Microsoft

MLOps for R with Azure Machine Learning

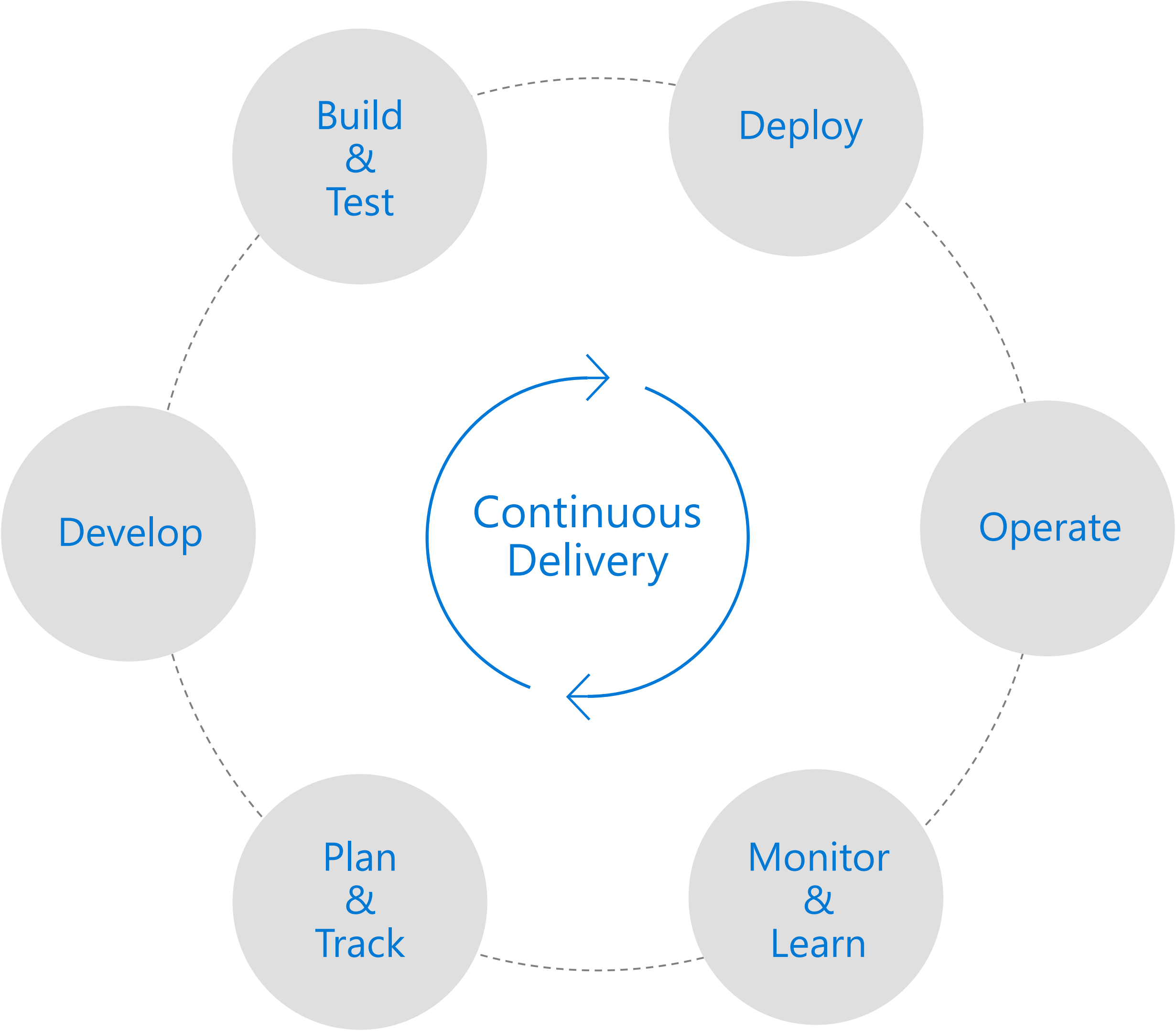
David Smith
Cloud Advocate, Microsoft
@revodavid

What is DevOps?

