Deploy Data Classification

NetApp Data Classification

NetApp
February 02, 2026

Table of Contents

Deploy Data Classification

Which NetApp Data Classification deployment should you use?

You can deploy NetApp Data Classification in different ways. Learn which method meets your needs.

Data Classification can be deployed in the following ways:

- Deploy in the cloud using the Console. The Console deploys the Data Classification instance in the same cloud provider network as the Console agent.
- Install on a Linux host with internet access. Install Data Classification on a Linux host in your network, or on a Linux host in the cloud, that has internet access. This type of installation may be a good option if you prefer to scan on-premises ONTAP systems using a Data Classification instance that's also located on premises, though this isn't a requirement.
- Install on a Linux host in an on-premises site without internet access, also known as *private mode.* This type of installation, which uses an installation script, has no connectivity to the Console SaaS layer.

> BlueXP private mode (legacy BlueXP interface) is typically used with on-premises environments that have no internet connection and with secure cloud regions, which includes AWS Secret Cloud, AWS Top Secret Cloud, and Azure IL6. NetApp continues to support these environments with the legacy BlueXP interface. For private mode documentation in the legacy BlueXP interface, see PDF documentation for BlueXP private mode.

Both the installation on a Linux host with internet access and the on-premises installation on a Linux host without internet access use an installation script. The script starts by checking if the system and environment meet the prerequisites. If the prerequisites are met, the installation starts. If you would like to verify the prerequisites independently of running the Data Classification installation, there is a separate software package you can download that only tests for the prerequisites.

Refer to Check that your Linux host is ready to install Data Classification.

Deploy NetApp Data Classification in the cloud using the NetApp Console

You can deploy NetApp Data Classification in the cloud with the NetApp Console. The Console deploys the Data Classification instance in the same cloud provider network as the Console agent.

Note that you can also install Data Classification on a Linux host that has internet access. This type of installation may be a good option if you prefer to scan on-premises ONTAP systems using a Data Classification 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 Console agent

If you don't already have a Console agent, create one. See creating a Console agent in AWS, creating a Console agent in Azure, or creating a Console agent in GCP.

You can also install the Console agent on-premises on a Linux host in your network or on a Linux host in the cloud.

Prerequisites

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

Deploy Data Classification

Launch the installation wizard to deploy the Data Classification instance in the cloud.

Create a Console agent

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

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

- When scanning data in Cloud Volumes ONTAP in AWS or Amazon FSx for ONTAP buckets, you use a Console agent in AWS.
- When scanning data in Cloud Volumes ONTAP in Azure or in Azure NetApp Files, you use a Console agent in Azure.
  - For Azure NetApp Files, it must be deployed in the same region as the volumes you wish to scan.
- When scanning data in Cloud Volumes ONTAP in GCP, you use a Console agent in GCP.

On-prem ONTAP systems, NetApp file shares, and databases can be scanned when using any of these cloud Console agents.

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

There might be situations where you need to use multiple Console agents.

Data Classification does not impose a limit on the amount of data it can scan. Each Console agent supports scanning and displaying 500 TiB of data. To scan more than 500 TiB of data, install another Console agent then deploy another Data Classification instance.
The Console UI displays data from a single connector. For tips on viewing data from multiple Console agents, see Work with multiple Console agents.

Government region support

Data Classification is supported when the Console agent is deployed in a Government region (AWS GovCloud, Azure Gov, or Azure DoD). When deployed in this manner, Data Classification has the following restrictions:

Learn about deploying the Console agent in a Government region.

Prerequisites

Review the following prerequisites to make sure that you have a supported configuration before you deploy Data Classification in the cloud. When you deploy Data Classification in the cloud, it's located in the same subnet as the Console agent.

Enable outbound internet access from Data Classification

Data Classification requires outbound internet access. If your virtual or physical network uses a proxy server for internet access, ensure that the Data Classification instance has outbound internet access to contact the following endpoints. The proxy must be non-transparent. Transparent proxies are not currently supported.

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

Required endpoints for AWS

Endpoints	Purpose
https://api.console.netapp.com	Communication with the Console service, which includes NetApp accounts.
https://netapp-cloud-account.auth0.com https://auth0.com	Communication with the Console website 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.
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 Data Classification to access and download manifests and templates, and to send logs and metrics.

Required endpoints for Azure

Endpoints	Purpose
https://api.console.netapp.com	Communication with the Console service, which includes NetApp accounts.
https://netapp-cloud-account.auth0.com https://auth0.com	Communication with the Console website for centralized user authentication.
https://support.compliance.api.console.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.api.console.netapp.com/	Enables NetApp to stream data from audit records.

Required endpoints for GCP

Endpoints	Purpose
https://api.console.netapp.com	Communication with the Console service, which includes NetApp accounts.

Endpoints	Purpose
https://netapp-cloud-account.auth0.com https://auth0.com	Communication with the Console website for centralized user authentication.
https://support.compliance.api.console.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.api.console.netapp.com/	Enables NetApp to stream data from audit records.

Ensure that Data Classification has the required permissions

Ensure that Data Classification has permissions to deploy resources and create security groups for the Data Classification instance.

