Airflow for ML at Twitter

@Dan Davydov

June 7th, 2020

Agenda

1. How Twitter uses Airflow for ML
2. Airflow pain points (focus on ML)
3. Future of Airflow @ Twitter

Airflow @ Twitter

~400 DAG Files

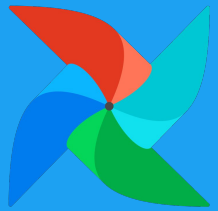
~30 Customer Teams

Mostly ML Use-Cases

Airflow ML Stack

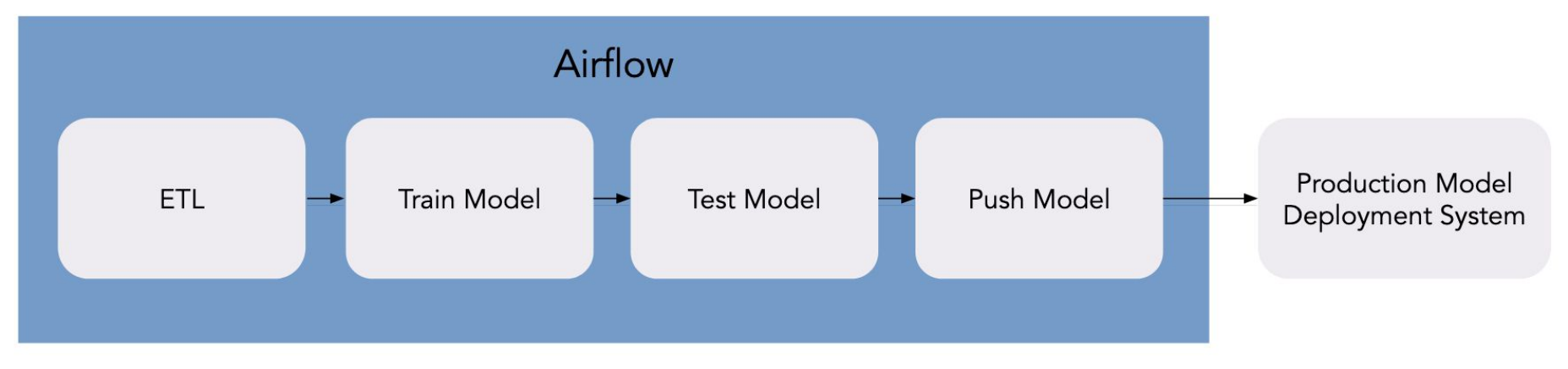
Google Cloud Platform

TensorFlow



