Google Cloud

Cloud Composer Airflow Summit 2022 Workshop

May 27, 2022

Before we begin - download slides from

https://bit.ly/composer-workshop-2022-airflow-summit



Agenda	01
Introduction to Cloud Composer	02
Setting up	03
Implementation of DAGs	04
DAG troubleshooting on Cloud Composer	05



Hi! It's nice to meet you!



Leah Cole Developer Relations Engineer, NYC



Bartosz Jankiewicz Engineering Manager, Warsaw



Filip Knapik Group Product Manager, Warsaw



Przemek Więch Software Engineer, Warsaw



01

Agenda Of workshop



Agenda

1h

Setting up Cloud Composer

- Setting up
- Architecture overview
- The first DAG running

1h

Implementation of DAGs.

- DAG development use case in Google Cloud Platform
- Data Analytics use case

1h

DAG Troubleshooting

- Autoscaling showcase
- Scheduling too fast
- Permissions
 troubleshooting
- Notifications on DAG failure
- Retries





Introduction to Cloud Composer





Create a Google Cloud account

https://cloud.google.com/

Troubleshooting: https://support.google.com/accounts/answer/27441

Create Google Cloud account associated with your email

≡ Google Account Help Q Describe your issue

makes it much more likely you'il get your account back.

Add a recovery phone number ☑

Add a recovery email address ☑





Update your information





Redeem your voucher

Prerequisite: Redeeming a voucher assumes that you have: a Google Account and you set up your first GCP project with a billing account.

- Acquire a GCP Voucher that you received via email • (if you didn't receive the email then please connect to your workshop instructor).
- Follow the link: https://cloud.google.com/redeem#<YOUR VOUCHER> •
- The following screen should be presented to you. • Press "Continue" button.

The following screen should be presented to you. ٠ Press "REDEEM" button.

Redeem promotion code

BBBB-AAAA-BBBB-AAAA	
Billing account Private Billing Account	•

The credit will be applied to the account you select here.

Redeem promotion code

\$300.00

Good toward the following service: all of Google Cloud Platform.

Expires May 19, 2023.

By redeeming this coupon, you, on behalf of yourself and the organization you represent ("You") agree to these terms and conditions.





CANCEL CONTINUE

Enable Cloud Composer API

https://cloud.google.com/composer/docs/composer-2/enable-comp oser-service#enable-api

Google Cloud Platform 🐉 Airflow Summit 2022 workshop 👻	
Cloud Composer API Google Enterprise API Manages Apache Airflow environments on Google Clour	d Platform.
ENABLE TRY THIS API	
OVERVIEW DOCUMENTATION	
Overview Manages Apache Airflow environments on Google Cloud Platform.	Additional details Type: SaaS & APIs Last updated: 4/30/22 Categor; Google Enterprise APIs Service name: composer.googleapis.com
Tutorials and documentation	
Learn more 2	
Terms of Service	

license: Google Cloud Platform.



03

Setting up Cloud Composer



GCP Projects used during Workshop

During this workshop you will be using two types of GCP projects:

 Main activities/exercises will be done in composer-workshop-X projects. These projects were pre-set up for you.
 Each of you should have access to one composer-workshop project
 These projects will be deleted after the workshop.

As part of this workshop you will receive a GCP credits voucher worth \$300.
 To be able to redeem the credits, in a addition to active Google Account you will need to set up your
 GCP project and associate it with <u>an active billing account</u>.

This project is owned by you. You can use it after the workshop for as long as you like. Workshop's GCP credits are valid for 1 year since activation.



