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

ETL → Train Model → Test Model → Push Model → Production Model Deployment System
Airflow Pain Points
Development Speed
Clunky Interfaces

```python
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

<table>
<thead>
<tr>
<th>Features</th>
<th>count</th>
<th>missing</th>
<th>mean</th>
<th>std dev</th>
<th>zeros</th>
<th>min</th>
<th>median</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>labels</td>
<td>35.6k</td>
<td>0%</td>
<td>0.5</td>
<td>0.5</td>
<td>50.01%</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>movie.genre.Action</td>
<td>35.6k</td>
<td>0%</td>
<td>0.28</td>
<td>0.45</td>
<td>71.93%</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>movie.genre.Adventure</td>
<td>35.6k</td>
<td>0%</td>
<td>0.22</td>
<td>0.41</td>
<td>78.11%</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>movie.genre.Animation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[Image of numeric feature statistics]
Lack of OSS ML Operators
No Multi-tenancy
Future of Airflow for ML @ Twitter
Airflow as Job Dispatch

Production Pipelines

jupyter

TensorFlow
Thanks For Watching!