



Apache 气流 NetApp Solutions

NetApp
May 10, 2024

目录

Apache 气流	1
Apache Airflow 部署	1
使用NetApp数据操作工具包和Airflow	4

Apache 气流

Apache Airflow 部署

本节介绍在 Kubernetes 集群中部署气流时必须完成的任务。



可以在 Kubernetes 以外的平台上部署 Airflow。在 Kubernetes 以外的平台上部署气流不在此解决方案的范围内。

前提条件

在执行本节所述的部署练习之前，我们假定您已执行以下任务：

1. 您已有一个工作正常的 Kubernetes 集群。
2. 您已在 Kubernetes 集群中安装并配置 NetApp Astra 三端存储。有关 Astra 三项功能的更多详细信息，请参见 "[Astra Trident 文档](#)"。

安装 Helm

Airflow 可使用 Kubernetes 常用的软件包管理器 Helm 进行部署。在部署气流之前，必须在部署跳转主机上安装 Helm。要在部署跳转主机上安装 Helm，请按照 "[安装说明](#)" 在官方 Helm 文档中。

设置默认 Kubernetes StorageClass

在部署 Airflow 之前，您必须在 Kubernetes 集群中指定一个默认 StorageClass。气流部署过程会尝试使用默认 StorageClass 配置新的永久性卷。如果未将任何 StorageClass 指定为默认 StorageClass，则部署将失败。要在集群中指定默认 StorageClass，请按照中所述的说明进行操作 "[Kubeflow 部署](#)" 部分。如果已在集群中指定默认 StorageClass，则可以跳过此步骤。

使用 Helm 部署气流

要使用 Helm 在 Kubernetes 集群中部署气流，请从部署跳转主机执行以下任务：

1. 按照说明使用 Helm 部署气流 "[部署说明](#)" 用于 Artifact Hub 上的官方气流图表。下面的示例命令显示了如何使用 Helm 部署气流。根据您的环境和所需配置，根据需要修改，添加和 / 或删除 custom-values.yaml 文件中的值。

```
$ 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. 确认所有气流 Pod 均已启动且正在运行。所有 POD 可能需要几分钟的时间才能启动。

```

$ 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-9d95fcdf9-clf4b	2/2	Running	2	2h
airflow-web-59c94db9c5-z7rg4	1/1	Running	0	2h
airflow-worker-0	2/2	Running	2	2h

3. 按照步骤 1 中使用 Helm 部署 Airflow 时控制台上印有的说明获取 Airflow Web 服务 URL。

```

$ 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. 确认您可以访问 Airflow Web 服务。

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, 0:00:00	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, 0:00:00	Airflow				
example_skip_dag	1 day, 0:00:00	Airflow				

使用NetApp数据操作工具包和Airflow

。 "适用于Kubernetes的NetApp DataOps工具包" 可与气流结合使用。通过将NetApp数据操作工具包与Airflow结合使用、您可以将NetApp数据管理操作(例如创建快照和克隆)整合到由Airflow编排的自动化 workflows 中。

请参见 "气流示例" 部分、NetApp了解有关在Airflow中使用此工具包的详细信息。

版权信息

版权所有 © 2024 NetApp, Inc.。保留所有权利。中国印刷。未经版权所有者事先书面许可，本档中受版权保护的任何部分不得以任何形式或通过任何手段（图片、电子或机械方式，包括影印、录音、录像或存储在电子检索系统中）进行复制。

从受版权保护的 NetApp 资料派生的软件受以下许可和免责声明的约束：

本软件由 NetApp 按“原样”提供，不含任何明示或暗示担保，包括但不限于适销性以及针对特定用途的适用性的隐含担保，特此声明不承担任何责任。在任何情况下，对于因使用本软件而以任何方式造成的任何直接性、间接性、偶然性、特殊性、惩罚性或后果性损失（包括但不限于购买替代商品或服务；使用、数据或利润方面的损失；或者业务中断），无论原因如何以及基于何种责任理论，无论出于合同、严格责任或侵权行为（包括疏忽或其他行为），NetApp 均不承担责任，即使已被告知存在上述损失的可能性。

NetApp 保留在不另行通知的情况下随时对本文档所述的任何产品进行更改的权利。除非 NetApp 以书面形式明确同意，否则 NetApp 不承担因使用本文档所述产品而产生的任何责任或义务。使用或购买本产品不表示获得 NetApp 的任何专利权、商标权或任何其他知识产权许可。

本手册中描述的产品可能受一项或多项美国专利、外国专利或正在申请的专利的保护。

有限权利说明：政府使用、复制或公开本文档受 DFARS 252.227-7013（2014 年 2 月）和 FAR 52.227-19（2007 年 12 月）中“技术数据权利 — 非商用”条款第 (b)(3) 条规定的限制条件的约束。

本文档中所含数据与商业产品和/或商业服务（定义见 FAR 2.101）相关，属于 NetApp, Inc. 的专有信息。根据本协议提供的所有 NetApp 技术数据和计算机软件具有商业性质，并完全由私人出资开发。美国政府对这些数据的使用权具有非排他性、全球性、受限且不可撤销的许可，该许可既不可转让，也不可再许可，但仅限在与交付数据所依据的美国政府合同有关且受合同支持的情况下使用。除本文档规定的情形外，未经 NetApp, Inc. 事先书面批准，不得使用、披露、复制、修改、操作或显示这些数据。美国政府对国防部的授权仅限于 DFARS 的第 252.227-7015(b)（2014 年 2 月）条款中明确的权利。

商标信息

NetApp、NetApp 标识和 <http://www.netapp.com/TM> 上所列的商标是 NetApp, Inc. 的商标。其他公司和产品名称可能是其各自所有者的商标。