Creating Cloud Composer

- 1. Create **service account**.
 - a. Goto IAM -> Service Accounts
 - b. Click + CREATE SERVICE ACCOUNT on top of the page
 - c. Add Composer worker permission to the account.

			v ocu
	IAM & Admin	Service accounts + CREATE SERVICE ACCOUN	NT 👕
	+ ● IAM	Service accounts for project "composer-worksh	hop-1"
•	Identity & Organization	A service account represents a Google Cloud service identity, such as	code runnin
	 Policy Troubleshooter 	Organization policies can be used to secure service accounts and bloc	ck risky serv
	Policy Analyzer	Filter Enter property name or value	
	Organization Policies	Email	
	Service Accounts	□ 🖻 composer-sa@composer-workshop-	
 Service accouncil Grant this ser (optional) 	unt details vice account access to	project	
Service accou Grant this ser (optional)	unt details rvice account access to count access to composer-worksh	project	
 Service accou Grant this service accouncil Grant this service accouncil<td>unt details vice account access to count access to composer-worksh ete specific actions on the resourc</td><td>project</td><td></td>	unt details vice account access to count access to composer-worksh ete specific actions on the resourc	project	
 Service accou Grant this ser (optional) Grant this service account permission to complement Role 	unt details vice account access to count access to composer-worksh ete specific actions on the resourc Condition	1.iam.gserviceaccount.com	
Service accou Grant this ser (optional) Grant this service act permission to complete nore Role Composer Worker	unt details rvice account access to count access to composer-worksh ete specific actions on the resourc Condition Add conditi	1.iam.gserviceaccount.com p project op-1 so that it has is in your project. Learn on	
Service accou Grant this ser (optional) Grant this service acc permission to complete permission to complete nore Role Composer Worker Worker access to Co for service accounts	unt details rvice account access to count access to composer-worksh ete specific actions on the resourc Condition Add conditi imposer. Intended	1.iam.gserviceaccount.com	
Service accou Grant this ser (optional) Grant this service acc permission to complete vore Role Composer Worker Worker access to Co for service accounts	unt details rvice account access to count access to composer-worksh ete specific actions on the resourc Condition Add conditi mposer. Intended	1.iam.gserviceaccount.com	
rvice accou ant this ser otional) nt this service acc nission to comple e e le mposer Worker rker access to Co service accounts - ADD ANOTHER	unt details rvice account access to count access to composer-worksh ete specific actions on the resourc Condition Add conditi mposer. Intended	1.iam.gserviceaccount.com	
 Service account Grant this service account Grant this service account Grant this service account Grant this service account Role Composer Worker Worker access to Conton the service accounts + ADD ANOTHER 	unt details rvice account access to count access to composer-worksh ete specific actions on the resourc Condition Add conditi mposer. Intended	1.iam.gserviceaccount.com	

CANCEL

DONE



Creating Cloud Composer



 When creating environment for the first time grant permissions to Composer product service account.

Grant required permissions to Cloud Composer service account

Cloud Composer relies on <u>Workload Identity</u> as Google API authentication mechanism for Airflow.

In order to support Workload Identity, Cloud Composer creates additional IAM role bindings which requires the <u>Cloud Composer v2</u> PI Service Agent Extension role.

Grant the Cloud Composer v2 API Service Agent Extension role to the service-544473252868@cloudcomposeraccounts.iam.gserviceaccount.com service account.

GRANT



Creating Cloud Composer



- In the Google Cloud console, go to the <u>Create environment</u> page for Cloud Composer 2.
- 4. In the **Name** field, enter a name for your environment. The environment name is used to create subcomponents for the environment, so you must provide a name that is also valid as a Cloud Storage <u>bucket name</u>.
- 5. In the Location drop-down list, choose a location for your environment.

A location is the region where the environment's GKE cluster is located.



Cloud Composer in a nutshell



Orchestrate work across Google Cloud, external SaaS services and proprietary APIs



Fully managed Apache Airflow environments





Cloud Composer benefits





Composer 2 Autoscaling

The challenge

- Sizing environments for variable workloads
- Task failures due to capacity constraints
- Cost control

Efficiency, Scalability, and Simplicity

Autoscaling, and usage-based pricing optimizes costs.



