

Deploying in Amazon Web Services

- Prerequisites and guidelines for deploying the Nexus Dashboard cluster in Amazon Web Services, on page 1
- Prepare Amazon Web Services for the Nexus Dashboard cluster, on page 1
- Deploying Nexus Dashboard in AWS, on page 3

Prerequisites and guidelines for deploying the Nexus Dashboard cluster in Amazon Web Services

Before you proceed with deploying the Nexus Dashboard cluster in Amazon Web Services (AWS), you must:

- Ensure that the AWS form factor supports your scale and services requirements.

  Scale and services support and co-hosting vary based on the cluster form factor. You can use the Nexus Dashboard Capacity Planning tool to verify that the cloud form factor satisfies your deployment requirements.

- Review and complete the general prerequisites described in the Deployment Overview.
- Review and complete any additional prerequisites described in the *Release Notes* for the services you plan to deploy.
- Have appropriate access privileges for your AWS account.

  You must be able to launch multiple instances of Elastic Compute Cloud (`m5.2xlarge`) to host the Nexus Dashboard cluster.

- Ensure that the CPU family used for the Nexus Dashboard VMs supports AVX instruction set.
- Perform the Prepare Amazon Web Services for the Nexus Dashboard cluster, on page 1 procedure.

Prepare Amazon Web Services for the Nexus Dashboard cluster

Before you deploy the Nexus Dashboard cluster in Amazon Web Services (AWS), you must perform several preliminary steps.

Follow these steps to prepare Amazon Web Services for the Nexus Dashboard cluster.

Procedure

Step 1 If your AWS account has fewer than 6 unused AWS Elastic IP addresses, request an Elastic IP address limit increase.

A typical Nexus Dashboard deployment consists of 3 nodes with each node requiring 2 AWS Elastic IP addresses for the management and data networks. Therefore, you need 6 unused AWS Elastic IP addresses for Nexus Dashboard.

a) In your AWS console, navigate to **Computer** > **EC2**.
b) In the EC2 Dashboard, click **Network & Security** > **Elastic IPs** and note how many Elastic IP addresses are already being used.
c) In the EC2 Dashboard, click **Limits** and note the maximum number of **EC2-VPC Elastic IPs** allowed.
d) Subtract the number of Elastic IP addresses already being used from the limit to determine the number of unused Elastic IP addresses that you have. If the number is less than 6, click **Request limit increase** and specify a number that will give you a total of 6 unused IP addresses.

Step 2 Create a Virtual Private Cloud (VPC).

A VPC is an isolated portion of the AWS cloud for AWS objects, such as Amazon EC2 instances.

a) In your AWS console, navigate to **Networking & Content Delivery Tools** > **VPC**.
b) In the VPC Dashboard, click **Your VPCs** and choose **Create VPC**, then, provide the **Name Tag** and **IPv4 CIDR block**.

The CIDR block is a range of IPv4 addresses for your VPC and must be in the `/16` to `/24` range. For example, `10.9.0.0/16`.

Step 3 Create an Internet Gateway and attach it to the VPC.

An Internet Gateway is a virtual router that allows a VPC to connect to the Internet.

a) In the VPC Dashboard, click **Internet Gateways** and choose **Create internet gateway**, then provide the **Name Tag**.
b) In the **Internet Gateways** screen, choose the Internet Gateway you created, then choose **Actions** > **Attach to VPC**.
c) From the **Available VPCs** drop-down menu, choose the VPC you created and click **Attach internet gateway**.

Step 4 Create a routes table.

Routes table is used for connecting the subnets within your VPC and Internet Gateway to your Nexus Dashboard cluster.

a) In the VPC Dashboard, click **Route Tables**, choose the **Routes** tab, and click **Edit routes**.
b) In the **Edit routes** screen, click **Add route** and create a `0.0.0.0/0` destination.
c) From the **Target** drop-down menu, choose `Internet Gateway` and choose the gateway you created.
d) Click **Save routes**.

Step 5 Create a key pair.

A key pair consists of a private key and a public key, which are used as security credentials to verify your identity when connecting to an EC2 instance.

a) Navigate to **All services** > **Compute** > **EC2**.
b) In the EC2 Dashboard, click **Network & Security** > **Key Pairs**, then click **Create Key Pairs**.
c) Provide a name for your key pair, choose the **pem** file format, and click **Create key pair**.

This will download the `.pem` private key file to your system. Move the file to a safe location; you will need to use it the first time you log in to an EC2 instance's console.

