



Flusso d'aria Apache

NetApp Solutions

NetApp
May 10, 2024

Sommario

- Flusso d'aria Apache 1
- Implementazione di Apache Airflow 1
- USA il toolkit DataOps di NetApp con Airflow 4

Flusso d'aria Apache

Implementazione di Apache Airflow

Questa sezione descrive le attività da completare per implementare il flusso d'aria nel cluster Kubernetes.



È possibile implementare il flusso d'aria su piattaforme diverse da Kubernetes. L'implementazione del flusso d'aria su piattaforme diverse da Kubernetes non rientra nell'ambito di questa soluzione.

Prerequisiti

Prima di eseguire l'esercizio di implementazione descritto in questa sezione, si presuppone che siano già state eseguite le seguenti attività:

1. Hai già un cluster Kubernetes funzionante.
2. Hai già installato e configurato NetApp Astra Trident nel tuo cluster Kubernetes. Per ulteriori informazioni su Astra Trident, fare riferimento alla ["Documentazione di Astra Trident"](#).

Installare Helm

Il flusso d'aria viene implementato utilizzando Helm, un popolare gestore di pacchetti per Kubernetes. Prima di implementare il flusso d'aria, è necessario installare Helm sull'host di distribuzione jump. Per installare Helm sull'host di distribuzione jump, seguire la ["istruzioni per l'installazione"](#) Nella documentazione ufficiale di Helm.

Impostare la classe di storage Kubernetes predefinita

Prima di implementare il flusso d'aria, è necessario specificare un StorageClass predefinito all'interno del cluster Kubernetes. Il processo di implementazione del flusso d'aria tenta di eseguire il provisioning di nuovi volumi persistenti utilizzando la classe di storage predefinita. Se non viene indicato StorageClass come StorageClass predefinito, l'implementazione non riesce. Per designare una StorageClass predefinita all'interno del cluster, segui le istruzioni riportate nella ["Implementazione di Kubeflow"](#) sezione. Se è già stata designata una StorageClass predefinita all'interno del cluster, è possibile saltare questo passaggio.

USA Helm per implementare il flusso d'aria

Per implementare il flusso d'aria nel cluster Kubernetes utilizzando Helm, eseguire le seguenti operazioni dall'host di distribuzione jump:

1. Implementare il flusso d'aria utilizzando Helm seguendo il ["istruzioni per l'implementazione"](#) Per il diagramma ufficiale del flusso d'aria sull'Artifact Hub. I comandi di esempio che seguono mostrano l'implementazione del flusso d'aria con Helm. Modificare, aggiungere e/o rimuovere i valori in `custom-values.yaml` file in base alle necessità, a seconda dell'ambiente e della configurazione desiderata.

```
$ cat << EOF > custom-values.yaml
#####
# Airflow - Common Configs
#####
```

```

airflow:
  ## the airflow executor type to use
  ##
  executor: "CeleryExecutor"
  ## environment variables for the web/scheduler/worker Pods (for
  airflow configs)
  ##
  #
  #####
  # Airflow - WebUI Configs
  #####
web:
  ## configs for the Service of the web Pods
  ##
  service:
    type: NodePort
  #####
  # Airflow - Logs Configs
  #####
  logs:
    persistence:
      enabled: true
  #####
  # Airflow - DAGs Configs
  #####
dags:
  ## configs for the DAG git repository & sync container
  ##
  gitSync:
    enabled: true
    ## url of the git repository
    ##
    repo: "git@github.com:mboglesby/airflow-dev.git"
    ## the branch/tag/sha1 which we clone
    ##
    branch: master
    revision: HEAD
    ## the name of a pre-created secret containing files for ~/.ssh/
    ##
    ## NOTE:
    ## - this is ONLY RELEVANT for SSH git repos
    ## - the secret commonly includes files: id_rsa, id_rsa.pub,
    known_hosts
    ## - known_hosts is NOT NEEDED if `git.sshKeyscan` is true
    ##
    sshSecret: "airflow-ssh-git-secret"

