Microsoft

DevOps for Data Science

KC Munnings and Jack Xue
Microsoft Cloud Solution Architect

1

Azure Indy User Group Meeting

January 24, 2019

What does "DevOps for DS" mean?

DevOps is the standard way to manage application lifecycles through a pipeline of code, test, build, deploy (CI/CD cycle).

**Infusing DS** into this lifecycle brings new challenges and changes to the DevOps pipeline.

A CI/CD solution for ML requires supporting:
- **Reproducibility** of data → model
- **Validation** of model (does it meet quality bar, A/B comparison)
- **Storage, versioning** (track lineage and evolution of model over time)
- **Deployment, tracking, data collection** (across intelligent cloud + edge)

The journey

App Developers and Data Scientists
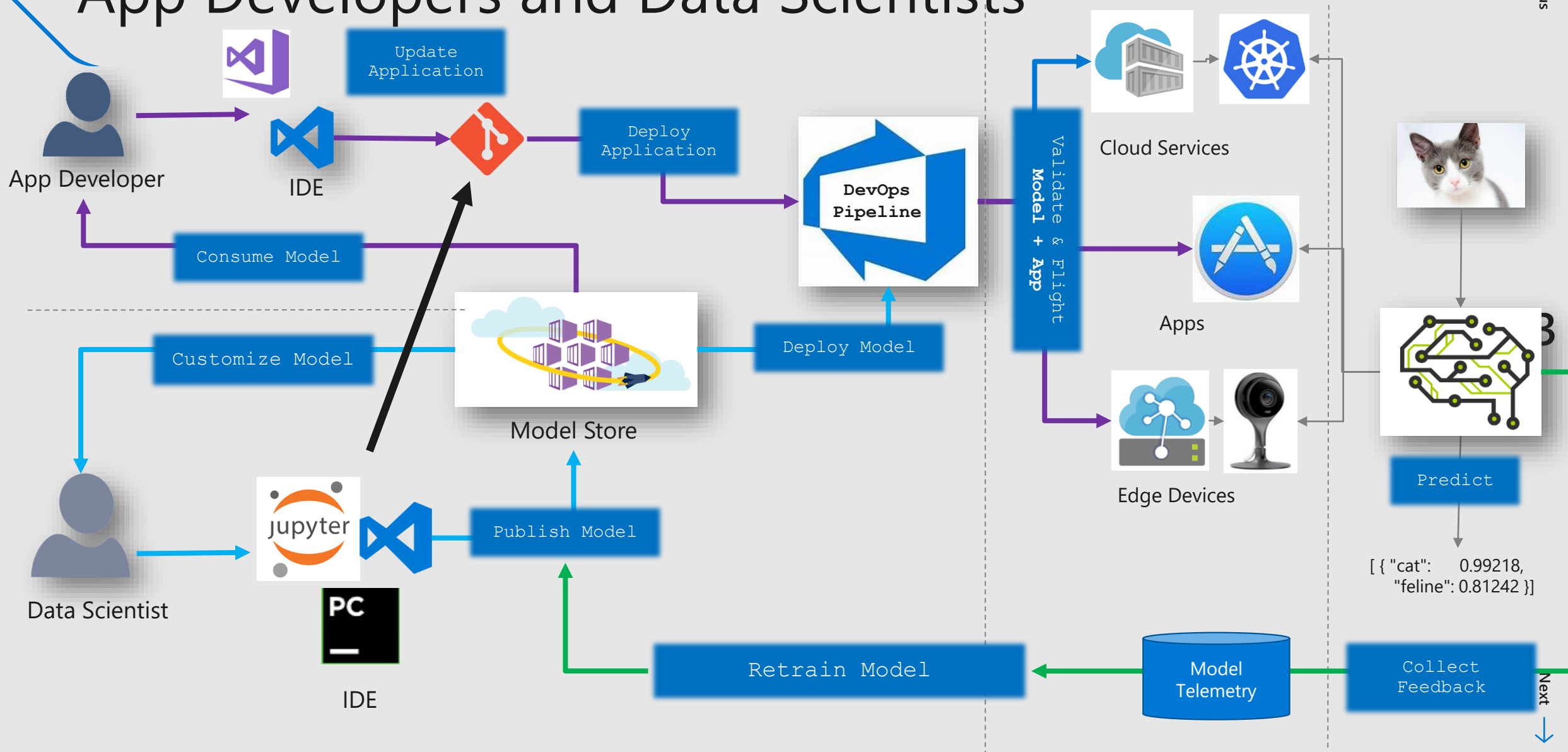
Azure DevOps

Better together

Azure Boards

Azure Repos

Azure Pipelines

Azure Test Plans

Azure Artifacts

An end-to-end solution for organizations looking for an enterprise-grade toolchain

