innovation
hub sessions

Organizado por:

pue

MESOSPHERE

16:10 a 16:50h

From DevOps to DataOps

Cristina Álvarez

Sales Engineer de EMEA en Mesosphere

Mesosphere DC/OS

From DevOps to DataOps

MESOSPHERE

Cristina Álvarez– Sales Engineer EMEA

The value of the data

Navigating the next industrial revolution

Revolution	Year	What happened?
1	1784	Steam, water, mechanical production equipment
2	1870	Division of labour, electricity, mass production
3	1969	The computer, electronics and the internet
4	?	The barriers between man and machine dissolve

Machine Learning: Why it matters?

Building precise models, an organization has a better chance of identifying profitable opportunities or avoiding unknown risks.

Health Care
Patient Diagnosis

Finance
Fraud Detection

Manufacturing
Anomaly Detection

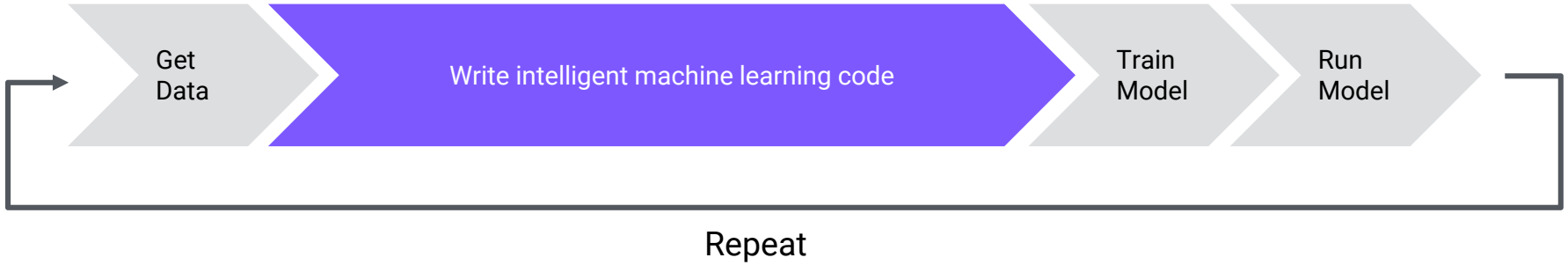
Retail
Inventory Optimization

Media
Interaction and Speed

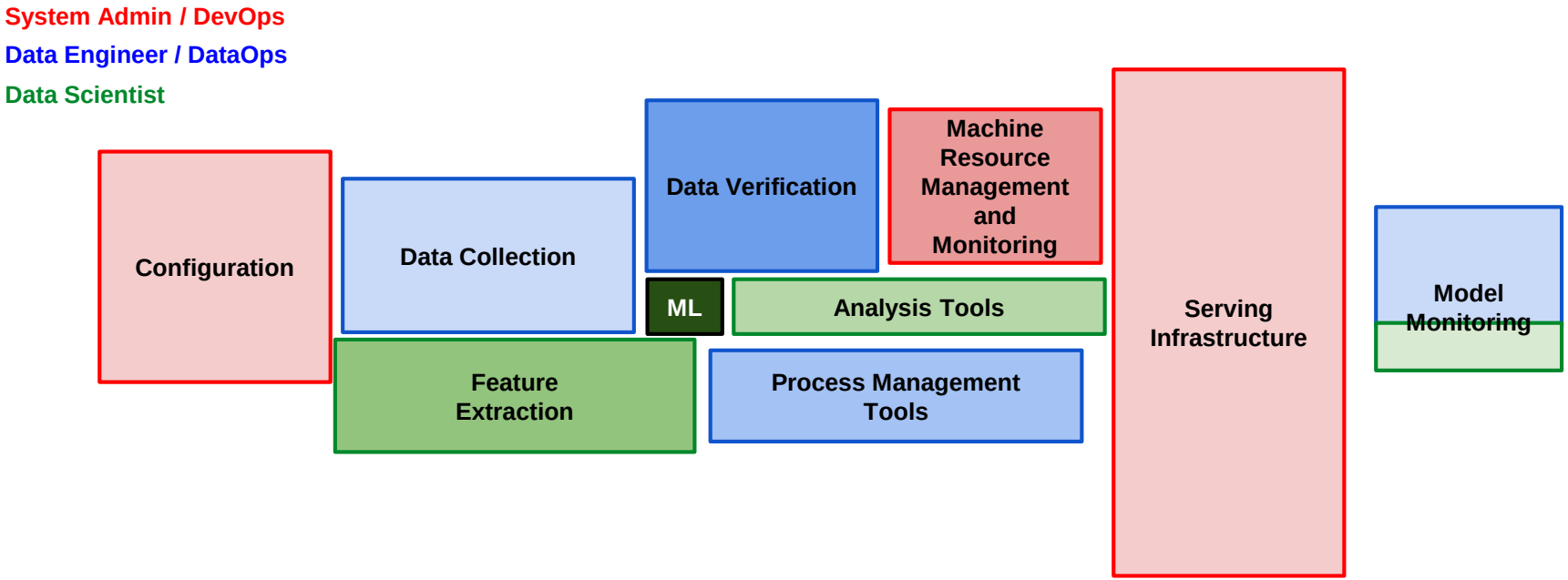
Transportation
Demand Forecasting

eCommerce
Recommender System

What you want to be doing

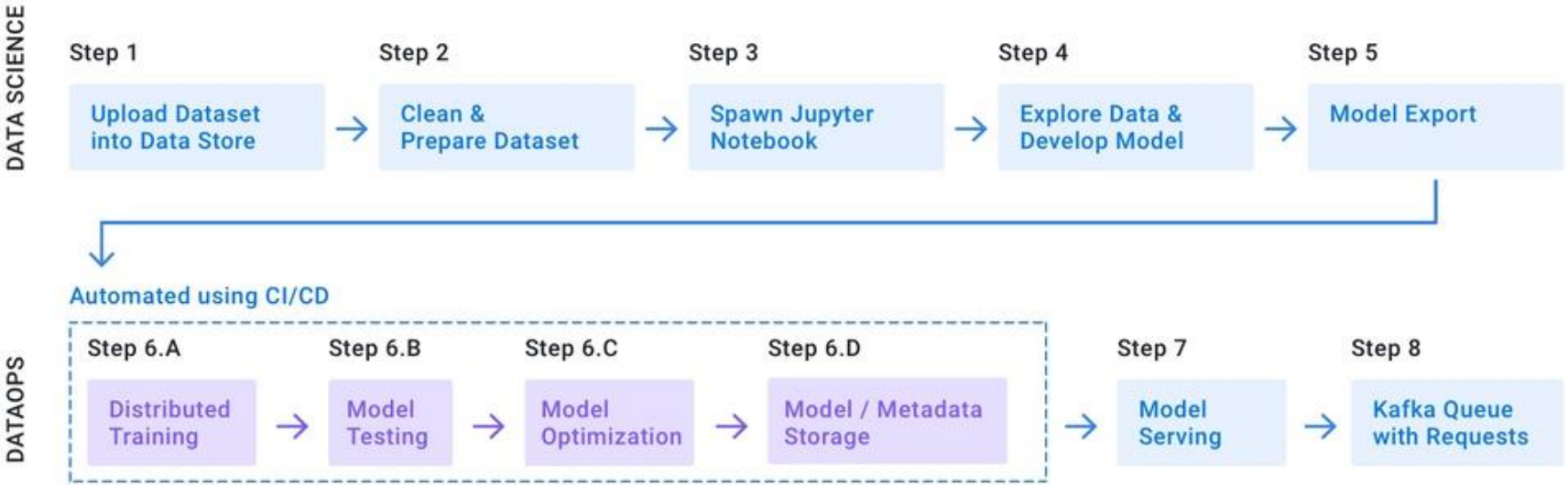


ML Code isn’t the hardest part of Machine Learning Lifecycle

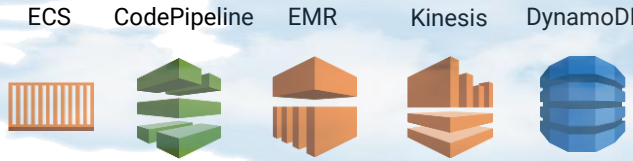


Sculley, D., Holt, G., Golovin, D. et al. Hidden Technical Debt in Machine Learning Systems