- Google Cloud permissions
- AWS permissions
- Azure permissions

Ensure that the Console agent can access Data Classification

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

Additional inbound and outbound security group rules are required for AWS and AWS GovCloud deployments. See Rules for the Console agent in AWS for details.

Additional inbound and outbound security group rules are required for Azure and Azure Government deployments. See Rules for the Console agent in Azure for details.

Ensure you can keep Data Classification running

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

Ensure web browser connectivity to Data Classification

After Data Classification is enabled, ensure that users access the Console interface from a host that has a connection to the Data Classification instance.

The Data Classification 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 the Console 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 Classification instance.

Check your vCPU limits

Ensure that your cloud provider's vCPU limit allows for the deployment of an instance with the necessary number of cores. You'll need to verify the vCPU limit for the relevant instance family in the region where the

Console 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

Deploy Data Classification in the cloud

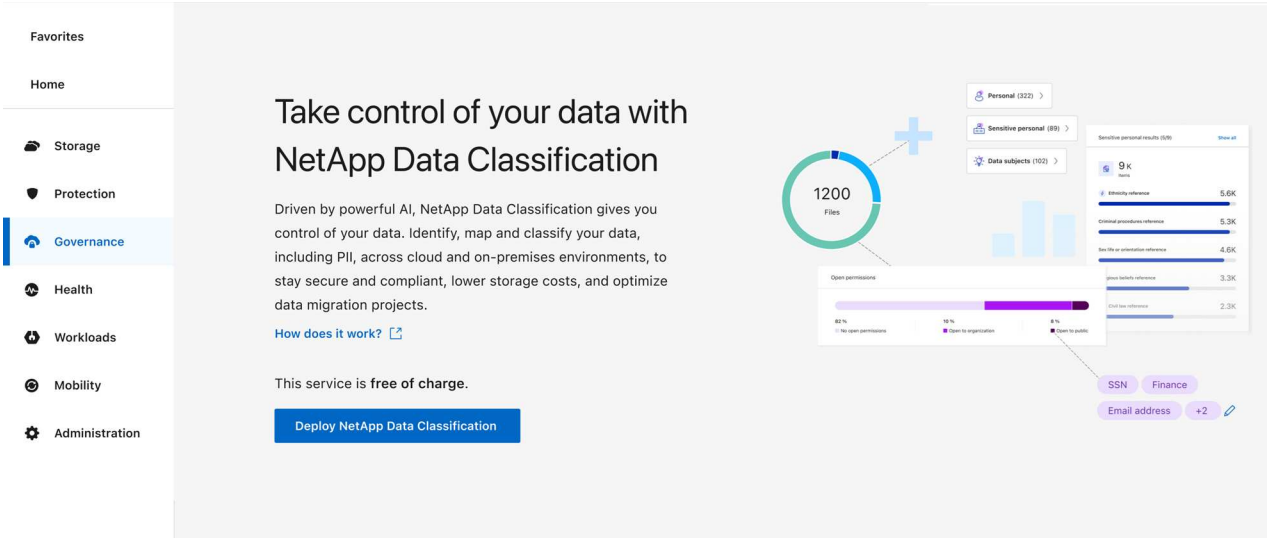
Follow these steps to deploy an instance of Data Classification in the cloud. The Console agent will deploy the instance in the cloud, and then install Data Classification software on that instance.

In regions where the default instance type isn't available, Data Classification runs on an alternate instance type.