```

```

## the name of the private key file in your `git.secret`
##
## NOTE:
## - this is ONLY RELEVANT for PRIVATE SSH git repos
##
sshSecretKey: id_rsa
## the git sync interval in seconds
##
syncWait: 60
EOF
$ helm install airflow airflow-stable/airflow -n airflow --version 8.0.8
--values ./custom-values.yaml
...
Congratulations. You have just deployed Apache Airflow!
1. Get the Airflow Service URL by running these commands:
  export NODE_PORT=$(kubectl get --namespace airflow -o
  jsonpath="{.spec.ports[0].nodePort}" services airflow-web)
  export NODE_IP=$(kubectl get nodes --namespace airflow -o
  jsonpath="{.items[0].status.addresses[0].address}")
  echo http://$NODE_IP:$NODE_PORT/
2. Open Airflow in your web browser

```

2. Verificare che tutti i pod del flusso d'aria siano in funzione. L'avvio di tutti i pod potrebbe richiedere alcuni minuti.

```

$ kubectl -n airflow get pod

```

NAME	READY	STATUS	RESTARTS	AGE
airflow-flower-b5656d44f-h8qjk	1/1	Running	0	2h
airflow-postgresql-0	1/1	Running	0	2h
airflow-redis-master-0	1/1	Running	0	2h
airflow-scheduler-9d95fcd9-clf4b	2/2	Running	2	2h
airflow-web-59c94db9c5-z7rg4	1/1	Running	0	2h
airflow-worker-0	2/2	Running	2	2h

3. Ottenere l'URL del servizio Web Airflow seguendo le istruzioni stampate sulla console quando si implementa Airflow utilizzando Helm nel passaggio 1.

```

$ export NODE_PORT=$(kubectl get --namespace airflow -o
  jsonpath="{.spec.ports[0].nodePort}" services airflow-web)
$ export NODE_IP=$(kubectl get nodes --namespace airflow -o
  jsonpath="{.items[0].status.addresses[0].address}")
$ echo http://$NODE_IP:$NODE_PORT/

```

4. Verificare che sia possibile accedere al servizio Web Airflow.

The screenshot shows the Airflow web interface with the 'DAGs' tab selected. The interface includes a search bar and a table listing various DAGs. The table columns are: DAG (with an info icon), Schedule, Owner, Recent Tasks (with an info icon), Last Run (with an info icon), DAG Runs (with an info icon), and Links. The DAGs listed include 'ai_training_run', 'create_data_scientist_workspace', 'example_bash_operator', 'example_branch_dop_operator_v3', 'example_branch_operator', 'example_complex', 'example_external_task_marker_child', 'example_external_task_marker_parent', 'example_http_operator', 'example_kubernetes_executor_config', 'example_nested_branch_dag', 'example_passing_params_via_test_command', 'example_pig_operator', 'example_python_operator', 'example_short_circuit_operator', and 'example_skip_dag'.

DAG	Schedule	Owner	Recent Tasks	Last Run	DAG Runs	Links
ai_training_run	None	NetApp				
create_data_scientist_workspace	None	NetApp				
example_bash_operator	0 0 * * *	Airflow				
example_branch_dop_operator_v3	* * * * *	Airflow				
example_branch_operator	@daily	Airflow				
example_complex	None	airflow				
example_external_task_marker_child	None	airflow				
example_external_task_marker_parent	None	airflow				
example_http_operator	1 day, 00:00	Airflow				
example_kubernetes_executor_config	None	Airflow				
example_nested_branch_dag	@daily	airflow				
example_passing_params_via_test_command	* * * * *	airflow				
example_pig_operator	None	Airflow				
example_python_operator	None	Airflow				
example_short_circuit_operator	1 day, 00:00	Airflow				
example_skip_dag	1 day, 00:00	Airflow				

USA il toolkit DataOps di NetApp con Airflow

Il ["NetApp DataOps Toolkit per Kubernetes"](#) Utilizzabile in combinazione con il flusso d'aria. L'utilizzo del toolkit NetApp DataOps con flusso d'aria consente di integrare le operazioni di gestione dei dati NetApp, come la creazione di snapshot e cloni, in workflow automatizzati orchestrati da flussi d'aria.

Fare riferimento a ["Esempi di flusso d'aria"](#) Sezione all'interno del repository GitHub del toolkit DataOps di NetApp per i dettagli sull'utilizzo del toolkit con flusso d'aria.

Informazioni sul copyright

Copyright © 2024 NetApp, Inc. Tutti i diritti riservati. Stampato negli Stati Uniti d'America. Nessuna porzione di questo documento soggetta a copyright può essere riprodotta in qualsiasi formato o mezzo (grafico, elettronico o meccanico, inclusi fotocopie, registrazione, nastri o storage in un sistema elettronico) senza previo consenso scritto da parte del detentore del copyright.

