Distribuisci la classificazione dei dati

NetApp Data Classification

NetApp
February 11, 2026

Sommario

Distribuisci la classificazione dei dati

Quale distribuzione NetApp Data Classification dovresti utilizzare?

È possibile distribuire NetApp Data Classification in diversi modi. Scopri quale metodo soddisfa le tue esigenze.

La classificazione dei dati può essere implementata nei seguenti modi:

- "Distribuisci nel cloud utilizzando la console" . La Console distribuisce l'istanza di classificazione dei dati nella stessa rete del provider cloud dell'agente della Console.
- "Installa su un host Linux con accesso a Internet" . Installa Data Classification su un host Linux nella tua rete o su un host Linux nel cloud che abbia accesso a Internet. Questo tipo di installazione può essere una buona opzione se si preferisce eseguire la scansione dei sistemi ONTAP locali utilizzando un'istanza di classificazione dei dati anch'essa ubicata in sede, sebbene non sia un requisito.
- "Installa su un host Linux in un sito locale senza accesso a Internet", nota anche come *modalità privata*. Questo tipo di installazione, che utilizza uno script di installazione, non ha connettività con il livello SaaS della console.

> La modalità privata BlueXP (interfaccia BlueXP legacy) viene in genere utilizzata con ambienti locali privi di connessione Internet e con regioni cloud sicure, tra cui AWS Secret Cloud, AWS Top Secret Cloud e Azure IL6. NetApp continua a supportare questi ambienti con l'interfaccia legacy BlueXP . Per la documentazione sulla modalità privata nell'interfaccia legacy BlueXP , vedere"Documentazione PDF per la modalità privata BlueXP" .

Sia l'installazione su un host Linux con accesso a Internet sia l'installazione in locale su un host Linux senza accesso a Internet utilizzano uno script di installazione. Lo script inizia verificando se il sistema e l'ambiente soddisfano i prerequisiti. Se i prerequisiti sono soddisfatti, l'installazione inizia. Se si desidera verificare i prerequisiti indipendentemente dall'esecuzione dell'installazione di Data Classification, è possibile scaricare un pacchetto software separato che verifica solo i prerequisiti.

Fare riferimento a "Verifica che il tuo host Linux sia pronto per installare Data Classification" .

Distribuisci NetApp Data Classification nel cloud utilizzando la NetApp Console

È possibile distribuire NetApp Data Classification nel cloud con NetApp Console. La Console distribuisce l'istanza di classificazione dei dati nella stessa rete del provider cloud dell'agente della Console.

Nota che puoi anche"installare Data Classification su un host Linux con accesso a Internet" . Questo tipo di installazione può essere una buona opzione se si preferisce eseguire la scansione dei sistemi ONTAP locali utilizzando un'istanza di classificazione dei dati anch'essa ubicata in sede, ma non è un requisito. Il software funziona esattamente allo stesso modo, indipendentemente dal metodo di installazione scelto.

Avvio rapido

Inizia subito seguendo questi passaggi oppure scorri verso il basso fino alle sezioni rimanenti per i dettagli completi.

1 Creare un agente Console

Se non si dispone già di un agente Console, crearne uno. Vedere"creazione di un agente Console in AWS" ,"creazione di un agente Console in Azure" , O"creazione di un agente Console in GCP" .

Puoi anche"installare l'agente Console in locale" su un host Linux nella tua rete o su un host Linux nel cloud.

2 Prerequisiti

Assicurati che il tuo ambiente possa soddisfare i prerequisiti. Questi includono outbound accesso a Internet per l'istanza, connettività tra l'agente Console e Data Classification sulla porta 443 e altro ancora. Vedi l'elenco completo.

3 Distribuisci la classificazione dei dati

Avviare la procedura guidata di installazione per distribuire l'istanza di Data Classification nel cloud.

Creare un agente Console

Se non disponi già di un agente Console, creane uno nel tuo provider cloud. Vedere "creazione di un agente Console in AWS" O "creazione di un agente Console in Azure" , O "creazione di un agente Console in GCP" . Nella maggior parte dei casi sarà probabilmente configurato un agente Console prima di tentare di attivare la classificazione dei dati perché la maggior parte "Le funzionalità della console richiedono un agente della console" , ma ci sono casi in cui sarà necessario impostarne uno ora.

Esistono alcuni scenari in cui è necessario utilizzare un agente Console distribuito in uno specifico provider cloud:

- Quando si esegue la scansione dei dati in Cloud Volumes ONTAP in AWS o Amazon FSx per i bucket ONTAP , si utilizza un agente Console in AWS.
- Quando si esegue la scansione dei dati in Cloud Volumes ONTAP in Azure o in Azure NetApp Files, si utilizza un agente Console in Azure.
  - Per Azure NetApp Files, è necessario distribuirlo nella stessa area dei volumi che si desidera analizzare.
- Quando si esegue la scansione dei dati in Cloud Volumes ONTAP in GCP, si utilizza un agente Console in GCP.

È possibile eseguire la scansione dei sistemi ONTAP on-premise, delle condivisioni file NetApp e dei database utilizzando uno qualsiasi di questi agenti della console cloud.

Nota che puoi anche "installare l'agente Console in locale" su un host Linux nella tua rete o nel cloud. Alcuni utenti che intendono installare Data Classification in locale potrebbero anche scegliere di installare l'agente Console in locale.

Potrebbero esserci situazioni in cui è necessario utilizzare "più agenti della console" .

La classificazione dei dati non impone limiti alla quantità di dati che può analizzare. Ogni agente della console supporta la scansione e la visualizzazione di 500 TiB di dati. Per scansionare più di 500 TiB di dati,"installare un altro agente Console" Poi"distribuire un'altra istanza di classificazione dei dati" . + L'interfaccia utente della console visualizza i dati da un singolo connettore. Per suggerimenti sulla visualizzazione dei dati da più agenti della console, vedere"Lavora con più agenti della console" .

Supporto regionale del governo

La classificazione dei dati è supportata quando l'agente della console viene distribuito in una regione governativa (AWS GovCloud, Azure Gov o Azure DoD). Quando implementata in questo modo, la classificazione dei dati presenta le seguenti restrizioni:

"Scopri come distribuire l'agente Console in una regione governativa".

Prerequisiti

Esaminare i seguenti prerequisiti per assicurarsi di disporre di una configurazione supportata prima di distribuire Data Classification nel cloud. Quando si distribuisce Data Classification nel cloud, questa si trova nella stessa subnet dell'agente Console.

Abilita l'accesso a Internet in uscita dalla classificazione dei dati

La classificazione dei dati richiede l'accesso a Internet in uscita. Se la rete virtuale o fisica utilizza un server proxy per l'accesso a Internet, assicurarsi che l'istanza di Data Classification disponga di accesso a Internet in uscita per contattare i seguenti endpoint. La delega non deve essere trasparente. I proxy trasparenti non sono attualmente supportati.

Consultare la tabella appropriata qui sotto a seconda che si stia distribuendo la classificazione dei dati in AWS, Azure o GCP.

Endpoint richiesti per AWS

Punti finali	Scopo
\ https://api.console.netapp.com	Comunicazione con il servizio Console, che include gli account NetApp .
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	Comunicazione con il sito web della Console per l'autenticazione centralizzata degli utenti.
\ 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/	Fornisce accesso a immagini software, manifesti e modelli.
\ https://kinesis.us-east-1.amazonaws.com	Consente a NetApp di trasmettere in streaming i dati dai record di audit.
\ 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	Consente a Data Classification di accedere e scaricare manifesti e modelli e di inviare registri e metriche.