Deploy in AWS

Steps

1. From the main page of Data Classification, select **Deploy Classification On-Premises or Cloud**.

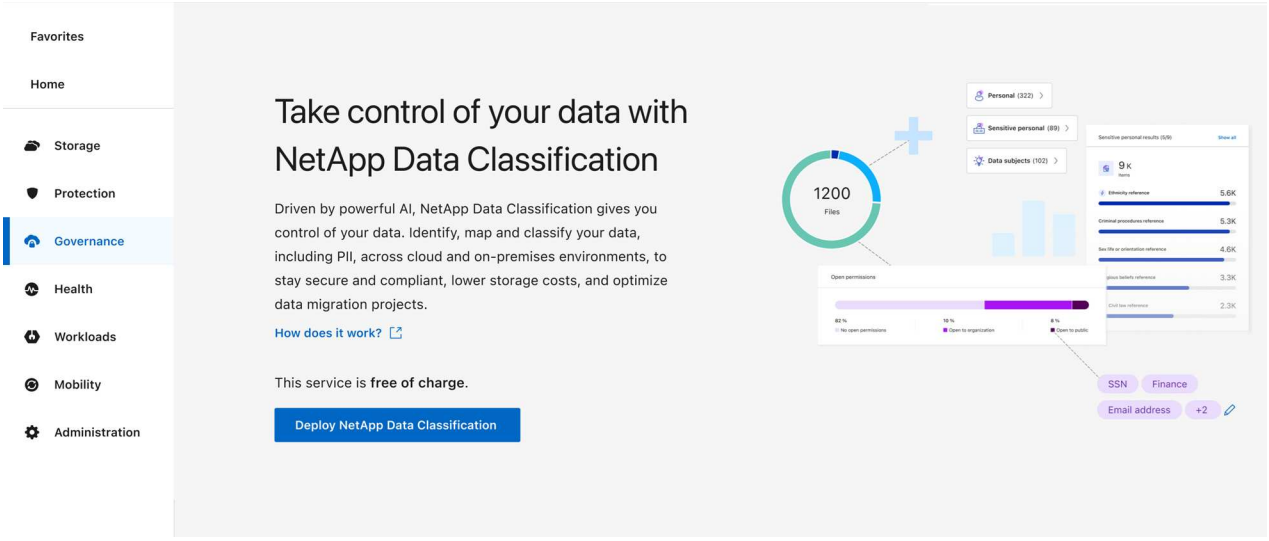


2. From the *Installation* page, select **Deploy > Deploy** to use the "Large" instance size and start the cloud deployment wizard.
3. The wizard displays progress as it goes through the deployment steps. When inputs are required or if it encounters issues, you are prompted.
4. When the instance is deployed and Data Classification is installed, select **Continue to configuration** to go to the *Configuration* page.

Deploy in Azure

Steps

1. From the main page of Data Classification, select **Deploy Classification On-Premises or Cloud**.



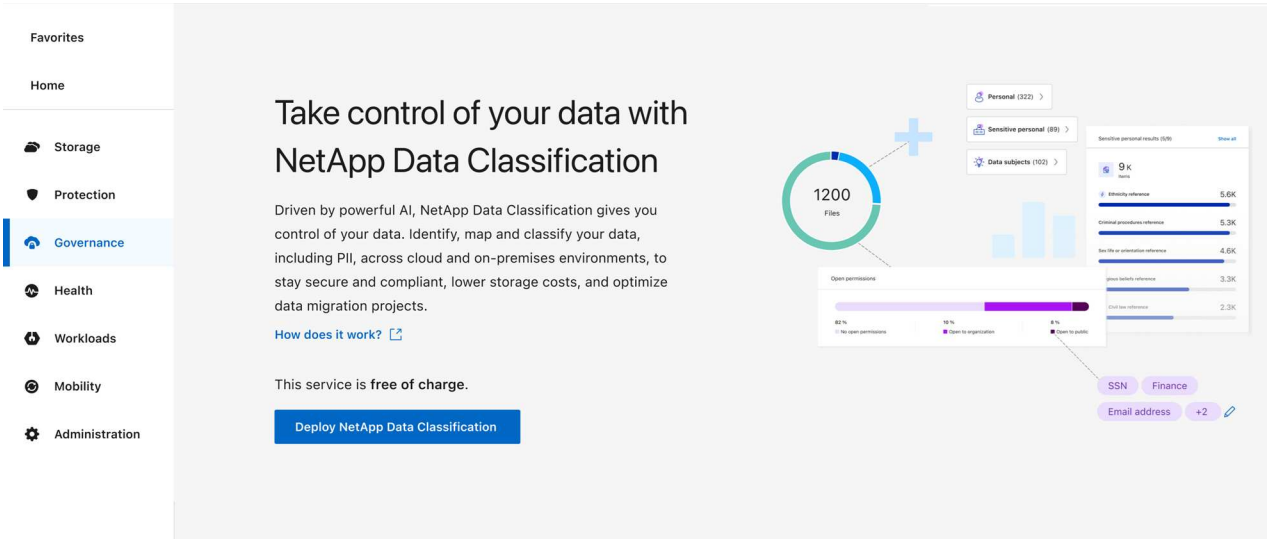
2. Select **Deploy** to start the cloud deployment wizard.
3. The wizard displays progress as it goes through the deployment steps. It will stop and prompt for input if it runs into any issues.

4. When the instance is deployed and Data Classification is installed, select **Continue to configuration** to go to the *Configuration* page.

Deploy in Google Cloud

Steps

1. From the main page of Data Classification, select **Governance > Classification**.
2. Select **Deploy Classification On-Premises or Cloud**.



3. Select **Deploy** to start the cloud deployment wizard.
4. The wizard displays progress as it goes through the deployment steps. It will stop and prompt for input if it runs into any issues.
5. When the instance is deployed and Data Classification is installed, select **Continue to configuration** to go to the *Configuration* page.

Result

The Console deploys the Data Classification instance in your cloud provider.

Upgrades to the Console agent and Data Classification software is automated as long as the instances have internet connectivity.

What's Next

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

Install NetApp Data Classification on a host that has internet access

To deploy NetApp Data Classification on a Linux host in your network or on a Linux host in the cloud that has internet access, you need deploy the Linux host manually in your network or in the cloud.

The on-premises installation is a good option if you prefer to scan on-premises ONTAP systems using a Data Classification instance that's also located on premises. This is not a requirement. The software functions the same regardless of which installation method you choose.

The Data Classification installation script starts by checking if the system and environment meet the required prerequisites. If the prerequisites are all met, then the installation starts. If you would like to verify the prerequisites independently of running the Data Classification installation, there is a separate software package you can download that only tests for the prerequisites. See how to check if your Linux host is ready to install Data Classification.

The typical installation on a Linux host *in your premises* has the following components and connections.

The typical installation on a Linux host *in the cloud* has the following components and connections.

Quick start

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

1 Create a Console agent

If you don't already have a Console agent , deploy the Console agent on-premises on a Linux host in your network, or on a Linux host in the cloud.