kubernetes

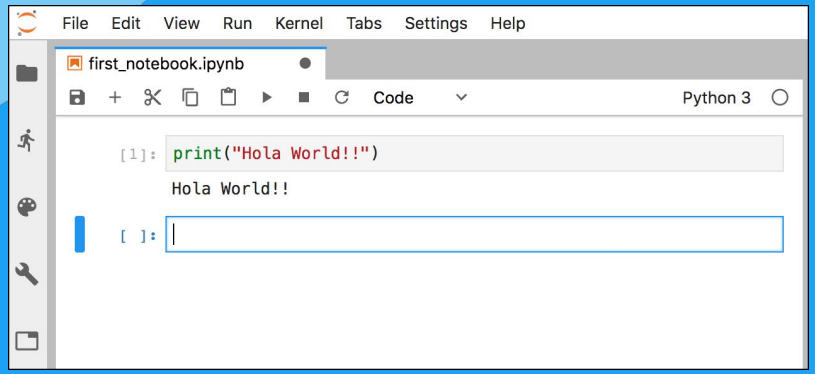
Typical ML Airflow Pipeline

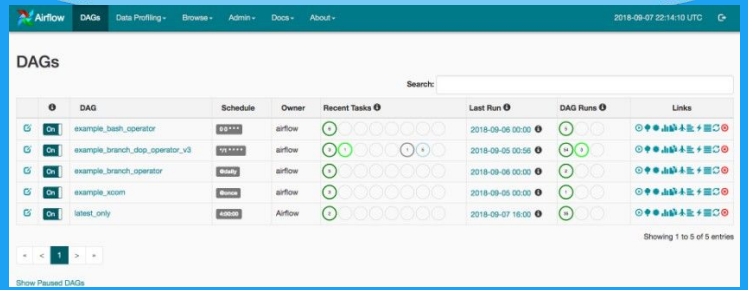


Airflow Pain Points

Development Speed

VS

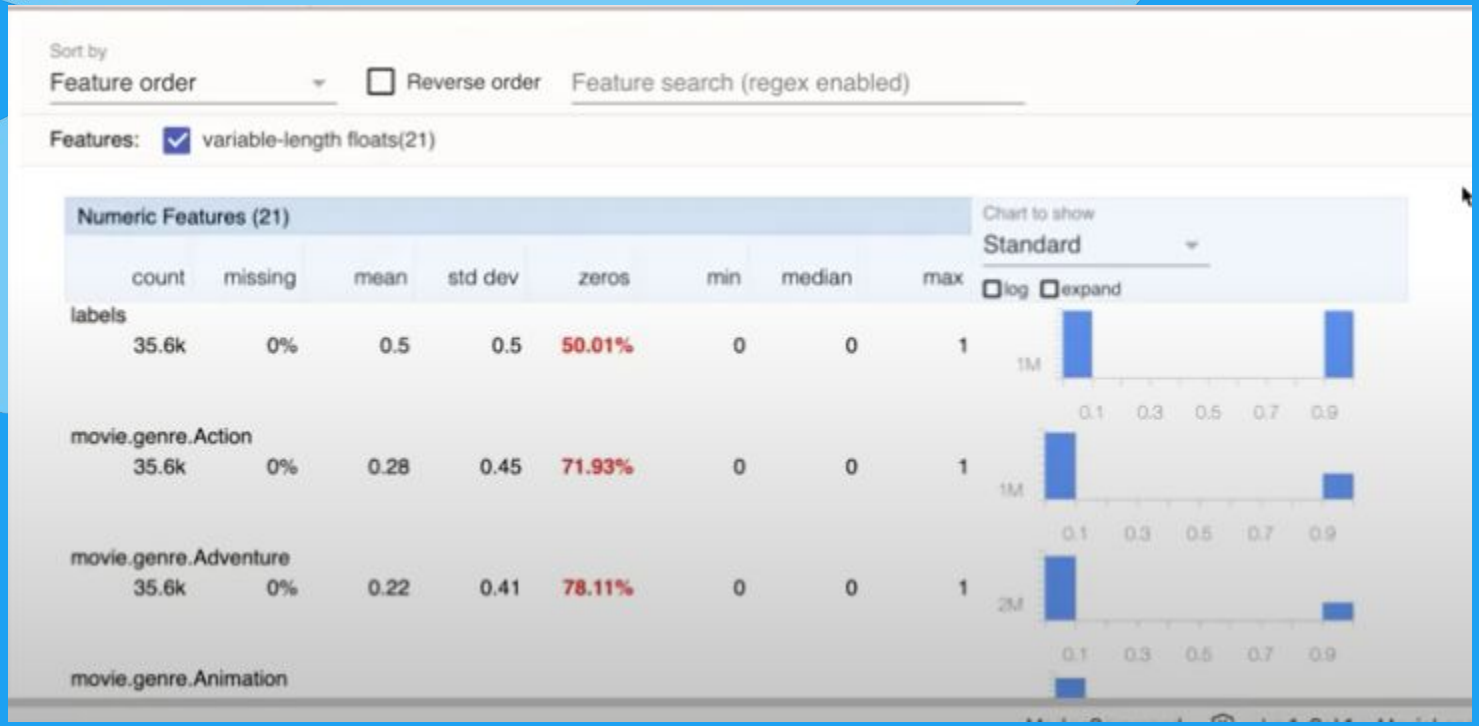




Clunky Interfaces

```
def push(**kwargs):
    # pushes an XCom without a specific target
    kwargs['ti'].xcom_push(key='value from pusher 1', value=value_1)
```