Note

By default, only PEM-based login is enabled for each node. To be able to SSH into the nodes using a password, as required by the GUI setup wizard, you must explicitly enable password-based logins by logging in to each node using the generated key and running the required command as described in Deploying Nexus Dashboard in AWS, on page 3.

Deploying Nexus Dashboard in AWS

This section describes how to deploy Cisco Nexus Dashboard cluster in Amazon Web Services (AWS).

Before you begin

- Ensure that you meet the requirements and guidelines described in Prerequisites and guidelines for deploying the Nexus Dashboard cluster in Amazon Web Services, on page 1.

Procedure

Step 1 Subscribe to Cisco Nexus Dashboard product in AWS Marketplace.

a) Log into your AWS account and navigate to the AWS Management Console

The Management Console is available at https://console.aws.amazon.com/.

b) Navigate to **Services** > **AWS Marketplace Subscriptions**.
c) Click **Manage Subscriptions**.
d) Click **Discover products**.
e) Search for **Cisco Nexus Dashboard** and click the result.
f) In the product page, click **Continue to Subscribe**.
g) Click **Accept Terms**.

It may take a couple of minutes for the subscription to be processed.

h) Finally click **Continue to Configuration**.

Step 2 Select software options and region.

a) From the **Fulfillment Option** dropdown, select `Nexus Dashboard - Cloud Deployment`
b) From the **Software Version** dropdown, select the version you want to deploy.
c) From the **Region** dropdown, select the regions where the template will be deployed.

This must be the same region where you created your VPC.

d) Click **Continue to Launch**.

The product page appears, which shows a summary of your configuration and enables you to launch the cloud formation template.

Step 3 From the **Choose Action**, select `Launch CloudFormation` and click **Launch**.

The **Create stack** page appears.

Step 4 Create stack.

a) In the **Prerequisite - Prepare template** area, select `Template is ready`.

b) In the **Specify Template** area, select `Amazon S3 URL` for the template source.

The template will be populated automatically.

c) Click **Next** to continue.

The **Specify stack details** page appears.

Step 5 Specify stack details.

a) Provide the **Stack name**.

b) From the **VPC identifier** dropdown, select the VPC you created.

For example, `vpc-038f83026b6a48e98(10.176.176.0/24)`.

c) In the **ND cluster Subnet block**, provide the VPC subnet CIDR block.

Choose a subnet from the VPC CIDR that you defined. You can provide a smaller subnet or use the whole CIDR. The CIDR can be a `/24` or `/25` subnet and will be segmented to be used across the availability zones.

For example, `10.176.176.0/24`.

d) From the **Availability Zones** dropdown, select one or more available zones.

We recommend you choose 3 availability zones. For regions that support only 2 availability zones, 2nd and 3rd nodes of the cluster will launch in the second availability zone.

e) From the **Number of Availability Zones** dropdown, select the number of zones you added in the previous substep.

Ensure that the number matches the number of availability zones you selected in the previous substep.

f) Enable **Data Interface EIP support**.

This field enables external connectivity for the node. External connectivity is required for communication with Cisco ACI fabrics outside AWS.

g) In the **Password** and **Confirm Password** fields, provide the password.

This password will be used for the Nexus Dashboard's `rescue-user` login, as well as the initial password for the GUI's `admin` user.

Note
You must provide the same password for all nodes or the cluster creation will fail.

h) From the **SSH key pair** dropdown, select the key pair you created.

i) In the **Access control** field, provide the external network allowed to access the cluster.

For example, `0.0.0.0/0` to be able to access the cluster from anywhere.

j) Click **Next** to continue.

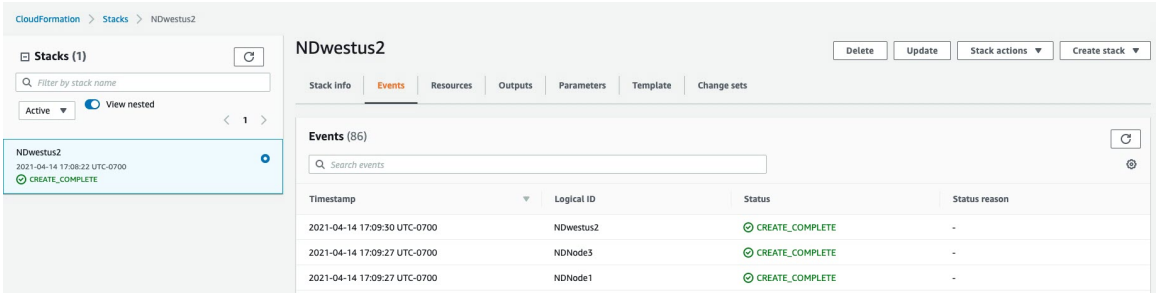
Step 6 In the **Advanced options** screen, simply click **Next**.