Endpoint richiesti per Azure

Punti finali	Scopo
\ https://api.console.netapp.com	Comunicazione con il servizio Console, che include gli account NetApp .
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	Comunicazione con il sito web della Console per l'autenticazione centralizzata degli utenti.
\ 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/	Fornisce accesso a immagini software, manifesti, modelli e consente di inviare log e metriche.
\ https://support.compliance.api.console.netapp.com/	Consente a NetApp di trasmettere in streaming i dati dai record di audit.

Endpoint richiesti per GCP

Punti finali	Scopo
\ https://api.console.netapp.com	Comunicazione con il servizio Console, che include gli account NetApp .
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	Comunicazione con il sito web della Console per l'autenticazione centralizzata degli utenti.

Punti finali	Scopo
\ 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/	Fornisce accesso a immagini software, manifesti, modelli e consente di inviare log e metriche.
\ https://support.compliance.api.console.netapp.com/	Consente a NetApp di trasmettere in streaming i dati dai record di audit.

Assicurarsi che la classificazione dei dati disponga delle autorizzazioni richieste

Assicurarsi che Data Classification disponga delle autorizzazioni per distribuire risorse e creare gruppi di sicurezza per l'istanza di Data Classification.

- "Autorizzazioni di Google Cloud"
- "Autorizzazioni AWS"
- "Autorizzazioni di Azure"

Assicurarsi che l'agente della console possa accedere alla classificazione dei dati

Garantire la connettività tra l'agente della console e l'istanza di classificazione dei dati. Il gruppo di sicurezza per l'agente Console deve consentire il traffico in entrata e in uscita sulla porta 443 da e verso l'istanza di classificazione dei dati. Questa connessione consente la distribuzione dell'istanza di classificazione dei dati e consente di visualizzare le informazioni nelle schede Conformità e Governance. La classificazione dei dati è supportata nelle regioni governative in AWS e Azure.

Per le distribuzioni AWS e AWS GovCloud sono necessarie regole aggiuntive per i gruppi di sicurezza in entrata e in uscita. Vedere "Regole per l'agente della console in AWS" per i dettagli.

Per le distribuzioni di Azure e Azure Government sono necessarie regole aggiuntive per i gruppi di sicurezza in entrata e in uscita. Vedere "Regole per l'agente Console in Azure" per i dettagli.

Assicurati di poter mantenere in esecuzione la classificazione dei dati

L'istanza di classificazione dei dati deve rimanere attiva per analizzare continuamente i dati.

Assicurare la connettività del browser Web alla classificazione dei dati

Dopo aver abilitato la classificazione dei dati, assicurarsi che gli utenti accedano all'interfaccia della console da un host che abbia una connessione all'istanza di classificazione dei dati.

L'istanza di classificazione dei dati utilizza un indirizzo IP privato per garantire che i dati indicizzati non siano accessibili da Internet. Di conseguenza, il browser Web utilizzato per accedere alla Console deve disporre di una connessione a tale indirizzo IP privato. Tale connessione può provenire da una connessione diretta al tuo provider cloud (ad esempio, una VPN) oppure da un host che si trova all'interno della stessa rete dell'istanza di classificazione dei dati.

Controlla i limiti della tua vCPU

Assicurati che il limite di vCPU del tuo provider cloud consenta la distribuzione di un'istanza con il numero necessario di core. Sarà necessario verificare il limite di vCPU per la famiglia di istanze pertinente nella regione in cui è in esecuzione la Console. "Visualizza i tipi di istanza richiesti" .

Per maggiori dettagli sui limiti vCPU, consultare i seguenti xref:./* "Documentazione AWS: quote di servizio Amazon EC2"
* "Documentazione di Azure: quote vCPU delle macchine virtuali"
* "Documentazione di Google Cloud: Quote di risorse"

Distribuisci la classificazione dei dati nel cloud

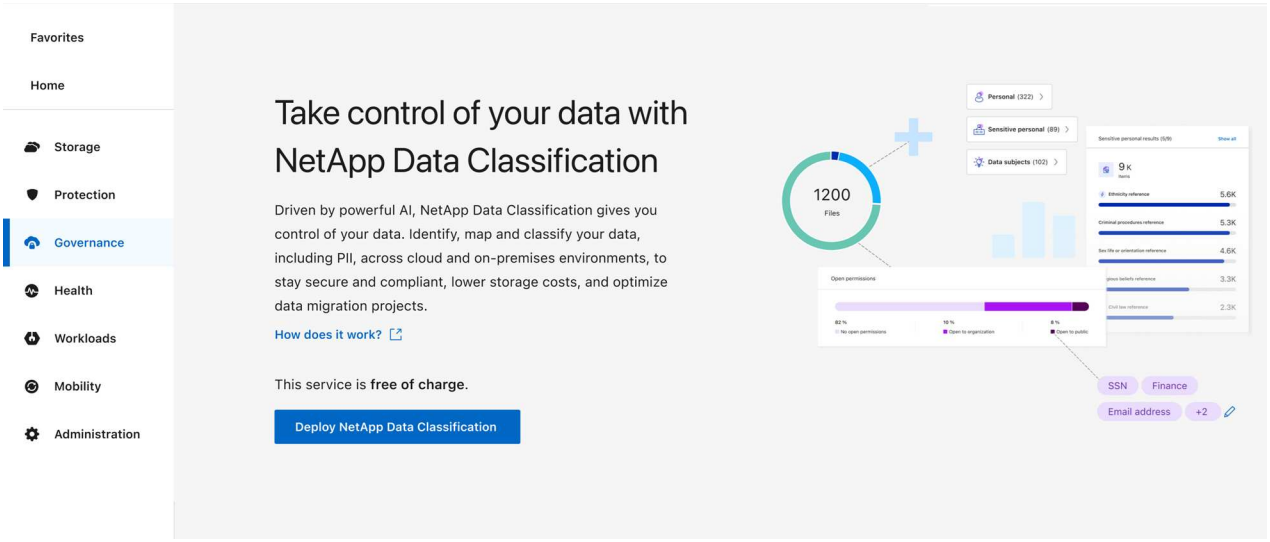
Per distribuire un'istanza di Data Classification nel cloud, seguire questi passaggi. L'agente della console distribuirà l'istanza nel cloud e quindi installerà il software di classificazione dei dati su tale istanza.

Nelle regioni in cui il tipo di istanza predefinito non è disponibile, la classificazione dei dati viene eseguita su un"tipo di istanza alternativo" .

