Implementar la clasificación de datos

NetApp Data Classification

NetApp
January 26, 2026

Tabla de contenidos

Implementar la clasificación de datos

¿Qué implementación de NetApp Data Classification debería utilizar?

Puede implementar NetApp Data Classification de diferentes maneras. Conozca qué método se adapta a sus necesidades.

La clasificación de datos se puede implementar de las siguientes maneras:

- "Implementar en la nube usando la consola" . La consola implementa la instancia de clasificación de datos en la misma red del proveedor de nube que el agente de la consola.
- "Instalar en un host Linux con acceso a Internet" . Instale Data Classification en un host Linux en su red, o en un host Linux en la nube, que tenga acceso a Internet. Este tipo de instalación puede ser una buena opción si prefiere escanear sistemas ONTAP locales utilizando una instancia de clasificación de datos que también se encuentre en las instalaciones, aunque esto no es un requisito.
- "Instalar en un host Linux en un sitio local sin acceso a Internet", también conocido como *modo privado*. Este tipo de instalación, que utiliza un script de instalación, no tiene conectividad con la capa SaaS de la consola.

> El modo privado de BlueXP (interfaz BlueXP heredada) generalmente se usa con entornos locales que no tienen conexión a Internet y con regiones de nube seguras, que incluyen AWS Secret Cloud, AWS Top Secret Cloud y Azure IL6. NetApp continúa brindando soporte a estos entornos con la interfaz BlueXP heredada. Para obtener documentación del modo privado en la interfaz heredada de BlueXP , consulte"Documentación en PDF para el modo privado de BlueXP" .

Tanto la instalación en un host Linux con acceso a Internet como la instalación local en un host Linux sin acceso a Internet utilizan un script de instalación. El script comienza verificando si el sistema y el entorno cumplen los requisitos previos. Si se cumplen los requisitos previos, se inicia la instalación. Si desea verificar los requisitos previos independientemente de ejecutar la instalación de Clasificación de datos, hay un paquete de software separado que puede descargar que solo prueba los requisitos previos.

Consulte "Compruebe que su host Linux esté listo para instalar Data Classification" .

Implemente la NetApp Data Classification en la nube mediante la NetApp Console

Puede implementar NetApp Data Classification en la nube con la NetApp Console. La consola implementa la instancia de clasificación de datos en la misma red del proveedor de nube que el agente de la consola.

Tenga en cuenta que también puede"Instalar Data Classification en un host Linux que tenga acceso a Internet" . Este tipo de instalación puede ser una buena opción si prefiere escanear sistemas ONTAP locales utilizando una instancia de clasificación de datos que también se encuentra en las instalaciones, pero esto no es un requisito. El software funciona exactamente de la misma manera independientemente del método de instalación que elija.

Inicio rápido

Comience rápidamente siguiendo estos pasos o desplácese hacia abajo hasta las secciones restantes para obtener detalles completos.

Crear un agente de consola

Si aún no tiene un agente de consola, cree uno. Ver"Creación de un agente de consola en AWS" ,"Creación de un agente de consola en Azure" , o"Creación de un agente de consola en GCP" .

También puedes"Instalar el agente de consola local" en un host Linux de su red o en un host Linux en la nube.

Prerrequisitos

Asegúrate de que tu entorno puede cumplir los requisitos previos. Esto incluye acceso saliente a internet para la instancia, conectividad entre el agente de la Console y Data Classification por el puerto 443 y más. Ver la lista completa.

Implementar la clasificación de datos

Inicie el asistente de instalación para implementar la instancia de Clasificación de datos en la nube.

Crear un agente de consola

Si aún no tiene un agente de consola, cree uno en su proveedor de nube. Ver "Creación de un agente de consola en AWS" o "Creación de un agente de consola en Azure" , o "Creación de un agente de consola en GCP" . En la mayoría de los casos, probablemente ya tendrá configurado un agente de consola antes de intentar activar la clasificación de datos, ya que la mayoría "Las funciones de la consola requieren un agente de consola" , pero hay casos en los que necesitarás configurarlo ahora.

Hay algunos escenarios en los que es necesario utilizar un agente de consola implementado en un proveedor de nube específico:

- Al escanear datos en Cloud Volumes ONTAP en AWS o Amazon FSx para buckets de ONTAP , se utiliza un agente de consola en AWS.
- Al escanear datos en Cloud Volumes ONTAP en Azure o en Azure NetApp Files, se utiliza un agente de consola en Azure.
  - Para Azure NetApp Files, debe implementarse en la misma región que los volúmenes que desea escanear.
- Al escanear datos en Cloud Volumes ONTAP en GCP, se utiliza un agente de consola en GCP.

Los sistemas ONTAP locales, los recursos compartidos de archivos de NetApp y las bases de datos se pueden escanear al usar cualquiera de estos agentes de consola en la nube.

Ten en cuenta que también puedes "Instalar el agente de consola local" en un host Linux de su red o en la nube. Algunos usuarios que planean instalar Data Classification localmente también pueden optar por instalar el agente de consola localmente.

Puede haber situaciones en las que necesites usar "varios agentes de consola" .

La clasificación de datos no impone un límite en la cantidad de datos que puede escanear. Cada agente de consola admite el escaneo y la visualización de 500 TiB de datos. Para escanear más de 500 TiB de datos,"instalar otro agente de consola" entonces"Implementar otra instancia de clasificación de datos" . + La interfaz de usuario de la consola muestra datos de un solo conector. Para obtener sugerencias sobre cómo ver datos de varios agentes de la consola, consulte"Trabajar con múltiples agentes de consola" .

Apoyo de la región gubernamental

La clasificación de datos se admite cuando el agente de consola se implementa en una región gubernamental (AWS GovCloud, Azure Gov o Azure DoD). Cuando se implementa de esta manera, la clasificación de datos tiene las siguientes restricciones:

"Obtenga información sobre cómo implementar el agente de consola en una región gubernamental.".

Prerrequisitos

Revise los siguientes requisitos previos para asegurarse de tener una configuración compatible antes de implementar la clasificación de datos en la nube. Cuando implementa la clasificación de datos en la nube, se ubica en la misma subred que el agente de consola.

Habilitar el acceso a Internet saliente desde la Clasificación de datos

La clasificación de datos requiere acceso a Internet saliente. Si su red virtual o física utiliza un servidor proxy para el acceso a Internet, asegúrese de que la instancia de clasificación de datos tenga acceso a Internet saliente para contactar los siguientes puntos finales. El proxy no debe ser transparente. Los servidores proxy transparentes no son compatibles actualmente.

Revise la tabla correspondiente a continuación según si está implementando la clasificación de datos en AWS, Azure o GCP.

Puntos finales necesarios para AWS

Puntos finales	Objetivo
\ https://api.console.netapp.com	Comunicación con el servicio de consola, que incluye cuentas de NetApp .
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	Comunicación con el sitio web de la consola para la autenticación centralizada de usuarios.
\ 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/	Proporciona acceso a imágenes de software, manifiestos y plantillas.
\ https://kinesis.us-east-1.amazonaws.com	Permite a NetApp transmitir datos desde registros de auditoría.
\ 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	Permite que la clasificación de datos acceda y descargue manifiestos y plantillas, y envíe registros y métricas.

Puntos de conexión necesarios para Azure

Puntos finales	Objetivo
\ https://api.console.netapp.com	Comunicación con el servicio de consola, que incluye cuentas de NetApp .
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	Comunicación con el sitio web de la consola para la autenticación centralizada de usuarios.
\ 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/	Proporciona acceso a imágenes de software, manifiestos, plantillas y permite enviar registros y métricas.
\ https://support.compliance.api.console.netapp.com/	Permite a NetApp transmitir datos desde registros de auditoría.