Il software derivato dal materiale sottoposto a copyright di NetApp è soggetto alla seguente licenza e dichiarazione di non responsabilità:

IL PRESENTE SOFTWARE VIENE FORNITO DA NETAPP "COSÌ COM'È" E SENZA QUALSIVOGLIA TIPO DI GARANZIA IMPLICITA O ESPRESSA FRA CUI, A TITOLO ESEMPLIFICATIVO E NON ESAUSTIVO, GARANZIE IMPLICITE DI COMMERCIALIZZABILITÀ E IDONEITÀ PER UNO SCOPO SPECIFICO, CHE VENGONO DECLINATE DAL PRESENTE DOCUMENTO. NETAPP NON VERRÀ CONSIDERATA RESPONSABILE IN ALCUN CASO PER QUALSIVOGLIA DANNO DIRETTO, INDIRETTO, ACCIDENTALE, SPECIALE, ESEMPLARE E CONSEGUENZIALE (COMPRESI, A TITOLO ESEMPLIFICATIVO E NON ESAUSTIVO, PROCUREMENT O SOSTITUZIONE DI MERCI O SERVIZI, IMPOSSIBILITÀ DI UTILIZZO O PERDITA DI DATI O PROFITTI OPPURE INTERRUZIONE DELL'ATTIVITÀ AZIENDALE) CAUSATO IN QUALSIVOGLIA MODO O IN RELAZIONE A QUALUNQUE TEORIA DI RESPONSABILITÀ, SIA ESSA CONTRATTUALE, RIGOROSA O DOVUTA A INSOLVENZA (COMPRESA LA NEGLIGENZA O ALTRO) INSORTA IN QUALSIASI MODO ATTRAVERSO L'UTILIZZO DEL PRESENTE SOFTWARE ANCHE IN PRESENZA DI UN PREAVVISO CIRCA L'EVENTUALITÀ DI QUESTO TIPO DI DANNI.

NetApp si riserva il diritto di modificare in qualsiasi momento qualunque prodotto descritto nel presente documento senza fornire alcun preavviso. NetApp non si assume alcuna responsabilità circa l'utilizzo dei prodotti o materiali descritti nel presente documento, con l'eccezione di quanto concordato espressamente e per iscritto da NetApp. L'utilizzo o l'acquisto del presente prodotto non comporta il rilascio di una licenza nell'ambito di un qualche diritto di brevetto, marchio commerciale o altro diritto di proprietà intellettuale di NetApp.

Il prodotto descritto in questa guida può essere protetto da uno o più brevetti degli Stati Uniti, esteri o in attesa di approvazione.

LEGENDA PER I DIRITTI SOTTOPOSTI A LIMITAZIONE: l'utilizzo, la duplicazione o la divulgazione da parte degli enti governativi sono soggetti alle limitazioni indicate nel sottoparagrafo (b)(3) della clausola Rights in Technical Data and Computer Software del DFARS 252.227-7013 (FEB 2014) e FAR 52.227-19 (DIC 2007).

I dati contenuti nel presente documento riguardano un articolo commerciale (secondo la definizione data in FAR 2.101) e sono di proprietà di NetApp, Inc. Tutti i dati tecnici e il software NetApp forniti secondo i termini del presente Contratto sono articoli aventi natura commerciale, sviluppati con finanziamenti esclusivamente privati. Il governo statunitense ha una licenza irrevocabile limitata, non esclusiva, non trasferibile, non cedibile, mondiale, per l'utilizzo dei Dati esclusivamente in connessione con e a supporto di un contratto governativo statunitense in base al quale i Dati sono distribuiti. Con la sola esclusione di quanto indicato nel presente documento, i Dati non possono essere utilizzati, divulgati, riprodotti, modificati, visualizzati o mostrati senza la previa approvazione scritta di NetApp, Inc. I diritti di licenza del governo degli Stati Uniti per il Dipartimento della Difesa sono limitati ai diritti identificati nella clausola DFARS 252.227-7015(b) (FEB 2014).

Informazioni sul marchio commerciale

NETAPP, il logo NETAPP e i marchi elencati alla pagina <http://www.netapp.com/TM> sono marchi di NetApp, Inc. Gli altri nomi di aziende e prodotti potrebbero essere marchi dei rispettivi proprietari.