

Data Science **PIPELINES**

Favio André Vázquez
Samuel Hinton

DSGo 2020

A very brief **introduction**



Favio André Vázquez [linkedin.com/in/faviovazquez/](https://www.linkedin.com/in/faviovazquez/)

Data scientist, physicist.

Your host and presenter for the workshop.

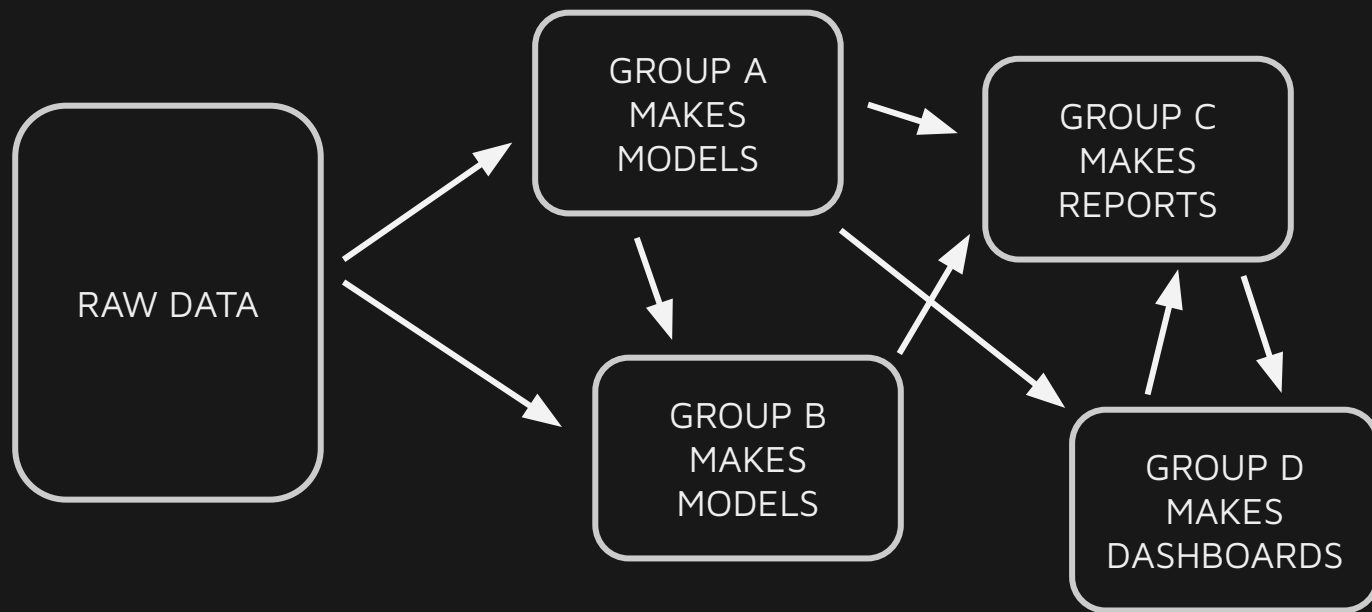


Samuel Hinton [linkedin.com/in/samuelreay/](https://www.linkedin.com/in/samuelreay/)