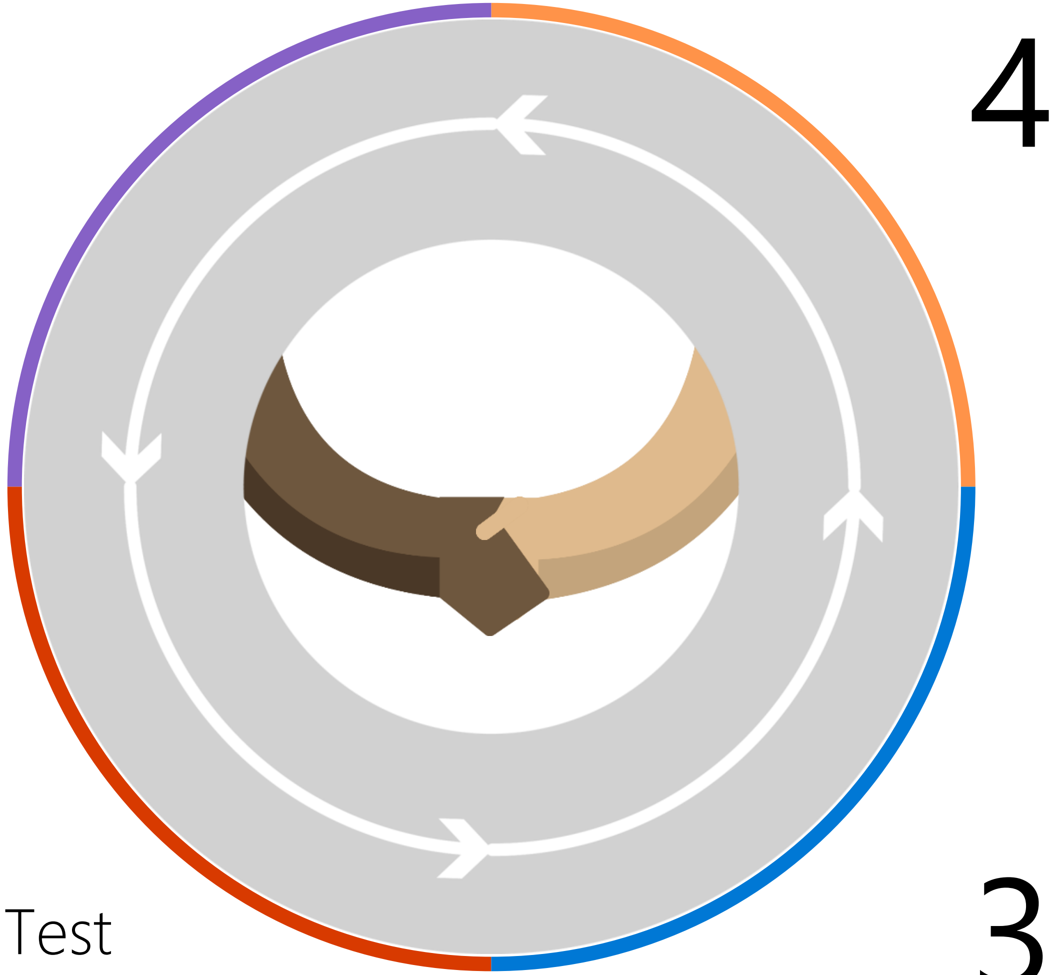
People. Process. Products.

