EARL London | 10-12 September, 2019

A DevOps Process for deploying R to Production

David Smith, Microsoft

Microsoft

A DevOps process for deploying R to production

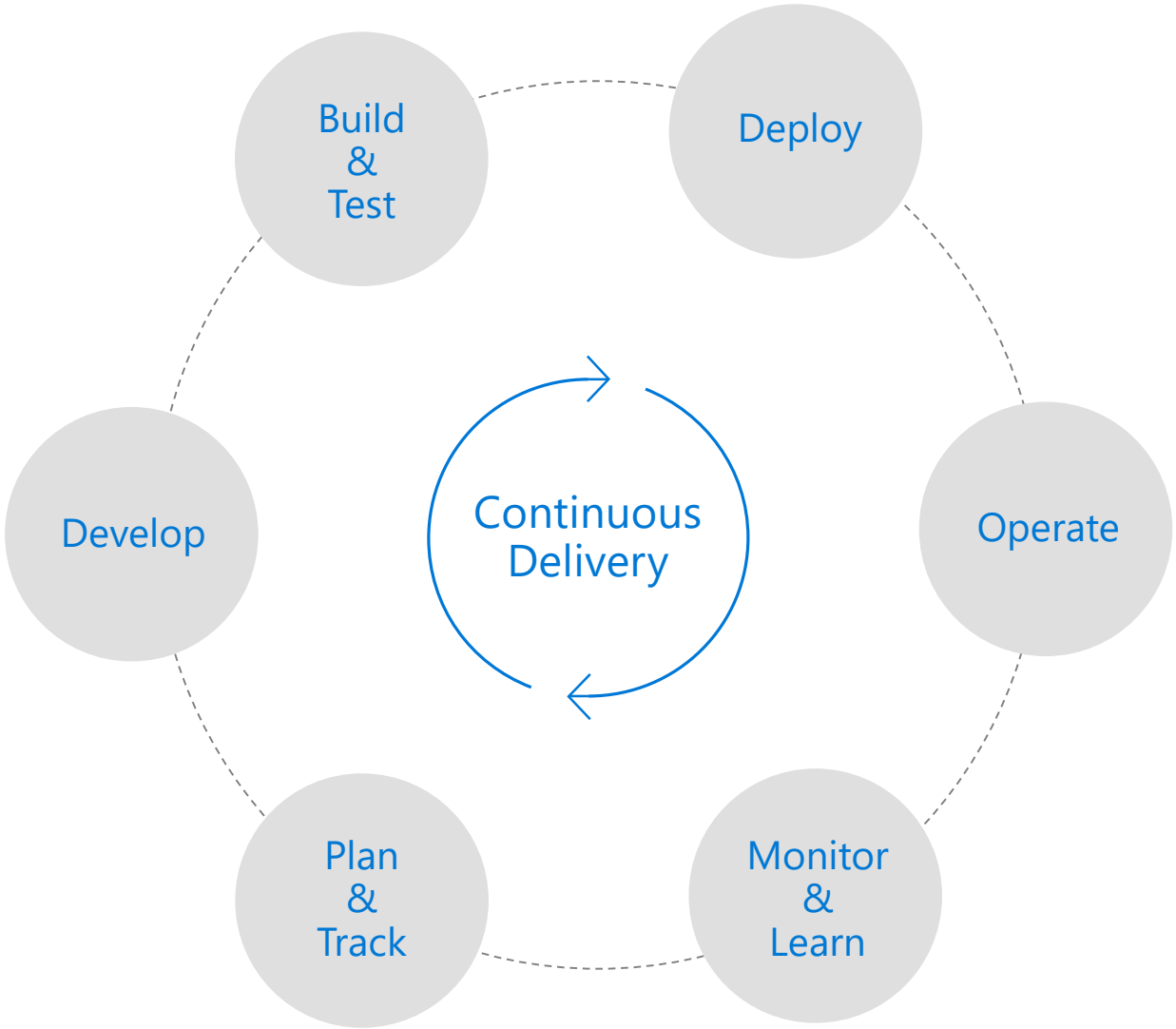
David Smith
Cloud Advocate, Microsoft
@revodavid

What is DevOps?