Applying Software Delivery Best Practices to Data Science



Digital Transformation Challenges



Locked to a cloud provider

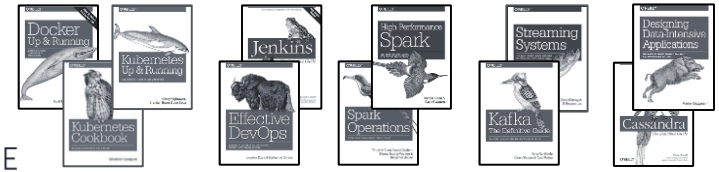
Expensive services

Access to the latest technology is needed

Container orchestration, data services, machine learning/AI tools, & CI/CD toolchains

High cost, low utilization

Cloud-native technologies are distributed systems that require dedicated clusters



Significant training & skills development

Complex IT projects to run containers and data services; developed talent poached by competitors

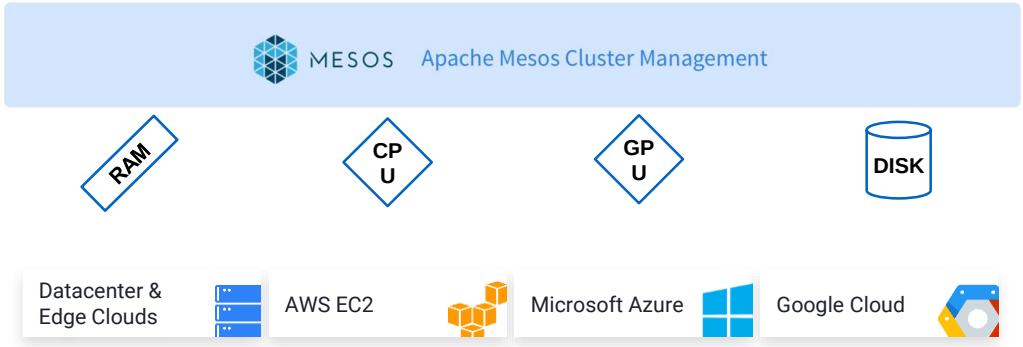
MESOSPHERE

Mesosphere DC/OS can be deployed anywhere

No lock-in

Moving working between clouds or even back on premise

Apache Mesos aggregates & manages all the resources

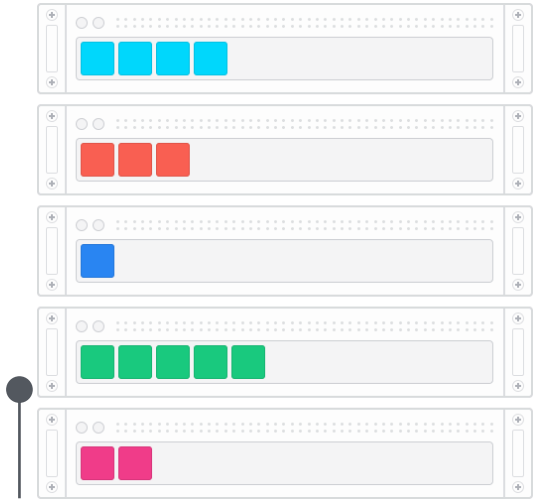


No lock-in

Moving working between clouds or even back on premise

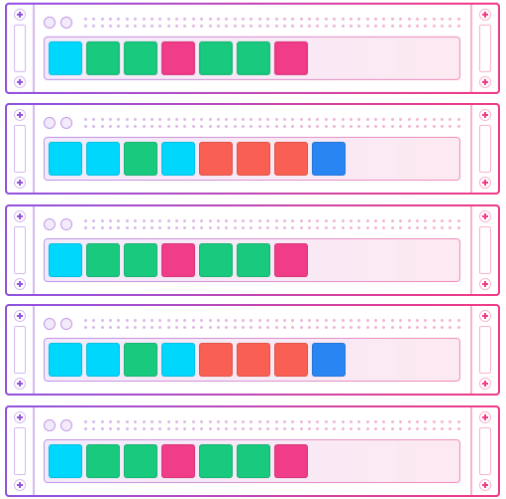
Run everything on the same cluster

- **Kubernetes**
- **Kafka**
- **Jenkins**
- **Cassandra**
- **Spark**



Industry Average
12-15% utilization

Typical Datacenter
siloed, over-provisioned servers, low utilization



DC/OS Multiplexing
30-40% utilization, up to 96% at some customers

4X

DC/OS Datacenter
automated schedulers, workload multiplexing onto the same machines