> “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



DevOps Process



1 Plan

4 Monitor + Learn

Development

Production

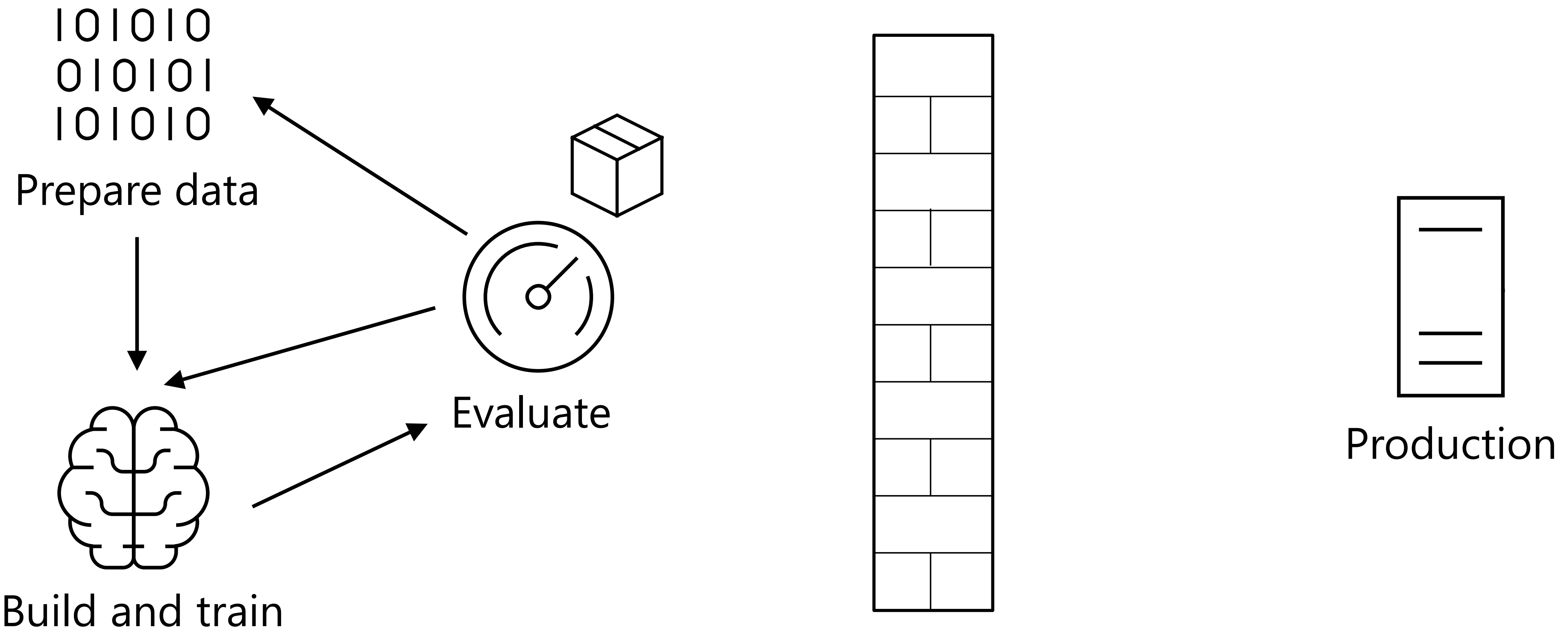
2 Develop + Test

3 Release

aka.ms/mlops-r

@revodavid at #rstudioconf

Machine Learning Process



Prepare data

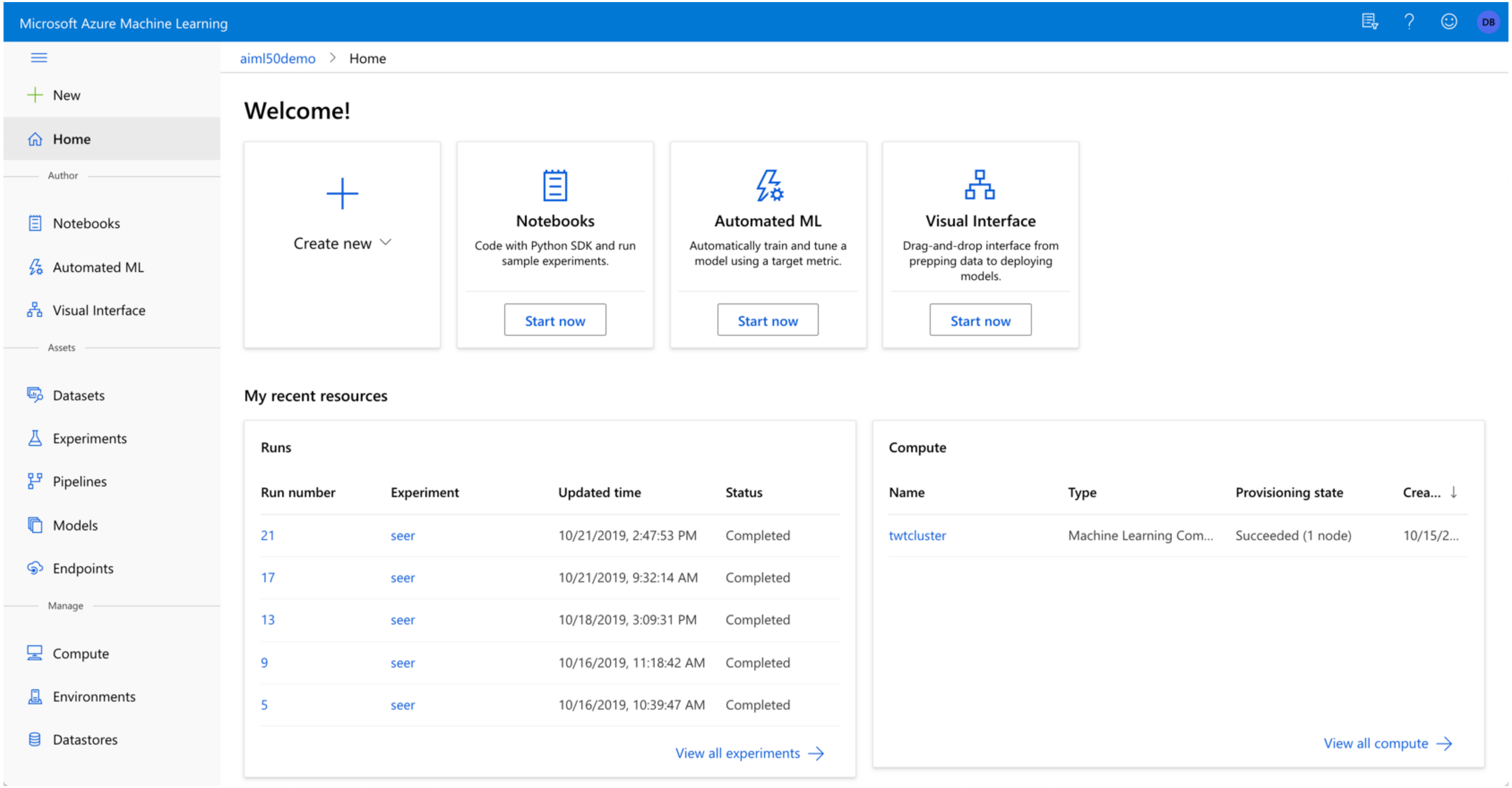
Evaluate

Build and train

Production

DEVOPS	MLOPS
Manage code (source files)	Manage code (source files) Manage data files, notebooks, Rmd docs
Manage infrastructure (as code)	Manage infrastructure (as code) Manage environments (as code)
Source code control	Source code control Track experiment outcomes Manage data sets
Build executables Builds take hours (mostly) commodity compute	Train models Model training may take weeks or months GPU compute
Manage build versions	Manage model versions Manage reproducible environments
Tests (deterministic) Fix bugs with code	Tests (probabilistic) Fix bugs with code and/or data Model drift / model retraining

Azure Machine Learning Service: ml.azure.com



Welcome!

Create new

Notebooks

Code with Python SDK and run sample experiments.