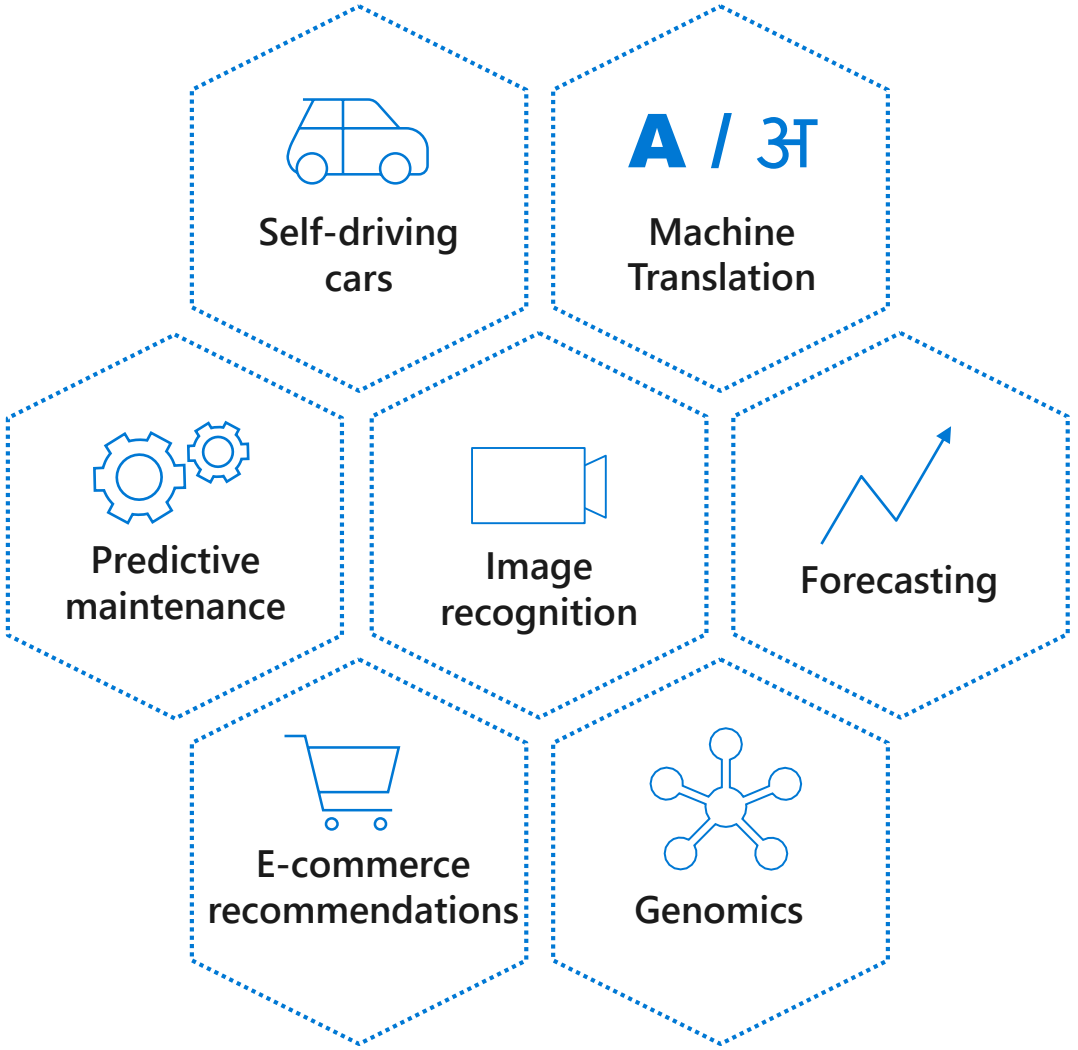
People. Process. Products.

> DevOps is the union of **people**, **process**, and **products** to enable continuous delivery of value to your end users.