You can also create a Console agent with your cloud provider. See creating a Console agent in AWS, creating a Console agent in Azure, or creating a Console agent in GCP.

2 Review prerequisites

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

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

3 Download and deploy Data Classification

Download the Cloud Data Classification 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 Classification instance.

Create a Console agent

A Console agent is required before you can install and use Data Classification. In most cases you'll probably have a Console agent set up before you attempt to activate Data Classification because most Console features require a Console agent, but there are cases where you'll you need to set one up now.

To create one in your cloud provider environment, see creating a Console agent in AWS, creating a Console agent in Azure, or creating a Console agent in GCP.

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

- When scanning data in Cloud Volumes ONTAP in AWS or Amazon FSx for ONTAP, you use a Console

agent in AWS.

- When scanning data in Cloud Volumes ONTAP in Azure or in Azure NetApp Files, you use a Console agent in Azure.

  For Azure NetApp Files, it must be deployed in the same region as the volumes you wish to scan.

- When scanning data in Cloud Volumes ONTAP in GCP, you use a Console agent in GCP.

On-prem ONTAP systems, NetApp file shares and database accounts can be scanned using any of these cloud Console agents.

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

You'll need the IP address or host name of the Console agent system when installing Data Classification. You'll have this information if you installed the Console agent in your premises. If the Console agent is deployed in the cloud, you can find this information from the Console: select the Help icon then **Support** then **Console agent**.

Prepare the Linux host system

Data Classification software must run on a host that meets specific operating system requirements, RAM requirements, software requirements, and so on. The Linux host can be in your network, or in the cloud.

Ensure that you can keep Data Classification running. The Data Classification machine needs to stay on to continuously scan your data.

- Data Classification must be on a dedicated host. The host can't be shared with other applications or third-party software such as antivirus.
- Choose the size that aligns with the data set you plan to scan with Data Classification.

System size	CPU	RAM (swap memory must be disabled)	Disk
Extra Large	32 CPUs	128 GB RAM	• 1 TiB SSD on /, or 100 GiB available on /opt • 895 GiB available on /var/lib/docker • 5 GiB on /tmp • **For Podman, 30 GB on /var/tmp**

System size	CPU	RAM (swap memory must be disabled)	Disk
Large	16 CPUs	64 GB RAM	• 500 GiB SSD on /, or 100 GiB available on /opt • 400 GiB available on /var/lib/docker or for Podman /var/lib/containers • 5 GiB on /tmp • **For Podman, 30 GB on /var/tmp**

- When deploying a compute instance in the cloud for your Data Classification installation, it's recommended you use a system that meets the "Large" system requirements above:
  - **Amazon Elastic Compute Cloud (Amazon EC2) instance type**: "m6i.4xlarge". See additional AWS instance types.
  - **Azure VM size**: "Standard_D16s_v3". See additional Azure instance types.
  - **GCP machine type**: "n2-standard-16". See additional GCP instance types.
- **UNIX folder permissions**: The following minimum UNIX permissions are required:

Folder	Minimum permissions
/tmp	`rwxrwxrwt`
/opt	`rwxr-xr-x`
/var/lib/docker	`rwx------`
/usr/lib/systemd/system	`rwxr-xr-x`

- **Operating system**:
  - The following operating systems require using the Docker container engine:
    - Red Hat Enterprise Linux version 7.8 and 7.9
    - Ubuntu 22.04 (requires Data Classification version 1.23 or greater)
    - Ubuntu 24.04 (requires Data Classification version 1.23 or greater)
  - The following operating systems require using the Podman container engine, and they require Data Classification version 1.30 or greater:
    - Red Hat Enterprise Linux version 8.8, 8.10, 9.0, 9.1, 9.2, 9.3, 9.4, 9.5, and 9.6.
  - Advanced Vector Extensions (AVX2) must be enabled on the host system.
- **Red Hat Subscription Management**: The host 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.
- **Additional software**: You must install the following software on the host before you install Data Classification:
  - Depending on the OS you are using, you need to install one of the container engines:
    - Docker Engine version 19.3.1 or greater. View installation instructions.

    - Podman version 4 or greater. To install Podman, enter (`sudo yum install podman netavark -y`).

- Python version 3.6 or greater. View installation instructions.
    - **NTP considerations**: NetApp recommends configuring the Data Classification system to use a Network Time Protocol (NTP) service. The time must be synchronized between the Data Classification system and the Console agent system.

- **Firewalld considerations**: If you are planning to use `firewalld`, we recommend that you enable it before installing Data Classification. Run the following commands to configure `firewalld` so that it is compatible with Data Classification:

```
firewall-cmd --permanent --add-service=http
firewall-cmd --permanent --add-service=https
firewall-cmd --permanent --add-port=80/tcp
firewall-cmd --permanent --add-port=8080/tcp
firewall-cmd --permanent --add-port=443/tcp
firewall-cmd --reload
```

If you're planning to use additional Data Classification hosts as scanner nodes, add these rules to your primary system at this time:

```
firewall-cmd --permanent --add-port=2377/tcp
firewall-cmd --permanent --add-port=7946/udp
firewall-cmd --permanent --add-port=7946/tcp
firewall-cmd --permanent --add-port=4789/udp
```

Note that you must restart Docker or Podman whenever you enable or update `firewalld` settings.