Puntos finales necesarios para GCP

Puntos finales	Objetivo
\ https://api.console.netapp.com	Comunicación con el servicio de consola, que incluye cuentas de NetApp .
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	Comunicación con el sitio web de la consola para la autenticación centralizada de usuarios.

Puntos finales	Objetivo
\ 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/	Proporciona acceso a imágenes de software, manifiestos, plantillas y permite enviar registros y métricas.
\ https://support.compliance.api.console.netapp.com/	Permite a NetApp transmitir datos desde registros de auditoría.

Asegúrese de que la clasificación de datos tenga los permisos necesarios

Asegúrese de que la clasificación de datos tenga permisos para implementar recursos y crear grupos de seguridad para la instancia de clasificación de datos.

- "Permisos de Google Cloud"
- "Permisos de AWS"
- "Permisos de Azure"

Asegúrese de que el agente de la consola pueda acceder a la clasificación de datos

Asegúrese de la conectividad entre el agente de la consola y la instancia de clasificación de datos. El grupo de seguridad del agente de consola debe permitir el tráfico entrante y saliente a través del puerto 443 hacia y desde la instancia de clasificación de datos. Esta conexión permite la implementación de la instancia de Clasificación de datos y le permite ver información en las pestañas Cumplimiento y Gobernanza. La clasificación de datos es compatible con las regiones gubernamentales en AWS y Azure.

Se requieren reglas de grupo de seguridad entrantes y salientes adicionales para las implementaciones de AWS y AWS GovCloud. Ver "Reglas para el agente de consola en AWS" Para más detalles.

Se requieren reglas de grupo de seguridad entrantes y salientes adicionales para las implementaciones de Azure y Azure Government. Ver "Reglas para el agente de consola en Azure" Para más detalles.

Asegúrese de poder mantener la clasificación de datos en funcionamiento

La instancia de Clasificación de datos debe permanecer activada para escanear continuamente sus datos.

Asegúrese de que el navegador web esté conectado a la clasificación de datos.

Una vez habilitada la clasificación de datos, asegúrese de que los usuarios accedan a la interfaz de la consola desde un host que tenga una conexión a la instancia de clasificación de datos.

La instancia de clasificación de datos utiliza una dirección IP privada para garantizar que los datos indexados no sean accesibles a través de Internet. Como resultado, el navegador web que utiliza para acceder a la Consola debe tener una conexión a esa dirección IP privada. Esa conexión puede provenir de una conexión directa a su proveedor de nube (por ejemplo, una VPN) o de un host que esté dentro de la misma red que la instancia de clasificación de datos.

Comprueba los límites de tu vCPU

Asegúrese de que el límite de vCPU de su proveedor de nube permita la implementación de una instancia con la cantidad necesaria de núcleos. Necesitará verificar el límite de vCPU para la familia de instancias relevante en la región donde se ejecuta la consola. "Ver los tipos de instancia requeridos" .

Consulte los siguientes enlaces para obtener más detalles sobre los límites de vCPU:

- "Documentación de AWS: Cuotas de servicio de Amazon EC2"
- "Documentación de Azure: Cuotas de vCPU de máquinas virtuales"
- "Documentación de Google Cloud: Cuotas de recursos"

Implementar la clasificación de datos en la nube

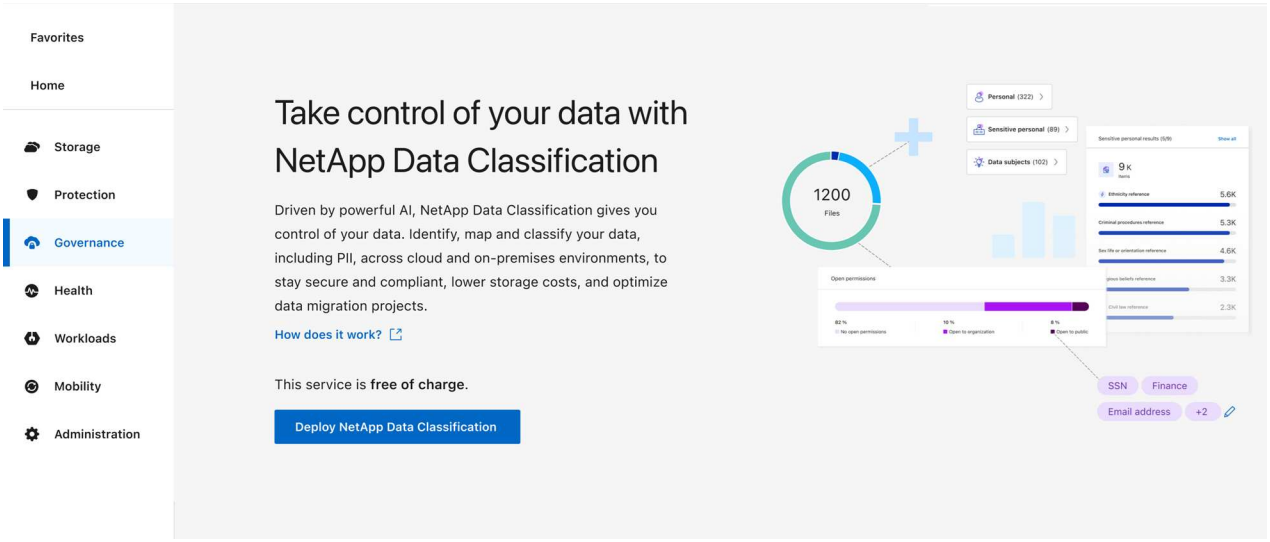
Siga estos pasos para implementar una instancia de Clasificación de datos en la nube. El agente de la consola implementará la instancia en la nube y luego instalará el software de clasificación de datos en esa instancia.

En las regiones donde el tipo de instancia predeterminado no está disponible, la clasificación de datos se ejecuta en un"tipo de instancia alternativo" .

Implementar en AWS

Pasos

1. Desde la página principal de Clasificación de datos, seleccione **Implementar clasificación en las instalaciones o en la nube**.

