Deploy Cloud Data Sense

Cloud Data Sense

NetApp
June 15, 2022

Table of Contents

Deploy Cloud Data Sense

Deploy Cloud Data Sense in the cloud

Complete a few steps to deploy Cloud Data Sense in the cloud.

Note that you can also deploy Data Sense on a Linux host that has internet access. The type of installation may be a good option if you prefer to scan on-premises ONTAP systems using a Data Sense instance that's also located on premises — but this is not a requirement. The software functions exactly the same way regardless of which installation method you choose.

Quick start

Get started quickly by following these steps, or scroll down to the remaining sections for full details.

Create a Connector

If you don't already have a Connector, create a Connector now. See creating a Connector in AWS, creating a Connector in Azure, or creating a Connector in GCP.

You can also deploy the Connector on-premises on a Linux host in your network or in the cloud.

Review prerequisites

Ensure that your environment can meet the prerequisites. This includes outbound internet access for the instance, connectivity between the Connector and Cloud Data Sense over port 443, and more. See the complete list.

The default configuration requires 16 vCPUs for the Cloud Data Sense instance. See more details about the instance type.

Deploy Cloud Data Sense

Launch the installation wizard to deploy the Cloud Data Sense instance in the cloud.

Subscribe to the Cloud Data Sense service

The first 1 TB of data that Cloud Data Sense scans in Cloud Manager is free. A Cloud Manager subscription through your cloud provider Marketplace, or a BYOL license from NetApp, is required to continue scanning data after that point.

Create a Connector

If you don't already have a Connector, create a Connector in your cloud provider. See creating a Connector in AWS or creating a Connector in Azure, or creating a Connector in GCP. In most cases you will probably have a Connector set up before you attempt to activate Cloud Data Sense because most Cloud Manager features require a Connector, but there are cases where you'll you need to set one up now.

There are some scenarios where you have to use a Connector that's deployed in a specific cloud provider:

- When scanning data in Cloud Volumes ONTAP in AWS, Amazon FSx for ONTAP, or in AWS S3 buckets, you use a connector in AWS.
- When scanning data in Cloud Volumes ONTAP in Azure or in Azure NetApp Files, you use a connector in Azure.
- When scanning data in Cloud Volumes ONTAP in GCP, you use a Connector in GCP.

On-prem ONTAP systems, non-NetApp file shares, generic S3 Object storage, databases, OneDrive folders, SharePoint accounts, and Google Drive accounts can be scanned when using any of these cloud Connectors.

Note that you can also deploy the Connector on-premises on a Linux host in your network or in the cloud. Some users planning to install Data Sense on-prem may also choose to install the Connector on-prem.

As you can see, there may be some situations where you need to use multiple Connectors.

If you're planning on scanning Azure NetApp Files volumes, you need to make sure you're deploying in the same region as the volumes you wish to scan.

Review prerequisites

Review the following prerequisites to make sure that you have a supported configuration before you deploy Cloud Data Sense in the cloud.

Enable outbound internet access from Cloud Data Sense

Cloud Data Sense requires outbound internet access. If your virtual or physical network uses a proxy server for internet access, ensure that the Data Sense instance has outbound internet access to contact the following endpoints. When you deploy Data Sense in the cloud, it's located in the same subnet as the Connector.

Review the appropriate table below depending on whether you are deploying Cloud Data Sense in AWS, Azure, or GCP.

Required endpoints for AWS deployments:

Endpoints	Purpose
https://cloudmanager.cloud.netapp.com	Communication with the Cloud Manager service, which includes NetApp accounts.
https://netapp-cloud-account.auth0.com https://auth0.com	Communication with NetApp Cloud Central for centralized user authentication.
https://cloud-compliance-support-netapp.s3.us-west-2.amazonaws.com https://hub.docker.com https://auth.docker.io https://registry-1.docker.io https://index.docker.io/ https://dseasb33srnrn.cloudfront.net/ https://production.cloudflare.docker.com/	Provides access to software images, manifests, and templates.
https://kinesis.us-east-1.amazonaws.com	Enables NetApp to stream data from audit records.

Endpoints	Purpose
https://cognito-idp.us-east-1.amazonaws.com https://cognito-identity.us-east-1.amazonaws.com https://user-feedback-store-prod.s3.us-west-2.amazonaws.com https://customer-data-production.s3.us-west-2.amazonaws.com	Enables Cloud Data Sense to access and download manifests and templates, and to send logs and metrics.

Required endpoints for Azure and GCP deployments:

Endpoints	Purpose
https://cloudmanager.cloud.netapp.com	Communication with the Cloud Manager service, which includes NetApp accounts.
https://netapp-cloud-account.auth0.com https://auth0.com	Communication with NetApp Cloud Central for centralized user authentication.
https://support.compliance.cloudmanager.cloud.netapp.com/ https://hub.docker.com https://auth.docker.io https://registry-1.docker.io https://index.docker.io/ https://dseasb33srnrn.cloudfront.net/ https://production.cloudflare.docker.com/	Provides access to software images, manifests, templates, and to send logs and metrics.
https://support.compliance.cloudmanager.cloud.netapp.com/	Enables NetApp to stream data from audit records.

Ensure that Cloud Manager has the required permissions

Ensure that Cloud Manager has permissions to deploy resources and create security groups for the Cloud Data Sense instance. You can find the latest Cloud Manager permissions in the policies provided by NetApp.

Check your vCPU limits

Ensure that your cloud provider's vCPU limit allows for the deployment of an instance with 16 cores. You'll need to verify the vCPU limit for the relevant instance family in the region where Cloud Manager is running. See the required instance types.

See the following links for more details on vCPU limits:

- AWS documentation: Amazon EC2 service quotas
- Azure documentation: Virtual machine vCPU quotas
- Google Cloud documentation: Resource quotas

Note that you can deploy Data Sense on a system with fewer CPUs and less RAM, but there are limitations when using these systems. See Using a smaller instance type for details.

