Airflow and Analytics Engineering
Dos and don´ts

Sergio Camilo Fandiño Hernández
Senior Analytics Engineer at Trade Republic

TRADE REPUBLIC

Airflow Summit 2021

1. **Why am I here?**

2. What does an Analytics Engineer do?

3. Why then Airflow?

4. The Dos

5. The Don'ts

6. Outro

TRADE
REPUBLIC

TRADE
REPUBLIC

What does an Analytics Engineer do?

"This job is neither data engineering, nor analysis. It's somewhere in the middle, and it needed a new title. Starting in 2018, we (dbt) and a few of our friends in the Locally Optimistic Community started calling this role the analytics engineer."

"Analytics engineers provide clean data sets to end users, modelling data in a way that empowers end users to answer their own questions. While a data analyst spends their time analysing data, an analytics engineer spends their time transforming, testing, deploying, and documenting data…."

TRADE
REPUBLIC

Why then Airflow?

Airflow is the Analytics Engineer's ally in any organisation, as it allows you to orchestrate the integration of external data sources by using different operators from different providers, the transformation of data using for example dbt. It can also help you monitor and alert the status of your data pipelines using notifications for example with Slack Operators.

TRADE
REPUBLIC

Use the Community's Slack Channel

– apache-airflow.slack.com

Have your own Local Airflow Environment

- Local System / Docker / Breeze ...

Define Together with your Team Best Practices for Writing DAGs

– Task Names

– Arguments

– Linters

– Dependency of Tasks

Always Include Notifications in your DAGs
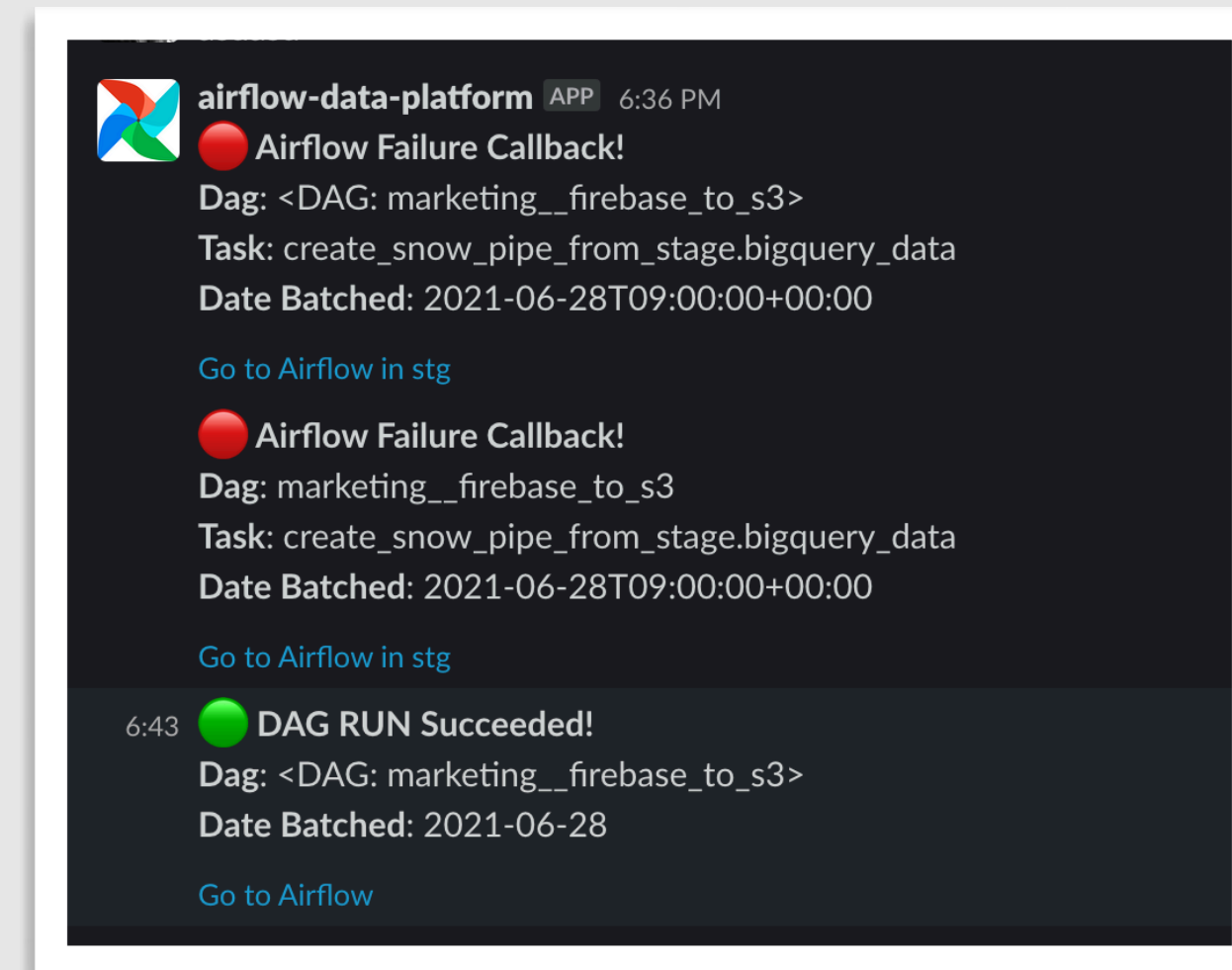

```
on_success_callback , on_failure_callback , or on_retry_callback
```