Step 7 In the **Review** screen, verify template configuration and click **Create stack**.

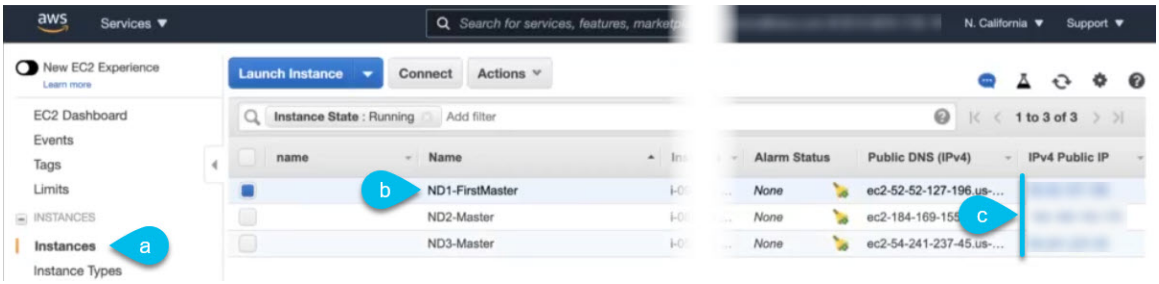
Step 8 Wait for the deployment to complete, then start the VMs.

You can view the status of the instance deployment in the **CloudFormation** page, for example `CREATE_IN_PROGRESS`. You can click the refresh button in the top right corner of the page to update the status.

When the status changes to `CREATE_COMPLETE`, you can proceed to the next step.



Step 9 Note down all nodes' public IP addresses.



a) After all instances are deployed, navigate to the AWS console's **EC2** > **Instances** page.

b) Note down which node is labeled as `ND1-Master`.

You will use this node's public IP address to complete cluster configuration.

c) Note down all nodes' public IP addresses.

You will provide this information to the GUI bootstrap wizard in the following steps.

Step 10 Enable password-based login on all nodes.

By default only PEM-based login is enabled for each node. To be able to SSH into the nodes using a password, as required by the GUI setup wizard, you will need to explicitly enable password-based logins.

Note
You must enable password-based login on all nodes before proceeding to cluster bootstrap described in the following steps or you will not be able to complete the cluster configuration.

a) SSH into one of the instances using its public IP address and the PEM file.

Use the PEM file you created for this as part of Prerequisites and guidelines for deploying the Nexus Dashboard cluster in Amazon Web Services, on page 1.

```
# ssh -i <pem-file-name>.pem rescue-user@<node-public-ip>
```

b) Enable password-based login.

On each node, run the following command:

```
# acs login-prompt enable
```

c) Repeat this step for the other two instances.

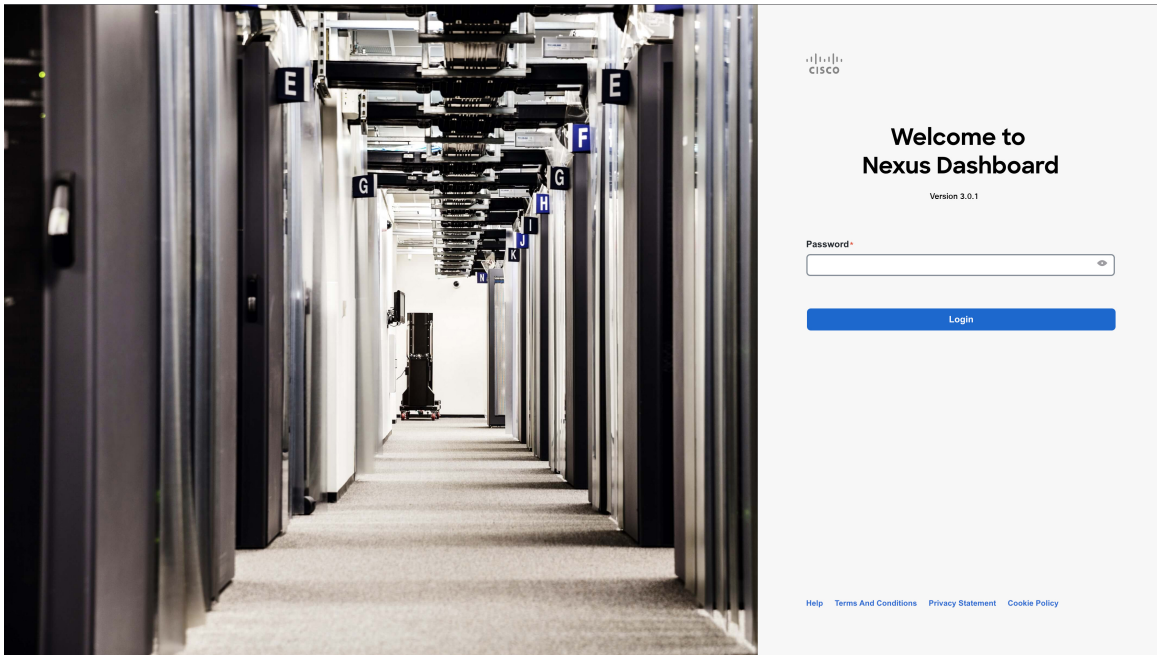
Step 11 Open your browser and navigate to `https://<first-node-public-ip>` to open the GUI.