Distribuisci in AWS

Passi

1. Dalla pagina principale di Data Classification, seleziona **Distribuisci classificazione in locale o nel cloud**.

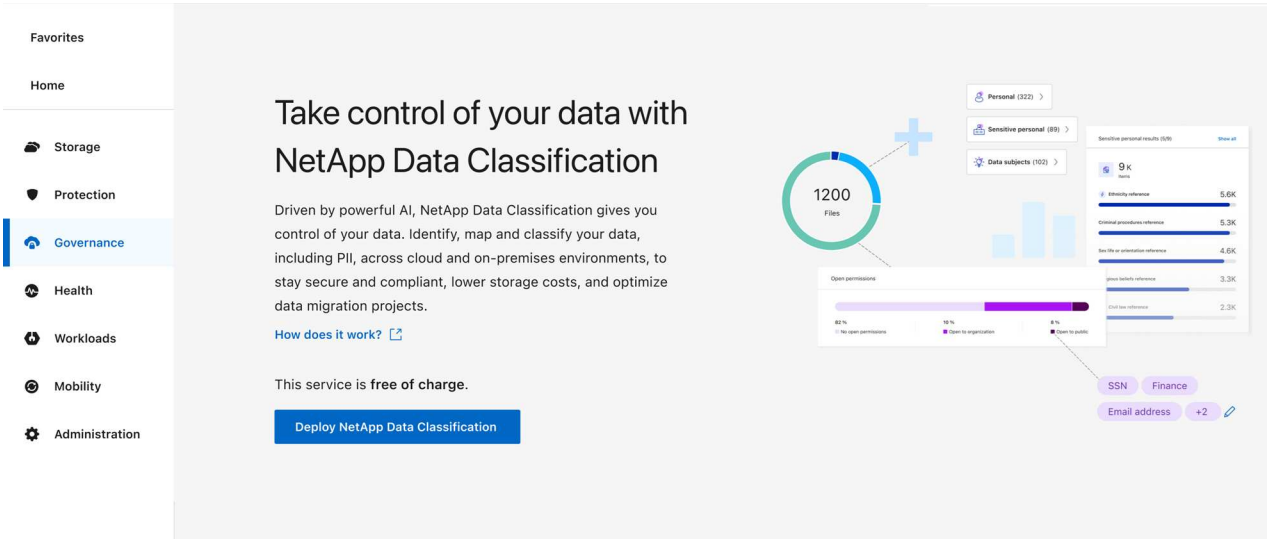


2. Dalla pagina *Installazione*, seleziona **Distribuisci > Distribuisci** per utilizzare la dimensione dell'istanza "Grande" e avviare la procedura guidata di distribuzione cloud.
3. La procedura guidata visualizza l'avanzamento dei passaggi di distribuzione. Quando sono richiesti input o se si verificano problemi, viene visualizzato un messaggio.
4. Una volta distribuita l'istanza e installata la classificazione dei dati, selezionare **Continua alla configurazione** per andare alla pagina *Configurazione*.

Distribuisci in Azure

Passi

1. Dalla pagina principale di Data Classification, seleziona **Distribuisci classificazione in locale o nel cloud**.



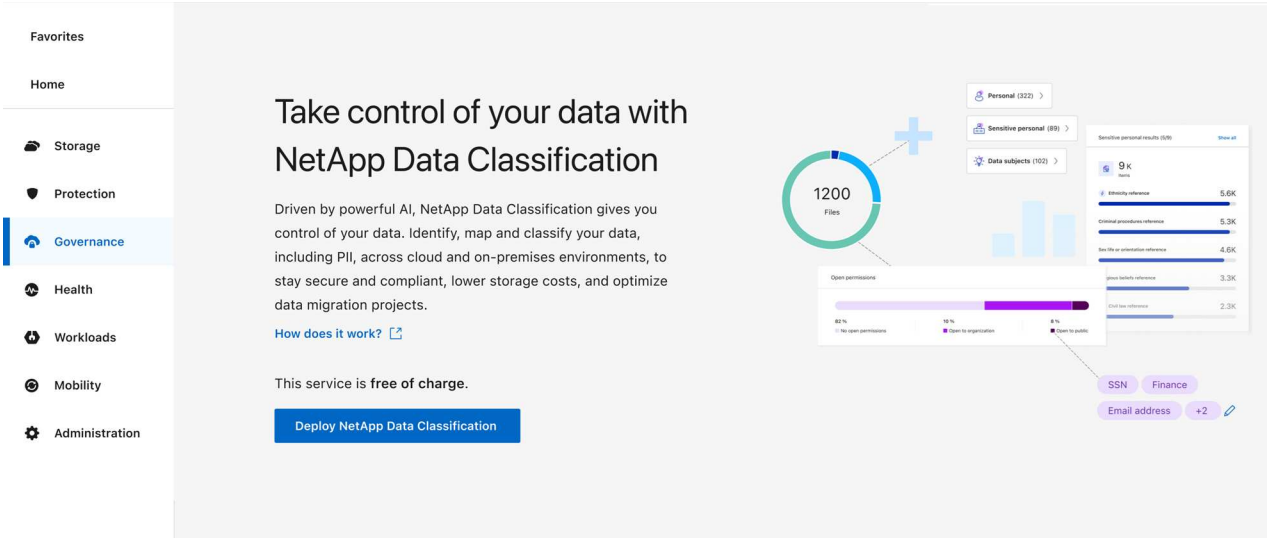
2. Selezionare **Distribuisci** per avviare la procedura guidata di distribuzione cloud.

3. La procedura guidata visualizza l'avanzamento dei passaggi di distribuzione. Se riscontra qualche problema, si fermerà e chiederà un input.
4. Una volta distribuita l'istanza e installata la classificazione dei dati, selezionare **Continua alla configurazione** per andare alla pagina *Configurazione*.

Distribuisci in Google Cloud

Passi

1. Dalla pagina principale di Data Classification, selezionare **Governance > Classificazione**.
2. Selezionare **Distribuisci classificazione in locale o nel cloud**.



3. Selezionare **Distribuisci** per avviare la procedura guidata di distribuzione cloud.
4. La procedura guidata visualizza l'avanzamento dei passaggi di distribuzione. Se riscontra qualche problema, si fermerà e chiederà un input.
5. Una volta distribuita l'istanza e installata la classificazione dei dati, selezionare **Continua alla configurazione** per andare alla pagina *Configurazione*.

Risultato

La Console distribuisce l'istanza di classificazione dei dati nel tuo provider cloud.

Gli aggiornamenti all'agente della console e al software di classificazione dei dati sono automatizzati, a condizione che le istanze dispongano di connettività Internet.

Cosa succederà dopo?

Dalla pagina Configurazione è possibile selezionare le origini dati che si desidera analizzare.

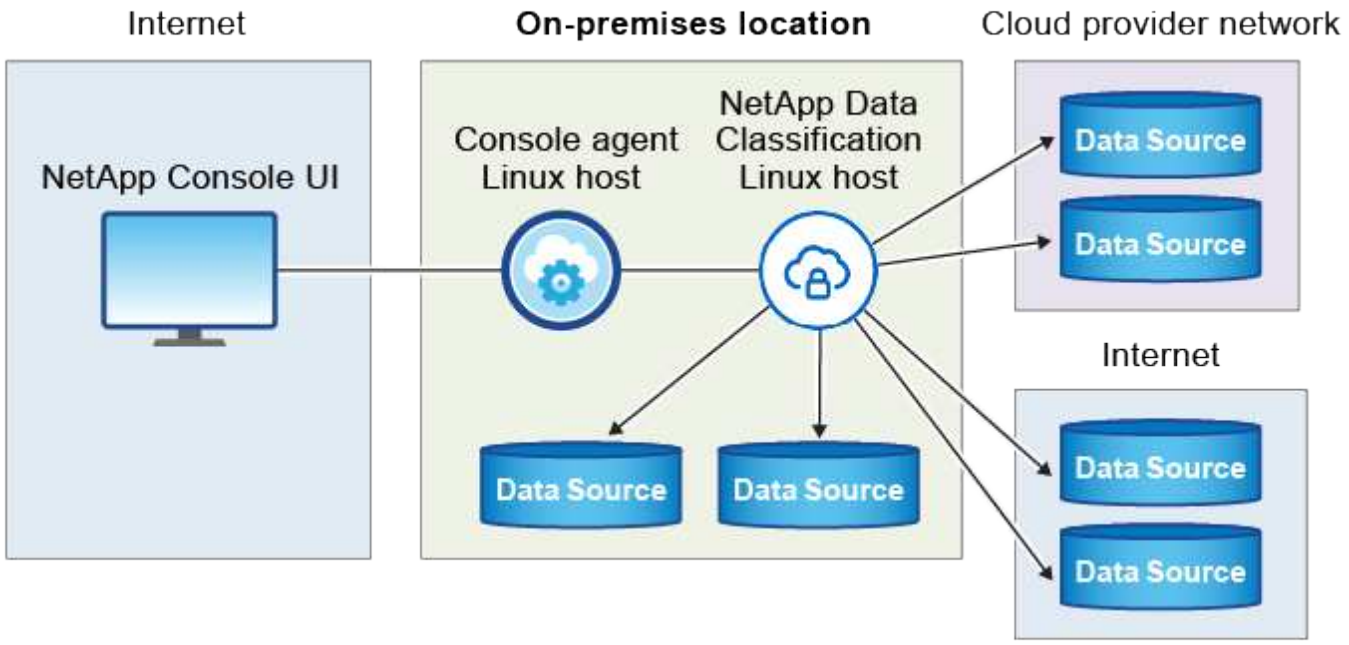
Installa NetApp Data Classification su un host con accesso a Internet

Per distribuire NetApp Data Classification su un host Linux nella tua rete o su un host Linux nel cloud con accesso a Internet, devi distribuire manualmente l'host Linux nella tua rete o nel cloud.