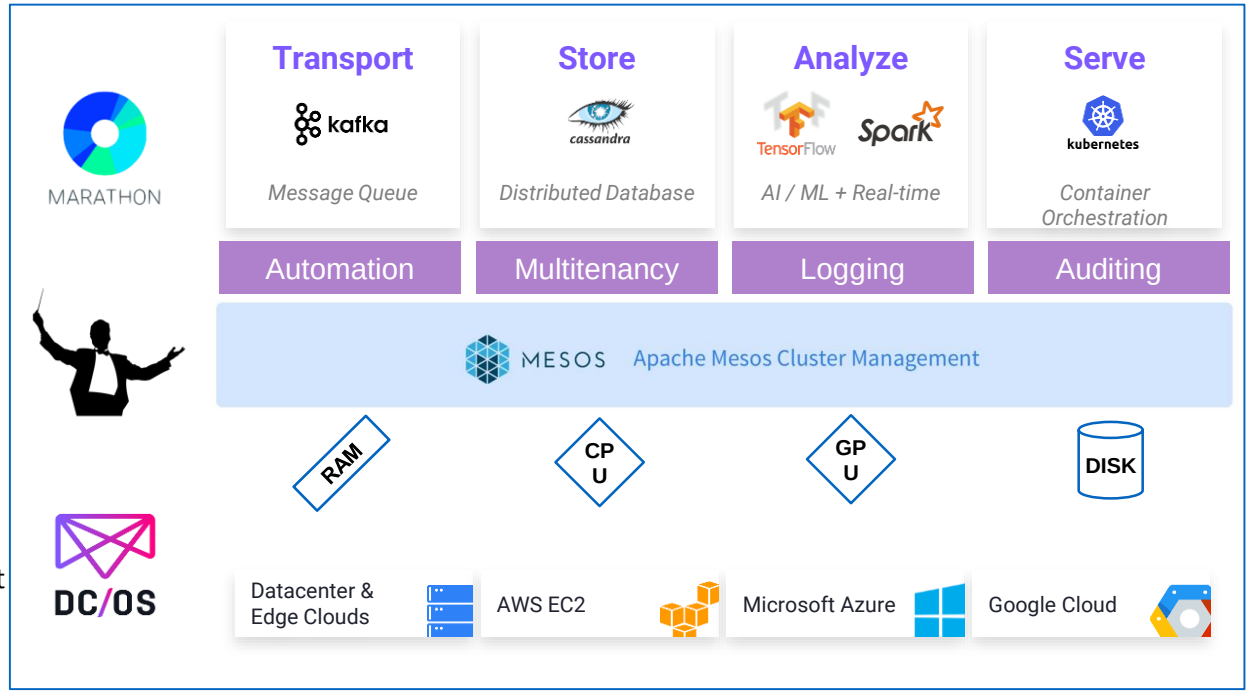
DC/OS provides everything you need as a Service

Reduce management cost

Automating the deployment, upgrade, recovery, …

Reduce infrastructure cost

Deploying all the services on the same hardware



Faster time to market

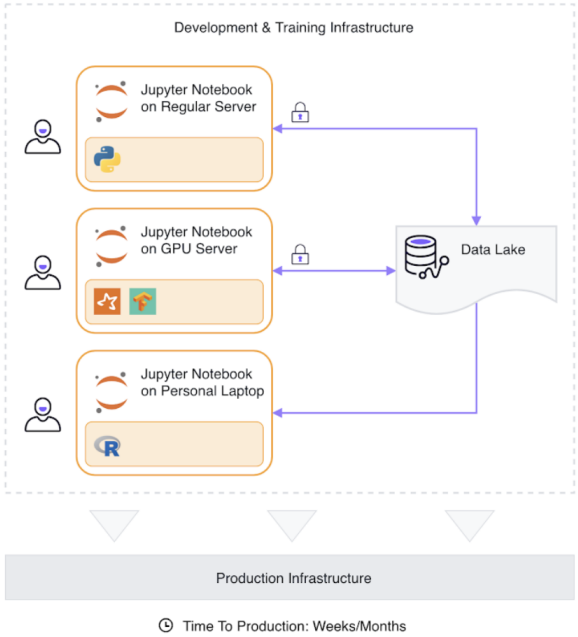
Providing the latest technology in few clicks