Ensure that the Cloud Manager Connector can access Cloud Data Sense

Ensure connectivity between the Connector and the Cloud Data Sense instance. The security group for the Connector must allow inbound and outbound traffic over port 443 to and from the Data Sense instance. This connection enables deployment of the Data Sense instance and enables you to view information in the Compliance and Governance tabs. Cloud Data Sense is supported in Government regions in AWS and Azure.

Additional inbound and outbound rules are required for AWS and AWS GovCloud deployments. See Rules for the Connector in AWS for details.

Additional inbound and outbound rules are required for Azure and Azure Government deployments. See Rules for the Connector in Azure for details.

Ensure that you can keep Cloud Data Sense running

The Cloud Data Sense instance needs to stay on to continuously scan your data.

Ensure web browser connectivity to Cloud Data Sense

After Cloud Data Sense is enabled, ensure that users access the Cloud Manager interface from a host that has a connection to the Data Sense instance.

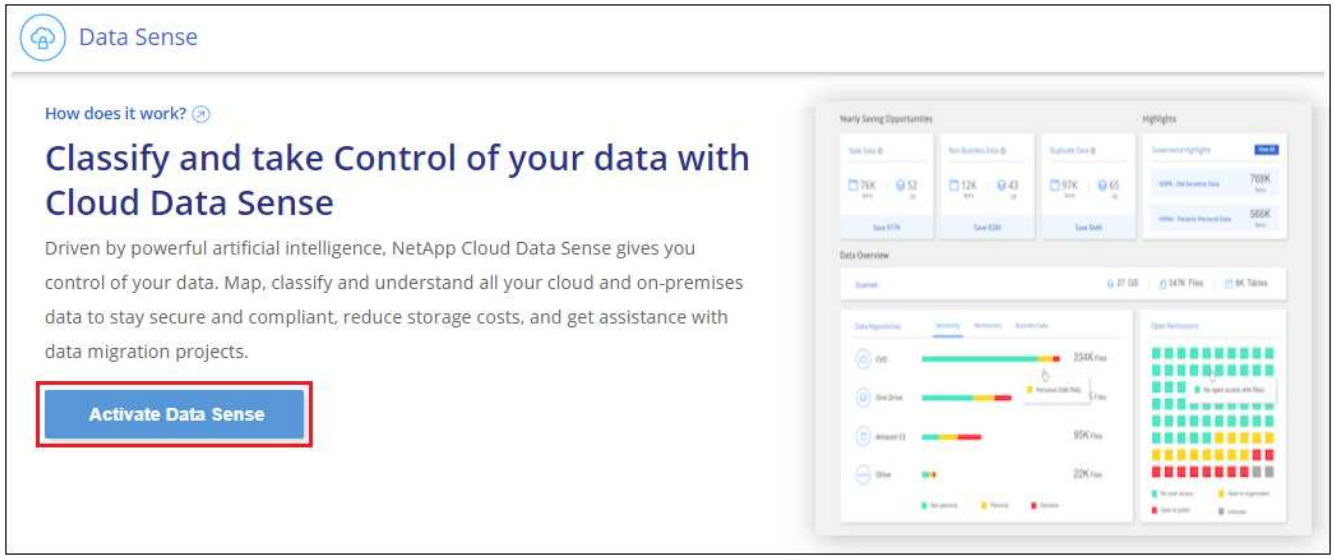
The Data Sense instance uses a private IP address to ensure that the indexed data isn't accessible to the internet. As a result, the web browser that you use to access Cloud Manager must have a connection to that private IP address. That connection can come from a direct connection to your cloud provider (for example, a VPN), or from a host that's inside the same network as the Data Sense instance.

Deploy Data Sense in the cloud

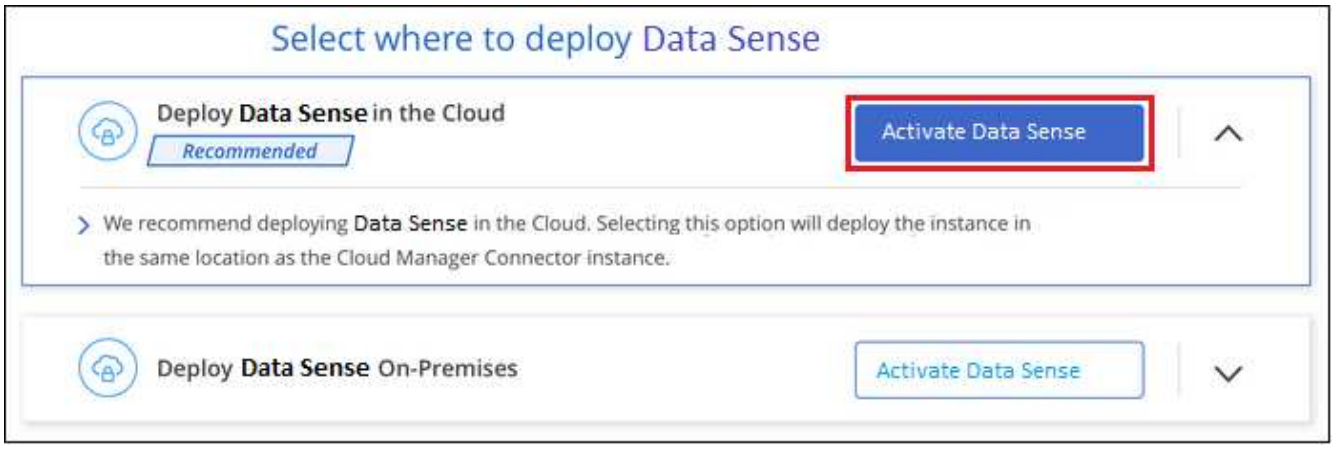
Follow these steps to deploy an instance of Cloud Data Sense in the cloud.

Steps

1. In Cloud Manager, click **Data Sense**.
2. Click **Activate Data Sense**.



3. Click **Activate Data Sense** to start the cloud deployment wizard.



4. The wizard displays progress as it goes through the deployment steps. It will stop and ask for input if it runs into any issues.