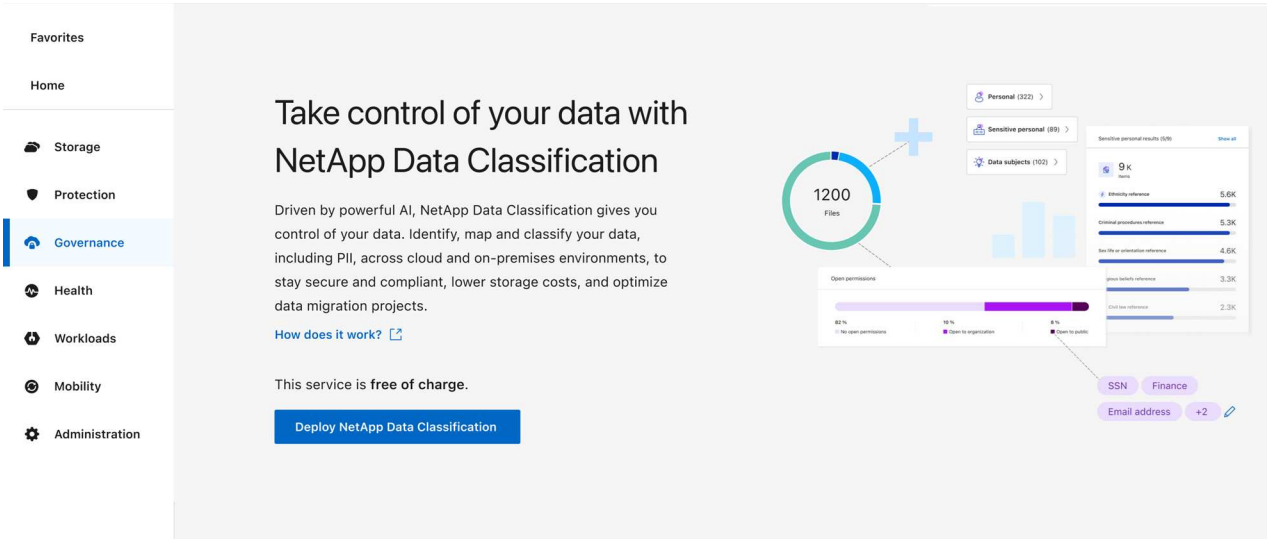


2. Desde la página *Instalación*, seleccione **Implementar > Implementar** para usar el tamaño de instancia "Grande" e iniciar el asistente de implementación en la nube.
3. El asistente muestra el progreso a medida que avanza en los pasos de implementación. Cuando se requieren entradas o si surgen problemas, se le solicitará información.
4. Cuando se implementa la instancia y se instala la Clasificación de datos, seleccione **Continuar con la configuración** para ir a la página *Configuración*.

Implementar en Azure

Pasos

1. Desde la página principal de Clasificación de datos, seleccione **Implementar clasificación en las instalaciones o en la nube**.



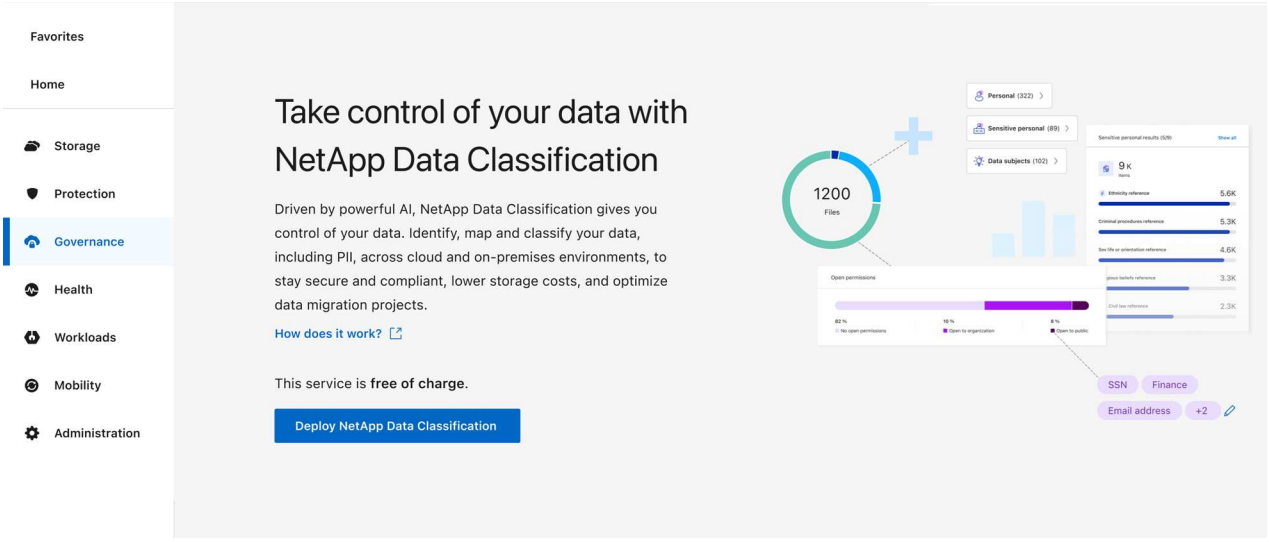
2. Seleccione **Implementar** para iniciar el asistente de implementación en la nube.

3. El asistente muestra el progreso a medida que avanza en los pasos de implementación. Se detendrá y solicitará información si surge algún problema.
4. Cuando se implementa la instancia y se instala la Clasificación de datos, seleccione **Continuar con la configuración** para ir a la página *Configuración*.

Implementar en Google Cloud

Pasos

1. Desde la página principal de Clasificación de datos, seleccione **Gobernanza > Clasificación**.
2. Seleccione **Implementar clasificación local o en la nube**.



3. Seleccione **Implementar** para iniciar el asistente de implementación en la nube.
4. El asistente muestra el progreso a medida que avanza en los pasos de implementación. Se detendrá y solicitará información si surge algún problema.
5. Cuando se implementa la instancia y se instala la Clasificación de datos, seleccione **Continuar con la configuración** para ir a la página *Configuración*.

Resultado

La consola implementa la instancia de clasificación de datos en su proveedor de nube.

Las actualizaciones del agente de consola y del software de clasificación de datos se automatizan siempre que las instancias tengan conectividad a Internet.

¿Qué sigue?

Desde la página de Configuración puede seleccionar las fuentes de datos que desea escanear.

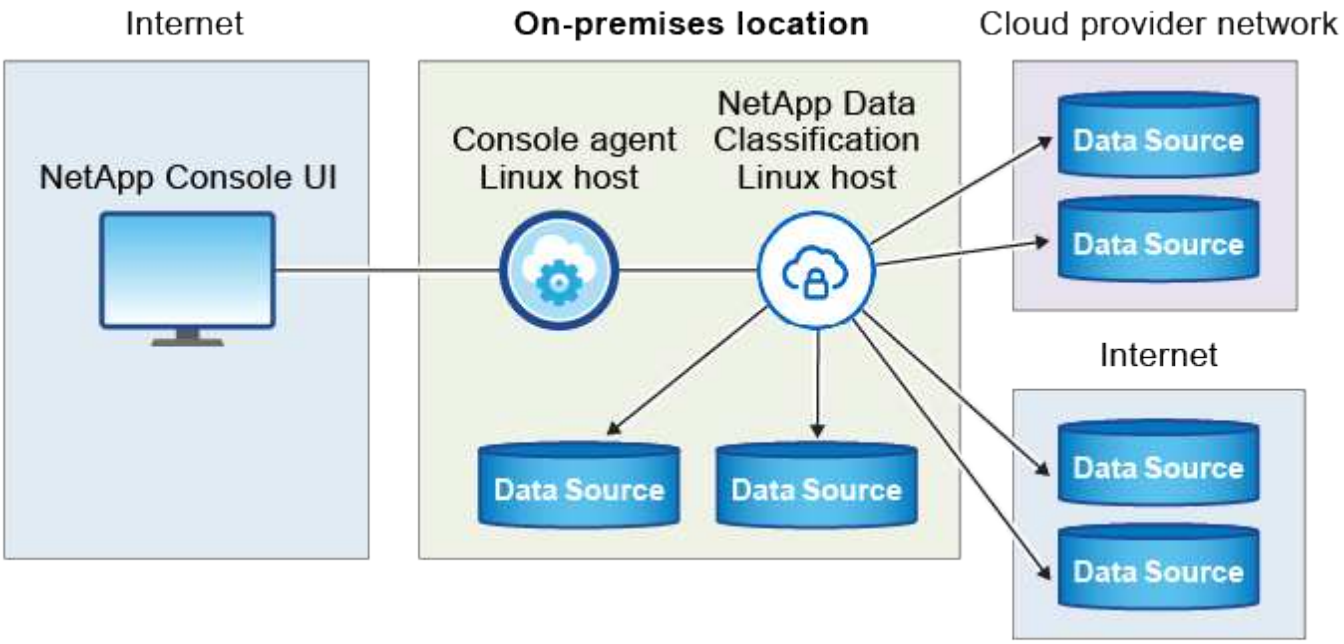
Instalar NetApp Data Classification en un host que tenga acceso a Internet

Para implementar NetApp Data Classification en un host Linux en su red o en un host Linux en la nube que tenga acceso a Internet, debe implementar el host Linux manualmente en su red o en la nube.