No lock-in

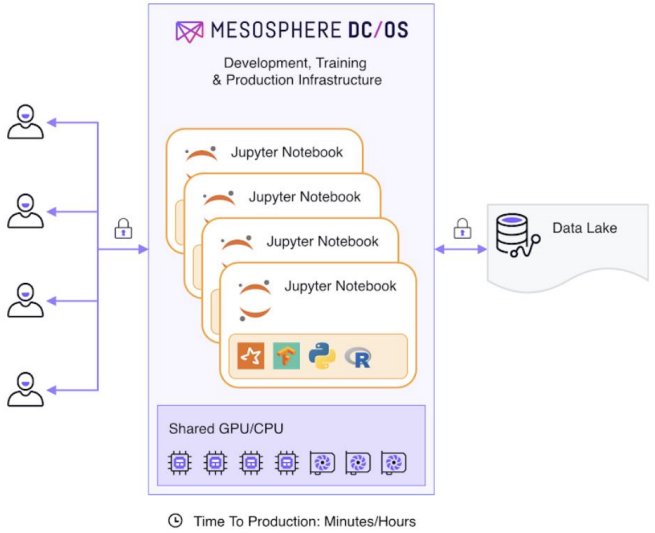
Moving working between clouds or even back on premise

All you need “as a Service” with Mesosphere support

Accelerate Machine Learning with Dramatically Lower Cost



- Ad-hoc provisioning
- Underutilized resource and data infrastructure
- Higher risk of data loss



- Jupyter Notebook as-a-Service
- Secure Collaboration
- Accelerated prototyping and deployment