5. When the instance is deployed, click **Continue to configuration** to go to the *Configuration* page.

Result

Cloud Manager deploys the Cloud Data Sense instance in your cloud provider.

What's Next

From the Configuration page you can select the data sources that you want to scan.

You can also set up licensing for Cloud Data Sense at this time. You will not be charged until the amount of data exceeds 1 TB.

Deploy Cloud Data Sense on a Linux host that has internet access

Complete a few steps to deploy Cloud Data Sense on a Linux host in your network, or in the cloud, that has internet access.

The on-prem installation may be a good option if you prefer to scan on-premises ONTAP systems using a Data Sense instance that's also located on premises — but this is not a requirement. The software functions exactly the same way regardless of which installation method you choose.

Note that you can also deploy Data Sense in an on-premises site that doesn't have internet access for

completely secure sites.

Quick start

Get started quickly by following these steps, or scroll down to the remaining sections for full details.

Create a Connector

If you don't already have a Connector, create a Connector now. See creating a Connector in AWS, creating a Connector in Azure, or creating a Connector in GCP.

You can also deploy the Connector on-premises on a Linux host in your network or in the cloud.

Review prerequisites

Ensure that your environment can meet the prerequisites. This includes outbound internet access for the instance, connectivity between the Connector and Cloud Data Sense over port 443, and more. See the complete list.

You also need a Linux system that meets the following requirements.

Deploy Cloud Data Sense

Download the Cloud Data Sense software from the NetApp Support Site and copy the installer file to the Linux host you plan to use. Then launch the installation wizard and follow the prompts to deploy the Data Sense instance.

Subscribe to the Cloud Data Sense service

The first 1 TB of data that Cloud Data Sense scans in Cloud Manager is free. A subscription to your cloud provider Marketplace, or a BYOL license from NetApp, is required to continue scanning data after that point.

Create a Connector

If you don't already have a Connector, create a Connector in your cloud provider. See creating a Connector in AWS or creating a Connector in Azure, or creating a Connector in GCP. In most cases you will probably have a Connector set up before you attempt to activate Cloud Data Sense because most Cloud Manager features require a Connector, but there are cases where you'll you need to set one up now.

There are some scenarios where you have to use a Connector that's deployed in a specific cloud provider:

- When scanning data in Cloud Volumes ONTAP in AWS, Amazon FSx for ONTAP, or in AWS S3 buckets, you use a connector in AWS.
- When scanning data in Cloud Volumes ONTAP in Azure or in Azure NetApp Files, you use a connector in Azure.
- When scanning data in Cloud Volumes ONTAP in GCP, you use a connector in GCP.

On-prem ONTAP systems, non-NetApp file shares, generic S3 Object storage, databases, OneDrive folders, SharePoint accounts, and Google Drive accounts can be scanned using any of these cloud Connectors.

Note that you can also deploy the Connector on-premises on a Linux host in your network or in the cloud. Some users planning to install Data Sense on-prem may also choose to install the Connector on-prem.

As you can see, there may be some situations where you need to use multiple Connectors.

If you're planning on scanning Azure NetApp Files volumes, you need to make sure you're deploying in the same region as the volumes you wish to scan.

Prepare the Linux host system

Data Sense software must run on a host that meets specific operating system requirements, RAM requirements, software requirements, and so on. Data Sense is not supported on a host that is shared with other applications - the host must be a dedicated host.

- Operating system: Red Hat Enterprise Linux or CentOS version 8.0, 8.1, or 8.4
  - The OS must be capable of installing the docker engine (for example, disable the *firewalld* service if needed)
- Disk: SSD with 500 GiB available on /, or
  - 100 GiB available on /opt
  - 400 GiB available on /var
  - 5 GiB on /tmp
- RAM: 64 GB (swap memory must be disabled on the host)
- CPU: 16 cores

  Note that you can deploy Data Sense on a system with fewer CPUs and less RAM, but there are limitations when using these systems. See Using a smaller instance type for details.

- A Red Hat Enterprise Linux system must be registered with Red Hat Subscription Management. If it's not registered, the system can't access repositories to update required 3rd party software during installation.
- The following software must be installed on the host. If it doesn't already exist on the host, then the installer will install the software for you:
  - Docker Engine version 19 or later. View installation instructions.
  - Python 3 version 3.6 or later. View installation instructions.

Verify Cloud Manager and Data Sense prerequisites

Review the following prerequisites to make sure that you have a supported configuration before you deploy Cloud Data Sense on a Linux system.

Enable outbound internet access from Cloud Data Sense

Cloud Data Sense requires outbound internet access. If your virtual or physical network uses a proxy server for internet access, ensure that the Data Sense instance has outbound internet access to contact the following endpoints.

Endpoints	Purpose
https://cloudmanager.cloud.netapp.com	Communication with the Cloud Manager service, which includes NetApp accounts.

Endpoints	Purpose
https://netapp-cloud-account.auth0.com https://auth0.com	Communication with NetApp Cloud Central for centralized user authentication.
https://support.compliance.cloudmanager.cloud.netapp.com/ https://hub.docker.com https://auth.docker.io https://registry-1.docker.io https://index.docker.io/ https://dseasb33srnrn.cloudfront.net/ https://production.cloudflare.docker.com/	Provides access to software images, manifests, templates, and to send logs and metrics.
https://support.compliance.cloudmanager.cloud.netapp.com/	Enables NetApp to stream data from audit records.
https://github.com/docker https://download.docker.com http://mirror.centos.org http://mirrorlist.centos.org http://mirror.centos.org/centos/7/extras/x86_64/Packages/container-selinux-2.107-3.el7.noarch.rpm	Provides prerequisite packages for installation.

Ensure that Cloud Manager has the required permissions

Ensure that Cloud Manager has permissions to deploy resources and create security groups for the Cloud Data Sense instance. You can find the latest Cloud Manager permissions in the policies provided by NetApp.