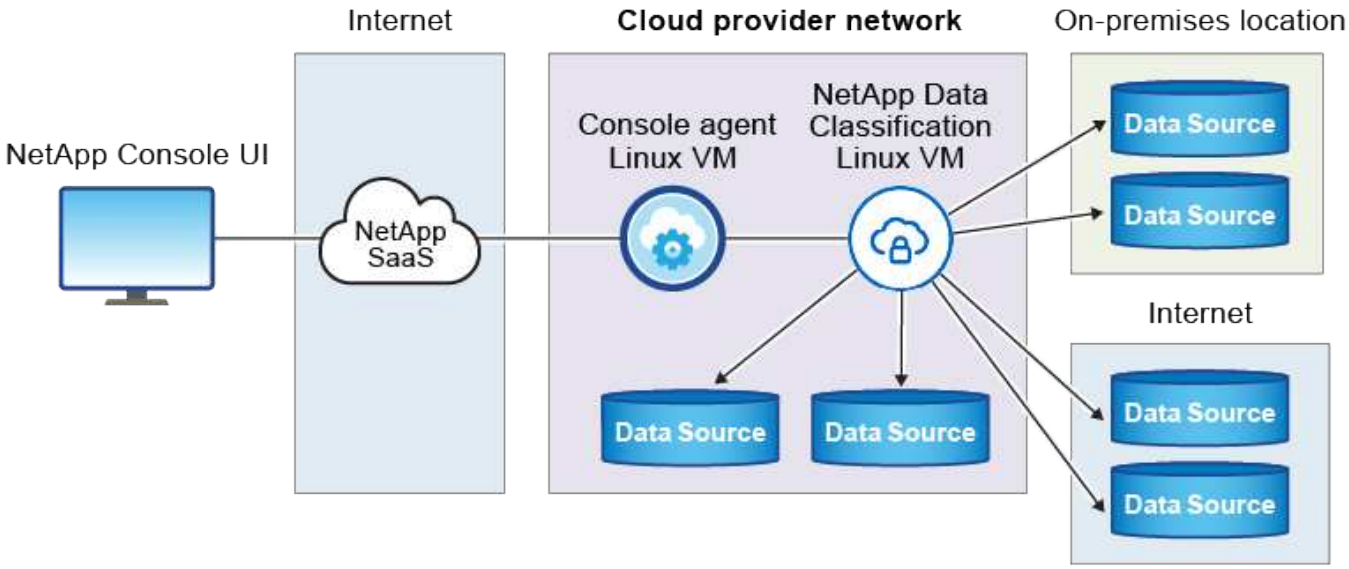
L'installazione in sede è una buona opzione se si preferisce eseguire la scansione dei sistemi ONTAP in sede utilizzando un'istanza di Data Classification anch'essa in sede. Questo non è un requisito. Il software funziona allo stesso modo indipendentemente dal metodo di installazione scelto.

Lo script di installazione di Data Classification inizia verificando se il sistema e l'ambiente soddisfano i prerequisiti richiesti. Se tutti i prerequisiti sono soddisfatti, l'installazione avrà inizio. Se si desidera verificare i prerequisiti indipendentemente dall'esecuzione dell'installazione di Data Classification, è possibile scaricare un pacchetto software separato che verifica solo i prerequisiti. "Scopri come verificare se il tuo host Linux è pronto per installare Data Classification" .

L'installazione tipica su un host Linux *nei tuoi locali* presenta i seguenti componenti e connessioni.



L'installazione tipica su un host Linux *nel cloud* presenta i seguenti componenti e connessioni.



Avvio rapido

Inizia subito seguendo questi passaggi oppure scorri verso il basso fino alle sezioni rimanenti per i dettagli completi.

1 Creare un agente Console

Se non hai ancora un agente Console, "distribuire l'agente della console in locale" su un host Linux nella tua rete o su un host Linux nel cloud.

Puoi anche creare un agente Console con il tuo provider cloud. Vedere "creazione di un agente Console in AWS" , "creazione di un agente Console in Azure" , O "creazione di un agente Console in GCP" .

2 Rivedere i prerequisiti

Assicurati che il tuo ambiente soddisfi i prerequisiti. Ciò include l'accesso a Internet in uscita per l'istanza, la connettività tra l'agente della console e la classificazione dei dati sulla porta 443 e altro ancora. Vedi l'elenco completo .

Hai anche bisogno di un sistema Linux che soddisfi i requisitiseguenti requisiti .

3 Scarica e distribuisci la classificazione dei dati

Scarica il software Cloud Data Classification dal sito di supporto NetApp e copia il file di installazione sull'host Linux che intendi utilizzare. Quindi avviare la procedura guidata di installazione e seguire le istruzioni per distribuire l'istanza di Data Classification.

Creare un agente Console

Per poter installare e utilizzare Data Classification è necessario un agente Console. Nella maggior parte dei casi, probabilmente avrai configurato un agente Console prima di tentare di attivare la classificazione dei dati perché la maggior parte "Le funzionalità della console richiedono un agente della console" , ma ci sono casi in cui sarà necessario impostarne uno ora.

Per crearne uno nell'ambiente del tuo provider cloud, vedi "creazione di un agente Console in AWS" , "creazione di un agente Console in Azure" , O "creazione di un agente Console in GCP" .

Esistono alcuni scenari in cui è necessario utilizzare un agente Console distribuito in uno specifico provider cloud:

- Quando si esegue la scansione dei dati in Cloud Volumes ONTAP in AWS o Amazon FSx per ONTAP, si utilizza un agente Console in AWS.
- Quando si esegue la scansione dei dati in Cloud Volumes ONTAP in Azure o in Azure NetApp Files, si utilizza un agente Console in Azure.

  Per Azure NetApp Files, è necessario distribuirlo nella stessa area dei volumi che si desidera analizzare.

- Quando si esegue la scansione dei dati in Cloud Volumes ONTAP in GCP, si utilizza un agente Console in GCP.

I sistemi ONTAP on-premise, le condivisioni file NetApp e gli account di database possono essere scansionati utilizzando uno qualsiasi di questi agenti della console cloud.

Nota che puoi anche "distribuire l'agente della console in locale" su un host Linux nella tua rete o su un host Linux nel cloud. Alcuni utenti che intendono installare Data Classification in locale potrebbero anche scegliere di installare l'agente Console in locale.

Durante l'installazione di Data Classification sarà necessario l'indirizzo IP o il nome host del sistema agente della console. Queste informazioni saranno disponibili se hai installato l'agente Console nella tua sede. Se l'agente della console è distribuito nel cloud, è possibile trovare queste informazioni nella console: selezionare l'icona della Guida, quindi **Supporto** e infine **Agente della console**.

Preparare il sistema host Linux

Il software di classificazione dei dati deve essere eseguito su un host che soddisfi specifici requisiti del sistema operativo, requisiti di RAM, requisiti software e così via. L'host Linux può trovarsi nella tua rete o nel cloud.

Assicurarsi di poter mantenere in esecuzione la classificazione dei dati. La macchina di classificazione dei dati deve rimanere accesa per analizzare continuamente i dati.