Note

You must use the public IP address of the first node (`ND1-Master`) or cluster configuration cannot be completed.

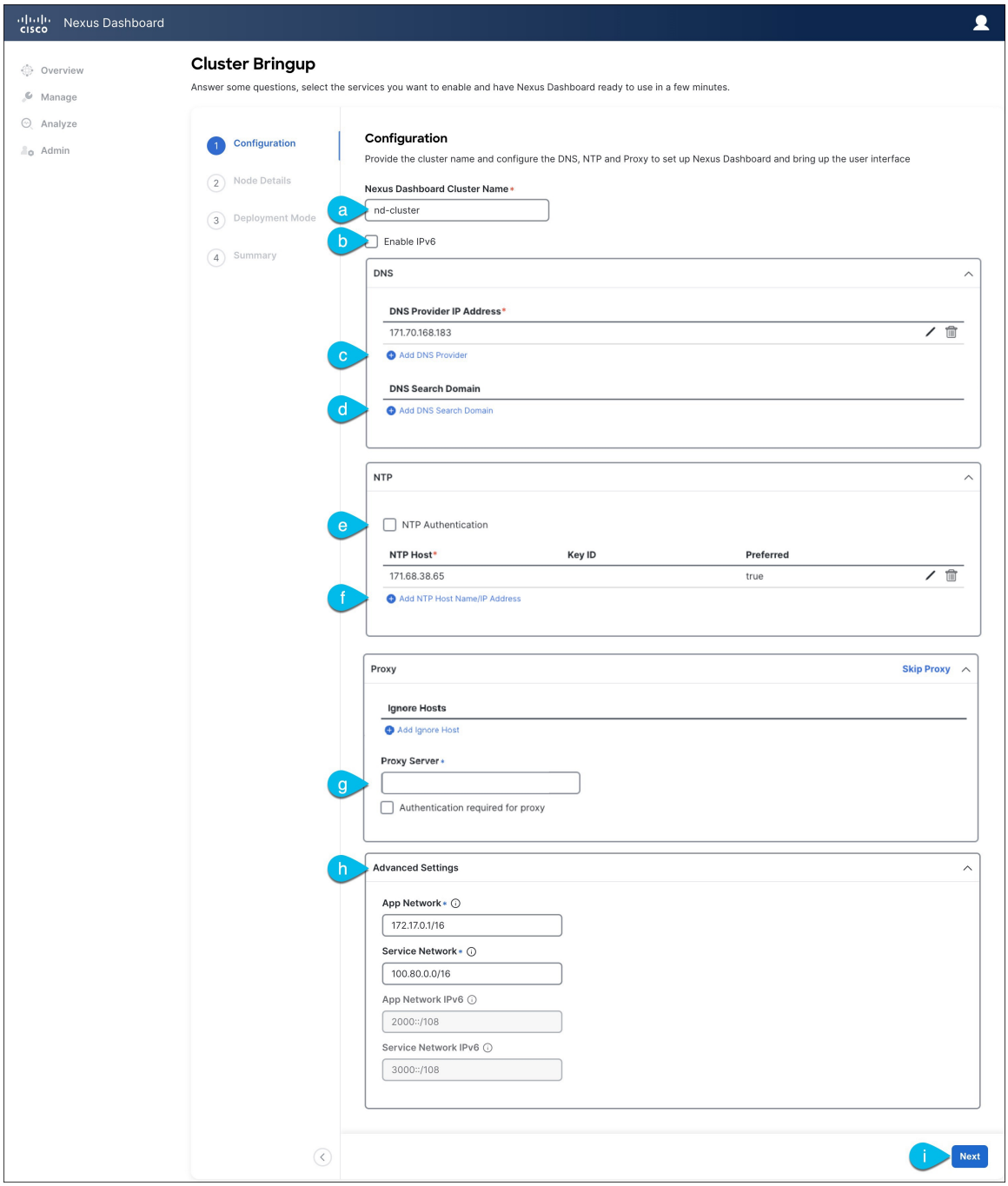
The rest of the configuration workflow takes place from the first node's GUI. You do not need to log in to or configure the other two nodes directly.