Donovan Brown, Microsoft
http://bit.ly/WhatIs-DevOps



Machine Learning Applications



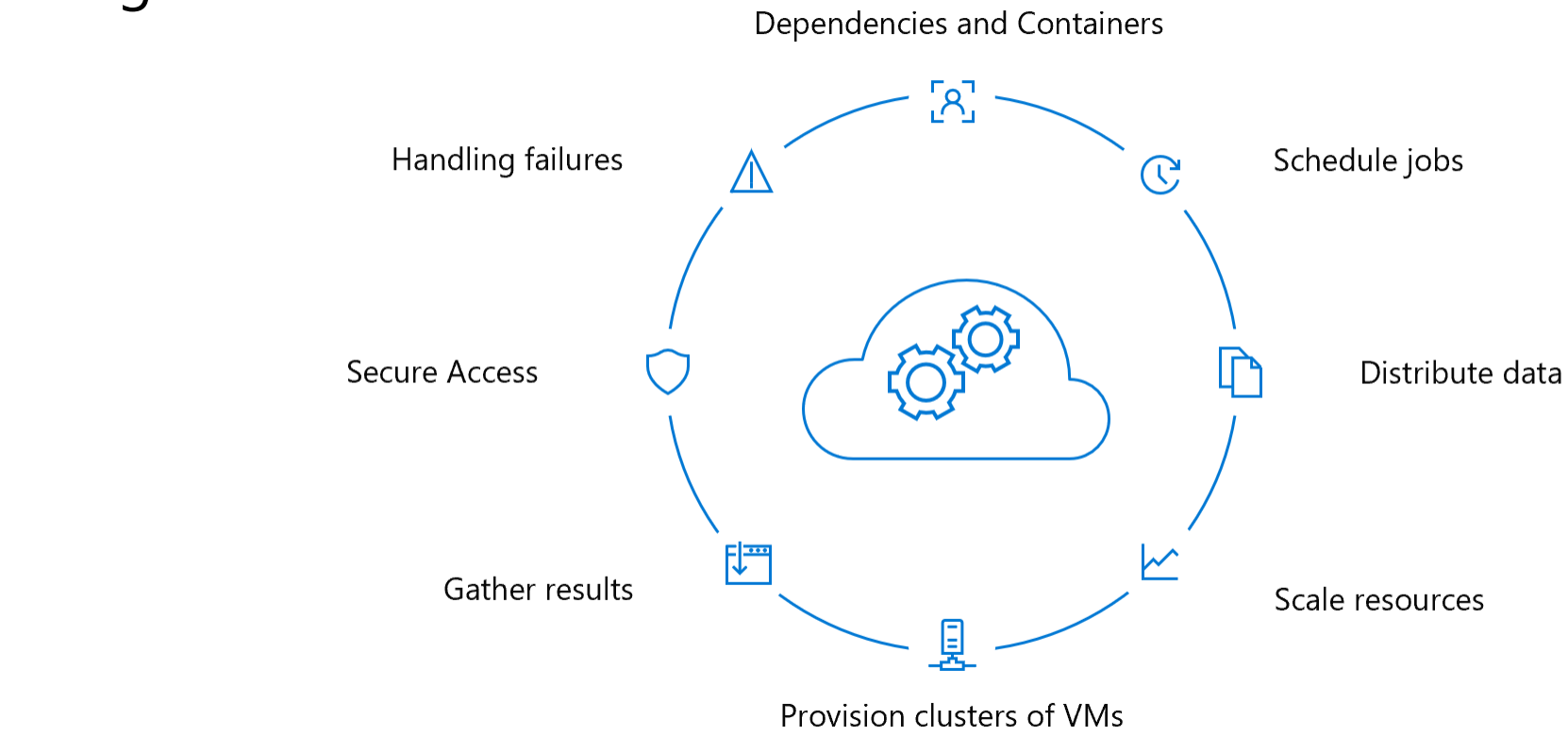
Example: Automatic windscreen wipers

Special considerations for MLOps

People: Data Scientists, ML Engineers

Process:

- Data as Code
- Model Lifecycle
- Tests
- Deployed APIs



Products:

- Hardware: Storage, GPUs
- Software: Datastores, Analytics Tools (R, Python, Tensorflow…)

Azure Pipelines

Cloud-hosted pipelines for Linux, Windows and macOS, with unlimited minutes for open source

Any language, any platform, any cloud

Build, test, and deploy Node.js, Python, Java, PHP, Ruby, C/C++, .NET, Android, and iOS apps. Run in parallel on Linux, macOS, and Windows. Deploy to Azure, AWS, GCP or on-premises

Extensible

Explore and implement a wide range of community-built build, test, and deployment tasks, along with hundreds of extensions from Slack to SonarCloud. Support for YAML, reporting and more