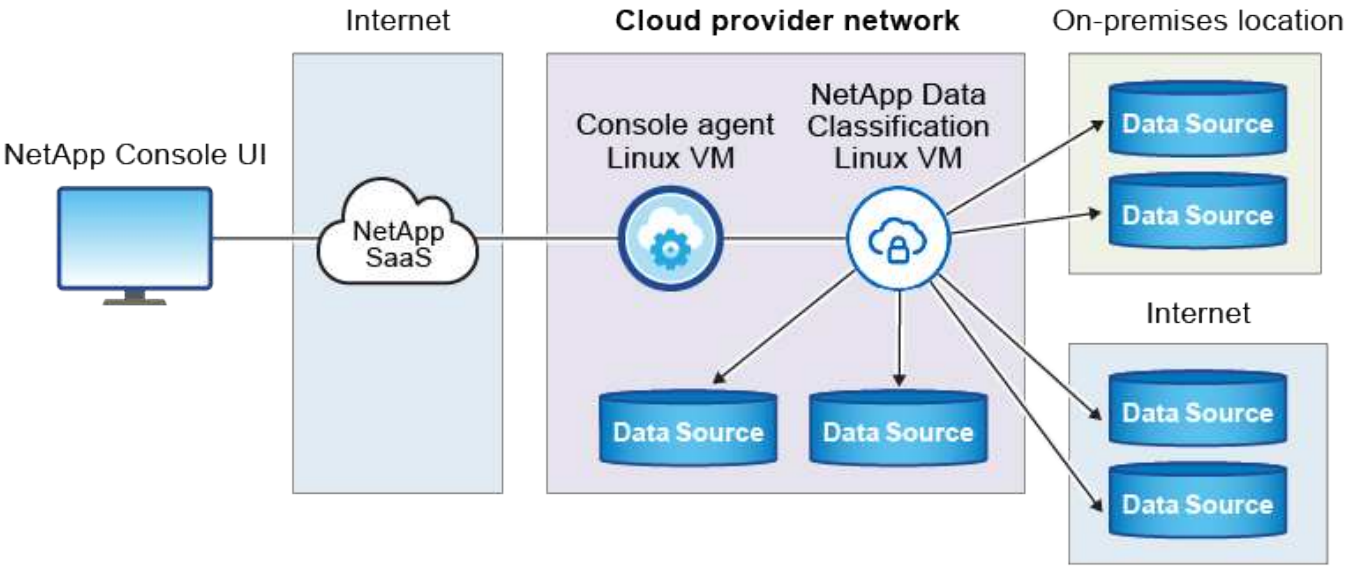
La instalación local es una buena opción si prefiere escanear sistemas ONTAP locales utilizando una instancia de clasificación de datos que también se encuentra en las instalaciones. Esto no es un requisito. El software funciona de la misma manera independientemente del método de instalación que elija.

El script de instalación de Clasificación de datos comienza verificando si el sistema y el entorno cumplen los requisitos previos requeridos. Si se cumplen todos los requisitos previos, se inicia la instalación. Si desea verificar los requisitos previos independientemente de ejecutar la instalación de Clasificación de datos, hay un paquete de software separado que puede descargar que solo prueba los requisitos previos. "Vea cómo comprobar si su host Linux está listo para instalar la Clasificación de Datos" .

La instalación típica en un host Linux *en sus instalaciones* tiene los siguientes componentes y conexiones.



La instalación típica en un host Linux *en la nube* tiene los siguientes componentes y conexiones.



Inicio rápido

Comience rápidamente siguiendo estos pasos o desplácese hacia abajo hasta las secciones restantes para obtener detalles completos.

1 Crear un agente de consola

Si aún no tienes un agente de consola, "Implementar el agente de consola localmente" en un host Linux en su red o en un host Linux en la nube.

También puedes crear un agente de consola con tu proveedor de nube. Ver "Creación de un agente de consola en AWS" , "Creación de un agente de consola en Azure" , o "Creación de un agente de consola en GCP" .

2 Revisar los prerrequisitos

Asegúrese de que su entorno pueda cumplir con los requisitos previos. Esto incluye acceso a Internet saliente para la instancia, conectividad entre el agente de la consola y la clasificación de datos a través del puerto 443 y más. Ver la lista completa .

También necesitas un sistema Linux que cumpla con los requisitossiguientes requisitos .

3 Descargar e implementar la clasificación de datos

Descargue el software Cloud Data Classification del sitio de soporte de NetApp y copie el archivo de instalación en el host Linux que planea utilizar. Luego, inicie el asistente de instalación y siga las instrucciones para implementar la instancia de Clasificación de datos.

Crear un agente de consola

Se requiere un agente de consola antes de poder instalar y utilizar la clasificación de datos. En la mayoría de los casos, probablemente tendrá un agente de consola configurado antes de intentar activar la clasificación de datos porque la mayoría "Las funciones de la consola requieren un agente de consola" , pero habrá casos en los que necesitarás configurar uno ahora.

Para crear uno en su entorno de proveedor de nube, consulte "Creación de un agente de consola en AWS" , "Creación de un agente de consola en Azure" , o "Creación de un agente de consola en GCP" .

Hay algunos escenarios en los que es necesario utilizar un agente de consola implementado en un proveedor de nube específico:

- Al escanear datos en Cloud Volumes ONTAP en AWS o Amazon FSx para ONTAP, se utiliza un agente de consola en AWS.
- Al escanear datos en Cloud Volumes ONTAP en Azure o en Azure NetApp Files, se utiliza un agente de consola en Azure.

  Para Azure NetApp Files, debe implementarse en la misma región que los volúmenes que desea escanear.

- Al escanear datos en Cloud Volumes ONTAP en GCP, se utiliza un agente de consola en GCP.

Los sistemas ONTAP locales, los recursos compartidos de archivos de NetApp y las cuentas de bases de

datos se pueden escanear utilizando cualquiera de estos agentes de consola en la nube.

Tenga en cuenta que también puede "Implementar el agente de consola localmente" en un host Linux en su red o en un host Linux en la nube. Algunos usuarios que planean instalar Data Classification en sus instalaciones también pueden optar por instalar el agente de consola en sus instalaciones.

Necesitará la dirección IP o el nombre de host del sistema del agente de consola al instalar Clasificación de datos. Tendrás esta información si instalaste el agente de consola en tus instalaciones. Si el agente de la consola está implementado en la nube, puede encontrar esta información en la consola: seleccione el ícono Ayuda, luego **Soporte** y luego **Agente de consola**.

Preparar el sistema host Linux

El software de clasificación de datos debe ejecutarse en un host que cumpla con los requisitos específicos del sistema operativo, requisitos de RAM, requisitos de software, etc. El host Linux puede estar en su red o en la nube.

Asegúrese de poder mantener la clasificación de datos en ejecución. La máquina de clasificación de datos debe permanecer encendida para escanear continuamente sus datos.

- La clasificación de datos debe realizarse en un host dedicado. El host no se puede compartir con otras aplicaciones o software de terceros, como antivirus.
- Elija el tamaño que se alinee con el conjunto de datos que planea escanear con Clasificación de datos.