Ensure that the Cloud Manager Connector can access Cloud Data Sense

Ensure connectivity between the Connector and the Cloud Data Sense instance. The security group for the Connector must allow inbound and outbound traffic over port 443 to and from the Data Sense instance.

This connection enables deployment of the Data Sense instance and enables you to view information in the Compliance and Governance tabs.

Make sure port 8080 is open so you can see the installation progress in Cloud Manager.

Ensure that you can keep Cloud Data Sense running

The Cloud Data Sense instance needs to stay on to continuously scan your data.

Ensure web browser connectivity to Cloud Data Sense

After Cloud Data Sense is enabled, ensure that users access the Cloud Manager interface from a host that has a connection to the Data Sense instance.

The Data Sense instance uses a private IP address to ensure that the indexed data isn't accessible to the internet. As a result, the web browser that you use to access Cloud Manager must have a connection to that private IP address. That connection can come from a direct connection to your cloud provider (for example, a VPN), or from a host that's inside the same network as the Data Sense instance.

Deploy Data Sense on premises

For typical configurations you'll install the software on a single host system. See those steps here.

For very large configurations where you'll be scanning petabytes of data, you can include multiple hosts to provide additional processing power. See those steps here.

See Preparing the Linux host system and Reviewing prerequisites for the full list of requirements before you deploy Cloud Data Sense.

Upgrades to Data Sense software is automated as long as the instance has internet connectivity.

> Cloud Data Sense is currently unable to scan S3 buckets, Azure NetApp Files, or FSx for ONTAP when the software is installed on premises. In these cases you'll need to deploy a separate Connector and instance of Data Sense in the cloud and switch between Connectors for your different data sources.

Single-host installation for typical configurations

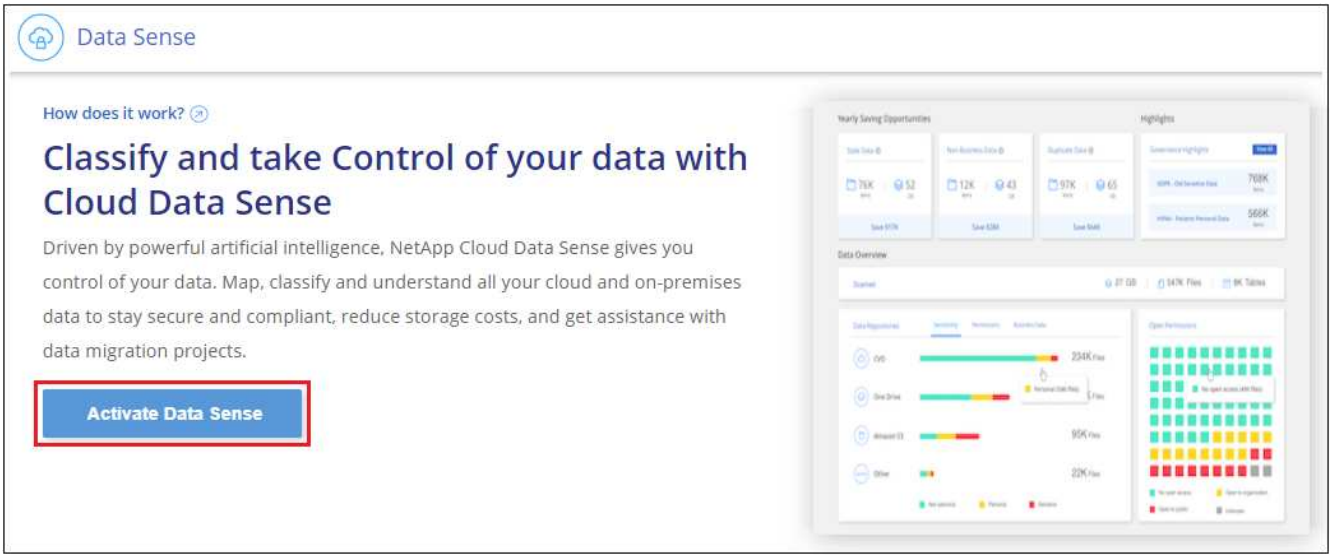
Follow these steps when installing Data Sense software on a single on-premises host.

What you'll need

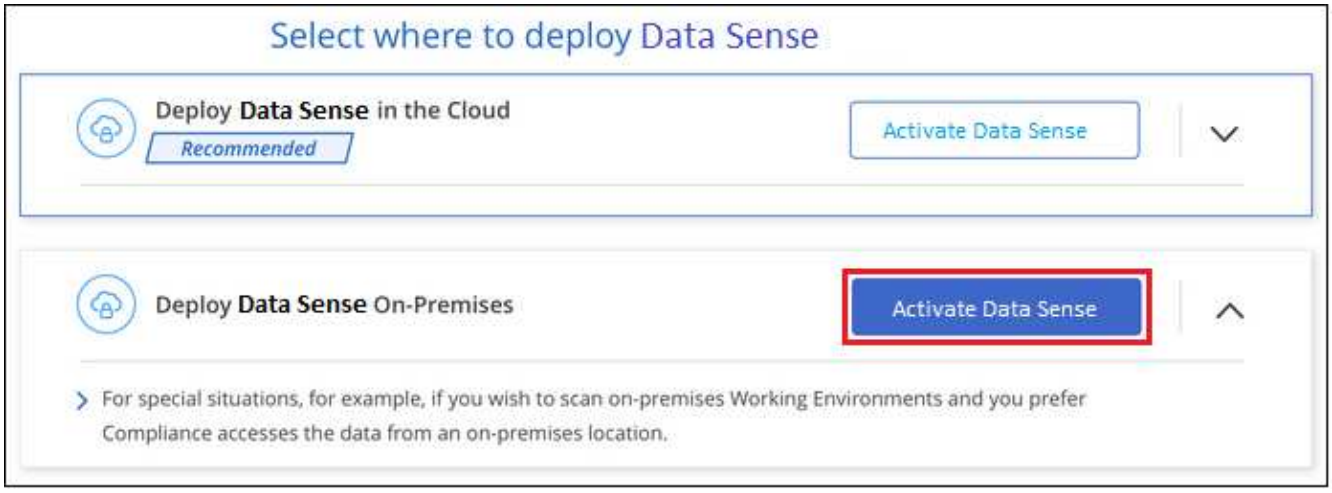
- Verify that your Linux system meets the host requirements.
- (Optional) Verify that the system has the two prerequisite software packages installed (Docker Engine and Python 3). The installer will install this software if it is not already on the system.
- Make sure you have root privileges on the Linux system.
- If you are using a proxy, and it is performing TLS interception, you'll need to know the path on the Data Sense Linux system where the TLS CA certificates are stored.
- Verify that your offline environment meets the required permissions and connectivity.

