



KubeCon



CloudNativeCon

North America 2019

# Kubeflow in the Financial Sector

## Kubeflow Contributors

Josh Bottum, Arrikto

Laura Schornack, JPMorgan Chase

Jeff Fogarty, USBank @jtfogarty

Thea Lamkin, Google @taylamkin



# Kubeflow for FinTech

FinTech Journeys to the Kubeflow Summit,  
*Acclimating ML teams with new DNA*

Kubeflow 1.0



2019 - Base Camp 1



2016 - Start





# Early investing in ML / Kubeflow

- Algorithmic innovation
  - ML increases productivity, margin, rev & mkt cap
- LinkedIn: Doubling productivity of ML Engineers
  - [Kubeflow Summit Presentation \(Mar '19\)](#)
- Spotify: 70% of ML models productized < 12 wks
  - [KF Summit \(Oct '19\)](#)
- GM: Buys self-driving start-up for \$1B (2016)
  - [Cruise's valuation: \\$19B++](#)

If software is the new oil, then  
ML software is the new gasoline



Company	Mkt Cap / worker
GM	\$300,000
Facebook	\$13,000,000

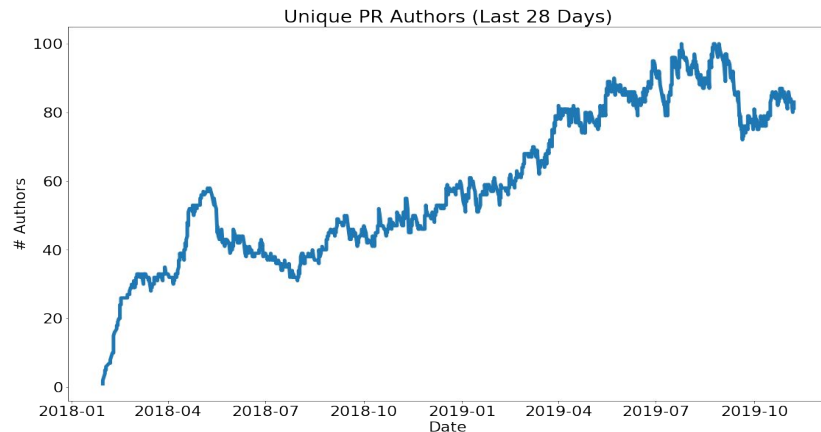
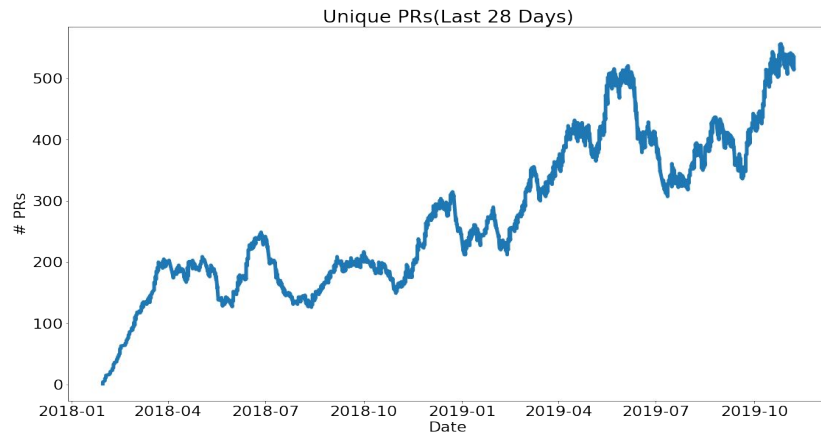


# Why is Kubeflow coming to a Fintech near you ?

- Building (and maintaining) your own ML stack is hard
- Kubeflow's growth is following path of successful mountain climbs
  - Ex. Kubernetes, TensorFlow
- Leveraging a thriving Community
  - provides a variety of support options
  - attracts Open Source contributors to bring their work to your team
- Financial Leaders already “acclimating” at the Kubeflow Summit
  - Bloomberg, Intuit, JPMorgan, USBank, Visa, ANT Financial