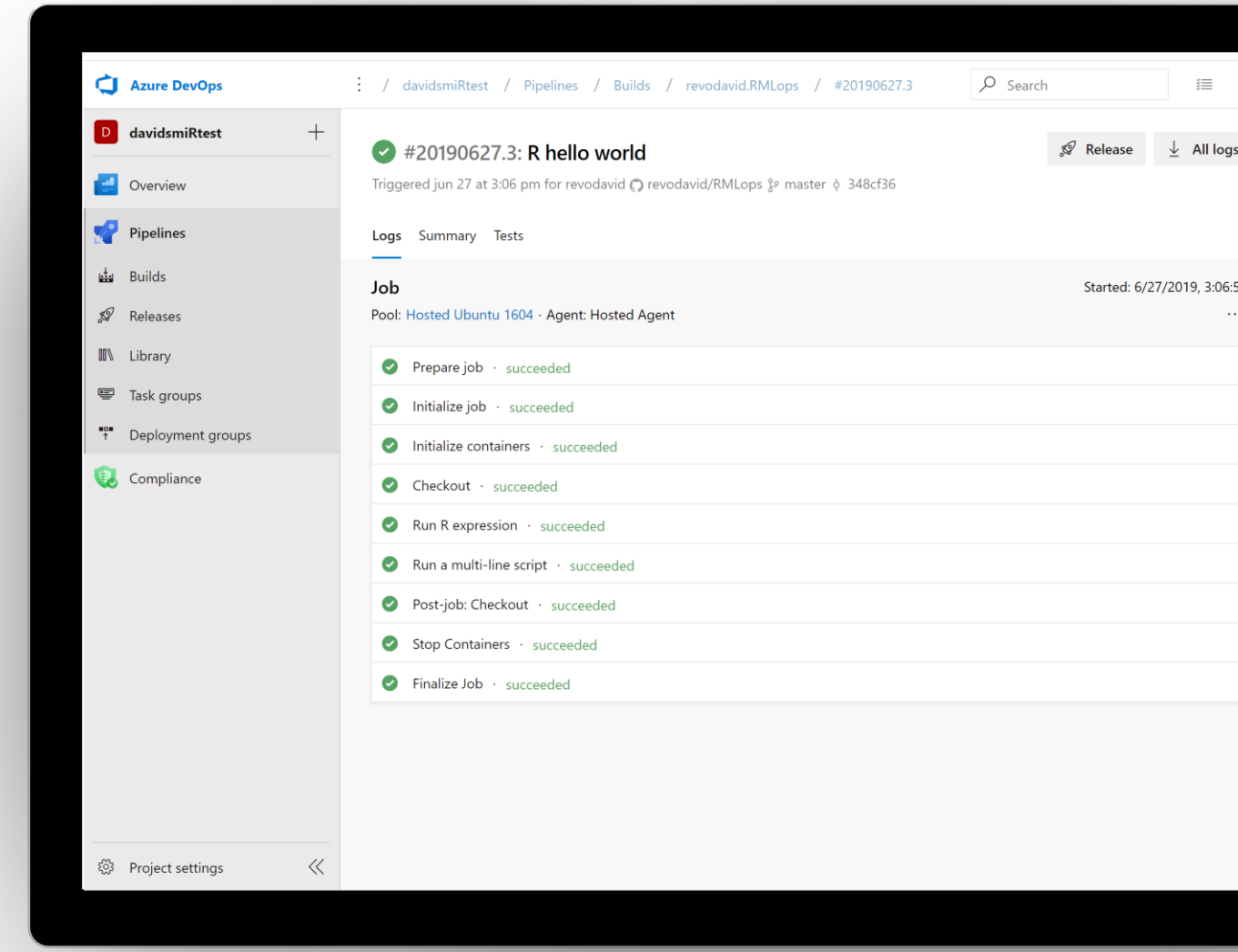
Containers and Kubernetes

Easily build and push images to container registries like Docker Hub and Azure Container Registry. Deploy containers to individual hosts or Kubernetes.

Best-in-class for open source

Ensure fast continuous integration/continuous delivery (CI/CD) pipelines for every open source project. Get unlimited build minutes for all open source projects with up to 10 free parallel jobs across Linux, macOS and Windows

https://azure.com/pipelines



Training R Models with Azure Pipelines

In Github Repo:

github.com/revodavid/RMLOps

Data files

R scripts

YAML Pipeline specification

Status Badge 

In Azure Pipelines:

dev.azure.com/
davidsmi0786/davidsmiRtest

Triggers

Build pipeline(s)

Release pipeline(s)

Azure Pipelines: Process

Create Azure Pipeline with Github repository

Link RStudio project to Github repository

Develop R code as R script files as usual

Add steps to azure-pipelines.yml with Rscript commands

Check in updates to trigger builds

Branch: master ▾ RMLops / azure-pipelines.yml

```
 9  pool:
10    vmImage: 'ubuntu-16.04'
11
12  container: 'rocker/r-ver:3.6.0'
```

```
14  steps:
15  - script: Rscript -e 'R.version'
16    displayName: 'Confirm R version'
17
18  - script: Rscript train-model.R
19    displayName: 'Train model'
```

```
6  trigger:
7  - master
```

D davidsmiRtest

Overview

Pipelines

Builds

Releases

Library

Task groups

Deployment groups

Compliance

Project settings

#20190711.5: save correct model

Triggered today at 9:57 am for revodavid revodavid/RMLops master 1b59b32

Release Artifacts

Logs Summary Tests

Job

Started: 7/11/2019, 9:57:22 AM

Pool: Hosted Ubuntu 1604 · Agent: Hosted Agent

··· 7m 27s

Prepare job · succeeded	<1s
Initialize job · succeeded	2s
Initialize containers · succeeded	31s
Checkout · succeeded	2s
Confirm R version · succeeded	1s
Install packages · succeeded	6m 14s
Train model · succeeded	1s
PublishPipelineArtifact · succeeded	5s
Run a multi-line script · succeeded	<1s
Component Detection (auto-injected by policy) · succeeded	24s
Post-job: Checkout · succeeded	<1s
Stop Containers · succeeded	1s
Finalize Job · succeeded	<1s

Containers and Azure Pipelines

Specify a container with your desired R version, packages, and any other software needed.