Enter the password you provided for the first node and click **Login**



Step 12 Provide the **Cluster Details**.

In the **Cluster Details** screen of the **Cluster Bringup** wizard, provide the following information:



a) Provide the **Cluster Name** for this Nexus Dashboard cluster.

- The cluster name must follow the RFC-1123 requirements.
- If you use multi-cluster connectivity to establish connectivity between multiple Nexus Dashboard clusters, the names of the clusters that you plan to connect together must be unique.

b) (Optional) If you want to enable IPv6 functionality for the cluster, check the **Enable IPv6** checkbox.

c) Click **+Add DNS Provider** to add one or more DNS servers.

After you've entered the information, click the checkmark icon to save it.

d) (Optional) Click **+Add DNS Search Domain** to add a search domain.

After you've entered the information, click the checkmark icon to save it.

e) (Optional) If you want to enable NTP server authentication, enable the **NTP Authentication** checkbox and click **Add NTP Key**.

In the additional fields, provide the following information:

- **NTP Key** – a cryptographic key that is used to authenticate the NTP traffic between the Nexus Dashboard and the NTP server(s). You will define the NTP servers in the following step, and multiple NTP servers can use the same NTP key.
- **Key ID** – each NTP key must be assigned a unique key ID, which is used to identify the appropriate key to use when verifying the NTP packet.
- **Auth Type** – this release supports `MD5`, `SHA`, and `AES128CMAC` authentication types.
- Choose whether this key is **Trusted**. Untrusted keys cannot be used for NTP authentication.

Note
After you've entered the information, click the checkmark icon to save it.

For the complete list of NTP authentication requirements and guidelines, see Prerequisites and guidelines for all enabled services.

f) Click **+Add NTP Host Name/IP Address** to add one or more NTP servers.

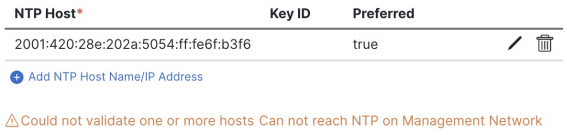
In the additional fields, provide the following information:

- **NTP Host** – you must provide an IP address; fully qualified domain name (FQDN) are not supported.
- **Key ID** – if you want to enable NTP authentication for this server, provide the key ID of the NTP key you defined in the previous step.

  If NTP authentication is disabled, this field is grayed out.
- Choose whether this NTP server is **Preferred**.

After you've entered the information, click the checkmark icon to save it.

Note
If the node into which you are logged in is configured with only an IPv4 address, but you have checked **Enable IPv6** in a previous step and provided an IPv6 address for an NTP server, you will get the following validation error:



This is because the node does not have an IPv6 address yet (you will provide it in the next step) and is unable to connect to an IPv6 address of the NTP server.

In this case, simply finish providing the other required information as described in the following steps and click **Next** to proceed to the next screen where you will provide IPv6 addresses for the nodes.

If you want to provide additional NTP servers, click **+Add NTP Host** again and repeat this substep.

g) Provide a **Proxy Server**, then click **Validate** it.

For clusters that do not have direct connectivity to Cisco cloud, we recommend configuring a proxy server to establish the connectivity. This allows you to mitigate risk from exposure to non-conformant hardware and software in your fabrics.

You can also choose to provide one or more IP addresses communication with which should skip proxy by clicking **+Add Ignore Host**.

The proxy server must have the following URLs enabled:

```
dcappcenter.cisco.com
svc.intersight.com
svc.ucs-connect.com
svc-static1.intersight.com
svc-static1.ucs-connect.com
```

If you want to skip proxy configuration, click **Skip Proxy**.

h) (Optional) If your proxy server required authentication, enable **Authentication required for Proxy**, provide the login credentials, then click **Validate**.

i) (Optional) Expand the **Advanced Settings** category and change the settings if required.

Under advanced settings, you can configure the following:

- Provide custom **App Network** and **Service Network**.