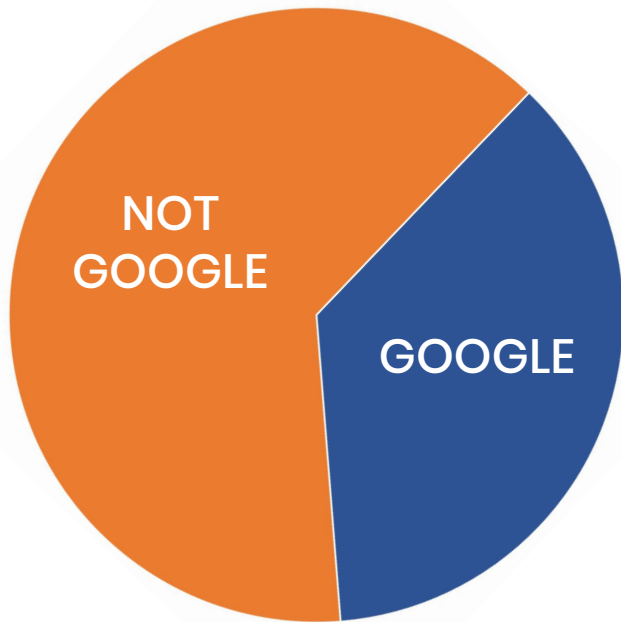


# Momentum

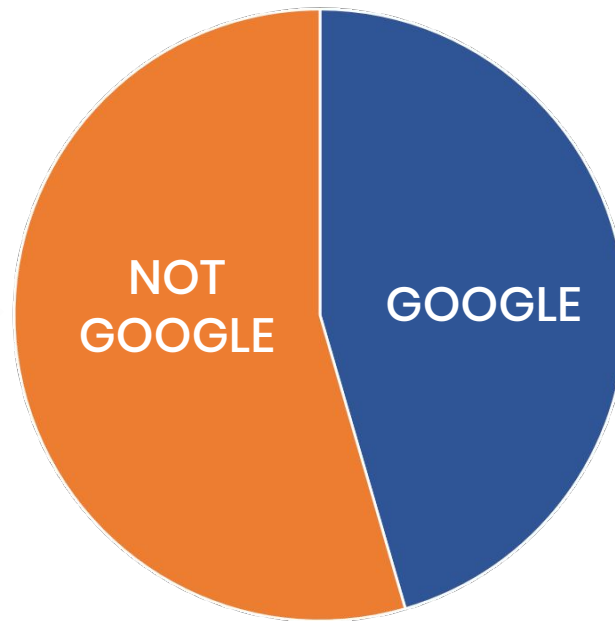




# Community Contributions



Kubernetes



Kubeflow



# Implementing Kubeflow for FinTech

- High Security  $\neq$  Rapid Innovation
  - Firewalls, Proxies, Authentication, Authorization
  - Sandboxes, Air-gap, Dataset access
  - Admin/Root access, Validating secure container builds
- Rapid Innovation = Self-Service Platform (for Data Scientists)
  - Deploy secure notebooks & datasets with isolation
  - Share ACL protected versions of models and datasets
  - Build Pipelines without K8s expertise & Pipeline DSL coding



## Gartner Report

Accessing the Open Source Enterprise  
Machine Learning Stack

Kubeflow has ~150 contributors.

*Supported by Technology Professionals at  
Google, Intel, Red Hat, Cisco and Arrikto*

# Why Arrikto?

- A leading Kubeflow contributor (nimble)
- Consistent (local disk) architecture that simplifies storage operations
- Materially simplify ML Pipeline workflows via Enterprise Kubeflow

Version	Date Released	Steps to create/run	Steps requiring K8 expertise	Resources Required
Kubeflow v0.3	Dec 18	26	14	Data Scientist + ML Engineer
Kubeflow v0.6	Jul 19	12	0	Data Scientist only