Tamaño del sistema	UPC	RAM (la memoria de intercambio debe estar deshabilitada)	Disco
Extra grande	32 CPU	128 GB de RAM	• SSD de 1 TiB en /, o 100 GiB disponibles en /opt • 895 GiB disponibles en /var/lib/docker • 5 GiB en /tmp • **Para Podman, 30 GB en /var/tmp**
Grande	16 CPU	64 GB de RAM	• SSD de 500 GiB en /, o 100 GiB disponibles en /opt • 400 GiB disponibles en /var/lib/docker o para Podman /var/lib/containers • 5 GiB en /tmp • **Para Podman, 30 GB en /var/tmp**

- Al implementar una instancia de cómputo en la nube para su instalación de Clasificación de datos, se recomienda utilizar un sistema que cumpla con los requisitos del sistema "Grande" mencionados anteriormente:
  - **Tipo de instancia de Amazon Elastic Compute Cloud (Amazon EC2)**: "m6i.4xlarge". "Ver tipos de instancias de AWS adicionales" .

- **Tamaño de máquina virtual de Azure**: "Standard_D16s_v3". "Ver tipos de instancias de Azure adicionales" .
- **Tipo de máquina GCP**: "n2-standard-16". "Ver tipos de instancias de GCP adicionales" .

- **Permisos de carpeta UNIX**: Se requieren los siguientes permisos mínimos de UNIX:

Carpeta	Permisos mínimos
/tmp	`rwxrwxrwt`
/optar	`rwxr-xr-x`
/var/lib/docker	`rwx------`
/usr/lib/systemd/sistema	`rwxr-xr-x`

- **Sistema operativo**:
  - Los siguientes sistemas operativos requieren el uso del motor de contenedores Docker:
    - Red Hat Enterprise Linux versión 7.8 y 7.9
    - Ubuntu 22.04 (requiere la versión 1.23 o superior de Data Classification)
    - Ubuntu 24.04 (requiere la versión 1.23 o superior de Data Classification)
  - Los siguientes sistemas operativos requieren el uso del motor de contenedores Podman y requieren la versión 1.30 o superior de Data Classification:
    - Red Hat Enterprise Linux versiones 8.8, 8.10, 9.0, 9.1, 9.2, 9.3, 9.4, 9.5 y 9.6.
  - Las extensiones vectoriales avanzadas (AVX2) deben estar habilitadas en el sistema host.
- **Gestión de suscripciones de Red Hat**: el host debe estar registrado en Gestión de suscripciones de Red Hat. Si no está registrado, el sistema no puede acceder a los repositorios para actualizar el software de terceros requerido durante la instalación.
- **Software adicional**: Debe instalar el siguiente software en el host antes de instalar Data Classification:
  - Dependiendo del sistema operativo que estés usando, necesitas instalar uno de los motores de contenedores:
    - Docker Engine versión 19.3.1 o superior. "Ver instrucciones de instalación" .
    - Podman versión 4 o superior. Para instalar Podman, ingrese(`sudo yum install podman netavark -y`).
- Versión de Python 3.6 o superior. "Ver instrucciones de instalación" .
  - **Consideraciones sobre NTP**: NetApp recomienda configurar el sistema de clasificación de datos para utilizar un servicio de Protocolo de tiempo de red (NTP). La hora debe estar sincronizada entre el sistema de clasificación de datos y el sistema del agente de consola.
- **Consideraciones sobre Firewalld**: Si planea utilizar `firewalld` Le recomendamos que lo habilite antes de instalar Data Classification. Ejecute los siguientes comandos para configurar `firewalld` para que sea compatible con la Clasificación de Datos:

```
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
```

Si planea utilizar hosts de clasificación de datos adicionales como nodos de escáner, agregue estas reglas a su sistema principal en este momento:

```
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
```

Tenga en cuenta que debe reiniciar Docker o Podman cada vez que habilite o actualice `firewalld` ajustes.

La dirección IP del sistema host de clasificación de datos no se puede cambiar después de la instalación.

Habilitar el acceso a Internet saliente desde la Clasificación de datos

La clasificación de datos requiere acceso a Internet saliente. Si su red virtual o física utiliza un servidor proxy para el acceso a Internet, asegúrese de que la instancia de clasificación de datos tenga acceso a Internet saliente para contactar los siguientes puntos finales.

Puntos finales	Objetivo
\ https://api.console.netapp.com	Comunicación con la consola, que incluye cuentas de NetApp .
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	Comunicación con el sitio web de la consola para la autenticación centralizada de usuarios.
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/	Proporciona acceso a imágenes de software, manifiestos, plantillas y permite enviar registros y métricas.
https://support.compliance.api.bluexp.netapp.com/	Permite a NetApp transmitir datos desde registros de auditoría.
\ https://github.com/docker \ https://download.docker.com	Proporciona paquetes de requisitos previos para la instalación de Docker.
\ http://packages.ubuntu.com/ \ http://archive.ubuntu.com	Proporciona paquetes de requisitos previos para la instalación de Ubuntu.

Verifique que todos los puertos requeridos estén habilitados

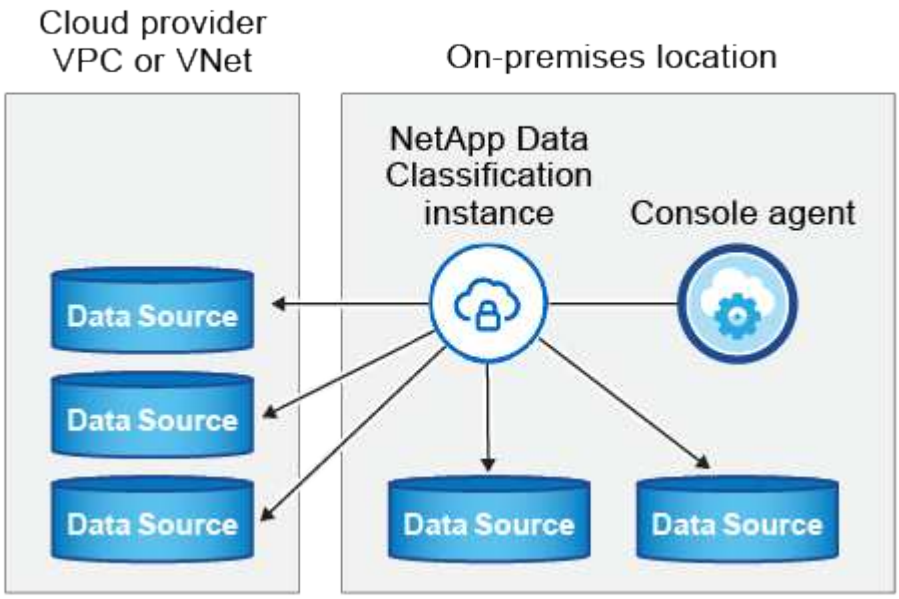
Debe asegurarse de que todos los puertos necesarios estén abiertos para la comunicación entre el agente de la consola, la clasificación de datos, Active Directory y sus fuentes de datos.