- La classificazione dei dati deve avvenire su un host dedicato. L'host non può essere condiviso con altre applicazioni o software di terze parti, come gli antivirus.
- Scegli la dimensione più adatta al set di dati che intendi analizzare con Data Classification.

Dimensioni del sistema	processore	RAM (la memoria di swap deve essere disabilitata)	Disco
Extra Large	32 CPU	128 GB di RAM	• SSD da 1 TiB su /, oppure 100 GiB disponibili su /opt • 895 GiB disponibili su /var/lib/docker • 5 GiB su /tmp • **Per Podman, 30 GB su /var/tmp**
Grande	16 CPU	64 GB di RAM	• SSD da 500 GiB su /, oppure 100 GiB disponibili su /opt • 400 GiB disponibili su /var/lib/docker o per Podman /var/lib/containers • 5 GiB su /tmp • **Per Podman, 30 GB su /var/tmp**

- Quando si distribuisce un'istanza di elaborazione nel cloud per l'installazione di Data Classification, si consiglia di utilizzare un sistema che soddisfi i requisiti di sistema "Large" sopra indicati:
  - **Tipo di istanza Amazon Elastic Compute Cloud (Amazon EC2)**: "m6i.4xlarge". "Vedi altri tipi di istanze AWS" .
  - **Dimensioni della VM di Azure**: "Standard_D16s_v3". "Visualizza altri tipi di istanze di Azure" .
  - **Tipo di macchina GCP**: "n2-standard-16". "Vedi altri tipi di istanza GCP" .
- **Autorizzazioni cartella UNIX**: sono richieste le seguenti autorizzazioni UNIX minime:

Cartella	Permessi minimi
/tmp	`rwxrwxrwt`
/optare	`rwxr-xr-x`
/var/lib/docker	`rwx------`
/usr/lib/systemd/sistema	`rwxr-xr-x`

- **Sistema operativo**:
  - I seguenti sistemi operativi richiedono l'utilizzo del motore container Docker:
    - Red Hat Enterprise Linux versione 7.8 e 7.9
    - Ubuntu 22.04 (richiede Data Classification versione 1.23 o successiva)
    - Ubuntu 24.04 (richiede Data Classification versione 1.23 o successiva)
  - I seguenti sistemi operativi richiedono l'utilizzo del motore contenitore Podman e la versione 1.30 o successiva di Data Classification:
    - Red Hat Enterprise Linux versione 8.8, 8.10, 9.0, 9.1, 9.2, 9.3, 9.4, 9.5 e 9.6.
  - Le estensioni vettoriali avanzate (AVX2) devono essere abilitate sul sistema host.
- **Red Hat Subscription Management**: l'host deve essere registrato presso Red Hat Subscription Management. Se non è registrato, il sistema non può accedere ai repository per aggiornare il software di terze parti richiesto durante l'installazione.
- **Software aggiuntivo**: è necessario installare il seguente software sull'host prima di installare Data Classification:
  - A seconda del sistema operativo utilizzato, è necessario installare uno dei seguenti motori container:
    - Docker Engine versione 19.3.1 o successiva. "Visualizza le istruzioni di installazione" .
    - Podman versione 4 o successiva. Per installare Podman, inserisci(`sudo yum install podman netavark -y`).
- Python versione 3.6 o successiva. "Visualizza le istruzioni di installazione" .
  - **Considerazioni su NTP**: NetApp consiglia di configurare il sistema di classificazione dei dati per utilizzare un servizio Network Time Protocol (NTP). L'ora deve essere sincronizzata tra il sistema di classificazione dei dati e il sistema agente della console.
- **Considerazioni su Firewalld**: se si prevede di utilizzare `firewalld` , ti consigliamo di abilitarlo prima di installare Data Classification. Eseguire i seguenti comandi per configurare `firewalld` in modo che sia compatibile con la classificazione dei dati:

```
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
```

Se si prevede di utilizzare host di classificazione dati aggiuntivi come nodi scanner, aggiungere subito queste regole al sistema primario:

```
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
```

Tieni presente che devi riavviare Docker o Podman ogni volta che abiliti o aggiorni `firewalld` impostazioni.

L'indirizzo IP del sistema host di classificazione dei dati non può essere modificato dopo l'installazione.

Abilita l'accesso a Internet in uscita dalla classificazione dei dati

La classificazione dei dati richiede l'accesso a Internet in uscita. Se la rete virtuale o fisica utilizza un server proxy per l'accesso a Internet, assicurarsi che l'istanza di Data Classification disponga di accesso a Internet in uscita per contattare i seguenti endpoint.

Punti finali	Scopo
\ https://api.console.netapp.com	Comunicazione con la Console, che include gli account NetApp .
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	Comunicazione con il sito web della Console per l'autenticazione centralizzata degli utenti.
\ 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/	Fornisce accesso a immagini software, manifesti, modelli e consente di inviare log e metriche.
\ https://support.compliance.api.bluexp.netapp.com/	Consente a NetApp di trasmettere in streaming i dati dai record di audit.
\ https://github.com/docker \ https://download.docker.com	Fornisce i pacchetti prerequisiti per l'installazione di Docker.
\ http://packages.ubuntu.com/ \ http://archive.ubuntu.com	Fornisce i pacchetti prerequisiti per l'installazione di Ubuntu.

Verificare che tutte le porte richieste siano abilitate

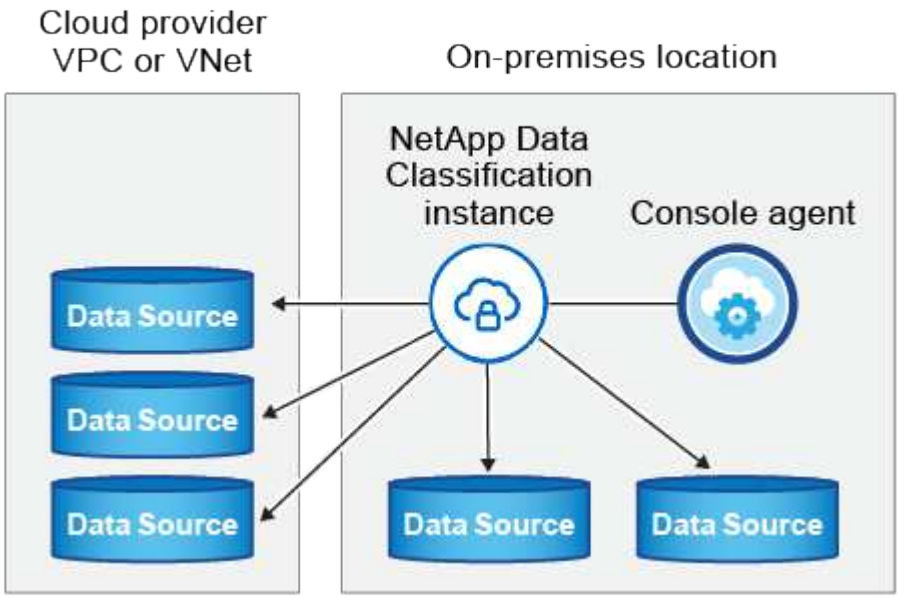
È necessario assicurarsi che tutte le porte necessarie siano aperte per la comunicazione tra l'agente della console, Data Classification, Active Directory e le origini dati.