# On Prem Kubeflow in FinTech

Laura Schornack, JPMorgan Chase

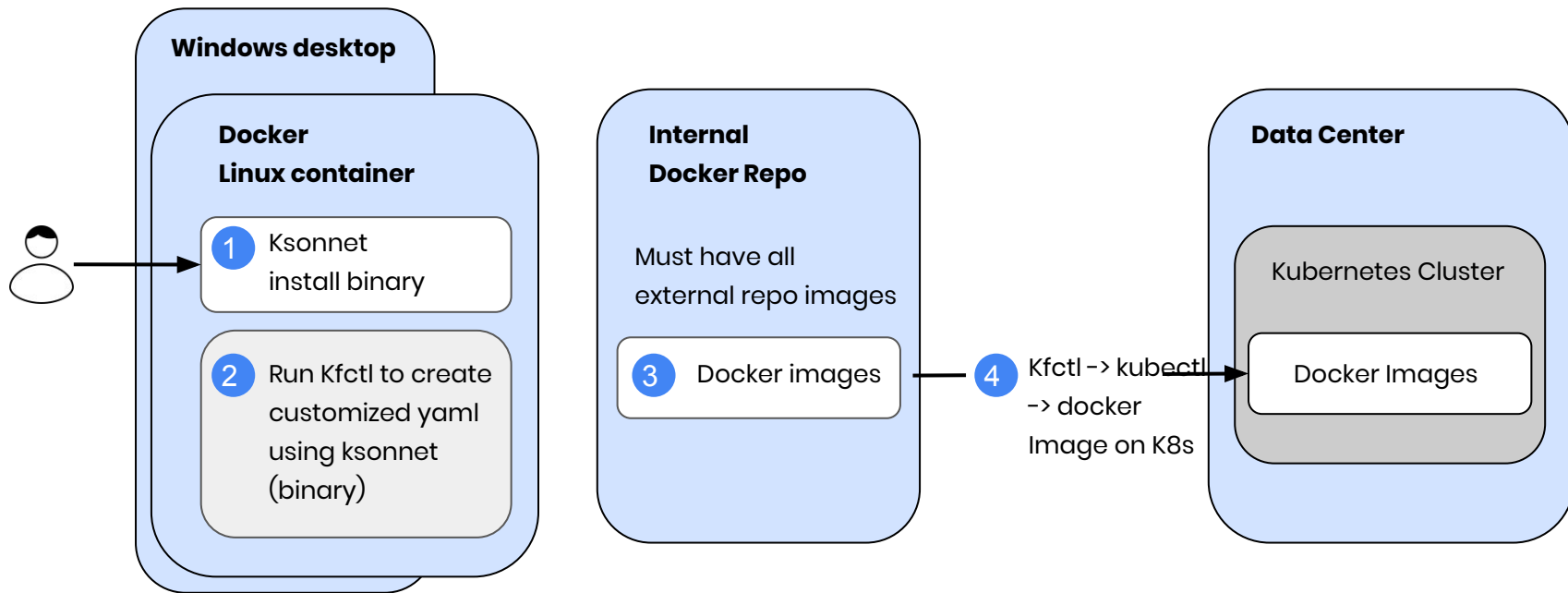




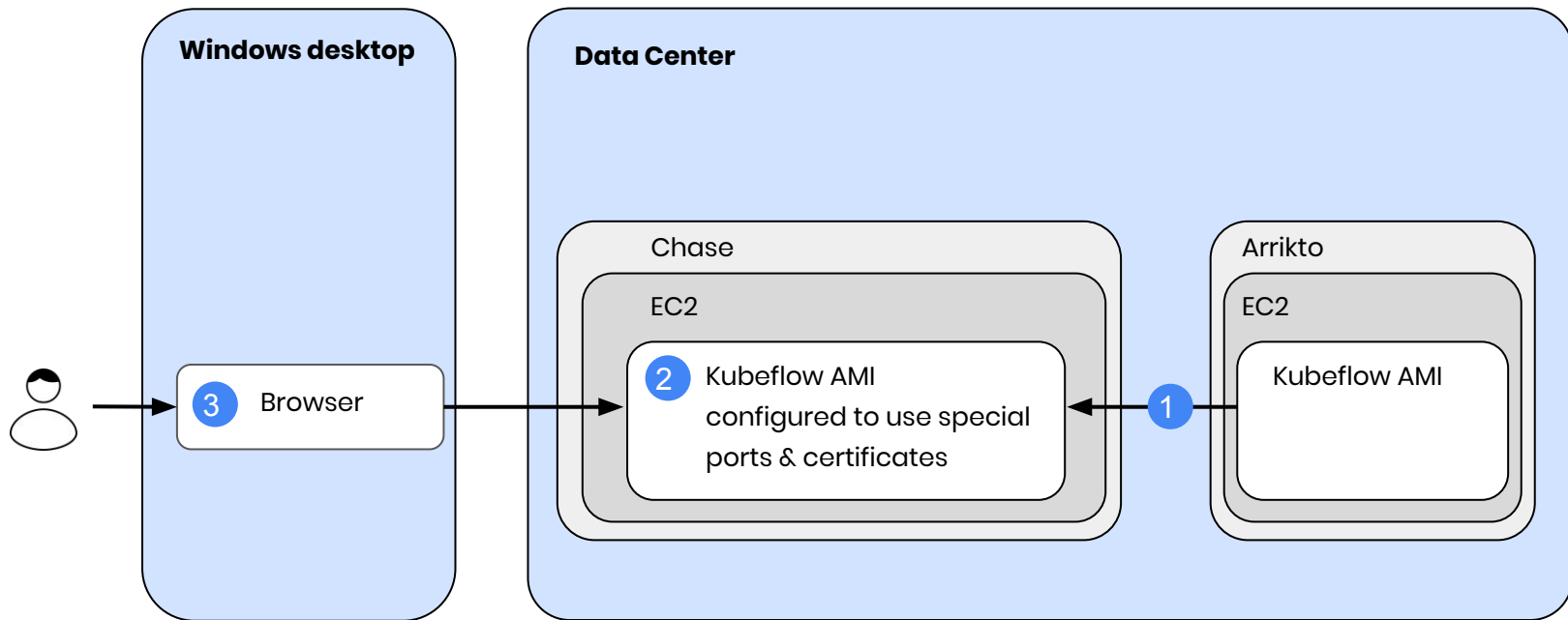
# How did we get here

- Orchestration
- Faster training
- Spin up Notebook Servers as necessary, and turn them off, self-service
- Code & Data version control
- Data sharing
- Model sharing
- Data transferring
- Security
- Analysis of new data