The IP address of the Data Classification host system can't be changed after installation.

Enable outbound internet access from Data Classification

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

Endpoints	Purpose
https://api.console.netapp.com	Communication with the Console, which includes NetApp accounts.
https://netapp-cloud-account.auth0.com https://auth0.com	Communication with the Console website for centralized user authentication.

Endpoints	Purpose
https://support.compliance.api.bluexp.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.api.bluexp.netapp.com/	Enables NetApp to stream data from audit records.
https://github.com/docker https://download.docker.com	Provides prerequisite packages for docker installation.
http://packages.ubuntu.com/ http://archive.ubuntu.com	Provides prerequisite packages for Ubuntu installation.

Verify that all required ports are enabled

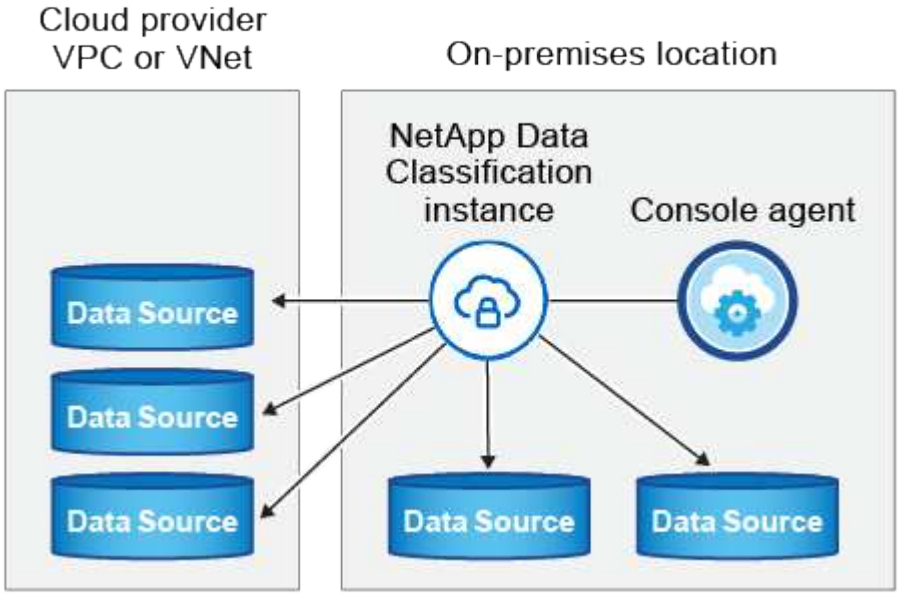
You must ensure that all required ports are open for communication between the Console agent, Data Classification, Active Directory, and your data sources.

Connection Type	Ports	Description
Console agent <> Data Classification	8080 (TCP), 443 (TCP), and 80. 9000	The firewall or routing rules for the Console agent must allow inbound and outbound traffic over port 443 to and from the Data Classification instance. Make sure port 8080 is open so you can see the installation progress in the Console. If a firewall is used on the Linux host, port 9000 is required for internal processes within an Ubuntu server.

Connection Type	Ports	Description
Console agent <> ONTAP cluster (NAS)	443 (TCP)	The Console discovers ONTAP clusters using HTTPS. If you use custom firewall policies, they must meet the following requirements: • The Console agent host must allow outbound HTTPS access through port 443. If the Console agent is in the cloud, all outbound communication is allowed by the predefined firewall or routing rules. • The ONTAP cluster must allow inbound HTTPS access through port 443. The default "mgmt" firewall policy allows inbound HTTPS access from all IP addresses. If you modified this default policy, or if you created your own firewall policy, you must associate the HTTPS protocol with that policy and enable access from the Console agent host.
Data Classification <> ONTAP cluster	• For NFS - 111 (TCP\UDP) and 2049 (TCP\UDP) • For CIFS - 139 (TCP\UDP) and 445 (TCP\UDP)	Data Classification needs a network connection to each Cloud Volumes ONTAP subnet or on-prem ONTAP system. Firewalls or routing rules for Cloud Volumes ONTAP must allow inbound connections from the Data Classification instance. Make sure these ports are open to the Data Classification instance: • For NFS - 111 and 2049 • For CIFS - 139 and 445 NFS volume export policies must allow access from the Data Classification instance.
Data Classification <> Active Directory	389 (TCP & UDP), 636 (TCP), 3268 (TCP), and 3269 (TCP)	You must have an Active Directory already set up for the users in your company. Additionally, Data Classification needs Active Directory credentials to scan CIFS volumes. You must have the information for the Active Directory: • DNS Server IP Address, or multiple IP Addresses • User Name and Password for the server • Domain Name (Active Directory Name) • Whether you are using secure LDAP (LDAPS) or not • LDAP Server Port (typically 389 for LDAP, and 636 for secure LDAP)

Install Data Classification on the Linux host

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



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

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

> Data Classification 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 Console agent and instance of Data Classification in the cloud and switch between Connectors for your different data sources.

Single-host installation for typical configurations

Review the requirements and follow these steps when installing Data Classification software on a single on-premises host.