Airflow queue

Google Cloud

Composer versions

	Composer 1	Composer 2
Supported Airflow versions	Airflow 1.10.* and 2.*	Airflow 2.*
<u>Cluster</u>	GKE Standard	GKE Autopilot
Autoscaled Workers	No	Yes
Pricing model	Composer 1 pricing + GKE Pricing	Composer 2 pricing
Private IP Networks	VPC Peering	Private Service Connect VPC Peering
Restore from Snapshots (e.g. for Disaster Recovery)	No	Yes
Web Server Plugins	Not supported with DAG Serialization	Supported



More details:

https://cloud.google.com/composer/docs/concepts/versioning/composer-versioning-overview#major-versions

Cloud Composer 2 architecture

Cloud Composer 2 interacts with the following services:

- CloudSQL running Airflow metadata storage
- Cloud Storage user uploaded content (DAGs, user data)
- Kubernetes runs Scheduler(s), WebServer, Redis queue, SQL proxy and Airflow workloads
- Cloud Logging stores and indexes components logs
- Cloud Monitoring searchable Cloud Composer metrics

... and many more that we are managing for you.





Implementation of DAGs



Quick <u>DAG</u> refresher



- A directed graph with no detected cycles
- In Airflow collection of tasks organized with their relationships and their schedule
- Made up of:
 - \circ Operators
 - Sensors
 - TaskFlow



"""An example DAG demonstrating simple Apache Airflow operators."""

import datetime

```
from airflow import models
from airflow.operators import bash_operator
from airflow.operators import python_operator
default dag args = {
    'start date': datetime.datetime(2022, 5, 24),
}
with models.DAG(
        'composer_sample_simple_greeting',
       schedule interval=datetime.timedelta(days=1),
       default_args=default_dag_args) as dag:
   def greeting():
       import logging
       logging.info('Hello World!')
   hello_python = python_operator.PythonOperator(
       task_id='hello',
       python callable=greeting)
   goodbye_bash = bash_operator.BashOperator(
        task_id='bye',
        bash_command='echo Goodbye.')
   hello python >> goodbye bash
```



Individual Exercise: Upload a DAG to Composer



- 1. <u>Open the DAG</u> using the VSCode GitHub editor
- 2. Download the DAG locally by right clicking the filename

and selecting "Download"

- 3. <u>Use the Console</u> to add your DAG to your environment
- 4. View your DAG in the <u>Airflow UI</u>
- 5. See that the DAG runs automatically



Cloud Composer Ul



Composer UI navigation

	dev 🕶	Search Products, resources, docs (/)
Composer - Environment der	ails 🛛 Open Airflow UI 🖿 Open dags folder 🕀 Save Snapshot 📩 Load Snapshot	T CREFRESH TO DELETE
The latest snapshot of this environment has been	successfully saved in gs://bjankiewicz-state-store/snapshots/bjankiewicz-dev_us-central1_dr-test-21-source_2022-05-17	T1001-23
S dr-test-21-source This environment is runn	ing	
MONITORING LOGS DAGS PREVIEW	ENVIRONMENT CONFIGURATION AIRFLOW CONFIGURATION OVERRIDES ENVIRONMENT VARIABLES	LABELS PYPI PACKAGES
Detail views		Environment actions
		bar
	Notifications	



Monitoring



Monitoring



Exercise:

- 1. Verify memory consumption of the Scheduler
- 2. Find the number of active workers
- 3. Find the CPU consumption of SQL database