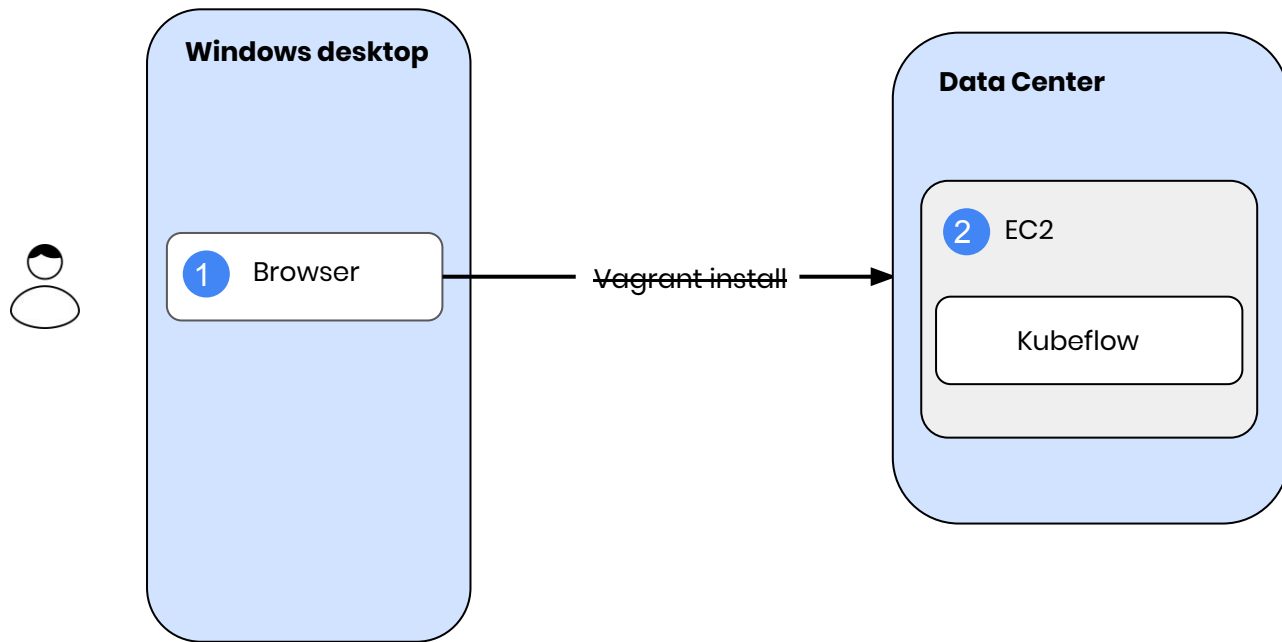
# POC – First attempt



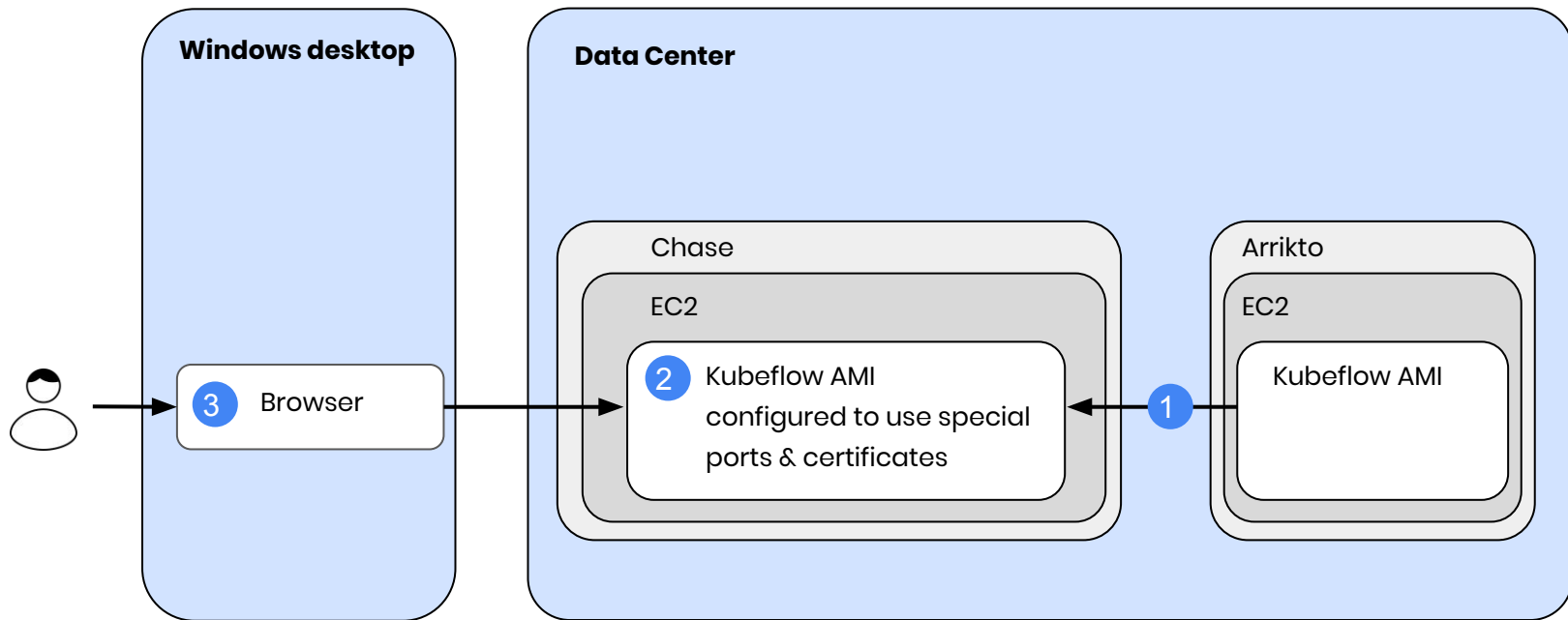
# POC – Second attempt



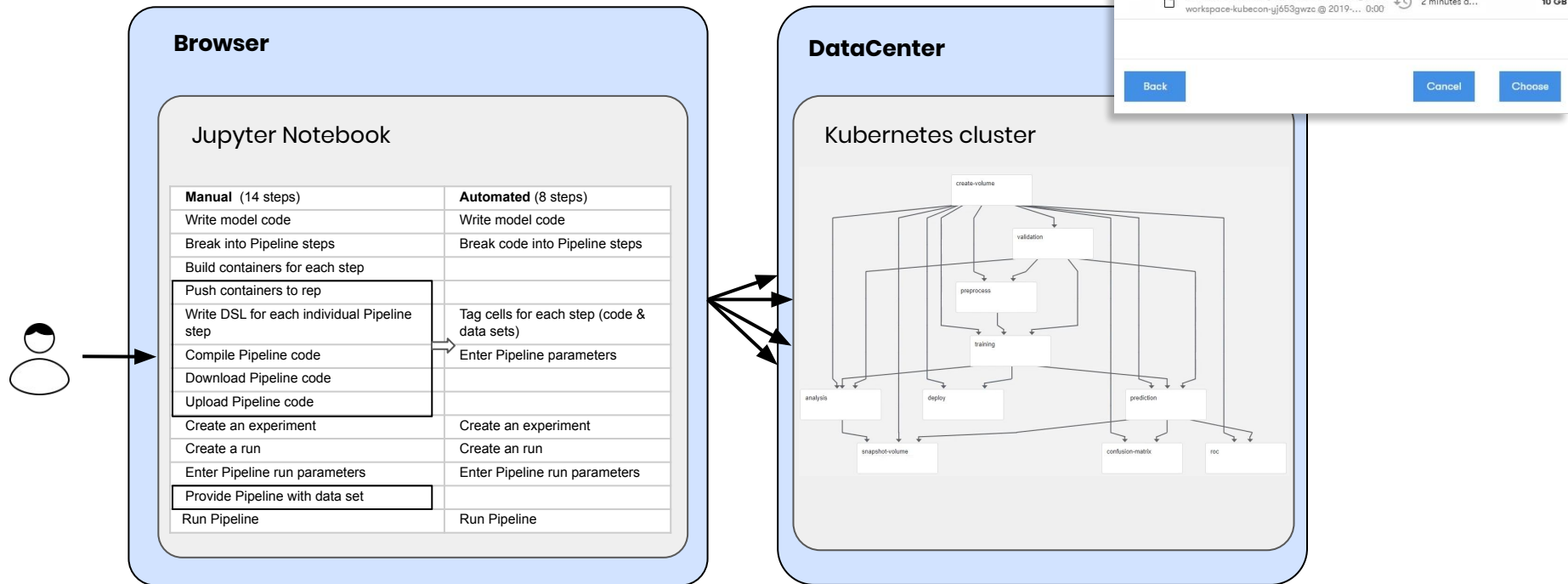
# POC – Third attempt



# POC – Fourth attempt

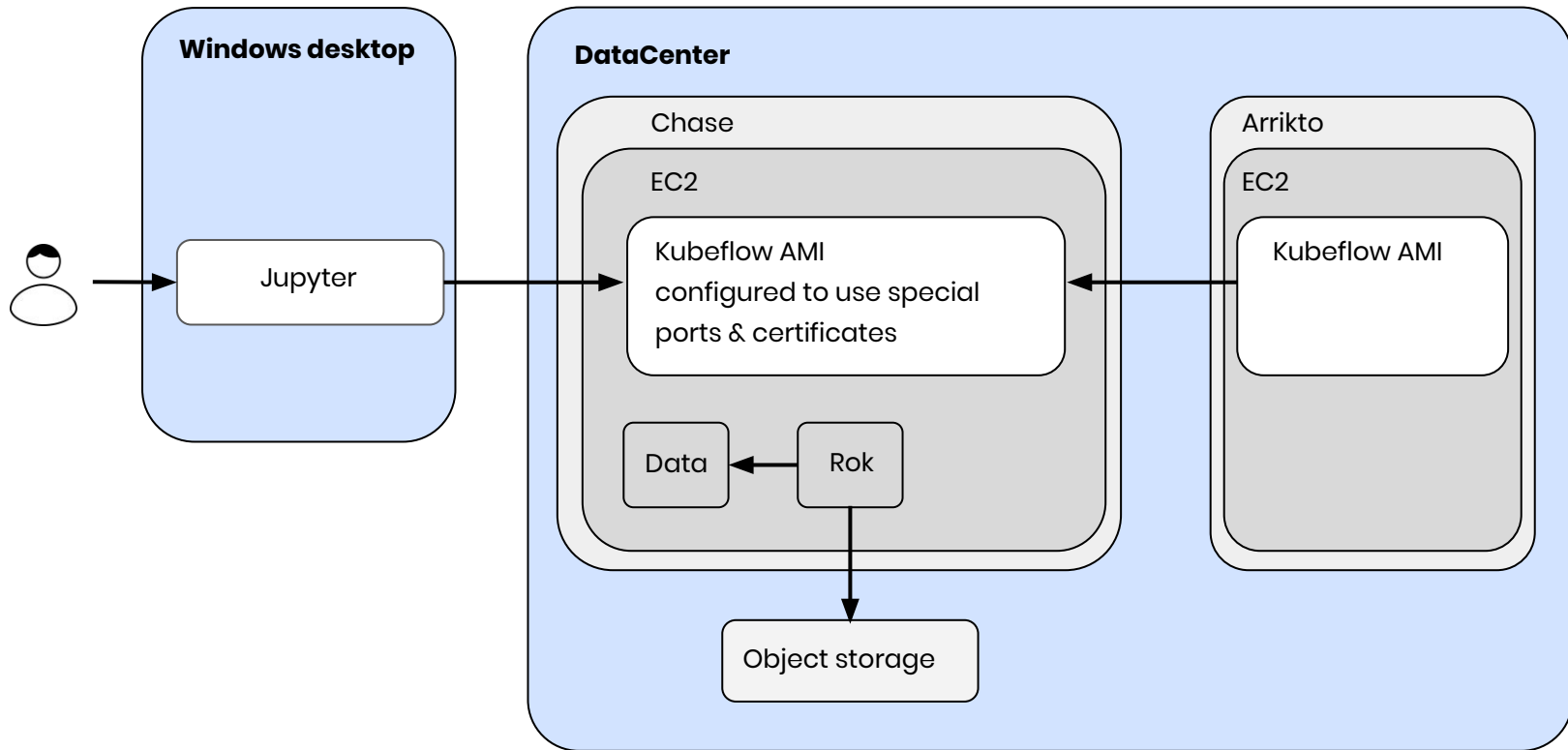


# Success!





# Storage





# On Prem Kubeflow in FinTech

*Jeff Fogarty, US Bank*



# An Enterprise Kubeflow Software Journey

- Define Project's Goals & Requirements
- Review Critical Features & Operations
- Review On-prem Hardware infrastructure
- Collaborate with the Kubeflow community