Watch this video to see how to install Data Classification.

Note that all installation activities are logged when installing Data Classification. If you run into any issues during installation, you can view the contents of the installation audit log. It is written to `/opt/netapp/install_logs/`.

Before you begin

- Verify that your Linux system meets the host requirements.
- Verify that the system has the two prerequisite software packages installed (Docker Engine or Podman, and Python 3).
- Make sure you have root privileges on the Linux system.
- If you're using a proxy for access to the internet:
  - You'll need the proxy server information (IP address or host name, connection port, connection scheme: https or http, user name and password).

- If the proxy is performing TLS interception, you'll need to know the path on the Data Classification Linux system where the TLS CA certificates are stored.
- The proxy must be non-transparent. Data Classification does not currently support transparent proxies.
- The user must be a local user. Domain users are not supported.

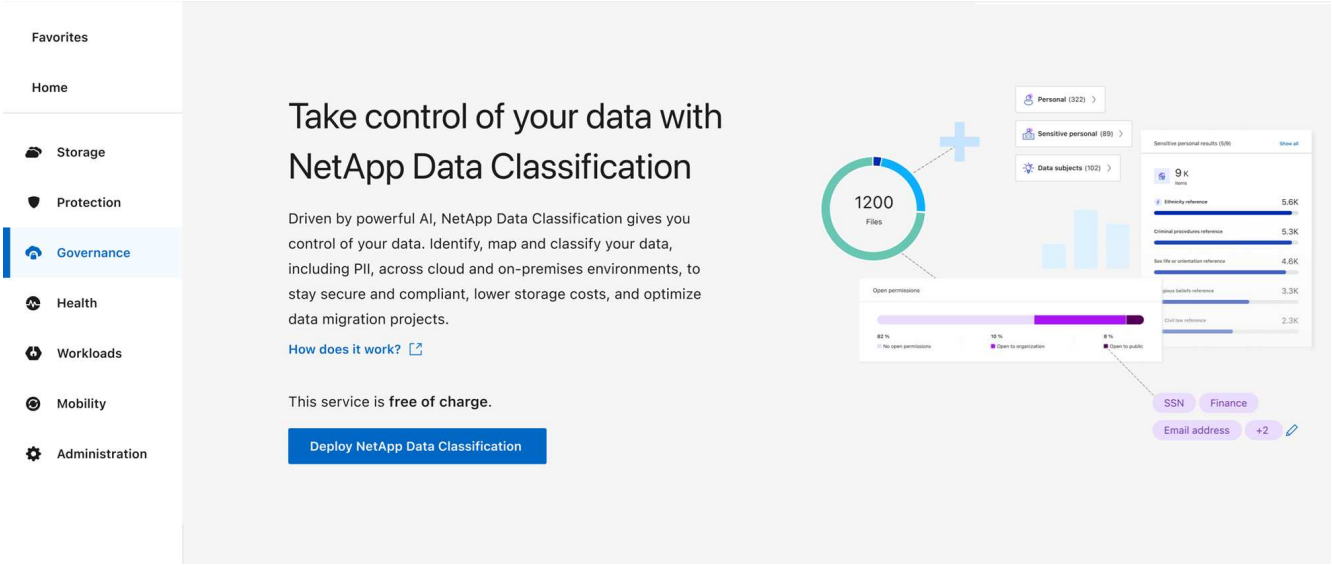
- Verify that your offline environment meets the required permissions and connectivity.

Steps

1. Download the Data Classification software from the NetApp Support Site. The file you should select is named **DATASENSE-INSTALLER-<version>.tar.gz**.
2. Copy the installer file to the Linux host you plan to use (using `scp` or some other method).
3. Unzip the installer file on the host machine, for example:

```
tar -xzf DATASENSE-INSTALLER-V1.25.0.tar.gz
```

4. In the Console, select **Governance > Classification**.
5. Select **Deploy Classification On-Premises or Cloud**.



6. Depending on whether you are installing Data Classification on an instance you prepared in the cloud or on an instance you prepared in your premises, select the appropriate **Deploy** option to start the Data Classification installation.
7. The *Deploy Data Classification On Premises* dialog is displayed. Copy the provided command (for example: `sudo ./install.sh -a 12345 -c 27AG75 -t 2198qq`) and paste it in a text file so you can use it later. Then select **Close** to dismiss the dialog.
8. On the host machine, enter the command you copied and then follow a series of prompts, or you can provide the full command including all required parameters as command line arguments.

   Note that the installer performs a pre-check to make sure your system and networking requirements are in place for a successful installation. Watch this video to understand the pre-check messages and implications.