Container images can be hosted anywhere

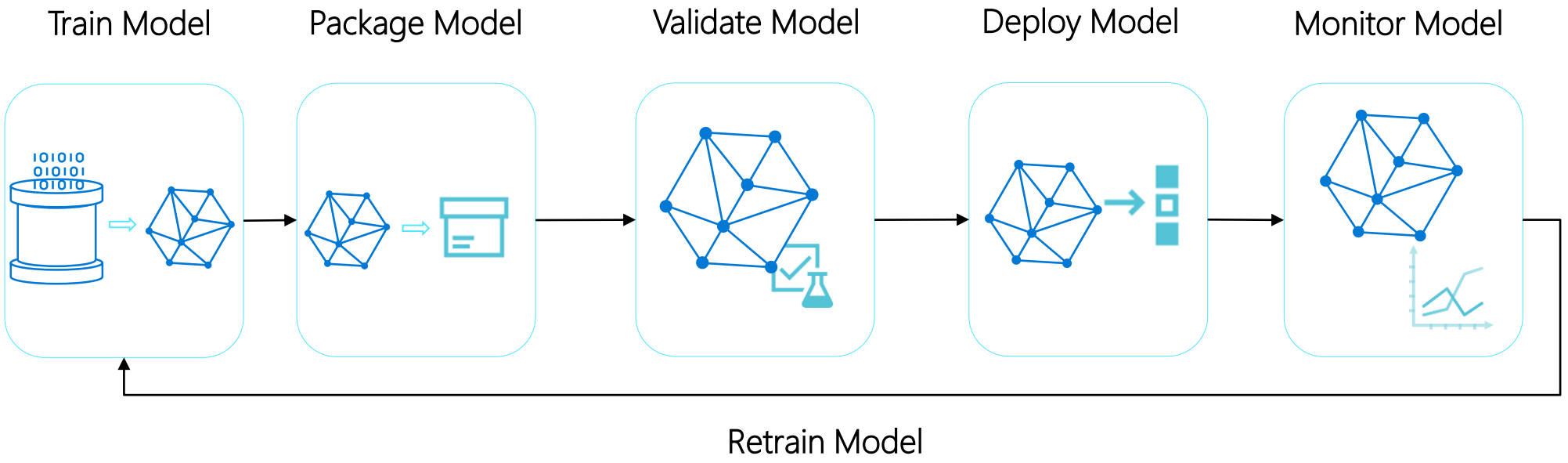
- Docker Hub, your own container registry, Azure Container Registry …

Tip 1: The "Rocker" project provides pre-defined R containers for training: www.rocker-project.org

Tip 2: You can also deploy with containers using the AzureContainers package: blog.revolutionanalytics.com/2018/12/azurecontainers.html

The End-to-End Machine Learning Model Lifecycle

- **Develop & train model** with reusable ML pipelines
- **Package model** using containers to capture runtime dependencies for inference
- **Validate model behavior—**functionally, in terms of responsiveness, in terms of regulatory compliance
- **Deploy model—**to cloud & edge, for use in real-time/streaming/batch processing
- **Monitor model** behavior & business value, know **when to replace/deprecate a stale model**



What is Azure Machine Learning service?

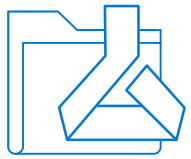
Set of Azure Cloud Services

Python SDK

That enables you to:

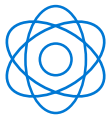
- ✓ Prepare Data
- ✓ Build Models
- ✓ Train Models
- ✓ Manage Models
- ✓ Track Experiments
- ✓ Deploy Models