Steps

1. Download the Cloud Data Sense software from the NetApp Support Site. The file you should select is named **cc_onprem_installer_<version>.tar.gz**.
2. Copy the installer file to the Linux host you plan to use (using `scp` or some other method).
3. In Cloud Manager, click **Data Sense**.
4. Click **Activate Data Sense**.



5. Click **Activate Data Sense** to start the on-prem deployment wizard.



6. In the *Deploy Data Sense On Premises* dialog, copy the provided command and paste it in a text file so you can use it later, and click **Close**. For example:

```
sudo ./install.sh -a 12345 -c 27AG75 -t 2198qq
```

7. Unzip the installer file on the host machine, for example:

```
tar -xzf cc_onprem_installer_1.13.1.tar.gz
```

8. When prompted by the installer, you can enter the required values in a series of prompts, or you can provide the required parameters as command line arguments to the installer:

Enter parameters as prompted:	Enter the full command:
1. Paste the information you copied from step 6: `sudo ./install.sh -a <account_id> -c <agent_id> -t <token>` 2. Enter the IP address or host name of the Data Sense host machine so it can be accessed by the Connector instance. 3. Enter the IP address or host name of the Cloud Manager Connector host machine so it can be accessed by the Data Sense instance. 4. Enter proxy details as prompted. If your Cloud Manager already uses a proxy, there is no need to enter this information again here since Data Sense will automatically use the proxy used by Cloud Manager.	Alternatively, you can create the whole command in advance, providing the necessary host and proxy parameters: `sudo ./install.sh -a <account_id> -c <agent_id> -t <token> --host <ds_host> --manager-host <cm_host> --proxy-host <proxy_host> --proxy-port <proxy_port> --proxy-scheme <proxy_scheme> --proxy-user <proxy_user> --proxy-password <proxy_password> --cacert-folder-path <ca_cert_dir>`

Variable values:

- *account_id* = NetApp Account ID

- *agent_id* = Connector ID
- *token* = jwt user token
- *ds_host* = IP address or host name of the Data Sense Linux system.
- *cm_host* = IP address or host name of the Cloud Manager Connector system.
- *proxy_host* = IP or host name of the proxy server if the host is behind a proxy server.
- *proxy_port* = Port to connect to the proxy server (default 80).
- *proxy_scheme* = Connection scheme: https or http (default http).
- *proxy_user* = Authenticated user to connect to the proxy server, if basic authentication is required.
- *proxy_password* = Password for the user name that you specified.
- *ca_cert_dir* = Path on the Data Sense Linux system containing additional TLS CA certificate bundles. Only required if the proxy is performing TLS interception.

Result

The Cloud Data Sense installer installs packages, installs docker, registers the installation, and installs Data Sense. Installation can take 10 to 20 minutes.

If there is connectivity over port 8080 between the host machine and the Connector instance, you will see the installation progress in the Data Sense tab in Cloud Manager.

What's Next

From the Configuration page you can select the data sources that you want to scan.

You can also set up licensing for Cloud Data Sense at this time. You will not be charged until the amount of data exceeds 1 TB.

Multi-host installation for large configurations

For very large configurations where you'll be scanning petabytes of data, you can include multiple hosts to provide additional processing power. When using multiple host systems, the primary system is called the *Manager node* and the additional systems that provide extra processing power are called *Scanner nodes*.

Follow these steps when installing Data Sense software on multiple on-premises hosts.

What you'll need

- Verify that all your Linux systems for the Manager and Scanner nodes meet the host requirements.
- (Optional) Verify that the systems have the two prerequisite software packages installed (Docker Engine and Python 3). The installer will install this software if it is not already on the systems.
- Make sure you have root privileges on the Linux systems.
- Verify that your environment meets the required permissions and connectivity.
- You must have the IP addresses of the scanner node hosts that you plan to use.
- The following ports and protocols must be enabled on all hosts:

Port	Protocols	Description
2377	TCP	Cluster management communications
7946	TCP, UDP	Inter-node communication

Port	Protocols	Description
4789	UDP	Overlay network traffic
50	ESP	Encrypted IPsec overlay network (ESP) traffic
111	TCP, UDP	NFS Server for sharing files between the hosts (needed from each scanner node to manager node)
2049	TCP, UDP	NFS Server for sharing files between the hosts (needed from each scanner node to manager node)

Steps

1. Follow steps 1 through 7 from the Single-host installation on the manager node.
2. As shown in step 8, when prompted by the installer, you can enter the required values in a series of prompts, or you can provide the required parameters as command line arguments to the installer.

   In addition to the variables available for a single-host installation, a new option **-n <node_ip>** is used to specify the IP addresses of the scanner nodes. Multiple scanner node IPs are separated by a comma.

   For example, this command adds 3 scanner nodes:
   ```
   sudo ./install.sh -a <account_id> -c <agent_id> -t <token> --host <ds_host>
   --manager-host <cm_host> -n <node_ip1>,<node_ip2>,<node_ip3> --proxy-host
   <proxy_host> --proxy-port <proxy_port> --proxy-scheme <proxy_scheme> --proxy
   -user <proxy_user> --proxy-password <proxy_password>
   ```
3. Before the manager node installation completes, a dialog displays the installation command needed for the scanner nodes. Copy the command and save it in a text file. For example:

   ```
   sudo ./node_install.sh -m 10.11.12.13 -t ABCDEF-1-3u69m1-1s35212
   ```
4. On **each** scanner node host:
   a. Copy the Data Sense installer file (**cc_onprem_installer_<version>.tar.gz**) to the host machine (using `scp` or some other method).
   b. Unzip the installer file.
   c. Paste and execute the command that you copied in step 3.