Enter parameters as prompted:	Enter the full command:
1. Paste the command you copied from step 7: `sudo ./install.sh -a <account_id> -c <client_id> -t <user_token>` If you are installing on a cloud instance (not on your premises), add `--manual-cloud-install <cloud_provider>`. 2. Enter the IP address or host name of the Data Classification host machine so it can be accessed by the Console agent system. 3. Enter the IP address or host name of the Console agent host machine so it can be accessed by the Data Classification system. 4. Enter proxy details as prompted. If your Console agent already uses a proxy, there is no need to enter this information again here since Data Classification will automatically use the proxy used by the Console agent.	Alternatively, you can create the whole command in advance, providing the necessary host and proxy parameters: `sudo ./install.sh -a <account_id> -c <client_id> -t <user_token> --host <ds_host> --manager-host <cm_host> --manual-cloud-install <cloud_provider> --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
- *client_id* = Console agent Client ID (add the suffix "clients" to the client ID if it not already there)
- *user_token* = JWT user access token
- *ds_host* = IP address or host name of the Data Classification Linux system.
- *cm_host* = IP address or host name of the Console agent system.
- *cloud_provider* = When installing on a cloud instance, enter "AWS", "Azure", or "Gcp" depending on cloud provider.
- *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. The user must be a local user - domain users are not supported.
- *proxy_password* = Password for the user name that you specified.
- *ca_cert_dir* = Path on the Data Classification Linux system containing additional TLS CA certificate bundles. Only required if the proxy is performing TLS interception.

Result

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

If there is connectivity over port 8080 between the host machine and the Console agent instance, you'll see the installation progress in the Data Classification tab in the Console.

What's Next

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

Install NetApp Data Classification on a Linux host with no internet access

Installing NetApp Data Classification on a Linux host in an on-premises site that doesn't have internet access is known as *private mode*. This type of installation, which uses an installation script, has no connectivity to the NetApp Console SaaS layer.

BlueXP private mode (legacy BlueXP interface) is typically used with on-premises environments that have no internet connection and with secure cloud regions, which includes AWS Secret Cloud, AWS Top Secret Cloud, and Azure IL6. NetApp continues to support these environments with the legacy BlueXP interface. For private mode documentation in the legacy BlueXP interface, see PDF documentation for BlueXP private mode.

Check that your Linux host is ready to install NetApp Data Classification

Before installing NetApp Data Classification manually on a Linux host, optionally run a script on the host to verify that all the prerequisites are in place for installing Data Classification. You can run this script on a Linux host in your network, or on a Linux host in the cloud. The host can be connected to the internet, or the host can reside in a site that doesn't have internet access (a *dark site*).

The Data Classification installation script encompasses a test script to ensure your environment meets the requirements. You can run this script separately to verify the Linux host's readiness before run the installation script.

Getting Started

You'll perform the following tasks.

- Optionally, install a Console agent if you don't already have one installed. You can run the test script without having a Console agent installed, but the script checks for connectivity between the Console agent and the Data Classification host machine - so it is recommended that you have a Console agent.
- Prepare the host machine and verify that it meets all the requirements.
- Enable outbound internet access from the Data Classification host machine.
- Verify that all required ports are enabled on all systems.
- Download and run the Prerequisite test script.

Create a Console agent

A Console agent is required before you can install and use Data Classification. You can, however, run the Prerequisites script without a Console agent.

You can install the Console agent on-premises on a Linux host in your network or on a Linux host in the cloud.

You can also install Data Classification on-premises if the Console agent is installed on-premises.

To create a Console agent in your cloud provider environment, see:

- creating a Console agent in AWS
- creating a Console agent in Azure
- creating a Console agent in GCP

You need the IP address or host name of the Console agent system when running the Prerequisites script. You have this information if you installed the Console agent in your premises. If the Console agent is deployed in the cloud, you can find this information from the Console: select the Help icon then **Support**; in the Agent and Audit section, select **Go to the agent**.

Verify host requirements

Data Classification software must run on a host that meets specific operating system requirements, RAM requirements, and software requirements.

- Data Classification must be on a dedicated host. The host can't be shared with other applications or third-party software such as antivirus.
- Choose the size that aligns with the data set you plan to scan with Data Classification.

System size	CPU	RAM (swap memory must be disabled)	Disk
Extra Large	32 CPUs	128 GB RAM	• 1 TiB SSD on /, or 100 GiB available on /opt • 895 GiB available on /var/lib/docker • 5 GiB on /tmp • **For Podman, 30 GB on /var/tmp**
Large	16 CPUs	64 GB RAM	• 500 GiB SSD on /, or 100 GiB available on /opt • 400 GiB available on /var/lib/docker or for Podman /var/lib/containers • 5 GiB on /tmp • **For Podman, 30 GB on /var/tmp**

- When deploying a compute instance in the cloud for your Data Classification installation, it's recommended you use a system that meets the "Large" system requirements above:
  - **Amazon Elastic Compute Cloud (Amazon EC2) instance type**: "m6i.4xlarge". See additional AWS instance types.
  - **Azure VM size**: "Standard_D16s_v3". See additional Azure instance types.
  - **GCP machine type**: "n2-standard-16". See additional GCP instance types.