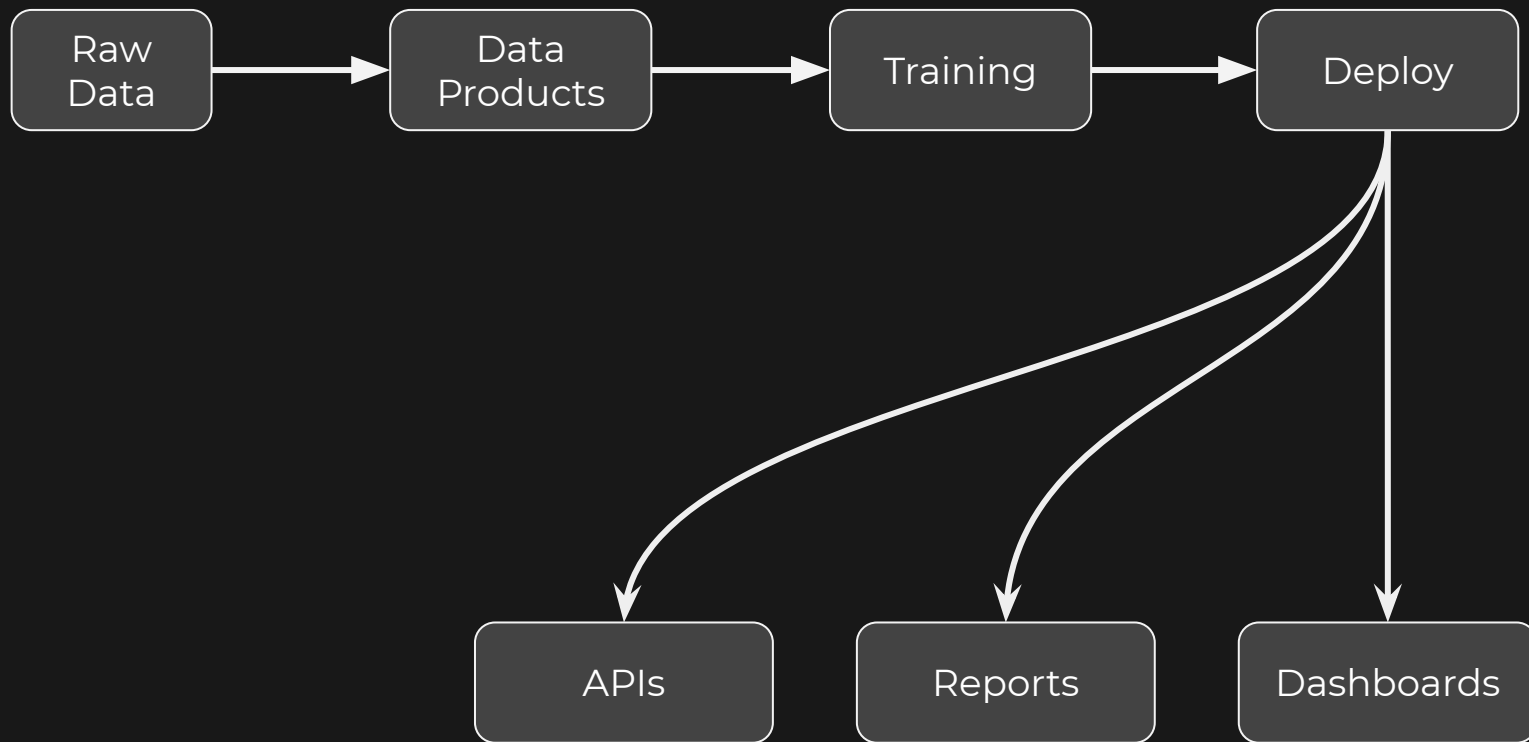
Astrophysicist, data scientist.

Currently recovering from throat surgery and will not be speaking! Might croak in the background.