Logs

\equiv Google Cloud Platform	• composer-workshop-1 🗸 🔍 Q Search Products, resources, docs (/) V 🔽 😨 💈 🔋
🖁 Composer 🔶 Envir	ronment details 🛛 OPEN AIRFLOW UI 🖿 OPEN DAGS FOLDER 🗄 SAVE SNAPSHOT 📩 LOAD SNAPSHOT 🥂 REFRESH 🍵 DELETE
Stest-bj This environment is runni	ng
MONITORING LOGS DAGS	REVIEW ENVIRONMENT CONFIGURATION AIRFLOW CONFIGURATION OVERRIDES ENVIRONMENT VARIABLES LABELS PYPI PACKAGES
6 HOURS V	Logs Showing 100 log entries Severity Default - Filter logs
✓ All logs	V No LOLL OU TOTTO.UL. MLTTOW HORKOT EXPORTING CHE TOILONING CHE TOILONING CHE TOILONING CHE TOILONING
✓ Airflow logs	2022-05-18T15:02:42.108487343Z airflow-worker AIRFLOW_CTX_DAG_OWNER=airflow
▼ Scheduler 😧	2022-05-18T15:02:42.108495767Z airflow-worker AIRFLOW_CTX_DAG_ID=airflow_monitoring
airflow-scheduler-68fcbb6868-	2022-05-18T15:02:42.108501778Z airflow-worker AIRFLOW_CTX_TASK_ID=echo
5mwzs	2022-05-18T15:02:42.108510664Z airflow-worker AIRFLOW_CTX_EXECUTION_DATE=2022-05-18T15:02:19+00:00
Workers 😮	3 2022-05-18T15:02:42.108522688Z airflow-worker AIRFLOW_CTX_DAG_RUN_ID=manual_2022-05-18T15:02:19+00:00
Web server ?	▶ 🚺 2022-05-18T15:02:42.110012602Z airflow-worker Tmp dir root location:
 DAG processor manager 	2022-05-18T15:02:42.110037291Z airflow-worker /tmp
airflow-scheduler-68fcbb6868-	3 2022-05-18T15:02:42.111084688Z airflow-worker Running command: ['bash', '-c', 'echo test']
5mwzs	3 2022-05-18T15:02:42.265166993Z airflow-worker Output:
 Composer logs 	▶ 3 2022-05-18T15:02:42.265723118Z airflow-worker test
Composer agent	▶ 1 2022-05-18T15:02:42.267144361Z airflow-worker Command exited with return code 0
Build 😮	1 2022-05-18T15:02:42.899141828Z airflow-worker Marking task as SUCCESS. dag_id=airflow_monitoring, task_id=echo, execution_date=20220518T150219, start_date=202205.
Database operations 2	▶ 1 2022-05-18T15:02:43.123801944Z airflow-worker Task exited with return code 0
Monitoring	3 2022-05-18T15:02:43.212301981Z airflow-worker 0 downstream tasks scheduled from follow-on schedule check
 Infrastructure ? 	2022-05-18T15:02:43.509973765Z airflow-worker Task airflow.executors.celery_executor.execute_command[0879a5b4-0760-4e77-a0c4-8100eb4f932f] succeeded in 4.0945245_

Google Cloud

Logs



Exercise:

- 1. Find logs of Webserver
- 2. Find logs of Scheduler
- 3. Find if there are any errors reported in Workers logs



DAG UI

≡ Google (Cloud Plat	form 🐉 composer-work	shop-1 🔻 🔍 🔍	Search Products, resour	ces, docs (/)		~	▶ ?
Compose	er <	Environment details	S 🖸 OPEN AIRFL	.OW UI 👘 OPEN DAGS FO	LDER 🗄 SAVE SNAPSHOT	🛓 LOAD SNAPSHOT	C REFRESH	T DELETE
Stest-bj	This environm	ent is running DAGS PREVIEW EN	VIRONMENT CONFIGURATI	ION AIRFLOW CONFIGU	RATION OVERRIDES ENVIR	ONMENT VARIABLES	LABELS P	YPI PACKAGES
= Filter Filter DAG	Gs					6 hours 12 hours 1 da	y 2 days 4 days	7 days 14 days
DAG id 🕇	State	Description	Schedule interval	Last completed run 💡	Active runs 🔞	Successful runs (1h) 💡	Failed runs (1h) 😧
airflow_monitoring	Active	liveness monitoring dag		1 hour ago	0	10		0

Exercise:

- 1. Navigate to the DAG you have uploaded recently
- 2. Find logs of the tasks for the most recent run of the DAG
- 3. Verify if all the tasks executed successfully



Configuration management



Please note: All configuration management operations are taking up to several minutes to execute (depending on the operation type, e.g. PYPI installation can take 6+ minutes)



Environment configuration

0.5 vCPUs, 1.875 GB memory, 1 GB storage

0.5 vCPUs, 1.875 GB memory, 1 GB storage 0.5 vCPUs, 1.875 GB memory, 1 GB storage