Azure Machine Learning Service

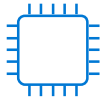


Workspace

Data Stores



Experiments



Compute Target

HyperDrive

Models

Images

Deployments

Use any IDE / compute framework via Python SDK

Share compute resources

Containerized deployments

Monitor performance

Reproducible cloud collaboration

ml.azure.com

Coming Soon: Azure ML Service SDK for R

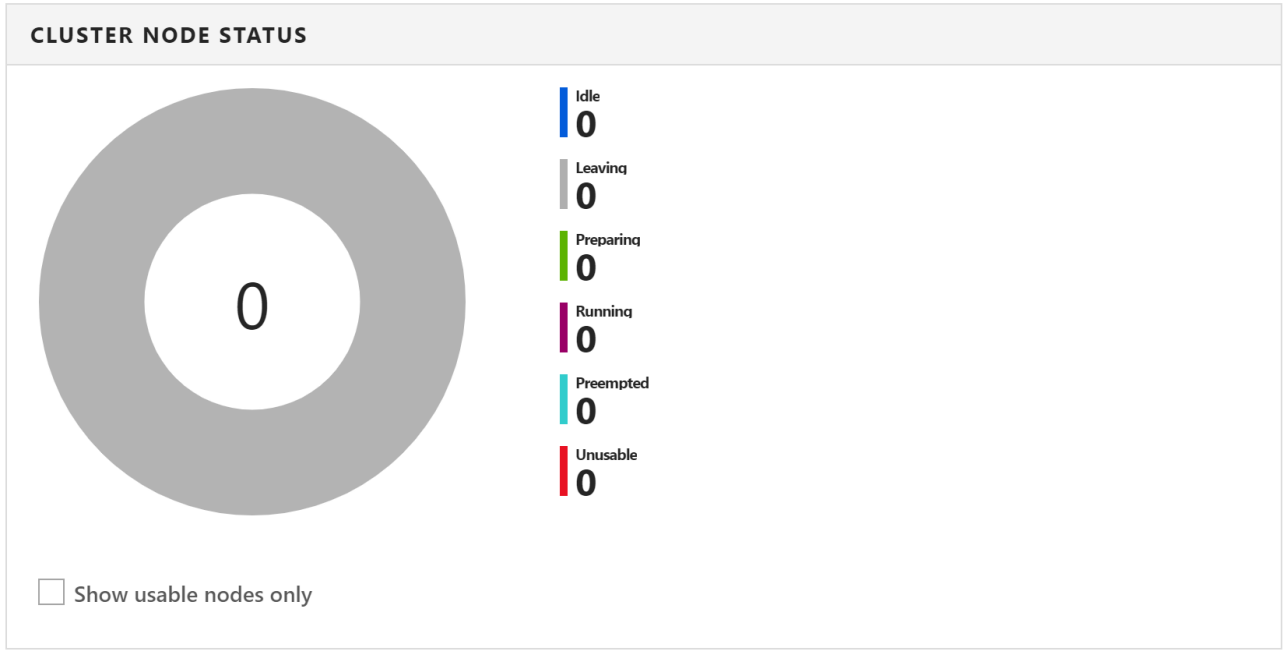
New open-source R package for use with CRAN R

- Create Workspaces, Experiments, Compute, and other artefacts with R commands.
- Use any R function/package, while tracking dependencies for deployment
- Package models into containers for deployment as web-services for predictions
- HyperDrive support: smart hyperparameter search with parallel compute
- GPU support (useful for tensorflow, keras packages)

Create Compute

Create max 4-node cluster in Azure for training:

```
compute_target <- create_aml_compute(
  workspace = ws, cluster_name = "caret-cluster",
  vm_size = "STANDARD_D2_V2", max_nodes = 4)
```

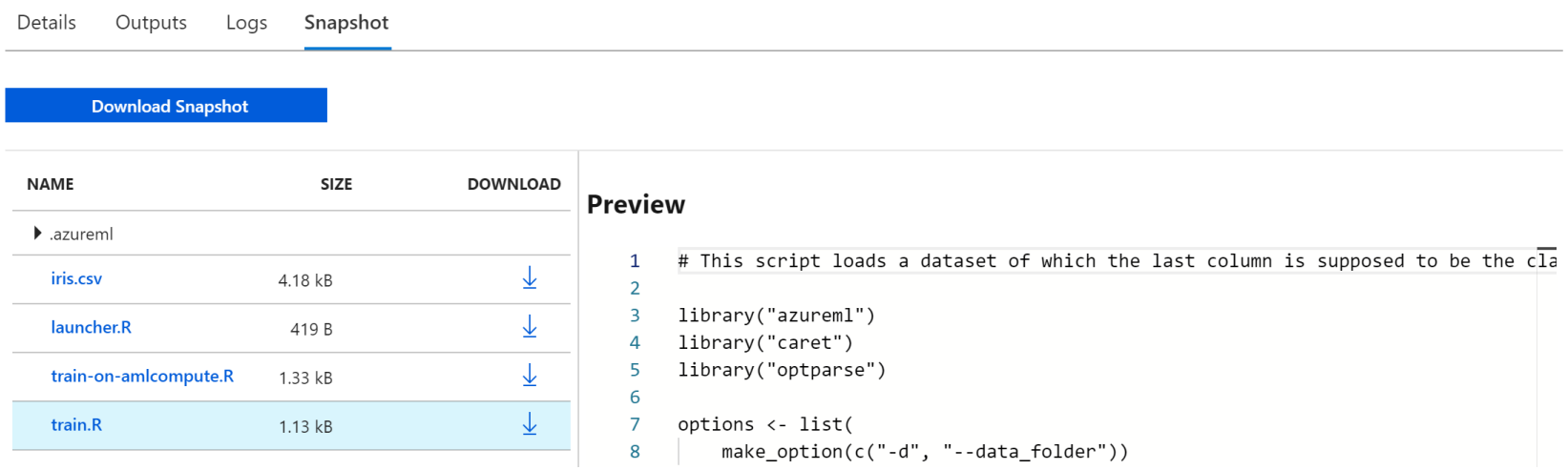


Monitor status in R or via the Azure portal

Experiments and Runs

Train model with caret, log accuracy, and save reproducible environment

```
control <- trainControl(method = "cv", number = 10)
metric <- "Accuracy"
model <- train(response ~ ., data = train_data, method = "lda",
               metric = metric, trControl = control)
predictions <- predict(model, test_data)

conf_matrix = confusionMatrix(predictions, test_data$response)
current_run <- get_current_run()
log_metric_to_run(metric, conf_matrix$overall["Accuracy"], current_run)
```