Tipo de conexión	Puertos	Descripción
Agente de consola <> Clasificación de datos	8080 (TCP), 443 (TCP) y 80. 9000	Las reglas de firewall o enrutamiento para el agente de la consola deben permitir el tráfico entrante y saliente a través del puerto 443 hacia y desde la instancia de clasificación de datos. Asegúrese de que el puerto 8080 esté abierto para que pueda ver el progreso de la instalación en la consola. Si se utiliza un firewall en el host Linux, se requiere el puerto 9000 para los procesos internos dentro de un servidor Ubuntu.
Agente de consola <> clúster ONTAP (NAS)	443 (TCP)	La consola descubre clústeres ONTAP mediante HTTPS. Si utiliza políticas de firewall personalizadas, deben cumplir los siguientes requisitos: • El host del agente de la consola debe permitir el acceso HTTPS saliente a través del puerto 443. Si el agente de la consola está en la nube, toda comunicación saliente está permitida por el firewall predefinido o las reglas de enrutamiento. • El clúster ONTAP debe permitir el acceso HTTPS entrante a través del puerto 443. La política de firewall predeterminada "mgmt" permite el acceso HTTPS entrante desde todas las direcciones IP. Si modificó esta política predeterminada o si creó su propia política de firewall, debe asociar el protocolo HTTPS con esa política y habilitar el acceso desde el host del agente de la Consola.
Clasificación de datos <> Clúster ONTAP	• Para NFS - 111 (TCP\UDP) y 2049 (TCP\UDP) • Para CIFS - 139 (TCP\UDP) y 445 (TCP\UDP)	La clasificación de datos necesita una conexión de red a cada subred de Cloud Volumes ONTAP o al sistema ONTAP local. Los firewalls o las reglas de enrutamiento para Cloud Volumes ONTAP deben permitir conexiones entrantes desde la instancia de clasificación de datos. Asegúrese de que estos puertos estén abiertos para la instancia de clasificación de datos: • Para NFS - 111 y 2049 • Para CIFS - 139 y 445 Las políticas de exportación de volumen NFS deben permitir el acceso desde la instancia de clasificación de datos.

Tipo de conexión	Puertos	Descripción
Clasificación de datos <> Active Directory	389 (TCP y UDP), 636 (TCP), 3268 (TCP) y 3269 (TCP)	Debe tener un Directorio Activo ya configurado para los usuarios de su empresa. Además, la clasificación de datos necesita credenciales de Active Directory para escanear volúmenes CIFS. Debes tener la información del Directorio Activo: • Dirección IP del servidor DNS o varias direcciones IP • Nombre de usuario y contraseña para el servidor • Nombre de dominio (nombre de Active Directory) • Ya sea que esté utilizando LDAP seguro (LDAPS) o no • Puerto del servidor LDAP (normalmente 389 para LDAP y 636 para LDAP seguro)

Instalar la clasificación de datos en el host Linux

Para configuraciones típicas, instalará el software en un solo sistema host. Vea esos pasos aquí .



VerPreparación del sistema host Linux yRevisión de prerrequisitos para obtener la lista completa de requisitos antes de implementar la clasificación de datos.

Las actualizaciones del software de clasificación de datos se automatizan siempre que la instancia tenga conectividad a Internet.

Actualmente, la clasificación de datos no puede escanear depósitos S3, Azure NetApp Files o FSx para ONTAP cuando el software está instalado en las instalaciones. En estos casos, necesitará implementar un agente de consola independiente y una instancia de clasificación de datos en la nube y "cambiar entre conectores" para sus diferentes fuentes de datos.

Instalación de un solo host para configuraciones típicas

Revise los requisitos y siga estos pasos al instalar el software de clasificación de datos en un solo host local.

"Mira este vídeo"para ver cómo instalar Clasificación de Datos.

Tenga en cuenta que todas las actividades de instalación se registran al instalar Data Classification. Si surge algún problema durante la instalación, puede ver el contenido del registro de auditoría de la instalación. Esta escrito para `/opt/netapp/install_logs/` .

Antes de empezar

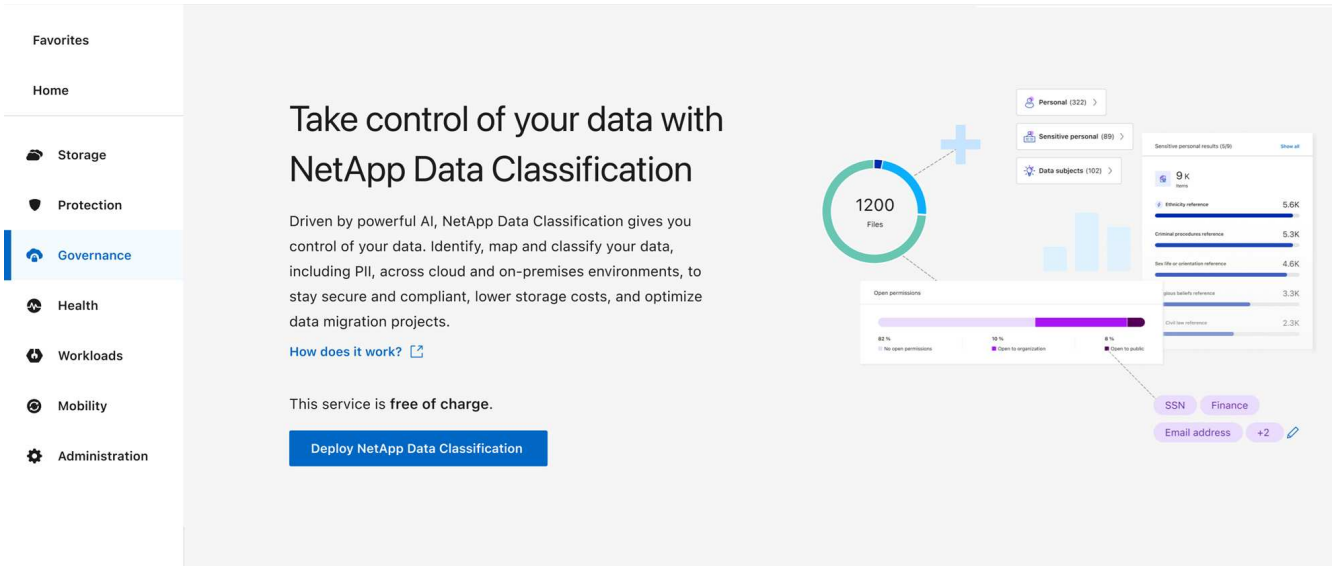
- Verifique que su sistema Linux cumpla con losrequisitos del anfitrión .
- Verifique que el sistema tenga instalados los dos paquetes de software necesarios (Docker Engine o Podman y Python 3).
- Asegúrese de tener privilegios de root en el sistema Linux.
- Si está utilizando un proxy para acceder a Internet:
  - Necesitará la información del servidor proxy (dirección IP o nombre de host, puerto de conexión, esquema de conexión: https o http, nombre de usuario y contraseña).
  - Si el proxy realiza la interceptación de TLS, necesitará saber la ruta en el sistema Linux de clasificación de datos donde se almacenan los certificados CA de TLS.
  - El proxy no debe ser transparente. Actualmente, la clasificación de datos no admite servidores proxy transparentes.
  - El usuario debe ser un usuario local. Los usuarios del dominio no son compatibles.
- Verifique que su entorno fuera de línea cumpla con los requisitospermisos y conectividad .

Pasos

1. Descargue el software de clasificación de datos desde "Sitio de soporte de NetApp" . El archivo que debe seleccionar se llama **DATASENSE-INSTALLER-<versión>.tar.gz**.
2. Copie el archivo de instalación en el host Linux que planea utilizar (usando `scp` o algún otro método).
3. Descomprima el archivo de instalación en la máquina host, por ejemplo:

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

4. En la consola, seleccione **Gobernanza > Clasificación**.
5. Seleccione **Implementar clasificación local o en la nube**.



6. Dependiendo de si está instalando Data Classification en una instancia que preparó en la nube o en una instancia que preparó en sus instalaciones, seleccione la opción **Implementar** adecuada para iniciar la instalación de Data Classification.
7. Se muestra el cuadro de diálogo *Implementar clasificación de datos en las instalaciones*. Copie el comando proporcionado (por ejemplo: `sudo ./install.sh -a 12345 -c 27AG75 -t 2198qq`) y pégalo en un archivo de texto para que puedas usarlo más tarde. Luego seleccione **Cerrar** para cerrar el cuadro de diálogo.
8. En la máquina host, ingrese el comando que copió y luego siga una serie de indicaciones, o puede proporcionar el comando completo incluidos todos los parámetros requeridos como argumentos de la línea de comando.