Autoscaling between 1 and 3 workers

1

Scheduler Number of schedulers

Web server

Worker Number of workers

≡ Google Cloud Plat	form 🛟 composer-workshop-1 👻 🔍 Q Search Products, resources, docs (/)	~	Þ. 9 2 i 🌘
Composer 🗧	Environment details 🛛 Open Airflow UI 🖿 Open dags folder 🕀 Save Snapshot 📩 load Snapshot	C REFRESH	👕 DELETE
Stest-bj This environm	ent is running		
MONITORING LOGS	DAGS PREVIEW ENVIRONMENT CONFIGURATION AIRFLOW CONFIGURATION OVERRIDES ENVIRONMENT VARIABLES	LABELS	PYPI PACKAGES
Name	test-bj		
Location	us-central1		
Service account	composer-sa@composer-workshop-1.iam.gserviceaccount.com		
Image version	composer-2.0.12-airflow-2.2.3 UPGRADE		
	Newest available version		
Python version	3		
DAGs folder	gs://us-central1-test-bj-f91dd597-bucket/dags		
Airflow web UI	https://3993596edd064e8e96c4bc2fc8b234b2-dot-us-central1.composer.googleusercontent.com		
Cloud Logging	view logs		
Maintenance windows	4 hours starting 12:00 AM UTC+2 on Friday, Saturday, and Sunday BETA EDIT		
Data encryption key	Google-managed		
Created	Wed May 18 2022 14:27:35 GMT+0200 (Central European Summer Time)		
Updated	Wed May 18 2022 14:44:39 GMT+0200 (Central European Summer Time)		
Resources	EDIT		



Airflow configuration overrides



The configuration properties are stored in a file called airflow.cfg the first time you run Apache Airflow. You can choose to change these properties and override the default values. Sections include: core, cli, api, operators, hive, webserver, email, smtp, celery, celery_broker_transport_options, dask, scheduler, Idap, mesos, kerberos, github_enterprise, admin.

No Airflow configuration overrides

Please take note of blocked configurations:

https://cloud.google.com/composer/docs/concepts/airflow-configurations



Environment variables



Please mind reserved variables:

https://cloud.google.com/composer/docs/how-to/managing/environment-variables#reserved_variables



Labels



Optional user-defined labels for this environment.

No labels added







Required libraries from the Python Package Index (PyPI)

No PyPI packages added



CICD + DAGs



CICD + DAGs

Use CI/CD to Keep DAGs in sync with Version Control





Group Exercise: CICD + DAGs



- Prerequisites
 - Cloud Composer environment created or available
 - GitHub account has been created
- Instructions
 - Fork Github <u>repository</u> <u>https://bit.ly/cicd-sample-repo</u>
 - Make a branch using VSCode in your fork (https://github.dev/{YOUR_GIT_NAME}/composer-workshop-cice
 - Work on this tutorial https://bit.ly/composer-cicd-tutorial
 - SKIP Before you Begin section
 - In "Prepare your environment" SKIP "Add unit tests" (they already exist)
 - When it says "Create a file" don't instead just open the file in VSCode to look at it
 - Skip any steps to "add gcloud" to your environment
 - SKIP THE CLEANUP STEPS



Data Analytics DAG



Data Analytics DAG



BigQuery - Serverless, highly scalable, and cost-effective multicloud data warehouse designed for business agility



Dataproc Serverless - Run Spark batch workloads on a managed compute infrastructure, autoscaling resources as needed



Cloud Storage - Object storage for companies of all sizes. Store any amount of data. Retrieve it as often as you'd like.



Cloud Composer - Fully managed workflow orchestration service built on Apache Airflow.



Data Analytics DAG

- Federal holidays data
- NOAA dataset

How warm was it in Chicago on Thanksgiving for the past 25 years?





Data analytics workflow

Data Analytics DAG		
Composer		DataprocCreateBatchOperator
GCSToBigQueryOperator Cloud Storage	BigQueryInsertJobOperator BigQuery	Dataproc Process the temperature column with a serverless PySpark job
Federal Holidays CSV dataset	Join datasets on date column	spark-bigquery-connector BigQuery Dataproc results output to BigQuery