Tipo di connessione	porti	Descrizione
Agente console <> Classificazione dati	8080 (TCP), 443 (TCP) e 80. 9000	Le regole del firewall o di routing per l'agente della console devono consentire il traffico in entrata e in uscita sulla porta 443 da e verso l'istanza di classificazione dei dati. Assicurati che la porta 8080 sia aperta in modo da poter visualizzare l'avanzamento dell'installazione nella Console. Se sull'host Linux viene utilizzato un firewall, per i processi interni di un server Ubuntu è richiesta la porta 9000.
Agente console <> cluster ONTAP (NAS)	443 (TCP)	La console rileva i cluster ONTAP tramite HTTPS. Se si utilizzano criteri firewall personalizzati, questi devono soddisfare i seguenti requisiti: • L'host dell'agente della console deve consentire l'accesso HTTPS in uscita tramite la porta 443. Se l'agente della console si trova nel cloud, tutte le comunicazioni in uscita sono consentite dalle regole di routing o dal firewall predefiniti. • Il cluster ONTAP deve consentire l'accesso HTTPS in entrata tramite la porta 443. Il criterio firewall predefinito "mgmt" consente l'accesso HTTPS in entrata da tutti gli indirizzi IP. Se hai modificato questa policy predefinita o se hai creato una policy firewall personalizzata, devi associare il protocollo HTTPS a tale policy e abilitare l'accesso dall'host dell'agente della console.
Classificazione dei dati <> cluster ONTAP	• Per NFS - 111 (TCP\UDP) e 2049 (TCP\UDP) • Per CIFS - 139 (TCP\UDP) e 445 (TCP\UDP)	La classificazione dei dati necessita di una connessione di rete a ciascuna subnet Cloud Volumes ONTAP o sistema ONTAP locale. I firewall o le regole di routing per Cloud Volumes ONTAP devono consentire le connessioni in ingresso dall'istanza di classificazione dei dati. Assicurarsi che queste porte siano aperte all'istanza di classificazione dei dati: • Per NFS - 111 e 2049 • Per CIFS - 139 e 445 I criteri di esportazione del volume NFS devono consentire l'accesso dall'istanza di classificazione dei dati.

Tipo di connessione	porti	Descrizione
Classificazione dei dati <> Active Directory	389 (TCP e UDP), 636 (TCP), 3268 (TCP) e 3269 (TCP)	È necessario che sia già stata configurata una Active Directory per gli utenti della propria azienda. Inoltre, la classificazione dei dati necessita delle credenziali di Active Directory per analizzare i volumi CIFS. È necessario disporre delle informazioni per Active Directory: • Indirizzo IP del server DNS o più indirizzi IP • Nome utente e password per il server • Nome di dominio (nome di Active Directory) • Se stai utilizzando LDAP sicuro (LDAPS) o meno • Porta del server LDAP (in genere 389 per LDAP e 636 per LDAP sicuro)

Installa Data Classification sull'host Linux

Nelle configurazioni tipiche, il software verrà installato su un singolo sistema host. Guarda i passaggi qui .



VederePreparazione del sistema host Linux ERevisione dei prerequisiti per l'elenco completo dei requisiti prima di implementare Data Classification.

Gli aggiornamenti al software di classificazione dei dati sono automatizzati, a condizione che l'istanza disponga di connettività Internet.

Al momento, Data Classification non è in grado di analizzare bucket S3, Azure NetApp Files o FSx per ONTAP quando il software è installato in locale. In questi casi sarà necessario distribuire un agente Console separato e un'istanza di Data Classification nel cloud e "passare da un connettore all'altro" per le tue diverse fonti di dati.

Installazione a host singolo per configurazioni tipiche

Esaminare i requisiti e seguire questi passaggi quando si installa il software di classificazione dei dati su un singolo host locale.

"Guarda questo video"per vedere come installare Data Classification.

Si noti che tutte le attività di installazione vengono registrate durante l'installazione di Data Classification. Se si verificano problemi durante l'installazione, è possibile visualizzare il contenuto del registro di controllo dell'installazione. È scritto a `/opt/netapp/install_logs/` .

Prima di iniziare

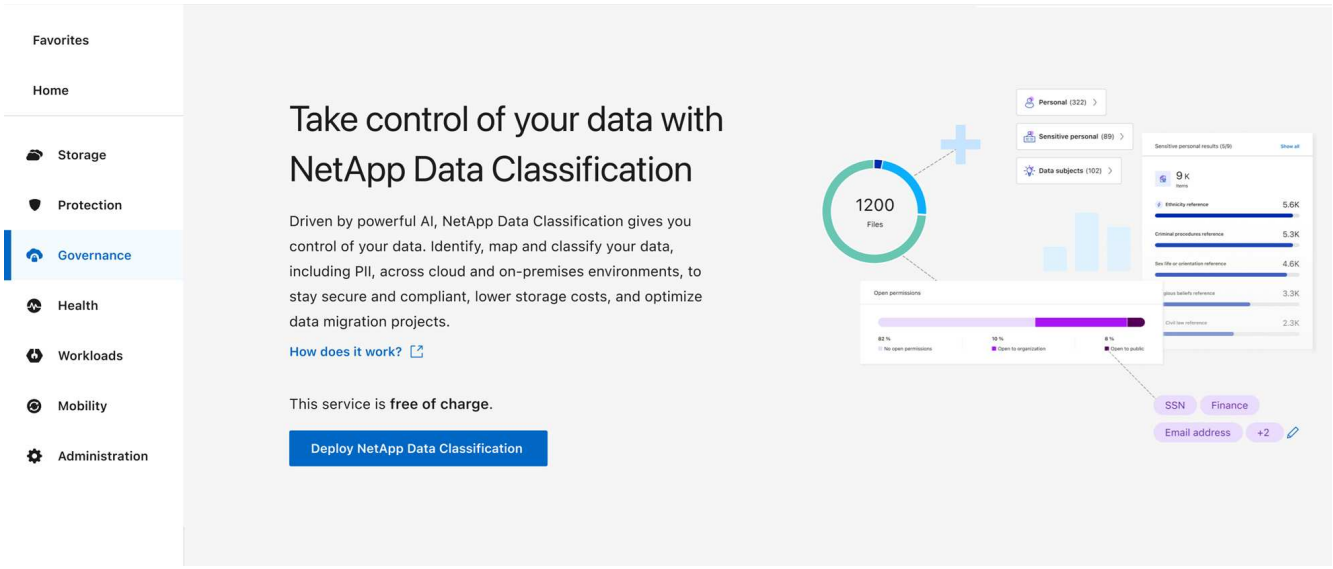
- Verifica che il tuo sistema Linux soddisfi i requisitirequisiti dell'host .
- Verificare che nel sistema siano installati i due pacchetti software prerequisiti (Docker Engine o Podman e Python 3).
- Assicurati di avere i privilegi di root sul sistema Linux.
- Se utilizzi un proxy per accedere a Internet:
  - Avrai bisogno delle informazioni sul server proxy (indirizzo IP o nome host, porta di connessione, schema di connessione: https o http, nome utente e password).
  - Se il proxy esegue l'intercettazione TLS, è necessario conoscere il percorso sul sistema Data Classification Linux in cui sono archiviati i certificati TLS CA.
  - La delega non deve essere trasparente. Attualmente la classificazione dei dati non supporta proxy trasparenti.
  - L'utente deve essere un utente locale. Gli utenti di dominio non sono supportati.
- Verifica che il tuo ambiente offline soddisfi i requisiti richiestipermessi e connettività .

Passi

1. Scarica il software di classificazione dei dati da "Sito di supporto NetApp" . Il file da selezionare si chiama **DATASENSE-INSTALLER-<versione>.tar.gz**.
2. Copia il file di installazione sull'host Linux che intendi utilizzare (utilizzando `scp` o qualche altro metodo).
3. Decomprimere il file di installazione sul computer host, ad esempio:

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

4. Nella Console, seleziona **Governance > Classificazione**.
5. Selezionare **Distribuisci classificazione in locale o nel cloud**.