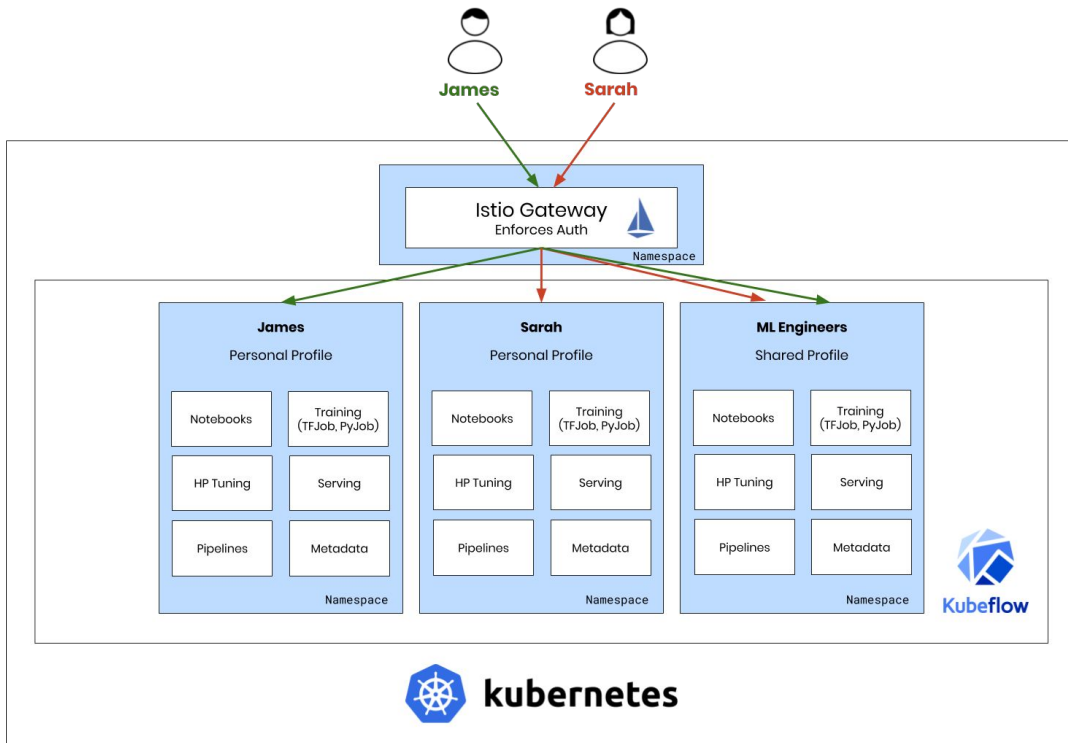
# An Enterprise's Kubeflow Journey

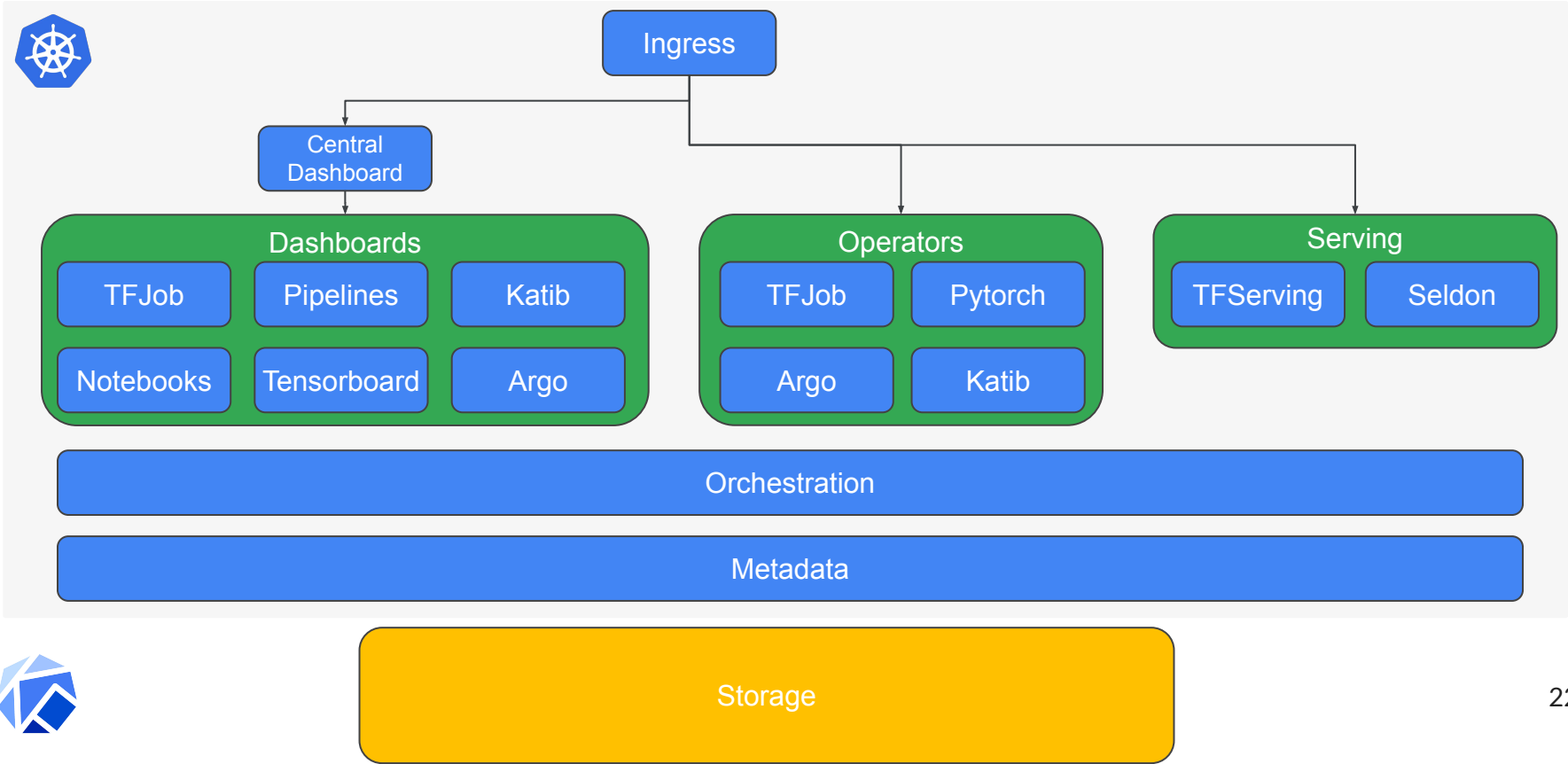
## Goals

- Improve Data Scientist productivity, via a self service ML Platform
  - reproducible model building processes
- Reduce Data Scientist's need to learn K8s, Kubeflow CLI, or KFP DSL



# Isolation – Namespaces/Profiles/Teams







# Data Management

- Tried several storage options
- Deployed Arrikto's Rok on locally-attached NVMe storage
- Versioning of Notebooks along with data
- Ran Pipelines efficiently, seeded with data prepared in Notebook
- Explored versioned Pipeline outputs in Notebooks



# Next Steps

- Explore archiving for storage
  - for code, models and data