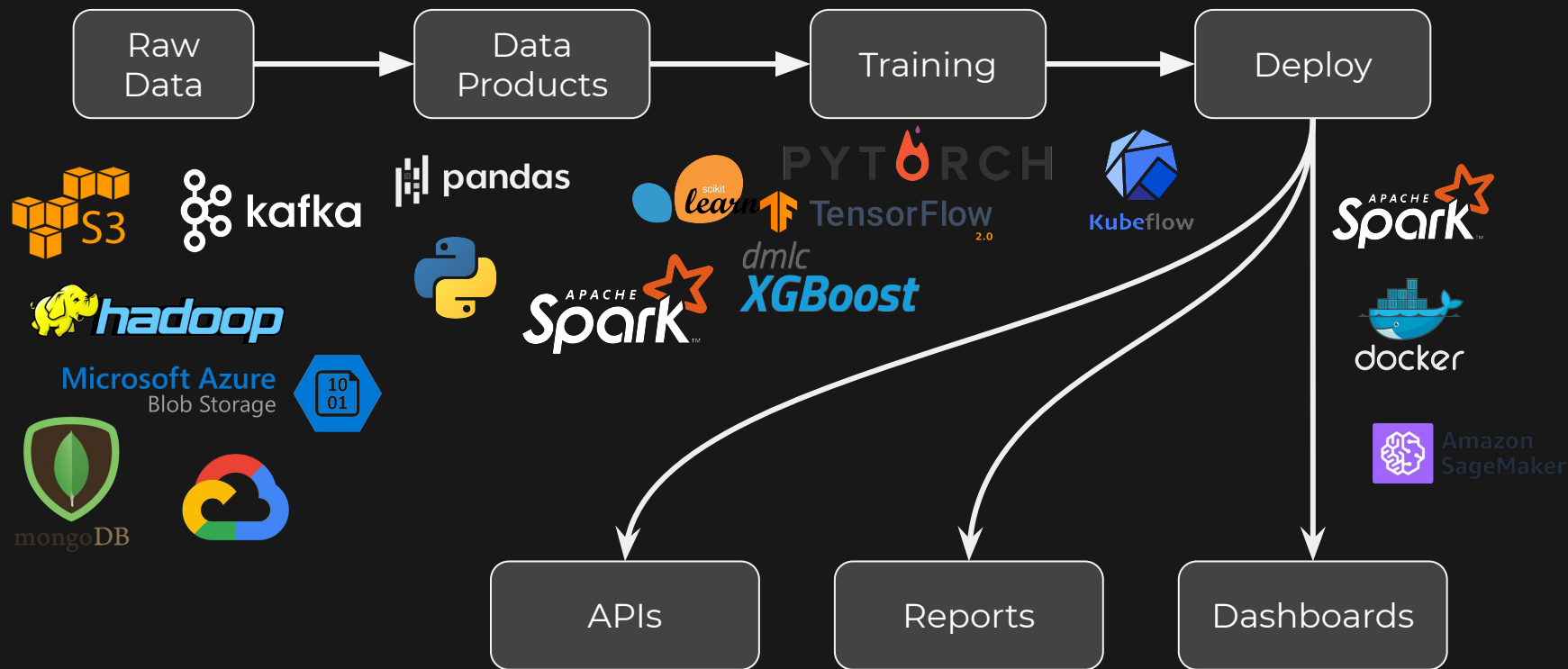
Data Science Products are **Complex**



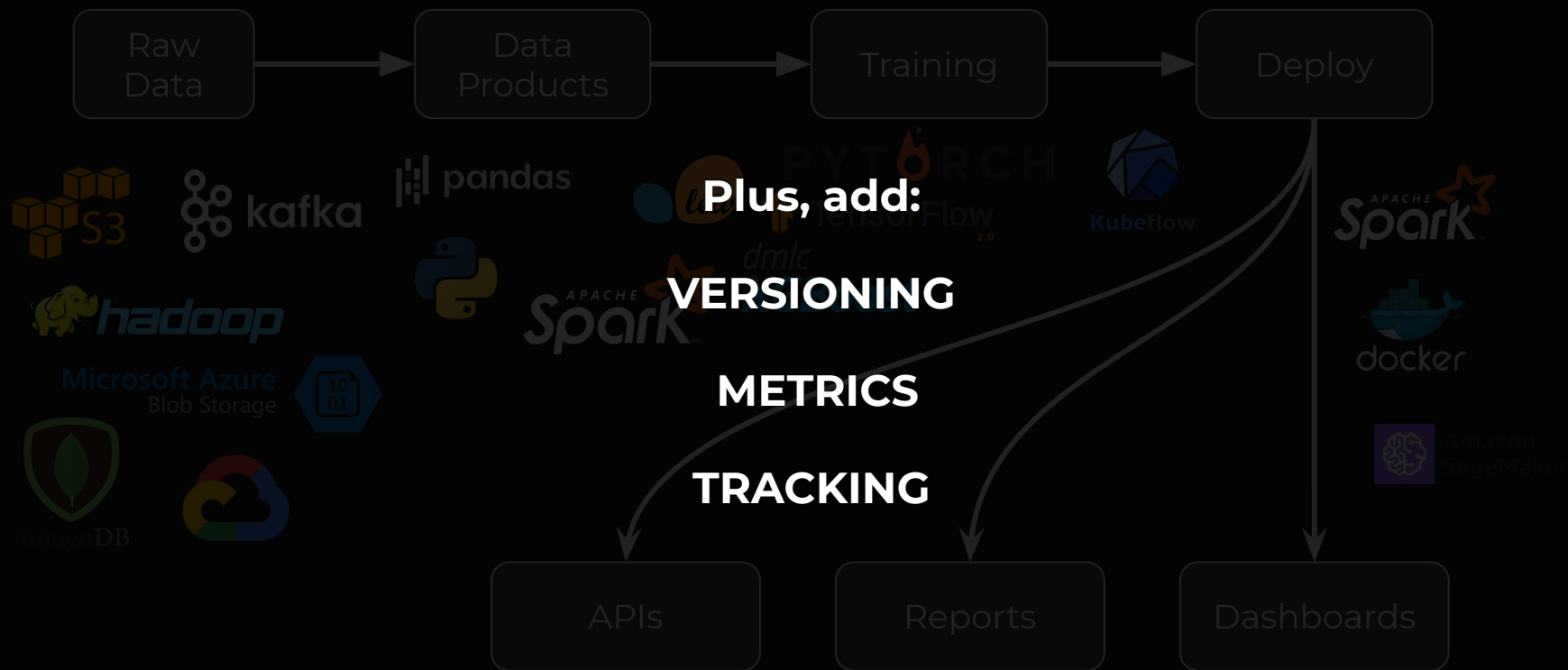
Data Science Products are **Complex**



Data Science Products are **Complex**



Data Science Products are **Complex**



Helping hand from **mlflow**

- **mlflow tracking**
 - Track, log and query models, metrics, hyperparameters, code versions
- **mlflow projects**
 - Package data + code in format to allow reproduction anywhere
- **mlflow models**
 - Deploy standardised ML models to diverse locations
- **mlflow model registry**
 - Store, annotate and manage models, staging and versions in a central repository.