6. A seconda che si stia installando Data Classification su un'istanza preparata nel cloud o su un'istanza preparata in sede, selezionare l'opzione **Distribuisci** appropriata per avviare l'installazione di Data Classification.
7. Viene visualizzata la finestra di dialogo *Distribuisci classificazione dati in locale*. Copia il comando fornito (ad esempio: `sudo ./install.sh -a 12345 -c 27AG75 -t 2198qq`) e incollalo in un file di testo in modo da poterlo utilizzare in seguito. Quindi seleziona **Chiudi** per chiudere la finestra di dialogo.
8. Sulla macchina host, immetti il comando che hai copiato e segui una serie di prompt, oppure puoi fornire il comando completo, inclusi tutti i parametri richiesti, come argomenti della riga di comando.

   Tieni presente che il programma di installazione esegue un controllo preliminare per assicurarsi che i requisiti di sistema e di rete siano soddisfatti per un'installazione corretta. "Guarda questo video" per comprendere i messaggi e le implicazioni del pre-controllo.

Inserire i parametri come richiesto:	Inserisci il comando completo:
a. Incolla il comando che hai copiato dal passaggio 7: `sudo ./install.sh -a <account_id> -c <client_id> -t <user_token>` Se stai installando su un'istanza cloud (non nei tuoi locali), aggiungi `--manual-cloud -install <cloud_provider>`. b. Immettere l'indirizzo IP o il nome host della macchina host di classificazione dei dati in modo che sia accessibile al sistema agente della console. c. Immettere l'indirizzo IP o il nome host della macchina host dell'agente Console in modo che sia accessibile al sistema di classificazione dei dati. d. Inserisci i dettagli del proxy come richiesto. Se l'agente della console utilizza già un proxy, non è necessario immettere nuovamente queste informazioni qui, poiché la classificazione dei dati utilizzerà automaticamente il proxy utilizzato dall'agente della console.	In alternativa, è possibile creare l'intero comando in anticipo, fornendo i parametri host e proxy necessari: `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>`

Valori variabili:

- *account_id* = ID account NetApp
- *client_id* = ID client dell'agente della console (aggiungere il suffisso "client" all'ID client se non è già presente)
- *user_token* = token di accesso utente JWT
- *ds_host* = Indirizzo IP o nome host del sistema Linux di classificazione dei dati.
- *cm_host* = Indirizzo IP o nome host del sistema agente della console.
- *cloud_provider* = Quando si esegue l'installazione su un'istanza cloud, immettere "AWS", "Azure" o "Gcp" a seconda del provider cloud.
- *proxy_host* = IP o nome host del server proxy se l'host si trova dietro un server proxy.
- *proxy_port* = Porta per connettersi al server proxy (predefinita 80).
- *proxy_scheme* = Schema di connessione: https o http (predefinito http).
- *proxy_user* = Utente autenticato per connettersi al server proxy, se è richiesta l'autenticazione di base. L'utente deve essere un utente locale: gli utenti di dominio non sono supportati.
- *proxy_password* = Password per il nome utente specificato.
- *ca_cert_dir* = Percorso sul sistema Linux di classificazione dei dati contenente bundle di certificati TLS CA aggiuntivi. Richiesto solo se il proxy esegue l'intercettazione TLS.

Risultato

Il programma di installazione di Data Classification installa i pacchetti, registra l'installazione e installa Data

Classification. L'installazione può richiedere dai 10 ai 20 minuti.

Se è presente connettività sulla porta 8080 tra la macchina host e l'istanza dell'agente Console, l'avanzamento dell'installazione verrà visualizzato nella scheda Classificazione dati nella Console.

Cosa succederà dopo?

Dalla pagina Configurazione è possibile selezionare le origini dati che si desidera analizzare.

Installa NetApp Data Classification su un host Linux senza accesso a Internet

L'installazione di NetApp Data Classification su un host Linux in un sito locale che non dispone di accesso a Internet è nota come *modalità privata*. Questo tipo di installazione, che utilizza uno script di installazione, non ha connettività con il livello SaaS NetApp Console .

> La modalità privata BlueXP (interfaccia BlueXP legacy) viene in genere utilizzata con ambienti locali privi di connessione Internet e con regioni cloud sicure, tra cui AWS Secret Cloud, AWS Top Secret Cloud e Azure IL6. NetApp continua a supportare questi ambienti con l'interfaccia legacy BlueXP . Per la documentazione sulla modalità privata nell'interfaccia legacy BlueXP , vedere"Documentazione PDF per la modalità privata BlueXP" .

Verifica che il tuo host Linux sia pronto per installare NetApp Data Classification

Prima di installare manualmente NetApp Data Classification su un host Linux, è possibile eseguire uno script sull'host per verificare che siano soddisfatti tutti i prerequisiti per l'installazione di Data Classification. Puoi eseguire questo script su un host Linux nella tua rete o su un host Linux nel cloud. L'host può essere connesso a Internet oppure risiedere in un sito che non ha accesso a Internet (un *dark site*).

Lo script di installazione di Data Classification comprende uno script di test per garantire che l'ambiente soddisfi i requisiti. È possibile eseguire questo script separatamente per verificare la disponibilità dell'host Linux prima di eseguire lo script di installazione.

Iniziare

Dovrai svolgere le seguenti attività.

- Facoltativamente, installa un agente Console se non ne hai già installato uno. È possibile eseguire lo script di test senza avere installato un agente Console, ma lo script verifica la connettività tra l'agente Console e la macchina host di Data Classification, pertanto è consigliabile disporre di un agente Console.
- Preparare la macchina host e verificare che soddisfi tutti i requisiti.
- Abilitare l'accesso a Internet in uscita dalla macchina host di classificazione dei dati.
- Verificare che tutte le porte richieste siano abilitate su tutti i sistemi.
- Scarica ed esegui lo script di test dei prerequisiti.

Creare un agente Console

Per poter installare e utilizzare Data Classification è necessario un agente Console. Tuttavia, è possibile eseguire lo script Prerequisiti senza un agente Console.

Puoi "installare l'agente Console in locale" su un host Linux nella tua rete o su un host Linux nel cloud. È anche possibile installare Data Classification in locale se l'agente Console è installato in locale.

Per creare un agente Console nell'ambiente del tuo provider cloud, consulta:

- "creazione di un agente Console in AWS"
- "creazione di un agente Console in Azure"
- "creazione di un agente Console in GCP"

Quando si esegue lo script dei prerequisiti, è necessario l'indirizzo IP o il nome host del sistema agente della console. Queste informazioni sono disponibili se hai installato l'agente Console nei tuoi locali. Se l'agente della Console è distribuito nel cloud, è possibile trovare queste informazioni nella Console: selezionare l'icona della Guida, quindi **Supporto**; nella sezione Agente e audit, selezionare **Vai all'agente**.

Verifica i requisiti dell'host

Il software di classificazione dei dati deve essere eseguito su un host che soddisfi specifici requisiti del sistema operativo, requisiti di RAM e requisiti software.

- La classificazione dei dati deve avvenire su un host dedicato. L'host non può essere condiviso con altre applicazioni o software di terze parti, come gli antivirus.
- Scegli la dimensione più adatta al set di dati che intendi analizzare con Data Classification.

Dimensioni del sistema	processore	RAM (la memoria di swap deve essere disabilitata)	Disco
Extra Large	32 CPU	128 GB di RAM	• SSD da 1 TiB su /, oppure 100 GiB disponibili su /opt • 895 GiB disponibili su /var/lib/docker • 5 GiB su /tmp • **Per Podman, 30 GB su /var/tmp**
Grande	16 CPU	64 GB di RAM	• SSD da 500 GiB su /, oppure 100 GiB disponibili su /opt • 400 GiB disponibili su /var/lib/docker o per Podman /var/lib/containers • 5 GiB su /tmp • **Per Podman, 30 GB su /var/tmp**

- Quando si distribuisce un'istanza di elaborazione nel cloud per l'installazione di Data Classification, si

consiglia di utilizzare un sistema che soddisfi i requisiti di sistema "Large" sopra indicati:

  - **Tipo di istanza Amazon Elastic Compute Cloud (Amazon EC2)**: "m6i.4xlarge". "Vedi altri tipi di istanze AWS" .
  - **Dimensioni della VM di Azure**: "Standard_D16s_v3". "Visualizza altri tipi di istanze di Azure" .
  - **Tipo di macchina GCP**: "n2-standard-16". "Vedi altri tipi di istanza GCP" .