4

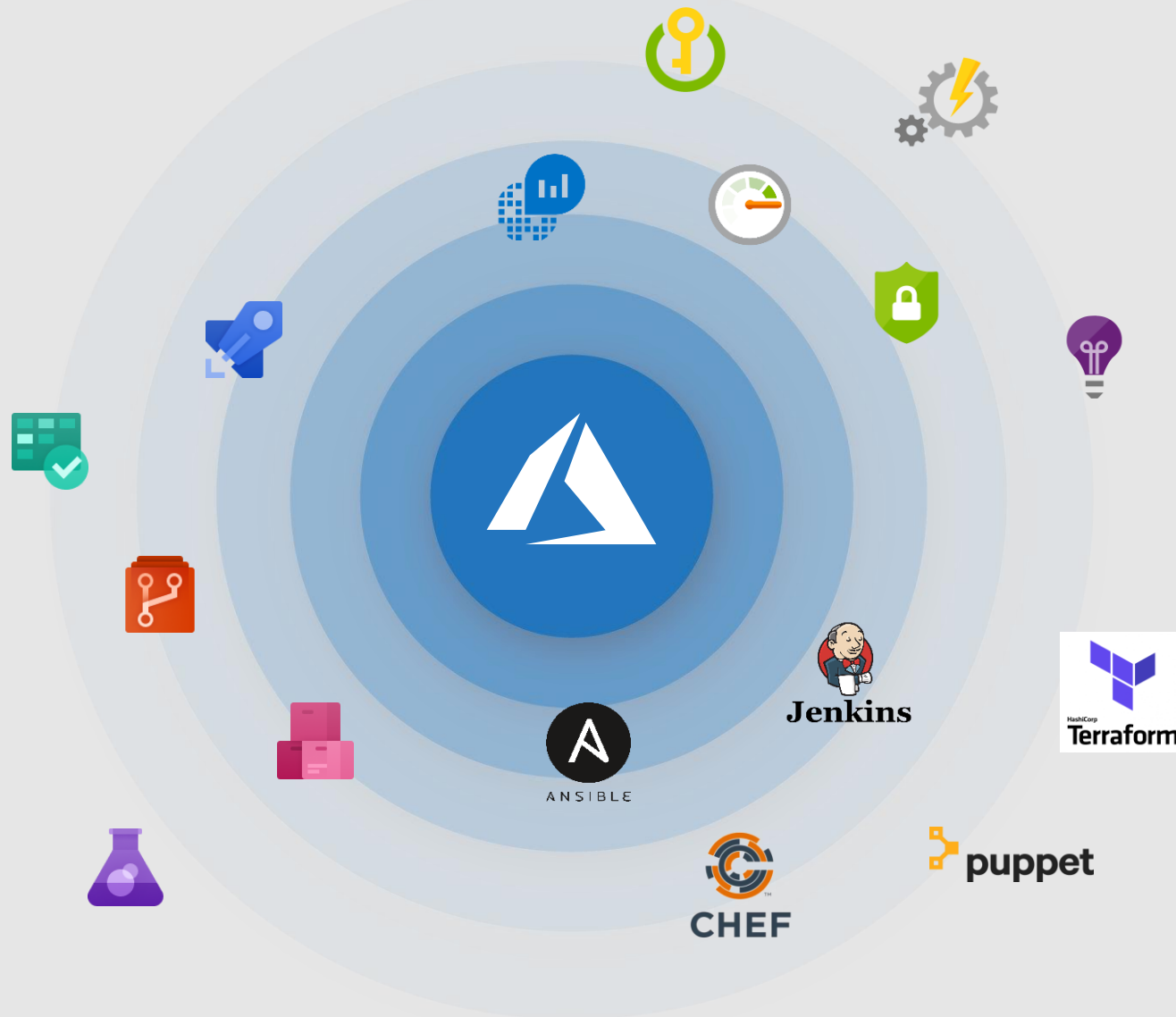| Fully Integrated with end to end traceability	Scalable to any team and project size	Highly available, multi region, hybrid cloud & on-prem	Customer Support	Consistent admin and access control

→ https://azure.com/devops

Broadening the Azure Ecosystem

Jenkins

Terraform

ANSIBLE

CHEF

puppet

DevOps at Microsoft

Azure DevOps is the toolchain of choice for Microsoft engineering with over 90,000 internal users

https://aka.ms/DevOpsAtMicrosoft

372k
Pull Requests per month

4.4m
Builds per month

5m
Work items viewed per day

6

2m
Git commits per month

500m
Test executions per day

500k
Work items updated per day

78,000
Deployments per day

Data: Internal Microsoft engineering system activity, August 2018

Demo

- Azure Machine Learning Service – Dec 2018
  - For Python developer only
  - For R developer is H1 2019
- This demo is targeted to a R shop
  - Batch mode
  - Online mode
    - Containerized and
    - Deployed in Kubernetes
- CI/CD pipeline for models in Azure DevOps