  The application overlay network defines the address space used by the application's services running in the Nexus Dashboard. The field is pre-populated with the default `172.17.0.1/16` value.

  The services network is an internal network used by the Nexus Dashboard and its processes. The field is pre-populated with the default `100.80.0.0/16` value.

  If you have checked the **Enable IPv6** option earlier, you can also define the IPv6 subnets for the App and Service networks.

  Application and Services networks are described in the Prerequisites and guidelines for all enabled services section earlier in this document.

j) Click **Next** to continue.

Step 13 In the **Node Details** screen, provide the node's information.

a) Click the **Edit** button next to the first node.

b) Provide the node's **Name**.

The **Management Network** and **Data Network** information will be already populated from the VPC subnet you have configured before deploying the cluster.

The cluster creates six subnets from the given VPC CIDR, from which the data and management networks will be allocated for the cluster's three nodes.

c) Leave IPv6 addresses and VLAN fields blank.

Cloud Nexus Dashboard clusters do not support these options.

d) Click **Save** to save the changes.

Step 14 Click **Add Node** to add the second node to the cluster.

The **Node Details** window opens.

a) Provide the node's **Name**.

b) In the **Credentials** section, provide the node's **Public IP Address** and the password you provided during template deployment, then click **Verify**.

The IP address and password are used to pull that node's **Management Network** and **Data Network** information, which will be populated in the fields below.

c) Ensure that you select `Primary` for the node type.

Only 3-node clusters are supported for cloud deployments, so all nodes must be `Primary`.

d) Click **Save** to save the changes.

Step 15 Repeat the previous step to add the 3rd node.

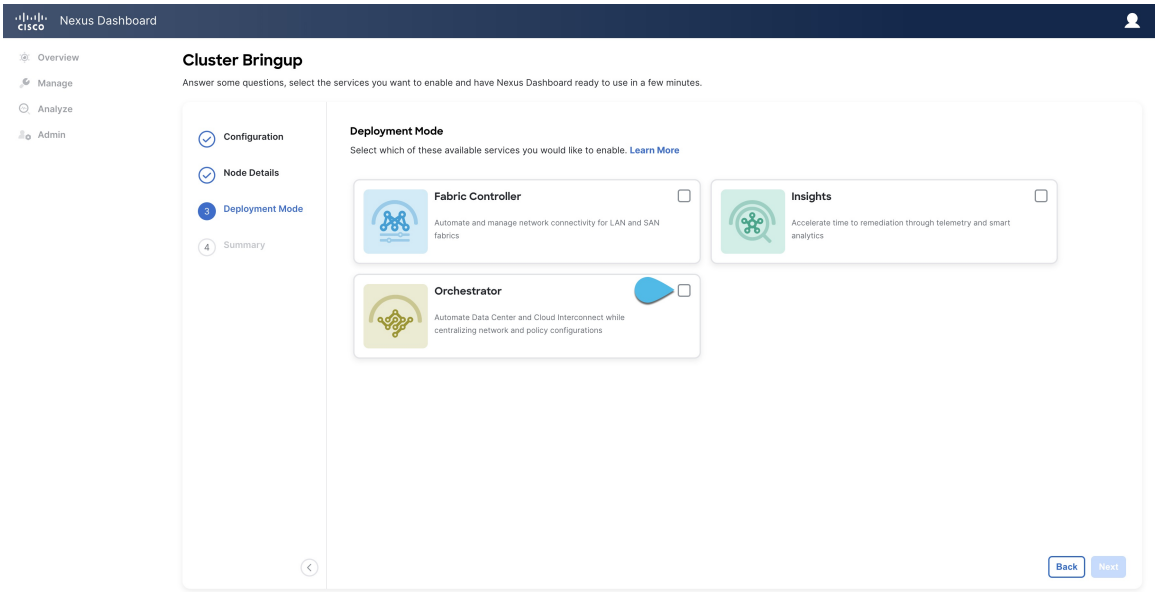
Step 16 In the **Node Details** page, click **Next** to continue.

Step 17 Choose the **Deployment Mode** for the cluster.

a) Choose the services you want to enable.

Prior to release 3.1(1), you had to download and install individual services after the initial cluster deployment was completed. Now you can choose to enable the services during the initial installation.

Note
Cloud deployments support the Orchestrator service, so you must not select any other deployment modes.



b) Click **Add Persistent Service IPs/Pools** to provide one or more persistent IPs required by Insights or Fabric Controller services.

For more information about persistent IPs, see the Prerequisites and guidelines for all enabled services section.

c) Click **Next** to proceed.