- **Autorizzazioni cartella UNIX**: sono richieste le seguenti autorizzazioni UNIX minime:

Cartella	Permessi minimi
/tmp	`rwxrwxrwt`
/optare	`rwxr-xr-x`
/var/lib/docker	`rwx------`
/usr/lib/systemd/sistema	`rwxr-xr-x`

- **Sistema operativo**:
  - I seguenti sistemi operativi richiedono l'utilizzo del motore container Docker:
    - Red Hat Enterprise Linux versione 7.8 e 7.9
    - Ubuntu 22.04 (richiede Data Classification versione 1.23 o successiva)
    - Ubuntu 24.04 (richiede Data Classification versione 1.23 o successiva)
  - I seguenti sistemi operativi richiedono l'utilizzo del motore contenitore Podman e la versione 1.30 o successiva di Data Classification:
    - Red Hat Enterprise Linux versione 8.8, 8.10, 9.0, 9.1, 9.2, 9.3, 9.4, 9.5 e 9.6.
  - Le estensioni vettoriali avanzate (AVX2) devono essere abilitate sul sistema host.
- **Red Hat Subscription Management**: l'host deve essere registrato presso Red Hat Subscription Management. Se non è registrato, il sistema non può accedere ai repository per aggiornare il software di terze parti richiesto durante l'installazione.
- **Software aggiuntivo**: è necessario installare il seguente software sull'host prima di installare Data Classification:
  - A seconda del sistema operativo utilizzato, è necessario installare uno dei seguenti motori container:
    - Docker Engine versione 19.3.1 o successiva. "Visualizza le istruzioni di installazione" .
    - Podman versione 4 o successiva. Per installare Podman, inserisci(`sudo yum install podman netavark -y`).
- Python versione 3.6 o successiva. "Visualizza le istruzioni di installazione" .
  - **Considerazioni su NTP**: NetApp consiglia di configurare il sistema di classificazione dei dati per utilizzare un servizio Network Time Protocol (NTP). L'ora deve essere sincronizzata tra il sistema di classificazione dei dati e il sistema agente della console.
- **Considerazioni su Firewalld**: se si prevede di utilizzare `firewalld` , ti consigliamo di abilitarlo prima di installare Data Classification. Eseguire i seguenti comandi per configurare `firewalld` in modo che sia compatibile con la classificazione dei dati:

```
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
```

Se si prevede di utilizzare host di classificazione dati aggiuntivi come nodi scanner (in un modello distribuito), aggiungere subito queste regole al sistema primario:

```
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
```

Tieni presente che devi riavviare Docker o Podman ogni volta che abiliti o aggiorni `firewalld` impostazioni.

Abilita l'accesso a Internet in uscita dalla classificazione dei dati

La classificazione dei dati richiede l'accesso a Internet in uscita. Se la rete virtuale o fisica utilizza un server proxy per l'accesso a Internet, assicurarsi che l'istanza di Data Classification disponga di accesso a Internet in uscita per contattare i seguenti endpoint.

Questa sezione non è richiesta per i sistemi host installati in siti senza connettività Internet.

Punti finali	Scopo
\ https://api.console.netapp.com	Comunicazione con il servizio Console, che include gli account NetApp .
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	Comunicazione con il sito web della Console per l'autenticazione centralizzata degli utenti.
\ 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/	Fornisce accesso a immagini software, manifesti, modelli e consente di inviare log e metriche.
\ https://support.compliance.api.console.netapp.com/	Consente a NetApp di trasmettere in streaming i dati dai record di audit.
\ https://github.com/docker \ https://download.docker.com	Fornisce i pacchetti prerequisiti per l'installazione di Docker.

Punti finali	Scopo
\ http://packages.ubuntu.com/ \ http://archive.ubuntu.com	Fornisce i pacchetti prerequisiti per l'installazione di Ubuntu.

Verificare che tutte le porte richieste siano abilitate

È necessario assicurarsi che tutte le porte necessarie siano aperte per la comunicazione tra l'agente della console, Data Classification, Active Directory e le origini dati.

Tipo di connessione	porti	Descrizione
Agente console <> Classificazione dati	8080 (TCP), 443 (TCP) e 80. 9000	Le regole del firewall o di routing per l'agente della console devono consentire il traffico in entrata e in uscita sulla porta 443 da e verso l'istanza di classificazione dei dati. Assicurati che la porta 8080 sia aperta in modo da poter visualizzare l'avanzamento dell'installazione nella Console. Se sull'host Linux viene utilizzato un firewall, per i processi interni di un server Ubuntu è richiesta la porta 9000.
Agente console <> cluster ONTAP (NAS)	443 (TCP)	La console rileva i cluster ONTAP tramite HTTPS. Se si utilizzano criteri firewall personalizzati, l'host dell'agente della console deve consentire l'accesso HTTPS in uscita tramite la porta 443. Se l'agente della console si trova nel cloud, tutte le comunicazioni in uscita sono consentite dalle regole di routing o dal firewall predefiniti.

Eseguire lo script dei prerequisiti per la classificazione dei dati

Per eseguire lo script dei prerequisiti per la classificazione dei dati, seguire questi passaggi.

"Guarda questo video"per vedere come eseguire lo script Prerequisiti e interpretare i risultati.

Prima di iniziare

- Verifica che il tuo sistema Linux soddisfi i requisitirequisiti dell'host .
- Verificare che nel sistema siano installati i due pacchetti software prerequisiti (Docker Engine o Podman e Python 3).
- Assicurati di avere i privilegi di root sul sistema Linux.

Passi

1. Scarica lo script dei prerequisiti per la classificazione dei dati da "Sito di supporto NetApp" . Il file da selezionare si chiama **standalone-pre-requisite-tester-<versione>**.
2. Copia il file sull'host Linux che intendi utilizzare (utilizzando `scp` o qualche altro metodo).
3. Assegnare le autorizzazioni per eseguire lo script.

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

4. Eseguire lo script utilizzando il seguente comando.

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

   Aggiungere l'opzione "--darksite" solo se si esegue lo script su un host che non ha accesso a Internet. Alcuni test preliminari vengono saltati quando l'host non è connesso a Internet.

5. Lo script richiede l'indirizzo IP della macchina host di classificazione dei dati.
   - Immettere l'indirizzo IP o il nome host.
6. Lo script chiede se è installato un agente Console.
   - Immettere **N** se non è installato un agente Console.
   - Inserisci **Y** se hai un agente Console installato. Quindi immettere l'indirizzo IP o il nome host dell'agente della console in modo che lo script di test possa testare questa connettività.
7. Lo script esegue una serie di test sul sistema e ne visualizza i risultati man mano che procede. Quando termina, scrive un registro della sessione in un file denominato `prerequisites-test-<timestamp>.log` nella directory `/opt/netapp/install_logs` .

Risultato

Se tutti i test dei prerequisiti sono stati eseguiti correttamente, puoi installare Data Classification sull'host quando sei pronto.

Se vengono rilevati problemi, questi vengono classificati come "Consigliati" o "Obbligatori" per essere risolti. I problemi consigliati sono in genere elementi che potrebbero rallentare le attività di scansione e categorizzazione della classificazione dei dati. Non è necessario correggere questi elementi, ma potresti volerli risolvere.

Se si verificano problemi "obbligatori", è necessario risolverli ed eseguire nuovamente lo script di test dei prerequisiti.