      When the installation finishes on all scanner nodes and they have been joined to the manager node, the manager node installation finishes as well.

Result

The Cloud Data Sense installer finishes installing packages, docker, and registers the installation. Installation can take 10 to 20 minutes.

What's Next

From the Configuration page you can select the data sources that you want to scan.

You can also set up licensing for Cloud Data Sense at this time. You will not be charged until the amount of data exceeds 1 TB.

Deploy Cloud Data Sense on prem without internet access

Complete a few steps to deploy Cloud Data Sense on a host in an on-premises site that doesn't have internet access. This type of installation is perfect for your secure sites.

Note that you can also deploy Data Sense in an on-premises site that has internet access.

Supported data sources

When installed in this manner (sometimes called an "offline" or "dark" site), Data Sense can only scan data from data sources that are also local to the on-premises site. At this time, Data Sense can scan the following local data sources:

- On-premises ONTAP systems
- Database schemas
- Non-NetApp NFS or CIFS file shares
- Object Storage that uses the Simple Storage Service (S3) protocol

For special situations where you need a very secure Cloud Manager installation, but you also want to scan local data from OneDrive accounts or SharePoint accounts, you can use the Data Sense offline installer and provide internet access to a few select endpoints. See SharePoint and OneDrive special requirements for details.

There is no support currently for scanning Cloud Volumes ONTAP, Azure NetApp Files, FSx for ONTAP, AWS S3, or Google Drive accounts when Data Sense is deployed in a dark site.

Limitations

Most Data Sense features work when it is deployed in a site with no internet access. However, certain features that require internet access are not supported, for example:

- Managing Microsoft Azure Information Protection (AIP) labels
- Sending email alerts to Cloud Manager users when certain critical Policies return results
- Setting Cloud Manager roles for different users (for example, Account Admin or Compliance Viewer)
- Copying and synchronizing source files using Cloud Sync
- Receiving user feedback
- Automated software upgrades from Cloud Manager

  Both the Cloud Manager Connector and Data Sense will require periodic manual upgrades to enable new features. You can see the Data Sense version at the bottom of the Data Sense UI pages. Check the Cloud Data Sense Release Notes to see the new features in each release and whether you want those features. Then you can follow the steps to upgrade your Data Sense software.

Quick start

Get started quickly by following these steps, or scroll down to the remaining sections for full details.

Install the Cloud Manager Connector

If you don't already have a Connector installed at your offline on-premises site, deploy the Connector on a Linux host now.

Review Data Sense prerequisites

Ensure that your Linux system meets the host requirements, that it has all required software installed, and that your offline environment meets the required permissions and connectivity.

Deploy Data Sense

Download the Cloud Data Sense software from the NetApp Support Site and copy the installer file to the Linux host you plan to use. Then launch the installation wizard and follow the prompts to deploy the Cloud Data Sense instance.

Subscribe to the Cloud Data Sense service

The first 1 TB of data that Cloud Data Sense scans in Cloud Manager is free. A BYOL license from NetApp is required to continue scanning data after that point.

Install the Cloud Manager Connector

If you don't already have a Cloud Manager Connector installed at your offline on-premises site, deploy the Connector on a Linux host in your offline site.

Prepare the Linux host system

Data Sense software must run on a host that meets specific operating system requirements, RAM requirements, software requirements, and so on. Data Sense is not supported on a host that is shared with other applications - the host must be a dedicated host.

- Operating system: Red Hat Enterprise Linux or CentOS version 8.0, 8.1, or 8.4
  - The OS must be capable of installing the Docker Engine (for example, disable the *firewalld* service if needed)
- Disk: SSD with 500 GiB available on /, or
  - 100 GiB available on /opt
  - 400 GiB available on /var
  - 5 GiB on /tmp
- RAM: 64 GB (swap memory must be disabled on the host)
- CPU: 16 cores

  Note that you can deploy Data Sense on a system with fewer CPUs and less RAM, but there are limitations when using these systems. See Using a smaller instance type for details.

You must install the following software on the host before you install Data Sense:

- Docker Engine version 19 or later. View installation instructions.
- Python 3 version 3.6 or later. View installation instructions.

Verify Cloud Manager and Data Sense prerequisites

Review the following prerequisites to make sure that you have a supported configuration before you deploy Cloud Data Sense.

- Ensure that Cloud Manager has permissions to deploy resources and create security groups for the Cloud Data Sense instance.
- Ensure that the Cloud Manager Connector can access the Data Sense instance. The security group for the Connector must allow inbound and outbound traffic over port 443 to and from the Data Sense instance.

  This connection enables deployment of the Data Sense instance and enables you to view compliance and governance information.

  Make sure port 8080 is open so you can see the installation progress in Cloud Manager.

- Ensure that you can keep Cloud Data Sense running. The Cloud Data Sense instance needs to stay on to continuously scan your data.
- Ensure web browser connectivity to Cloud Data Sense. After Cloud Data Sense is enabled, ensure that users access the Cloud Manager interface from a host that has a connection to the Data Sense instance.

  The Data Sense instance uses a private IP address to ensure that the indexed data isn't accessible to others. As a result, the web browser that you use to access Cloud Manager must have a connection to that private IP address. That connection can come from a host that's inside the same network as the Data Sense instance.

SharePoint and OneDrive special requirements

When Cloud Manager and Data Sense are deployed in a site with no internet access, you can scan local files in SharePoint and OneDrive accounts by providing internet access to a few select endpoints.

Endpoints	Purpose
\login.microsoft.com \graph.microsoft.com	Communication with Microsoft servers to log in to the selected online service.
https://cloudmanager.cloud.netapp.com	Communication with the Cloud Manager service, which includes NetApp accounts.