Data analytics workflow - exploring the DAG

Exploring the DAG



Group Exercise: Data Analytics DAG



Q

- 1. <u>Create an empty BigQuery dataset</u> using the Cloud Console called holiday_weather in the US region
- 2. <u>Create a new Cloud Storage bucket</u> in the same region as your dataset using the Cloud Console -
- 3. Enable the Dataproc API
- 4. <u>Add the following permissions</u> to the **Service Account** for your Composer environment: BigQuery User, BigQuery Data Owner, Service Account User, Dataproc Editor, Dataproc Worker
- 5. Enable private Google access on the default subnet (requirement for Dataproc Serverless)
 - a. Open Cloud Shell
 - b. Run the following command

gcloud compute networks subnets update default \ --region=us-central1 \

--enable-private-ip-google-access



Group Exercise: Data Analytics DAG



- 1. Download <u>data_analytics_process.py</u> this is the Dataproc PySpark file. <u>Add it to your</u> <u>bucket</u> from step 2 of the previous slide.
- 2. Download and add <u>holidays.csv</u> to that same bucket
- 3. Download the <u>dag file</u> add it to the dags/ directory of your fork from the CICD tutorial by dragging it into the VSCode online UI
- 4. <u>Use the Airflow UI to add variables</u> (go to Admin -> Variables)
 - a. gcp_project the name of your project
 - b. gcs_bucket the name of the bucket you created in step 2 of the previous slides without the "gs://" prefix
 - c. gce_region us-central1
 - d. dataproc_service_account the service account for your composer environment - full name with domain
- 5. Trigger the DAG if it has not already triggered. It will be named "summit_dag"



Group Exercise: Data Analytics DAG



- 1. <u>Go to BigQuery</u>
- 2. Open the holidays_weather dataset dropdown and choose holidays_weather_normalized
- 3. Click "Query" at the top to open the Query window
- 4. Run the following Query SELECT Date, value FROM `<your-project>.holiday_weather.holidays_weather_normalized` where Holiday="Thanksgiving Day";
- 5. In the "Query Results" tab click "Explore data" and choose "Explore with Data Studio"
- 6. Delete the default views by hovering over them, choosing the 3 dots, then clicking "Delete"
- 7. At the top, click "add a chart" and Choose "Line Chart"
- 8. In the bottom right for "Metric" choose "Value" and "average"
- 9. At the top of the chart, click "AZ" and choose "Date" and the arrow that indicates "Ascending"
- 10. This will show the temperature in Chicago on every Thanksgiving from 1997 to 2021



05

DAG troubleshooting on Cloud Composer



Notifications



Notification on DAG run failures

Goal: get email notification on DAG run failures for all Composer environments in a project





1. Create Notification Channel



Go to Monitoring - Alerting

	Q Search alerting	× ~
	PRODUCTS & PAGES	
ŝ.	Monitoring	
	Alerting Monitoring	
om > filiŗ	Create alerting policy Monitoring	
Project I	Budgets & alerts Billing	



1. Create Notification Channel



Edit Notification Channels

🕽 osshost 👻	Q Search alerting	×
Alerting + CREATE	POLICY POLICY EDIT NOTIFICATION CHANNELS	
Monitoring now suppo	rts both user-scoped and device-scoped Cloud Console Mobile notification channels	N
Cummony		

Create Email Notification Channel for yourself

Create Email Channel
Email addresses can be set to receive notifications from your alerting when a new incident is created.
Email Address * someone@somewhere.com
Display Name *
Someone Important
CANCEL SAVE





Create Policy



Find Composer Workflow

SELECT A METRIC V	
Select a metric	Reset X
╤ composer	
V POPULAR RESOURCES	
No results	
✓ ACTIVE RESOURCES	
V INACTIVE RESOURCES	
Cloud Composer Environme	nt 71 metrics >
Cloud Composer Workflow	28 metrics >
Show only active resour	rces & metrics 🖓



Find Workflow Runs Metric

Select a metric			Reset X
- composer	Logs-Based Metric	6 metrics >	✓ ACTIVE METRICS
	➤ INACTIVE METRIC CATEGORIES		
V POPULAR RESOURCES			
No results	Cloudvolume_so	7 metrics >	
➤ ACTIVE RESOURCES			
➤ INACTIVE RESOURCES	Cloudvolumepool	2 metrics >	
Cloud Composer Environment 71 metrics > Composite_indexes_per_database1 metric >			
Cloud Composer Workflow 28 metrics >	Vorkflow 28 metrics > Field_configurations_per_database1 metric >		
Show only active resources & metrics	Local_ssd_total_storage_per	_vm_family1 meti	Tasks composer.googleapis.com/workflow/task/run
	Quota	6 metrics >	Workflow Run Duration composer.googleapis.com/workflow/run_durati
	Workflow	4 metrics >	Workflow Runs composer.googleapis.com/workflow/run_count
Selection preview			