- **UNIX folder permissions**: The following minimum UNIX permissions are required:

Folder	Minimum permissions
/tmp	`rwxrwxrwt`
/opt	`rwxr-xr-x`
/var/lib/docker	`rwx------`
/usr/lib/systemd/system	`rwxr-xr-x`

- **Operating system**:
  - The following operating systems require using the Docker container engine:
    - Red Hat Enterprise Linux version 7.8 and 7.9
    - Ubuntu 22.04 (requires Data Classification version 1.23 or greater)
    - Ubuntu 24.04 (requires Data Classification version 1.23 or greater)
  - The following operating systems require using the Podman container engine, and they require Data Classification version 1.30 or greater:
    - Red Hat Enterprise Linux version 8.8, 8.10, 9.0, 9.1, 9.2, 9.3, 9.4, 9.5, and 9.6.
  - Advanced Vector Extensions (AVX2) must be enabled on the host system.
- **Red Hat Subscription Management**: The host 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.
- **Additional software**: You must install the following software on the host before you install Data Classification:
  - Depending on the OS you are using, you need to install one of the container engines:
    - Docker Engine version 19.3.1 or greater. View installation instructions.
    - Podman version 4 or greater. To install Podman, enter (`sudo yum install podman netavark -y`).
- Python version 3.6 or greater. View installation instructions.
  - **NTP considerations**: NetApp recommends configuring the Data Classification system to use a Network Time Protocol (NTP) service. The time must be synchronized between the Data Classification system and the Console agent system.
- **Firewalld considerations**: If you are planning to use `firewalld`, we recommend that you enable it before installing Data Classification. Run the following commands to configure `firewalld` so that it is compatible with Data Classification:

```
firewall-cmd --permanent --add-service=http
firewall-cmd --permanent --add-service=https
firewall-cmd --permanent --add-port=80/tcp
firewall-cmd --permanent --add-port=8080/tcp
firewall-cmd --permanent --add-port=443/tcp
firewall-cmd --reload
```

If you're planning to use additional Data Classification hosts as scanner nodes (in a distributed model), add these rules to your primary system at this time:

```
firewall-cmd --permanent --add-port=2377/tcp
firewall-cmd --permanent --add-port=7946/udp
firewall-cmd --permanent --add-port=7946/tcp
firewall-cmd --permanent --add-port=4789/udp
```

Note that you must restart Docker or Podman whenever you enable or update `firewalld` settings.

Enable outbound internet access from Data Classification

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

This section is not required for host systems installed in sites without internet connectivity.

Endpoints	Purpose
https://api.console.netapp.com	Communication with the Console service, which includes NetApp accounts.
https://netapp-cloud-account.auth0.com https://auth0.com	Communication with the Console website for centralized user authentication.
https://support.compliance.api.console.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.api.console.netapp.com/	Enables NetApp to stream data from audit records.
https://github.com/docker https://download.docker.com	Provides prerequisite packages for docker installation.
http://packages.ubuntu.com/ http://archive.ubuntu.com	Provides prerequisite packages for Ubuntu installation.

Verify that all required ports are enabled

You must ensure that all required ports are open for communication between the Console agent, Data Classification, Active Directory, and your data sources.

Connection Type	Ports	Description
Console agent <> Data Classification	8080 (TCP), 443 (TCP), and 80. 9000	The firewall or routing rules for the Console agent must allow inbound and outbound traffic over port 443 to and from the Data Classification instance. Make sure port 8080 is open so you can see the installation progress in the Console. If a firewall is used on the Linux host, port 9000 is required for internal processes within an Ubuntu server.
Console agent <> ONTAP cluster (NAS)	443 (TCP)	The Console discovers ONTAP clusters using HTTPS. If you use custom firewall policies, the Console agent host must allow outbound HTTPS access through port 443. If the Console agent is in the cloud, all outbound communication is allowed by the predefined firewall or routing rules.

Run the Data Classification prerequisites script

Follow these steps to run the Data Classification prerequisites script.

Watch this video to see how to run the Prerequisites script and interpret the results.

Before you begin

- Verify that your Linux system meets the host requirements.
- Verify that the system has the two prerequisite software packages installed (Docker Engine or Podman, and Python 3).
- Make sure you have root privileges on the Linux system.

Steps

1. Download the Data Classification Prerequisites script from the NetApp Support Site. The file you should select is named **standalone-pre-requisite-tester-<version>**.
2. Copy the file to the Linux host you plan to use (using `scp` or some other method).
3. Assign permissions to run the script.

   ```
   chmod +x standalone-pre-requisite-tester-v1.25.0
   ```

4. Run the script using the following command.

   ```
   ./standalone-pre-requisite-tester-v1.25.0 <--darksite>
   ```

   Add the option "--darksite" only if you are running the script on a host that doesn't have internet access. Certain prerequisite tests are skipped when the host is not connected to the internet.

5. The script prompts you for the IP address of the Data Classification host machine.

- Enter the IP address or host name.

6. The script prompts whether you have an installed Console agent.
   - Enter **N** if you do not have an installed Console agent.
   - Enter **Y** if you do have an installed Console agent. And then enter the IP address or host name of the Console agent so the test script can test this connectivity.
7. The script runs a variety of tests on the system and it displays results as it progresses. When it finishes it writes a log of the session to a file named `prerequisites-test-<timestamp>.log` in the directory `/opt/netapp/install_logs`.

Result

If all the prerequisites tests ran successfully, you can install Data Classification on the host when you are ready.

If any issues were discovered, they are categorized as "Recommended" or "Required" to be fixed. Recommended issues are typically items that would make the Data Classification scanning and categorizing tasks run slower. These items do not need to be corrected - but you may want to address them.

If you have any "Required" issues, you should fix the issues and run the Prerequisites test script again.

Copyright information

Trademark information