Step 18 In the **Summary** screen, review and verify the configuration information, click **Save**, and click **Continue** to confirm the correct deployment mode and proceed with building the cluster.

During the node bootstrap and cluster bring-up, the overall progress as well as each node's individual progress will be displayed in the UI. If you do not see the bootstrap progress advance, manually refresh the page in your browser to update the status.

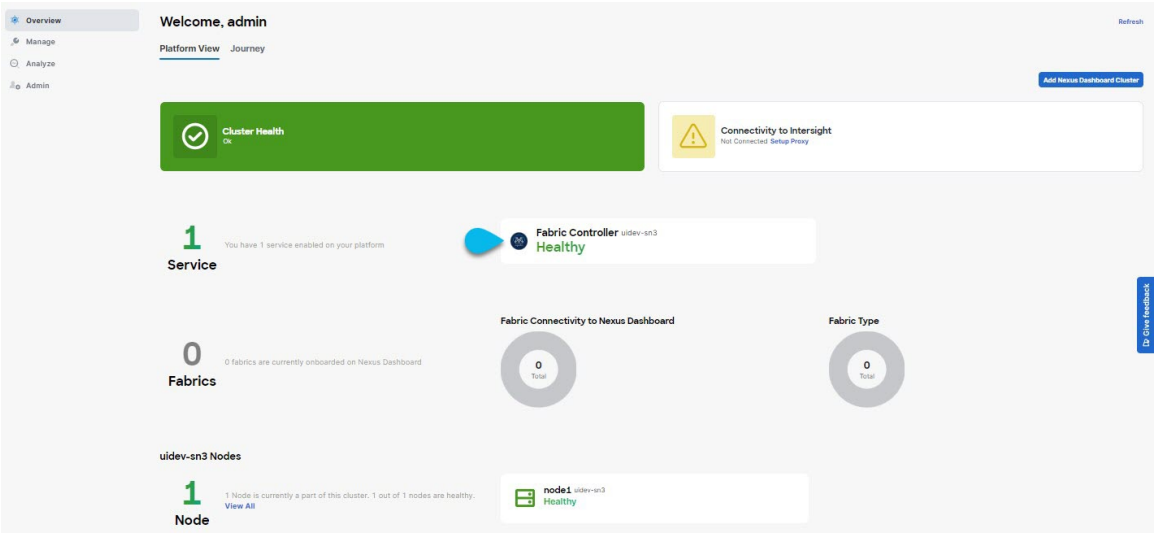
It may take up to 30 minutes for the cluster to form and all the services to start. When cluster configuration is complete, the page will reload to the Nexus Dashboard GUI.

Step 19 Verify that the cluster is healthy.

Depending of the deployment mode, it may take more than 30 minutes for the cluster to form and all the services to start.

After the cluster becomes available, you can access it by browsing to any one of your nodes' management IP addresses. The default password for the `admin` user is the same as the `rescue-user` password you chose for the first node. During this time, the UI will display a banner at the top stating "Service Installation is in progress, Nexus Dashboard configuration tasks are currently disabled":

After all the cluster is deployed and all services are started, you can check the **Overview** page to ensure the cluster is healthy:



Alternatively, you can log in to any one node via SSH as the `rescue-user` using the password you provided during node deployment and using the `acs health` command to check the status:

- While the cluster is converging, you may see the following outputs:

```
$ acs health
k8s install is in-progress
```

```
$ acs health
k8s services not in desired state - [...]
```

```
$ acs health
k8s: Etcd cluster is not ready
```

- When the cluster is up and running, the following output will be displayed:

```
$ acs health
All components are healthy
```

Note

In some situations, you might power cycle a node (power it off and then back on) and find it stuck in this stage:

```
deploy base system services
```