Lack of First-Class ML Tooling Integration

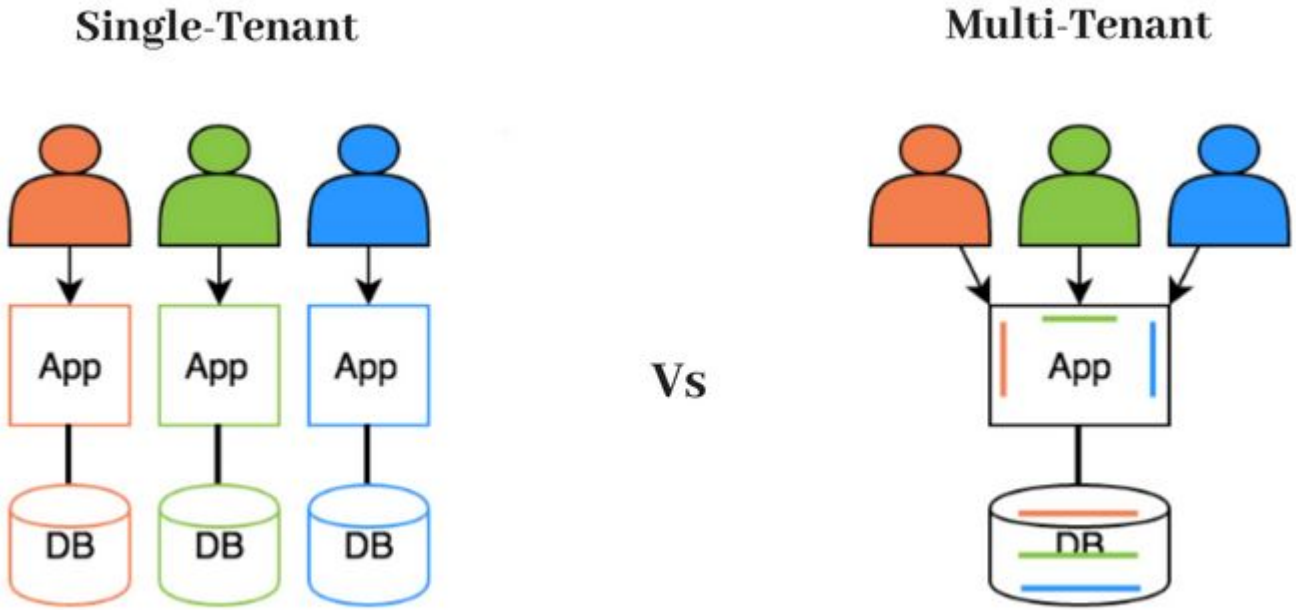


Lack of OSS ML Operators

?

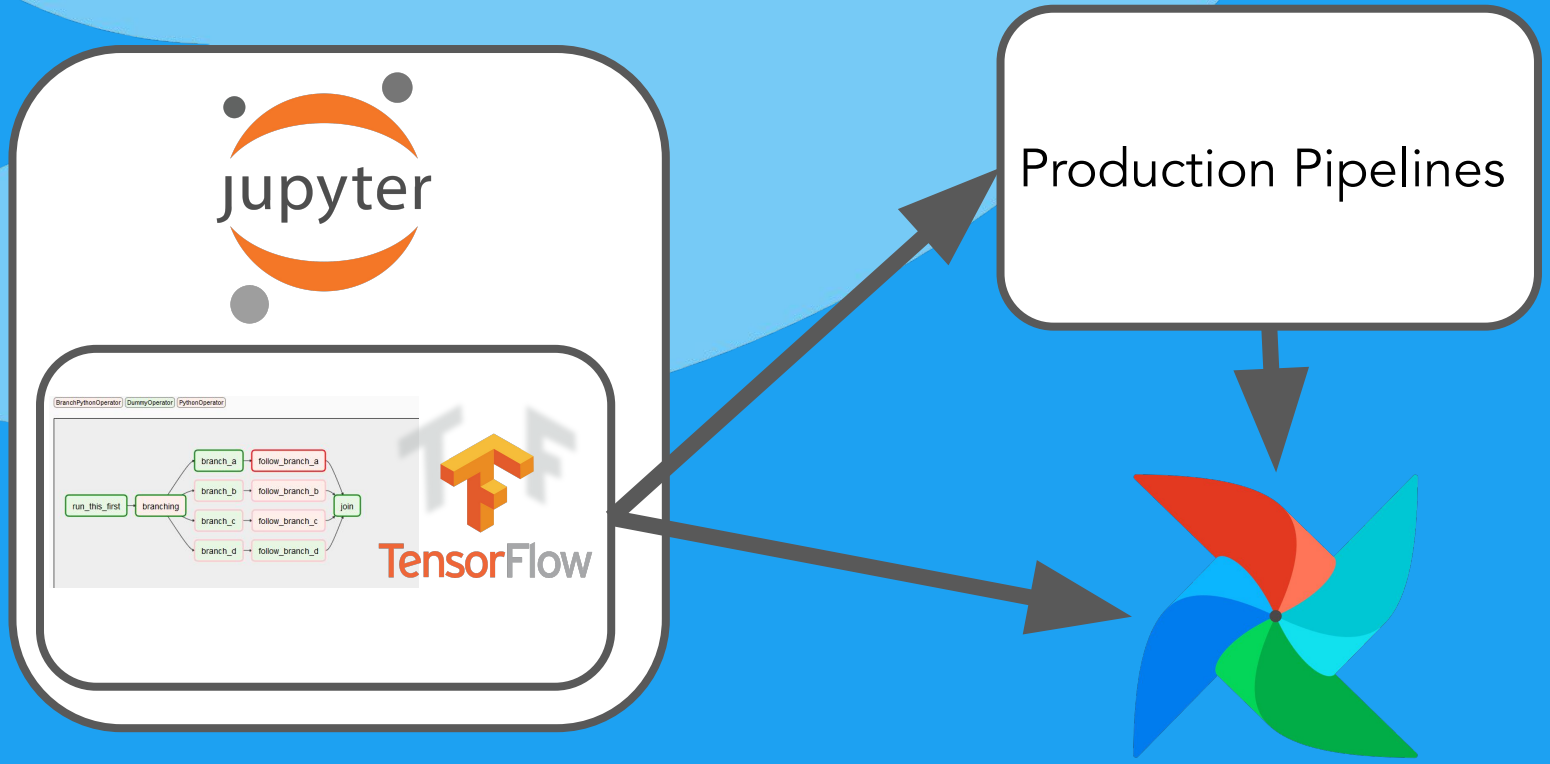
TensorFlow

No Multi-tenancy



Future of Airflow for ML @ Twitter

Airflow as Job Dispatch



Thanks For Watching!