Demo

flickr

1

Get Pictures on 2 categories using the Flickr API

hadoop HDFS

2

Persist some pictures in HDFS

Kerberos TLS

APACHE nifi

Produce other pictures in Kafka

3

Retrain an Image Classifier for New Categories

python

kafka

5

Configure a CI/CD pipeline to build and deploy a Docker image to serve the model

TensorFlow

jupyter

Provide a notebook experience to the end user

Consume pictures from Kafka

kubernetes

Jenkins

GitLab

4

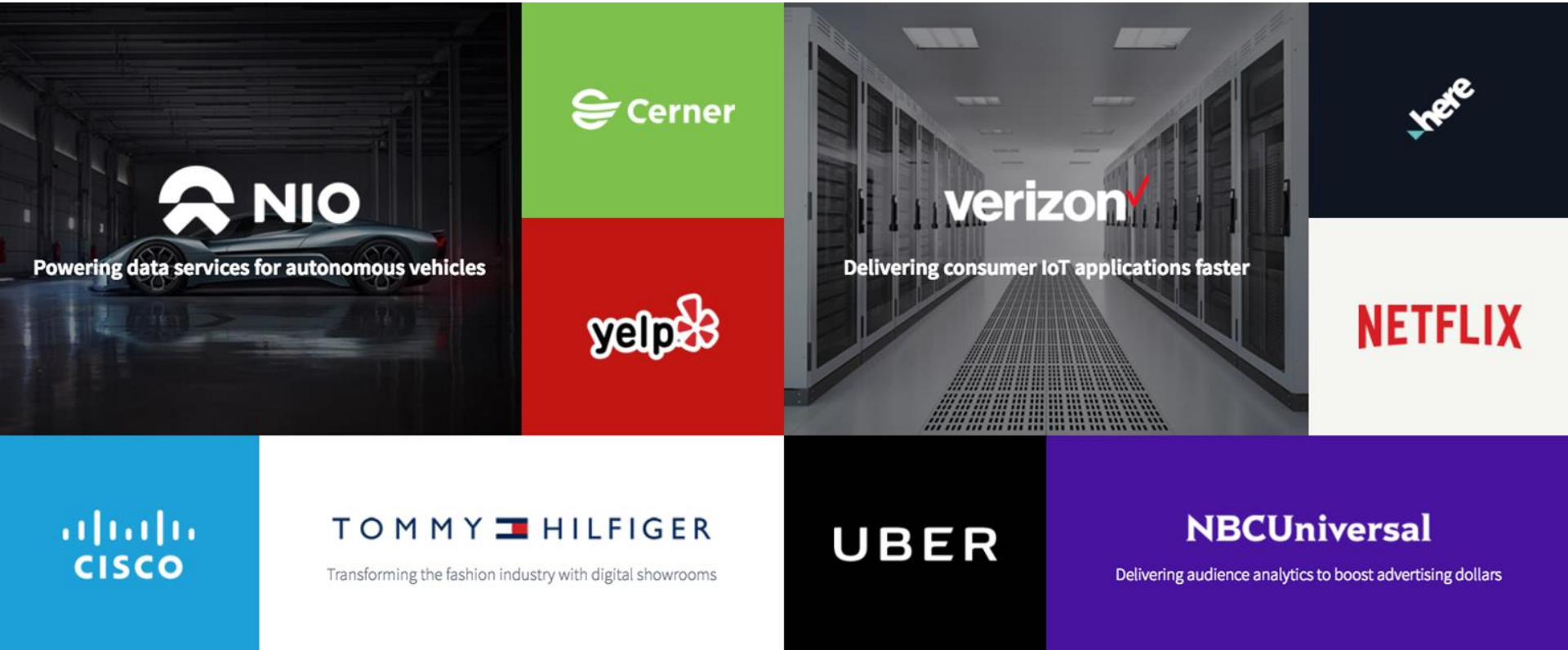
Push the model in a git repository

What have we demonstrated ?

- Mesosphere DC/OS provides all the software you need to create a Secure ML pipeline
- It can deploy each of them in few minutes
- It can secure all the communications with Kerberos and TLS
- It provides the same experience on premise or on any public cloud, at a predictable cost
- It provides a nice notebook experience for the data scientists
- It can leverage GPUs
- Mesos quotas can be leveraged to guarantee resources
- Kubernetes can be used to serve the models

Github: https://github.com/djannot/dcos-demos/tree/master/dcos-secure-pipeline

Mesosphere powers data-driven applications across industries



Transforming traditional carmakers

Goal:
One of the largest German auto manufacturers enables connected cars with Integrated data services like real-time traffic information and sensor data .

Challenges:

- Transporting, analyzing, and serving huge amounts of data across a distributed architecture
- Support for a vast number of different configurations of applications, technologies and car models
- Collaboration among many car brands, BUs and infrastructure types

Solution:

- By leveraging DC/OS as the foundation, this traditional carmaker gained an open platform that is free from vendor lock-in and supports an open ecosystem of essential data services.
- DC/OS's ability to support the delivery of software via containers also ensures that applications built using DC/OS can be installed on any type of vehicle, because containers abstract the underlying hardware away from the application.

Global Automotive Leader's Connected Car Infrastructure

Telecom Provider Local Edge Cloud

Microservices

Message Queue

MESOSPHERE DC/OS

- Localized route optimization stack
- Data Ingestion at scale
- Short latency response

Corporate Regional Datacenters

Container Orchestration

Analytics Engine

Distributed Database

Distributed File System

Message Queue

Search

CI/CD Automation

MESOSPHERE DC/OS

- Fast new service rollout with elastic CI/CD toolchain
- Telemetry analytics on vehicle usage for preventative maintenance
- Compliance with data privacy regulations (e.g., customer opt-out)
- Support mobile on-demand car club service

REST API

Third Party Mapping Provider Cloud

Container Orchestration

CI/CD Automation

Distributed Database

Message Queue

MESOSPHERE DC/OS

- Fast new service rollout with elastic CI/CD toolchain
- Container-based infrastructure

THANK YOU!

MESOSPHERE

Learn more at mesosphere.com

El gran reto del Big Data: la integración continua

Sergio Rodríguez de Guzmán, CTO PUE

Salón ROMA
16:55 – 17:55