HPE Private Cloud AI Documentation List

<https://www.hpe.com/support/Private-Cloud-AI-Doc-List>

HPE Private Cloud AI customer documentation

HPE Private Cloud AI Administration Guide

<https://www.hpe.com/support/PCAIAdminGuide>

HPE Private Cloud AI Release Notes

<https://www.hpe.com/support/PCAIReleaseNotes>

HPE AI Essentials Software Documentation

<https://www.hpe.com/support/AIEDocs>

HPE Private Cloud AI QuickSpecs

<https://www.hpe.com/support/PCAIQuickSpecs>

Storage websites

NVIDIA Network Hardware Compatibility List (HCL)

<https://www.nvidia.com/en-us/networking/ethernet-switching/hardware-compatibility-list/>

HPE Smart CID portal

<https://smartcid.itcs.hpe.com/xFrameworkCIDRestAPI/index-new.html>

Subtopics

Documentation feedback

Documentation feedback

Hewlett Packard Enterprise is committed to providing documentation that meets your needs. To help us improve the documentation, use the Feedback button and icons (at the bottom of an opened document) on the Hewlett Packard Enterprise Support Center portal (<https://www.hpe.com/support/hpesc>) to send any errors, suggestions, or comments. This process captures all document information.