   Tenga en cuenta que el instalador realiza una verificación previa para asegurarse de que los requisitos del sistema y de la red estén cumplidos para una instalación exitosa. "Mira este vídeo" Para comprender los mensajes previos a la verificación y sus implicaciones.

Introduzca los parámetros según se le solicite:	Introduzca el comando completo:
a. Pegue el comando que copió del paso 7: `sudo ./install.sh -a <account_id> -c <client_id> -t <user_token>` Si está instalando en una instancia de nube (no en sus instalaciones), agregue `--manual-cloud-install <cloud_provider>`. b. Introduzca la dirección IP o el nombre de host de la máquina host de clasificación de datos para que el sistema del agente de la consola pueda acceder a ella. c. Ingrese la dirección IP o el nombre de host de la máquina host del agente de consola para que el sistema de clasificación de datos pueda acceder a ella. d. Introduzca los detalles del proxy cuando se le solicite. Si su agente de consola ya utiliza un proxy, no es necesario ingresar esta información nuevamente aquí ya que la clasificación de datos utilizará automáticamente el proxy utilizado por el agente de consola.	Alternativamente, puede crear todo el comando por adelantado, proporcionando los parámetros de host y proxy necesarios: `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>`

Valores variables:

- *account_id* = ID de cuenta de NetApp
- *client_id* = ID de cliente del agente de consola (agregue el sufijo "clients" al ID de cliente si aún no está allí)
- *user_token* = token de acceso de usuario JWT
- *ds_host* = dirección IP o nombre de host del sistema Linux de clasificación de datos.
- *cm_host* = dirección IP o nombre de host del sistema del agente de consola.
- *cloud_provider* = Al instalar en una instancia de nube, ingrese "AWS", "Azure" o "Gcp" según el proveedor de nube.
- *proxy_host* = IP o nombre de host del servidor proxy si el host está detrás de un servidor proxy.
- *proxy_port* = Puerto para conectarse al servidor proxy (predeterminado 80).
- *proxy_scheme* = Esquema de conexión: https o http (predeterminado http).
- *proxy_user* = Usuario autenticado para conectarse al servidor proxy, si se requiere autenticación básica. El usuario debe ser un usuario local (no se admiten usuarios de dominio).
- *proxy_password* = Contraseña para el nombre de usuario que usted especificó.
- *ca_cert_dir* = Ruta en el sistema Linux de clasificación de datos que contiene paquetes de certificados CA TLS adicionales. Solo es necesario si el proxy está realizando intercepción TLS.

Resultado

El instalador de Data Classification instala paquetes, registra la instalación e instala Data Classification. La instalación puede tardar entre 10 y 20 minutos.

Si hay conectividad a través del puerto 8080 entre la máquina host y la instancia del agente de la consola, verá el progreso de la instalación en la pestaña Clasificación de datos en la consola.

¿Qué sigue?

Desde la página de Configuración puede seleccionar las fuentes de datos que desea escanear.

Instalar NetApp Data Classification en un host Linux sin acceso a Internet

La instalación de NetApp Data Classification en un host Linux en un sitio local que no tiene acceso a Internet se conoce como *modo privado*. Este tipo de instalación, que utiliza un script de instalación, no tiene conectividad con la capa SaaS de la NetApp Console .

> El modo privado de BlueXP (interfaz BlueXP heredada) generalmente se usa con entornos locales que no tienen conexión a Internet y con regiones de nube seguras, que incluyen AWS Secret Cloud, AWS Top Secret Cloud y Azure IL6. NetApp continúa brindando soporte a estos entornos con la interfaz BlueXP heredada. Para obtener documentación del modo privado en la interfaz heredada de BlueXP , consulte"Documentación en PDF para el modo privado de BlueXP" .

Compruebe que su host Linux esté listo para instalar NetApp Data Classification

Antes de instalar NetApp Data Classification manualmente en un host Linux, opcionalmente ejecute un script en el host para verificar que todos los requisitos previos estén cumplidos para instalar Data Classification. Puede ejecutar este script en un host Linux en su red o en un host Linux en la nube. El host puede estar conectado a Internet o puede residir en un sitio que no tenga acceso a Internet (un *sitio oscuro*).

El script de instalación de Clasificación de datos incluye un script de prueba para garantizar que su entorno cumpla con los requisitos. Puede ejecutar este script por separado para verificar la preparación del host Linux antes de ejecutar el script de instalación.

Empezando

Realizarás las siguientes tareas:

- Opcionalmente, instale un agente de consola si aún no tiene uno instalado. Puede ejecutar el script de prueba sin tener un agente de consola instalado, pero el script verifica la conectividad entre el agente de consola y la máquina host de clasificación de datos, por lo que se recomienda que tenga un agente de consola.
- Prepare la máquina host y verifique que cumpla con todos los requisitos.
- Habilitar el acceso a Internet saliente desde la máquina host de clasificación de datos.
- Verifique que todos los puertos necesarios estén habilitados en todos los sistemas.
- Descargue y ejecute el script de prueba de prerrequisitos.

Crear un agente de consola

Se requiere un agente de consola antes de poder instalar y utilizar la clasificación de datos. Sin embargo, puede ejecutar el script de Requisitos previos sin un agente de consola.

Puede "Instalar el agente de consola local" en un host Linux en su red o en un host Linux en la nube. También puede instalar Clasificación de datos localmente si el agente de Consola está instalado localmente.

Para crear un agente de consola en su entorno de proveedor de nube, consulte:

- "Creación de un agente de consola en AWS"
- "Creación de un agente de consola en Azure"
- "Creación de un agente de consola en GCP"

Necesita la dirección IP o el nombre de host del sistema del agente de la consola al ejecutar el script de requisitos previos. Tienes esta información si instalaste el agente de consola en tus instalaciones. Si el agente de la Consola está implementado en la nube, puede encontrar esta información desde la Consola: seleccione el ícono Ayuda y luego **Soporte**; en la sección Agente y Auditoría, seleccione **Ir al agente**.

Verificar los requisitos del host

El software de clasificación de datos debe ejecutarse en un host que cumpla con los requisitos específicos del sistema operativo, requisitos de RAM y requisitos de software.

- La clasificación de datos debe realizarse en un host dedicado. El host no se puede compartir con otras aplicaciones o software de terceros, como antivirus.
- Elija el tamaño que se alinee con el conjunto de datos que planea escanear con Clasificación de datos.