Share what you know about Airflow with your team

Sergio Camilo Fandiño Hernández
Senior Analytics Engineer

TRADE
REPUBLIC

Ensure that a Fail-Safe CI/CD is in Place

– Continuos Integration / Continuos Development

1. Why am I here?

2. What does an Analytics Engineer do?

3. Why then Airflow?

4. The Dos

5. **The Don'ts**

6. Outro

TRADE
REPUBLIC

Do not Ignore the Folder Structure
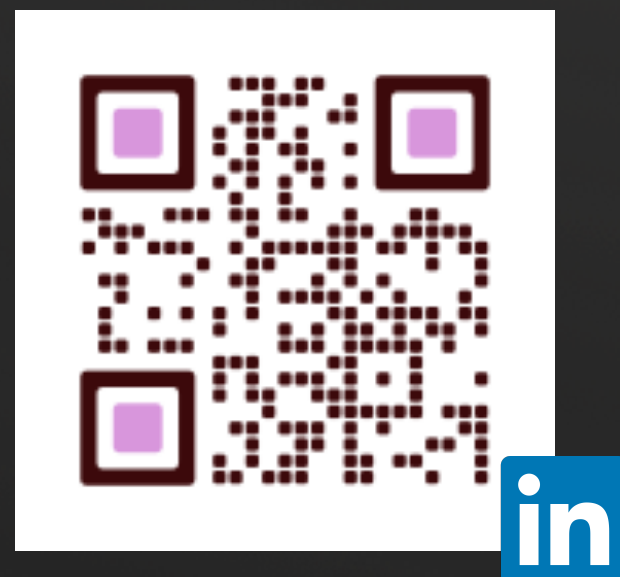

airflow / airflow / providers /

Do not put Tokens, Passwords or any Sensitive Data in Plain Text

```
key = 'pRmgMa8T0INjEAfks     uafzoZXEuwKI7wDe4c1F8AY='
```

Do not put Top-Level Statements Inside your DAGs

```
my_task = BigQueryOperator(
        task_id='bq-task',
        bql="""'SELECT 'Adams' as LastName, 51 as Oppone
               SELECT 'Buchanan', 77, 0 UNION ALL
               SELECT 'Coolidge', 77, 1 UNION ALL
               SELECT 'Adams', 52, 4 UNION ALL
               SELECT 'Buchanan', 50, 13)
               SELECT 'Coolidge', 77, 1 UNION ALL
               SELECT 'Adams', 52, 4 UNION ALL
               SELECT 'Buchanan', 50, 13)
               SELECT 'Coolidge', 77, 1 UNION ALL
               SELECT 'Adams', 52, 4 UNION ALL
               SELECT 'Buchanan', 50, 13)
               SELECT 'Coolidge', 77, 1 UNION ALL
               SELECT 'Adams', 52, 4 UNION ALL
               SELECT 'Buchanan', 50, 13)
               SELECT 'Coolidge', 77, 1 UNION ALL
               SELECT 'Adams', 52, 4 UNION ALL
               SELECT 'Buchanan', 50, 13)
               SELECT 'Coolidge', 77, 1 UNION ALL
               SELECT 'Adams', 52, 4 UNION ALL
               SELECT 'Buchanan', 50, 13)
               SELECT 'Coolidge', 77, 1 UNION ALL
               SELECT 'Adams', 52, 4 UNION ALL
               SELECT 'Buchanan', 50, 13)
               SELECT 'Coolidge', 77, 1 UNION ALL
```

Do not Write the same DAG Hundred of Times

– DRY

Do not Over Engineer your Data Pipelines

TRADE
REPUBLIC