Access to *cloudmanager.cloud.netapp.com* is required only during the initial connections to these external services.

Deploy Data Sense

For typical configurations you'll install the software on a single host system. See those steps here.

For very large configurations where you'll be scanning petabytes of data, you can include multiple hosts to provide additional processing power. See those steps here.

Single-host installation for typical configurations

Follow these steps when installing Data Sense software on a single on-premises host in an offline environment.

What you'll need

- Verify that your Linux system meets the host requirements.
- Verify that you have installed the two prerequisite software packages (Docker Engine and Python 3).
- Make sure you have root privileges on the Linux system.
- Verify that your offline environment meets the required permissions and connectivity.

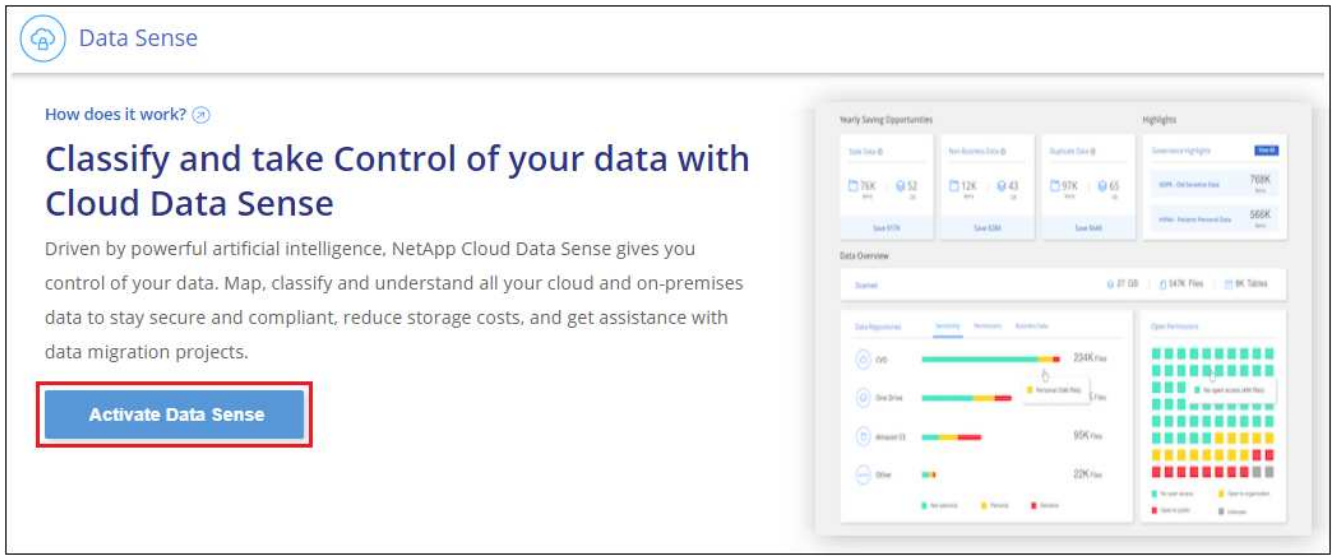
Steps

1. On an internet-configured system, download the Cloud Data Sense software from the NetApp Support Site. The file you should select is named **DataSense-offline-bundle-<version>.tar.gz**.
2. Copy the installer bundle to the Linux host you plan to use in the dark site.
3. Unzip the installer bundle on the host machine, for example:

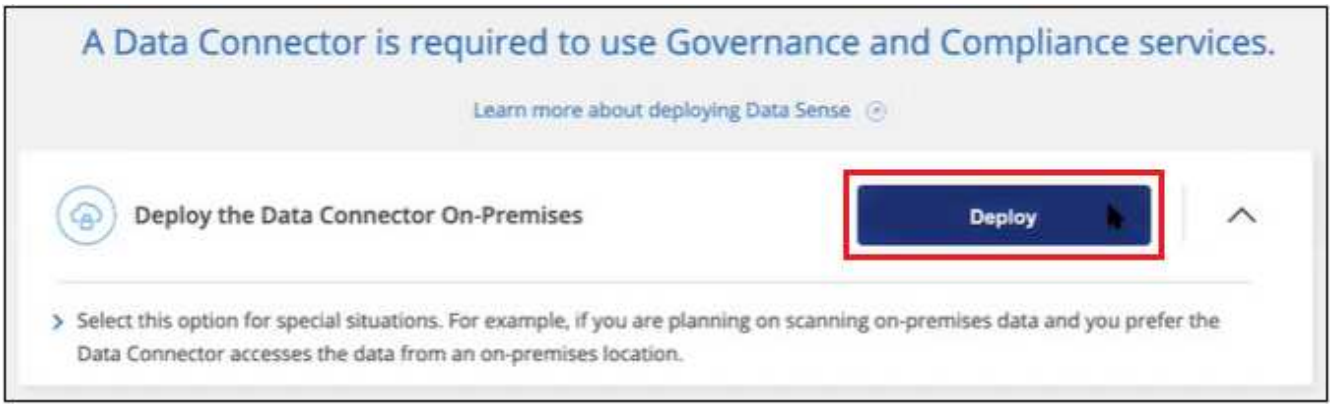
   ```
   tar -xzf DataSense-offline-bundle-v1.13.1.tar.gz
   ```

   This extracts required software and the actual installation file **cc_onprem_installer_<version>.tar.gz**.