Tamaño del sistema	UPC	RAM (la memoria de intercambio debe estar deshabilitada)	Disco
Extra grande	32 CPU	128 GB de RAM	• SSD de 1 TiB en /, o 100 GiB disponibles en /opt • 895 GiB disponibles en /var/lib/docker • 5 GiB en /tmp • **Para Podman, 30 GB en /var/tmp**
Grande	16 CPU	64 GB de RAM	• SSD de 500 GiB en /, o 100 GiB disponibles en /opt • 400 GiB disponibles en /var/lib/docker o para Podman /var/lib/containers • 5 GiB en /tmp • **Para Podman, 30 GB en /var/tmp**

- Al implementar una instancia de cómputo en la nube para su instalación de Clasificación de datos, se recomienda utilizar un sistema que cumpla con los requisitos del sistema "Grande" mencionados anteriormente:
  - **Tipo de instancia de Amazon Elastic Compute Cloud (Amazon EC2)**: "m6i.4xlarge". "Ver tipos de instancias de AWS adicionales" .
  - **Tamaño de máquina virtual de Azure**: "Standard_D16s_v3". "Ver tipos de instancias de Azure adicionales" .
  - **Tipo de máquina GCP**: "n2-standard-16". "Ver tipos de instancias de GCP adicionales" .
- **Permisos de carpeta UNIX**: Se requieren los siguientes permisos mínimos de UNIX:

Carpeta	Permisos mínimos
/tmp	`rwxrwxrwt`
/optar	`rwxr-xr-x`
/var/lib/docker	`rwx------`
/usr/lib/systemd/sistema	`rwxr-xr-x`

- **Sistema operativo**:
  - Los siguientes sistemas operativos requieren el uso del motor de contenedores Docker:
    - Red Hat Enterprise Linux versión 7.8 y 7.9
    - Ubuntu 22.04 (requiere la versión 1.23 o superior de Data Classification)
    - Ubuntu 24.04 (requiere la versión 1.23 o superior de Data Classification)
  - Los siguientes sistemas operativos requieren el uso del motor de contenedores Podman y requieren la versión 1.30 o superior de Data Classification:
    - Red Hat Enterprise Linux versiones 8.8, 8.10, 9.0, 9.1, 9.2, 9.3, 9.4, 9.5 y 9.6.
  - Las extensiones vectoriales avanzadas (AVX2) deben estar habilitadas en el sistema host.
- **Gestión de suscripciones de Red Hat**: el host debe estar registrado en Gestión de suscripciones de Red Hat. Si no está registrado, el sistema no puede acceder a los repositorios para actualizar el software de terceros requerido durante la instalación.
- **Software adicional**: Debe instalar el siguiente software en el host antes de instalar Data Classification:
  - Dependiendo del sistema operativo que estés usando, necesitas instalar uno de los motores de contenedores:
    - Docker Engine versión 19.3.1 o superior. "Ver instrucciones de instalación" .
    - Podman versión 4 o superior. Para instalar Podman, ingrese(`sudo yum install podman netavark -y`).
- Versión de Python 3.6 o superior. "Ver instrucciones de instalación" .
  - **Consideraciones sobre NTP**: NetApp recomienda configurar el sistema de clasificación de datos para utilizar un servicio de Protocolo de tiempo de red (NTP). La hora debe estar sincronizada entre el sistema de clasificación de datos y el sistema del agente de consola.
- **Consideraciones sobre Firewalld**: Si planea utilizar `firewalld` Le recomendamos que lo habilite antes de instalar Data Classification. Ejecute los siguientes comandos para configurar `firewalld` para que sea compatible con la Clasificación de Datos:

```
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
```

Si planea utilizar hosts de clasificación de datos adicionales como nodos de escáner (en un modelo distribuido), agregue estas reglas a su sistema principal en este momento:

```
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
```

Tenga en cuenta que debe reiniciar Docker o Podman cada vez que habilite o actualice `firewalld` ajustes.

Habilitar el acceso a Internet saliente desde la Clasificación de datos

La clasificación de datos requiere acceso a Internet saliente. Si su red virtual o física utiliza un servidor proxy para el acceso a Internet, asegúrese de que la instancia de clasificación de datos tenga acceso a Internet saliente para contactar los siguientes puntos finales.

Esta sección no es necesaria para los sistemas host instalados en sitios sin conectividad a Internet.

Puntos finales	Objetivo
\ https://api.console.netapp.com	Comunicación con el servicio de consola, que incluye cuentas de NetApp .
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	Comunicación con el sitio web de la consola para la autenticación centralizada de usuarios.
\ 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/	Proporciona acceso a imágenes de software, manifiestos, plantillas y permite enviar registros y métricas.
\ https://support.compliance.api.console.netapp.com/	Permite a NetApp transmitir datos desde registros de auditoría.
\ https://github.com/docker \ https://download.docker.com	Proporciona paquetes de requisitos previos para la instalación de Docker.

Puntos finales	Objetivo
\ http://packages.ubuntu.com/ \ http://archive.ubuntu.com	Proporciona paquetes de requisitos previos para la instalación de Ubuntu.

Verifique que todos los puertos requeridos estén habilitados

Debe asegurarse de que todos los puertos necesarios estén abiertos para la comunicación entre el agente de la consola, la clasificación de datos, Active Directory y sus fuentes de datos.

Tipo de conexión	Puertos	Descripción
Agente de consola <> Clasificación de datos	8080 (TCP), 443 (TCP) y 80. 9000	Las reglas de firewall o enrutamiento para el agente de la consola deben permitir el tráfico entrante y saliente a través del puerto 443 hacia y desde la instancia de clasificación de datos. Asegúrese de que el puerto 8080 esté abierto para que pueda ver el progreso de la instalación en la consola. Si se utiliza un firewall en el host Linux, se requiere el puerto 9000 para los procesos internos dentro de un servidor Ubuntu.
Agente de consola <> clúster ONTAP (NAS)	443 (TCP)	La consola descubre clústeres ONTAP mediante HTTPS. Si utiliza políticas de firewall personalizadas, el host del agente de la consola debe permitir el acceso HTTPS saliente a través del puerto 443. Si el agente de la consola está en la nube, toda comunicación saliente está permitida por el firewall predefinido o las reglas de enrutamiento.

Ejecute el script de requisitos previos de clasificación de datos

Siga estos pasos para ejecutar el script de requisitos previos de clasificación de datos.

"Mira este vídeo"para ver cómo ejecutar el script de requisitos previos e interpretar los resultados.

Antes de empezar

- Verifique que su sistema Linux cumpla con losrequisitos del anfitrión .
- Verifique que el sistema tenga instalados los dos paquetes de software necesarios (Docker Engine o Podman y Python 3).
- Asegúrese de tener privilegios de root en el sistema Linux.

Pasos

1. Descargue el script de Requisitos previos de clasificación de datos desde "Sitio de soporte de NetApp" . El archivo que debe seleccionar se llama **standalone-pre-requisite-tester-<version>**.
2. Copie el archivo al host Linux que planea utilizar (usando `scp` o algún otro método).
3. Asignar permisos para ejecutar el script.

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

4. Ejecute el script utilizando el siguiente comando.

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

   Agregue la opción "--darksite" solo si está ejecutando el script en un host que no tiene acceso a Internet. Se omiten ciertas pruebas de requisitos previos cuando el host no está conectado a Internet.

5. El script le solicita la dirección IP de la máquina host de clasificación de datos.
   - Introduzca la dirección IP o el nombre de host.
6. El script le preguntará si tiene un agente de consola instalado.
   - Ingrese **N** si no tiene un agente de consola instalado.
   - Ingrese **Y** si tiene un agente de consola instalado. Y luego ingrese la dirección IP o el nombre de host del agente de la consola para que el script de prueba pueda probar esta conectividad.
7. El script ejecuta una variedad de pruebas en el sistema y muestra resultados a medida que avanza. Cuando termina, escribe un registro de la sesión en un archivo llamado `prerequisites-test-<timestamp>.log` en el directorio `/opt/netapp/install_logs` .

Resultado

Si todas las pruebas de requisitos previos se ejecutaron correctamente, puede instalar Data Classification en el host cuando esté listo.

Si se descubre algún problema, se clasifica como "Recomendado" o "Obligatorio" para su solución. Los problemas recomendados suelen ser elementos que harían que las tareas de categorización y escaneo de clasificación de datos se ejecuten más lentamente. No es necesario corregir estos elementos, pero es posible que quieras abordarlos.

Si tiene algún problema "Obligatorio", debe solucionarlo y ejecutar nuevamente el script de prueba de requisitos previos.