- Use Kubeflow to build a reproducible patterns for Pipelines
  - Use this pattern as an example for fast-following teams
- New Kubeflow components (e.g., KFServing)





# Heading towards Kubeflow 1.0

FinTech perspectives on the Kubeflow User Survey

*Thea Lamkin, Google*

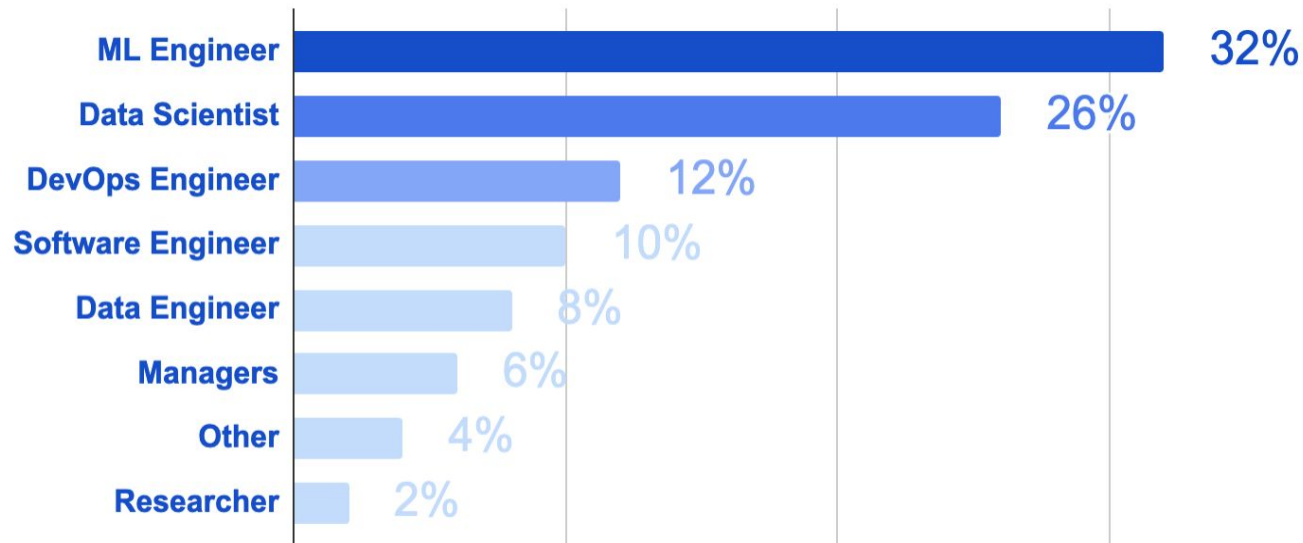
**Kubeflow 1.0 arriving January 2020**

**Provide a stable release of the Core  
CUJ enabling developing, training and  
deploying models from Notebooks**

**More Info: [Kubeflow Roadmap](#)**



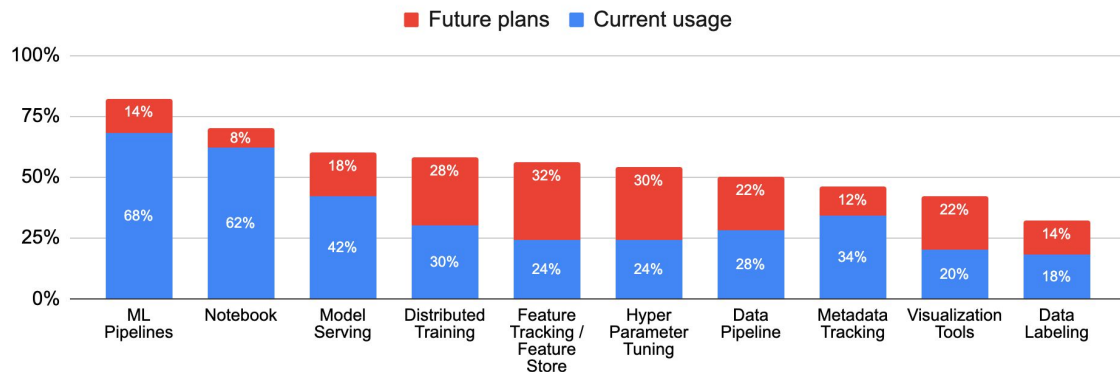
# Kubeflow Survey – Titles





# Kubeflow Survey – Component usage

Current usage and Future plans for Kubeflow components



Top 5 components:

82% Pipelines

70% Notebooks

60% Model Serving

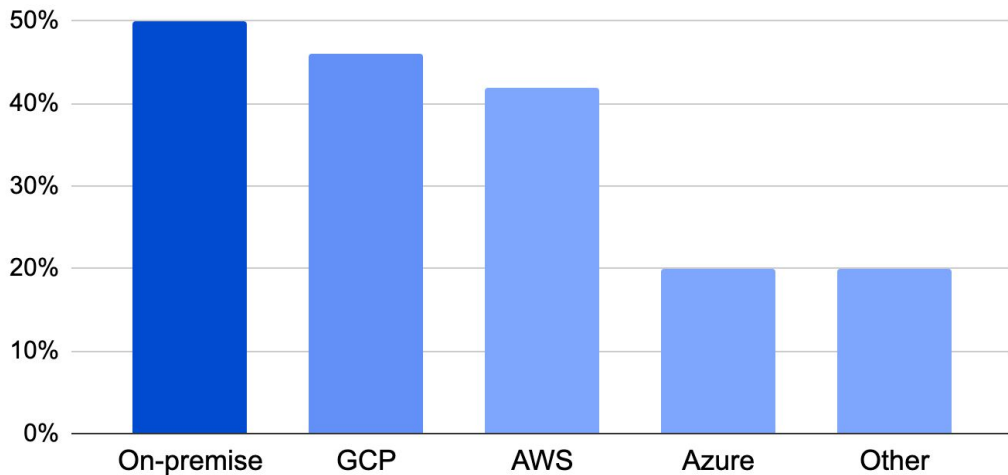
58% Distributed Training

56% Feature Tracking/Store



# Platforms, Processor, Components

Infrastructure type



## Hardware accelerators

78% CPU

72% GPU

## Istio usage

40% Deployed

30% Planning

20% Not Planning

## Identity providers

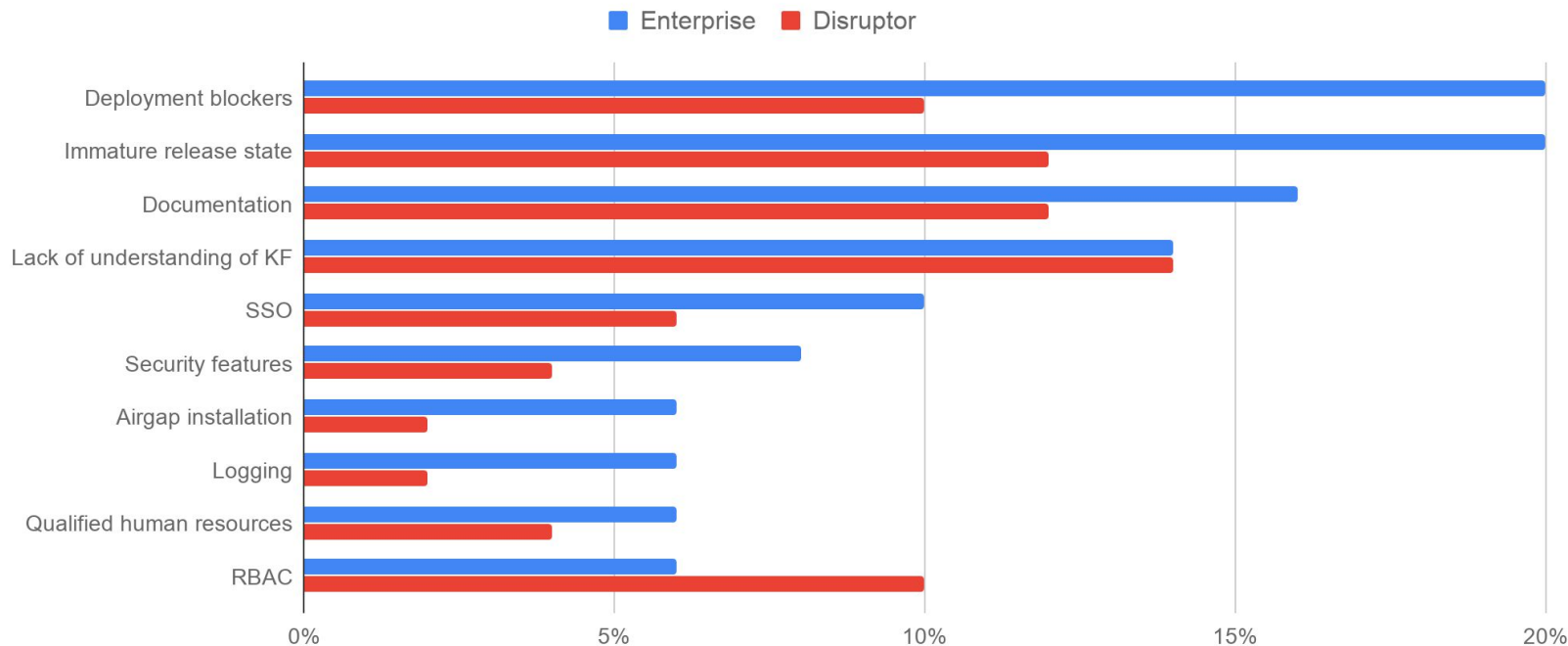
48% Github

42% AD/LDAP



# Do you have advice on these Adoption Blockers ?

## Enterprise vs Disruptor Adoption Blockers





Thank you!