4. Launch Cloud Manager and click the **Data Sense** tab.
5. Click **Activate Data Sense**.



6. Click **Deploy** to start the on-prem deployment wizard.



7. In the *Deploy Data Sense On Premises* dialog, copy the provided command and paste it in a text file so you can use it later, and click **Close**. For example:

   ```
   sudo ./install.sh -a 12345 -c 27AG75 -t 2198qq --darksite
   ```

8. Unzip the installation file on the host machine, for example:

   ```
   tar -xzf cc_onprem_installer_1.13.1.tar.gz
   ```

9. When prompted by the installer, you can enter the required values in a series of prompts, or you can provide the required parameters as command line arguments to the installer:

Enter parameters as prompted:	Enter the full command:
1. Paste the information you copied from step 7: `sudo ./install.sh -a <account_id> -c <agent_id> -t <token> --darksite` 2. Enter the IP address or host name of the Data Sense host machine so it can be accessed by the Connector instance. 3. Enter the IP address or host name of the Cloud Manager Connector host machine so it can be accessed by the Data Sense instance.	Alternatively, you can create the whole command in advance, providing the necessary host parameters: `sudo ./install.sh -a <account_id> -c <agent_id> -t <token> --host <ds_host> --manager-host <cm_host> --no-proxy --darksite`

   Variable values:

   - *account_id* = NetApp Account ID
   - *agent_id* = Connector ID
   - *token* = jwt user token
   - *ds_host* = IP address or host name of the Data Sense Linux system.
   - *cm_host* = IP address or host name of the Cloud Manager Connector system.

Result

The Data Sense installer installs packages, registers the installation, and installs Data Sense. Installation can take 10 to 20 minutes.

If there is connectivity over port 8080 between the host machine and the Connector instance, you will see the installation progress in the Data Sense tab in Cloud Manager.

What's Next

From the Configuration page you can select the local on-prem ONTAP clusters and databases that you want to scan.

You can also set up BYOL licensing for Cloud Data Sense from the Digital Wallet page at this time. You will not be charged until the amount of data exceeds 1 TB.

Multi-host installation for large configurations

For very large configurations where you'll be scanning petabytes of data, you can include multiple hosts to provide additional processing power. When using multiple host systems, the primary system is called the *Manager node* and the additional systems that provide extra processing power are called *Scanner nodes*.

Follow these steps when installing Data Sense software on multiple on-premises hosts in an offline environment.

What you'll need

- Verify that all your Linux systems for the Manager and Scanner nodes meet the host requirements.
- Verify that you have installed the two prerequisite software packages (Docker Engine and Python 3).
- Make sure you have root privileges on the Linux systems.
- Verify that your offline environment meets the required permissions and connectivity.
- You must have the IP addresses of the scanner node hosts that you plan to use.
- The following ports and protocols must be enabled on all hosts:

Port	Protocols	Description
2377	TCP	Cluster management communications
7946	TCP, UDP	Inter-node communication
4789	UDP	Overlay network traffic
50	ESP	Encrypted IPsec overlay network (ESP) traffic
111	TCP, UDP	NFS Server for sharing files between the hosts (needed from each scanner node to manager node)
2049	TCP, UDP	NFS Server for sharing files between the hosts (needed from each scanner node to manager node)

Steps

1. Follow steps 1 through 8 from the Single-host installation on the manager node.
2. As shown in step 9, when prompted by the installer, you can enter the required values in a series of prompts, or you can provide the required parameters as command line arguments to the installer.

   In addition to the variables available for a single-host installation, a new option **-n <node_ip>** is used to specify the IP addresses of the scanner nodes. Multiple node IPs are separated by a comma.

   For example, this command adds 3 scanner nodes:
   ```
   sudo ./install.sh -a <account_id> -c <agent_id> -t <token> --host <ds_host>
   ```

```
--manager-host <cm_host> -n <node_ip1>,<node_ip2>,<node_ip3> --no-proxy
--darksite
```

3. Before the manager node installation completes, a dialog displays the installation command needed for the scanner nodes. Copy the command and save it in a text file. For example:

```
sudo ./node_install.sh -m 10.11.12.13 -t ABCDEF-1-3u69m1-1s35212
```

4. On **each** scanner node host:
   a. Copy the Data Sense installer file (**cc_onprem_installer_<version>.tar.gz**) to the host machine.
   b. Unzip the installer file.
   c. Paste and run the command that you copied in step 3.

      When the installation finishes on all scanner nodes and they have been joined to the manager node, the manager node installation finishes as well.

Result

The Cloud Data Sense installer finishes installing packages, and registers the installation. Installation can take 15 to 25 minutes.

What's Next

From the Configuration page you can select the local on-prem ONTAP clusters and local databases that you want to scan.

You can also set up BYOL licensing for Cloud Data Sense from the Digital Wallet page at this time. You will not be charged until the amount of data exceeds 1 TB.

Upgrade Data Sense software

Since Data Sense software is updated with new features on a regular basis, you should get into a routine to check for new versions periodically to make sure you're using the newest software and features. You'll need to upgrade Data Sense software manually because there's no internet connectivity to perform the upgrade automatically.

Before you begin

- Data Sense software can be upgraded one major version at a time. For example, if you have version 1.11.x installed, you can upgrade only to 1.12.x. If you are a few major versions behind, you'll need to upgrade the software multiple times.
- Verify that your on-prem Connector software has been upgraded to the newest available version. See the Connector upgrade steps.

Steps

1. On an internet-configured system, download the Cloud Data Sense software from the NetApp Support Site. The file you should select is named **DataSense-offline-bundle-<version>.tar.gz**.
2. Copy the software bundle to the Linux host where Data Sense is installed in the dark site.
3. Unzip the software bundle on the host machine, for example:

```
tar -xvf DataSense-offline-bundle-v1.13.1.tar.gz
```

This extracts the installation file **cc_onprem_installer_<version>.tar.gz**.

4. Unzip the installation file on the host machine, for example:

```
tar -xzf cc_onprem_installer_1.13.1.tar.gz
```

This extracts the upgrade script **start_darksite_upgrade.sh** and any required third-party software.

5. Run the upgrade script on the host machine, for example:

```
start_darksite_upgrade.sh
```

Result

The Data Sense software is upgraded on your host. The update can take 5 to 10 minutes.

Note that no upgrade is required on scanner nodes if you have deployed Data Sense on multiple hosts systems for scanning very large configurations.

You can verify that the software has been updated by checking the version at the bottom of the Data Sense UI pages.