

**4**

**tips that help you  
leverage your  
Airflow pipelines**

# About me

- Aliaksandr Sheliutsin
- 6+ years in Data Engineering
- 1.7 - 2.7 Airflow experience



1

**You can create  
typical DAGs  
based on a CSV file**

# CASE:

- // Catchup pipelines for historical data download
- // Triggers for support dags

name	snapshot_id	blockchain_name	dvpc_start_date	delegate_changes_start_date	delegate_histories_start_date
balancer	balancer.eth	ethereum	2020-03-17	2020-03-11	2020-05-04
ens	ens.eth	ethereum	2020-03-17	2020-03-11	2020-05-04
uniswap	uniswap	ethereum	2020-03-17	2020-03-11	2020-05-04
ampleforth	ampleforthorg.eth	ethereum	2020-03-17	2020-03-11	2020-05-04
pooltogether	pooltogether.eth	ethereum	2020-03-17	2020-03-11	2020-05-04
nexusmutual	community.nexusmutual.eth	ethereum	2020-03-17	2020-03-11	2020-05-04
hop	hop.eth	ethereum	2020-03-17	2020-03-11	2020-05-04
optimism	opcollective.eth	optimism	2022-05-26	2022-05-26	2022-04-26
1inch	1inch.eth	ethereum	2020-03-17	2020-03-11	2020-05-04
orangedao	orangedaoxyz.eth	ethereum	2020-03-17	2020-03-11	2020-05-04

# PROS:

Easy to add a new entity

Less code

Flexible

Isolation/Can control start date

Also covered with data validation tests

# CONS:

Advanced understanding of Airflow


Kind of similar pipelines

2




**You can create  
typical DAGs based  
on a sharable  
Config** (like Google Sheets)

# CASE:





// Catchup pipelines for historical GA pipelines







CustomerConfig



File Edit View Insert Format Data Tools Extensions Help [Last edit was seconds ago](#)







100%










123

Default (Ari...


10







G15



	A	B	C	D	E	F	G	
1	Customer	ViewID	PullDataFrom	Logo	Account Id	Web Property	Customer ID	
2	Customer 1	123456789	10/01/2019		123456789	UA-1234567-2	1234	
3	Customer 2	123456790	10/01/2019		123456790	UA-123456789-1	1235	
4	Customer 3	123456791	10/01/2019		123456791	UA-123456790-1	1236	
5	Customer 4	123456792	10/01/2019		123456792	UA-123456791-1	1237	
6								
7								



## PROS:

Other people can create Airflow DAGs

Changes are available without deployments

# CONS:

Performance

You need to handle user experience

// Documentation

// Input data validation





// Error Notifications

3

**Autogenerate  
DAGs based  
on the Airflow  
Entities** (Connection, Variables)

# CASE:

// Catchup pipelines for different databases

<input type="checkbox"/>	 	processing_neo	mysql	Historical data from processing
<input type="checkbox"/>	 	processing_trinity	mysql	Historical data from processing

## PROS:

No need to handle External API calls

# CONS:

Configurations can be lost

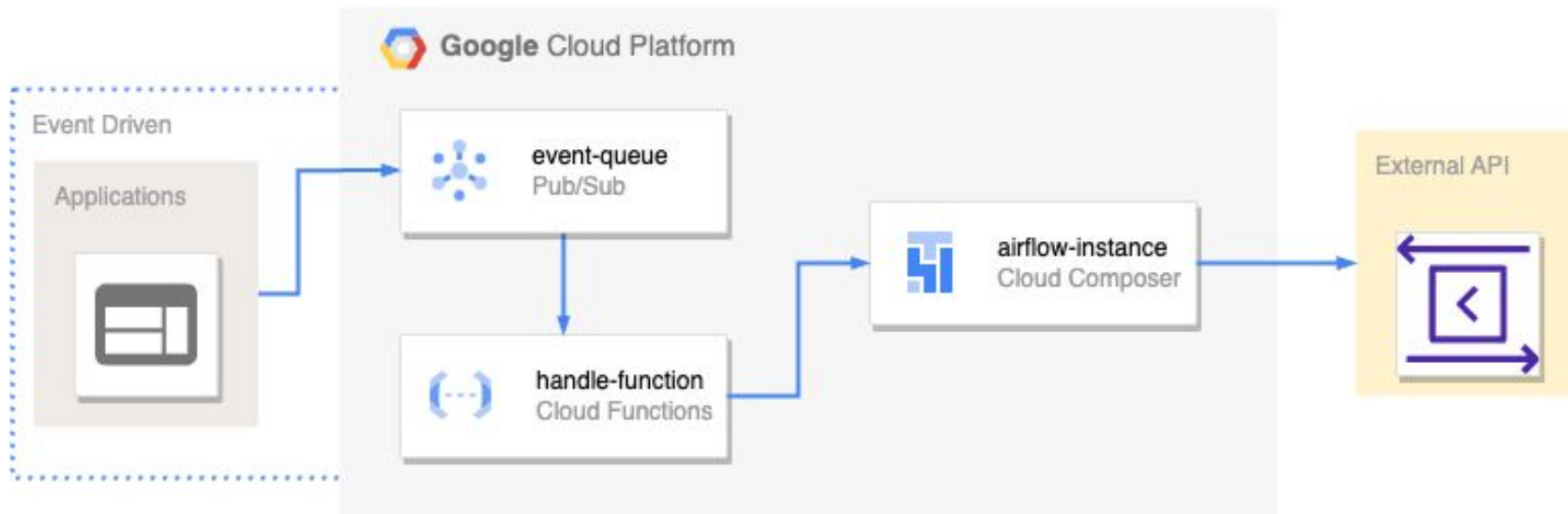
Higher load on Metadata DB

4

**Make your  
Airflow pipeline  
event-driven**

# CASE:

// Domain security tests execution via PubSub - Cloud Function





## PROS:

Results of execution do not rely on a schedule

Don't have tons of one-minute Dag Runs

## CONS:

This way will not work with a high volume of events

# Have questions?

// **Linkedin:** <https://www.linkedin.com/in/aliaksandr-sheliutsin/>

// **Telegram:** @asheliutsin