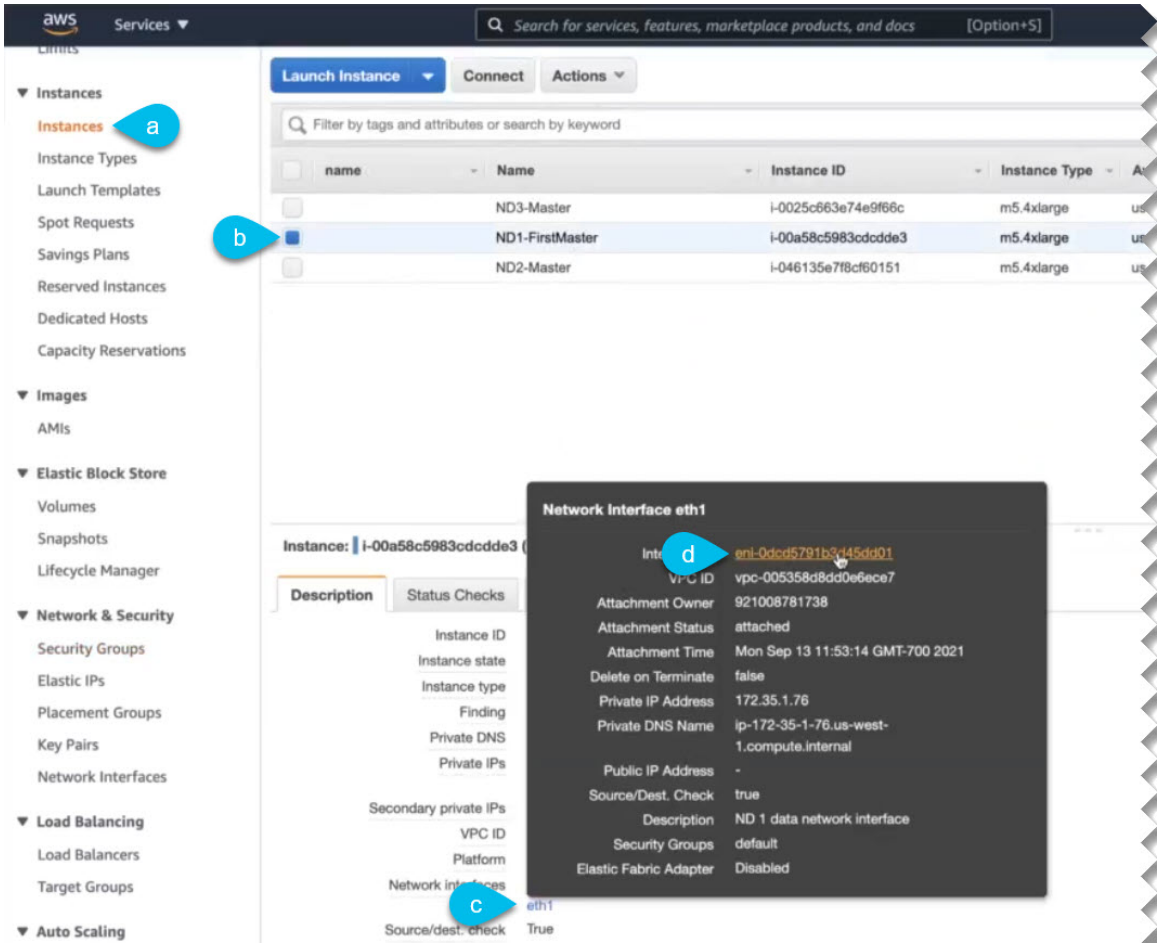
This is due to an issue with `etcd` on the node after a reboot of the pND (Physical Nexus Dashboard) cluster.

To resolve the issue, enter the `acs reboot clean` command on the affected node.

Step 20 Update the nodes' security group with required ports.

This step describes how to update the Nexus Dashboard nodes' instances with the required port configuration for on-boarding Cisco NDFC fabrics. If you do not plan on on-boarding any NDFC fabrics to your Nexus Dashboard cluster, you can skip this step.

Navigate to one of the nodes' data interface:



a) In the AWS console, navigate to **Instances**.
b) Select one of the Nexus Dashboard instances.

You will make changes to the default security group, so you only need to select one of the nodes.

c) Click the data interface (`eth1`).
d) Click the **Interface ID**.

The **Network Interface** properties page opens.

e) In the **Network Interface** page, click `default` in the **Security groups** column of the interface.

Add the new rules:

a) In the default security group's page, select the **Inbound rules** tab.
b) Click **Edit inbound rules**.
c) In the **Edit inbound rules** page, click **Add rule** to add a new inbound security rule, then provide the details to allow inbound communication on port 443.

   Provide the following information for the new rule:

   - For **Type**, select `Custom TCP`.
   - For **Port range**, enter `443`.
   - For **Source**, provide the IP addresses of the NDFC controllers which you plan to onboard to your Nexus Dashboard.

d) Still in the **Edit inbound rules** page, click **Add rule** to add another inbound security rule, then provide the details to allow inbound communication on port 9092.

   Provide the following information for the new rule:

   - For **Type**, select `Custom TCP`.
   - For **Port range**, enter `9092`.
   - For **Source**, provide the IP addresses of the NDFC controllers which you plan to onboard to your Nexus Dashboard.