#### **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

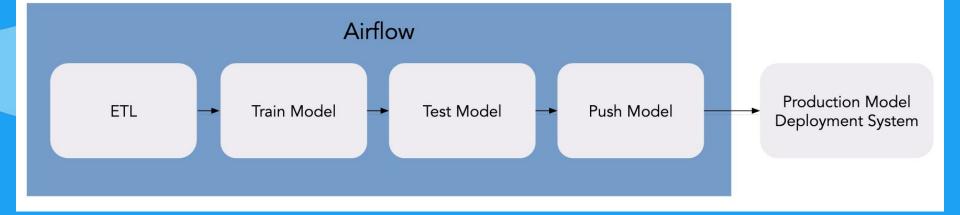








## **Typical ML Airflow Pipeline**



## Airflow Pain Points

## Development Speed

AG	S							
					Search:			
3	0	DAG	Schedule	Owner	Recent Tasks 0	Last Run	DAG Runs O	Links
	2n	example_bash_operator	00***	airflow	00000000	2018-09-06 00:00 0	000	⊙♥●.hbh本社+≣C⊙
	2n	example_branch_dop_operator_v3		airflow	00000000	2018-09-05 00:56 0	00	<b>○</b> ♥●.hbì木主+≡C <b>⊙</b>
	2n	example_branch_operator	Ostally	airflow	00000000	2018-09-06 00:00 0	000	<b>○ ♥ ● .ht</b> ★ ± + ≡ <b>C  ③</b>
	2n	example_xcom	Burca	airflow	00000000	2018-09-05 00:00 0	000	<b>○</b> ♥●.hb}+E+EC <b>⊙</b>
	n	latest_only	49000	Airflow	00000000	2018-09-07 16:00	0	<b>○</b> ♥●.hb}★E+≣C <b>⊙</b>
								Showing 1 to 5 of 5 entries

()	File Edit V	View Run Kernel	Tabs Settings	Help				
	First_notebook.ipynb							
	<b>B</b> + %		C Code ~		Python 3	0		
Ŕ	[1]:	<pre>print("Hola Worl</pre>	.d!!")					
۲	с. Г. 1:	Hola World!!						
عر								

## Clunky Interfaces

def push(\*\*kwargs):
# pushes an XCom without a specific target
kwargs['ti'].xcom\_push(key='value from pusher 1', value=value\_1)

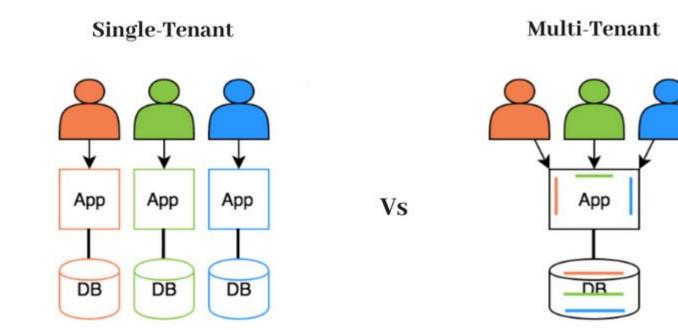
### Lack of First-Class ML Tooling Integration

Feature order - Reverse order			Feature s	Feature search (regex enabled)				
eatures: 🔽	variable-leng	th floats(21	)					
Numeric Feat	ures (21)							Chart to show Standard ~
count	missing	mean	std dev	zeros	min	median	max	log expand
labels								
35.6k	0%	0.5	0.5	50.01%	0	0	1	зм
	ation							0.1 0.3 0.5 0.7 0.9
movie.genre.A 35.6k	0%	0.28	0.45	71.93%	0	0	1	
55.0K	070	0.20	0.40	11.3376	v	U		1M
								0.1 0.3 0.5 0.7 0.9
movie.genre.A	dventure							-
35.6k	0%	0.22	0.41	78.11%	0	0	1	1 2M
								and the second se
movie.genre.A	almation							0.1 0.3 0.5 0.7 0.9

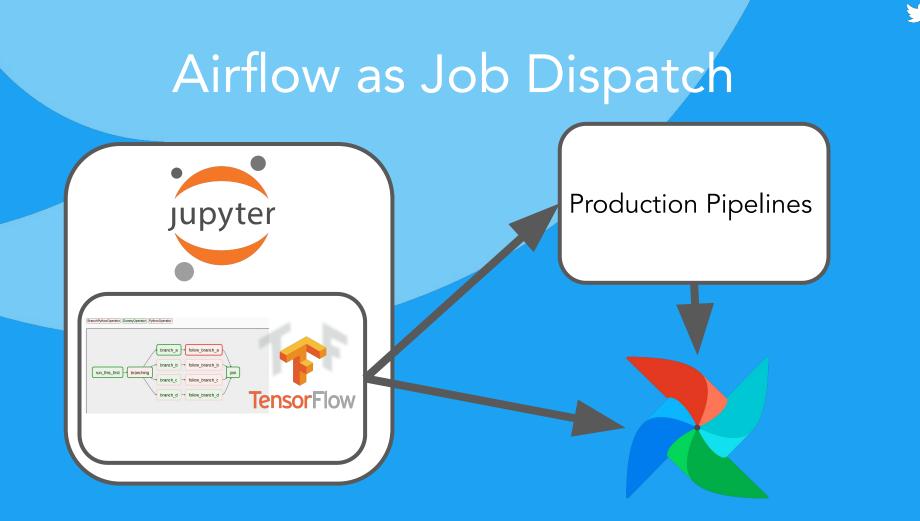
### Lack of OSS ML Operators

### **Tensor**Flow

### No Multi-tenancy



### Future of Airflow for ML @ Twitter



# Thanks For Watching!