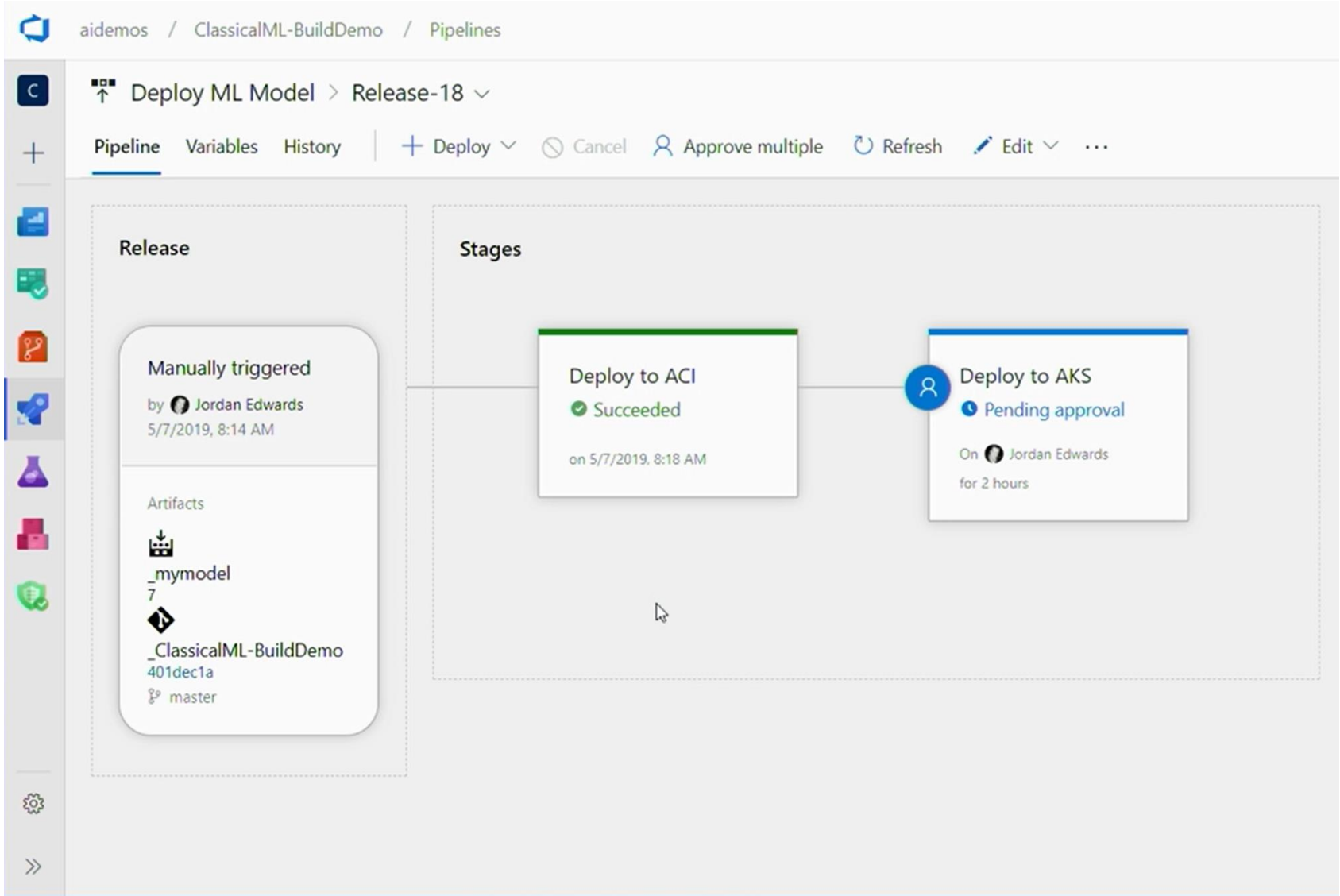
Hyperparameter optimization with keras

Efficiently search batch size, epoch, learning rate and decay

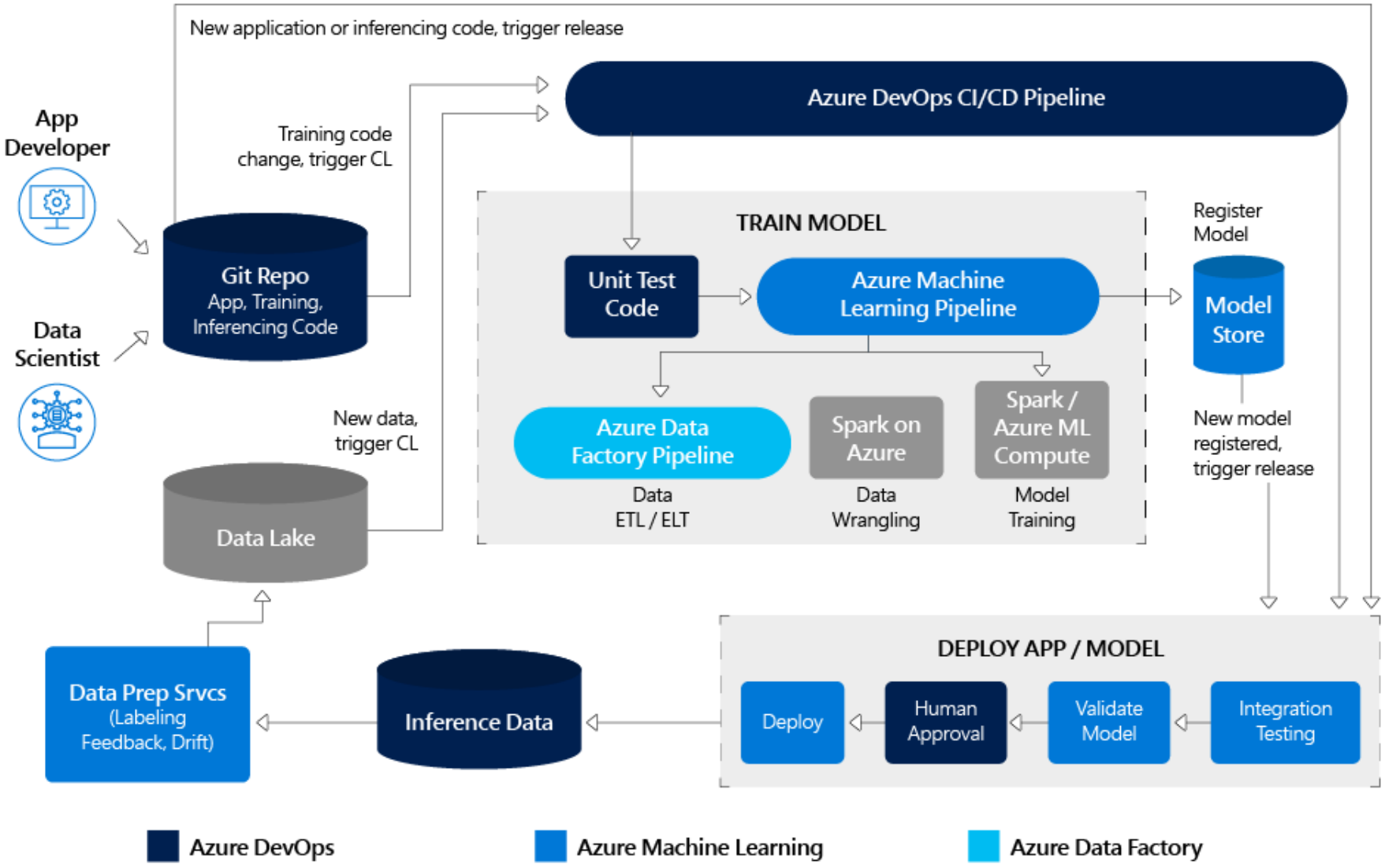
```
sampling <- random_parameter_sampling(list(batch_size = choice(c(16, 32, 64)),
                                           epochs = choice(c(200, 350, 500)),
                                           lr = normal(0.0001, 0.005),
                                           decay = uniform(1e-6, 3e-6)))

policy <- bandit_policy(slack_factor = 0.15)
hyperdrive_config <- hyperdrive_config(sampling, "Loss", primary_metric_goal("MINIMIZE"),
                                       4, policy = policy, estimator = est)

# submit hyperdrive run
hyperdrive_run <- submit_experiment(hyperdrive_config, exp)
wait_for_run_completion(hyperdrive_run, show_output = TRUE)

# find best-performing run
best_run <- get_best_run_by_primary_metric(hyperdrive_run)
```

Pipelines with Azure MLOPS extension

Trigger builds/releases on model registration in Azure ML Services



Complete MLOPS workflow



Integrating the Data Science and App Development Cycles

Francesca Lazzeri, Microsoft

Medium: https://aka.ms/AA5ib6c

Azure Pipelines

Free **unlimited** build minutes for public projects

Up to 10 free parallel jobs across Windows, Linux and macOS

https://azure.com/pipelines

Azure ML Service

Free worskpaces, experiments, model registry

Standard Azure rates for compute and deployment **or** use your own servers for free

https://azure.com/ml

Resources

MLOps: Manage, deploy, and monitor models with Azure Machine Learning Service

Microsoft Docs: https://aka.ms/mlopsdoc

Train and deploy ML models with Azure Pipelines

Microsoft Docs: https://aka.ms/azpipe

Azure Machine Learning Service

Microsoft Docs: https://aka.ms/amlsvc

These Slides

github.com/revodavid/RMLops/

$200 free credits for new Azure users

aka.ms/azure-free-credits

David Smith
Cloud Advocate, Microsoft
@revodavid

Thank you

David Smith
Cloud Advocate, Microsoft
@revodavid