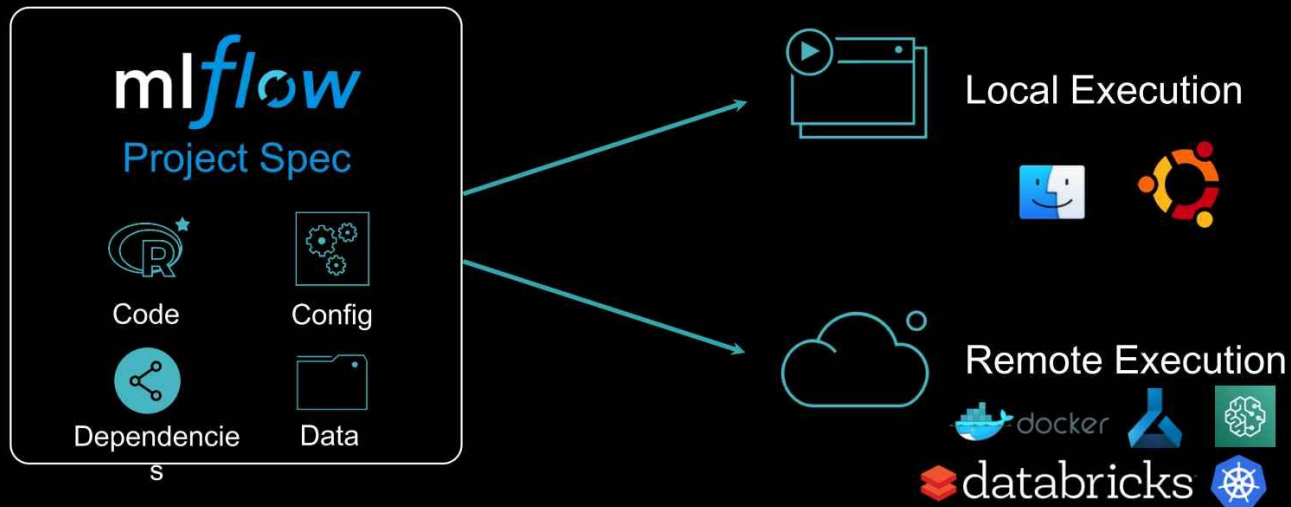
Helping hand from **mlflow**

MLflow Tracking



Helping hand from **mlflow**

MLflow Projects

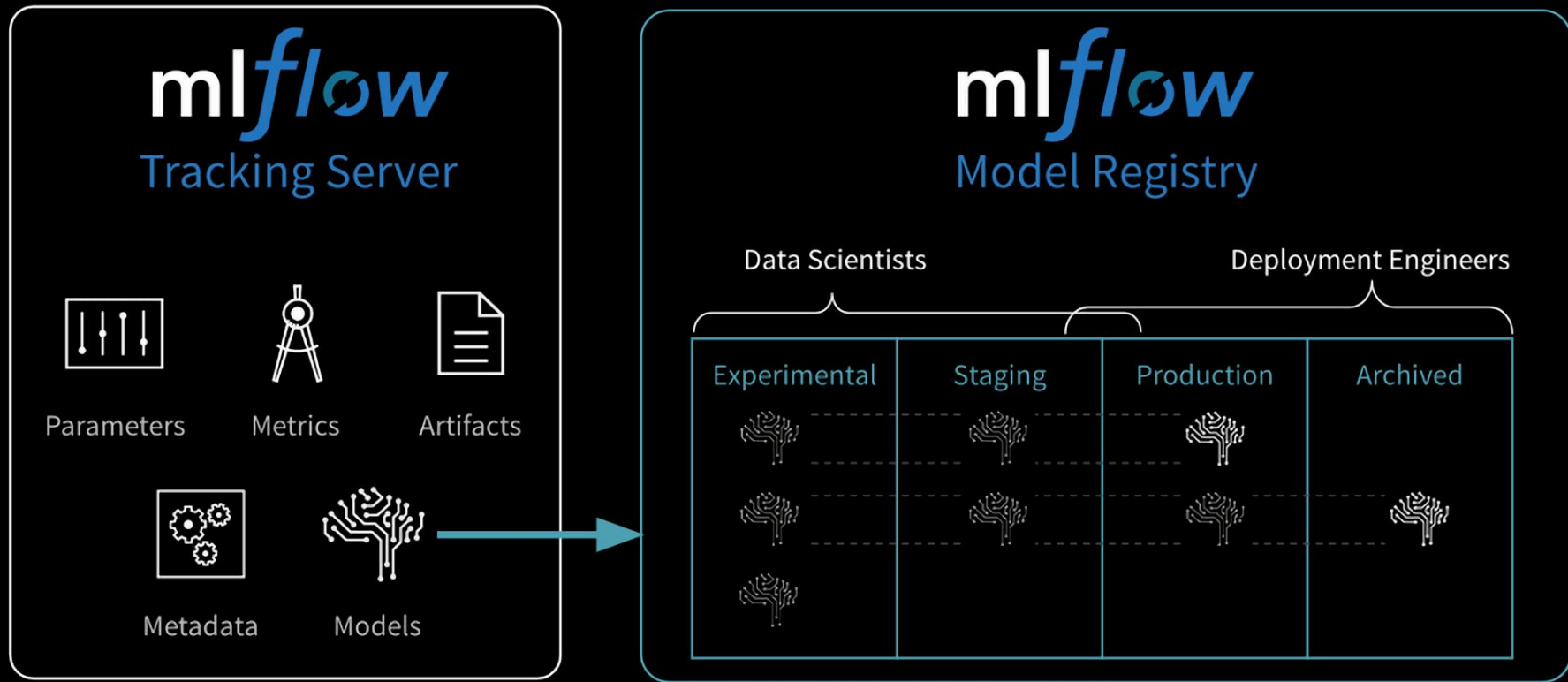


Helping hand from *mlflow*

MLflow Models



Helping hand from **mlflow**



Utilise Airflow along with **mlflow** for a pipeline

What we'll be making at the end of the workshop