Cancel Apply







Add Regex filter for failed Runs (we want notifications for failed runs only)

Add filter				^
Filter * state	•	Comparato	Value *	
			CANCEL	DONE

Change function to count - we want to be notified for count of DAG Run failures







•

Set trigger condition to >0 (i.e. when count of failed DAG runs exceeds 0)

Configure alert trigger

Condition type

O Threshold

Condition triggers if a time series rises above or falls below a value for a specific duration window

Metric absence

Condition triggers if any time series in the metric has no data for a specific duration window

Alert trigger —

Any time series violates

- Threshold position - Above threshold	•
Threshold value — 0	



Use your Email Notification Channel



And save it all under with a desired Alert Policy

Nar	ne the alert
Ale My	rt name *Alert
NE	хт







Troubleshooting



DAG Troubleshooting

- Autoscaling showcase
- Retries
- Notifications on DAG failure
- Permissions troubleshooting



Exercise: Autoscaling



The exercise shows how autoscaling in Cloud Composer 2 adds and removes computing resources automatically.

Prerequisites

- Running Cloud Composer 2 environment (small)
 - Minimum number of workers: 1
 - Maximum number of workers: 3
- Exercise DAG:

https://github.dev/GoogleCloudPlatform/python-docs-samples/blob/main/composer/2022_airflow_summit/parallel_work.py

- Related configuration:

[celery] worker_concurency = 6

work-0-0	• work-0-1	🔶 🔶 work-0.2	→ 🛱 work0.3	→ 🕸 work0.4 → 🕸 work0.5
work-1-0	→ Ø work-1-1	→ ♥ work-1-2	→ ♥ work-13	work-1-5
🔹 work-2-0	• work-2-1	→ ♥ work22	→ 🛊 work.2.3	work24 work25
work-3-0	work3-1	→ ♥ work32	→ 🛊 work33	\$ work3.5
S work-4-0	• werk-4-1	→ ♥ work+2	→ 🛊 work-4-3	work-44 work-4-5
work-5-0	work-5-1	→ ♥ work-52	• • • • • • • • • • • • • • • • • • •	work5-5
work-6-0	🔶 🗘 werk-6-1	• • • • • • • • • • • • • • • • • • •	• • • • • • • • • • • • • • • • • • •	work6-5
work-7-0	• work-7-1	• • • • • • • • • • • • • • • • • • •	• • • • • • • • • • • • • • • • • • •	work-7-4 work-7-5
work-8-0	• • • work-1	• • • • • • • • • • • • • • • • • • •	• • • • • • • • • • • • • • • • • • •	work84 work84
work-9-0		→ 🕸 work+92	→ Ø work93	work94 work95
🖨 work-10-0		• • • • • • • • • • • • • • • • • • •	→ 🛊 work-10-3	work-10-4 work-10-5
work-11-0	→ ♦ work-11-1	• work-11-2	• • • • • • • • • • • • • • • • • • •	◆ work-11-4. ◆ ♦ work-11-5



Google Cloud

Exercise: Retries



The exercise shows what happens when an API fails and how to prevent failures of the whole pipeline.

Prerequisites

- Running Cloud Composer environment
- Exercise DAG:

https://github.dev/GoogleCloudPlatform/python-docs-samples/blob/main/composer/2022_airflow_summit/retries.py





Exercise: Debugging Permissions



The exercise shows how to find permission problems and how to fix permissions for DAGs.

Prerequisites

- Newly created Cloud Composer environment (the one you created today)
- BigQuery table with some data
 - E.g. holiday_weather.holidays
 - Or your own BigQuery table in a different project
- Exercise DAG:

https://github.dev/GoogleCloudPlatform/python-docs-samples/blob/main/composer/2022_airflow_summit/bigquery_permissions.py





Thank you.

Google Cloud