Start now

Automated ML

Automatically train and tune a model using a target metric.

Start now

Visual Interface

Drag-and-drop interface from prepping data to deploying models.

Start now

My recent resources

Runs

Run number	Experiment	Updated time	Status
21	seer	10/21/2019, 2:47:53 PM	Completed
17	seer	10/21/2019, 9:32:14 AM	Completed
13	seer	10/18/2019, 3:09:31 PM	Completed
9	seer	10/16/2019, 11:18:42 AM	Completed
5	seer	10/16/2019, 10:39:47 AM	Completed

View all experiments →

Compute

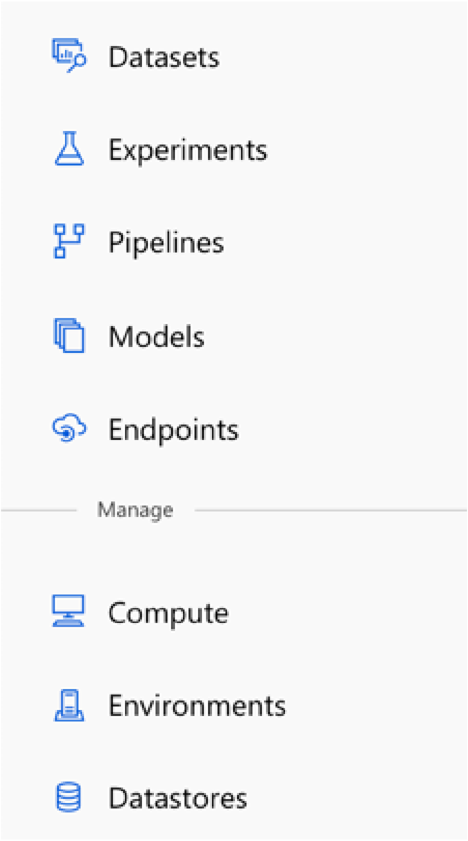
Name	Type	Provisioning state	Crea... ↓
twtcluster	Machine Learning Com...	Succeeded (1 node)	10/15/2...

View all compute →

aka.ms/mlops-r

@revodavid at #rstudioconf

Azure Machine Learning



Datasets – registered, known data sets

Experiments – Training runs

Pipelines – Training workflows

Models – Registered, versioned models

Endpoints:

- **Real-time Endpoints** – Deployed model endpoints
- **Pipeline Endpoints** – Training workflow endpoints

Compute – Managed compute

Environments – defined training and inference environments

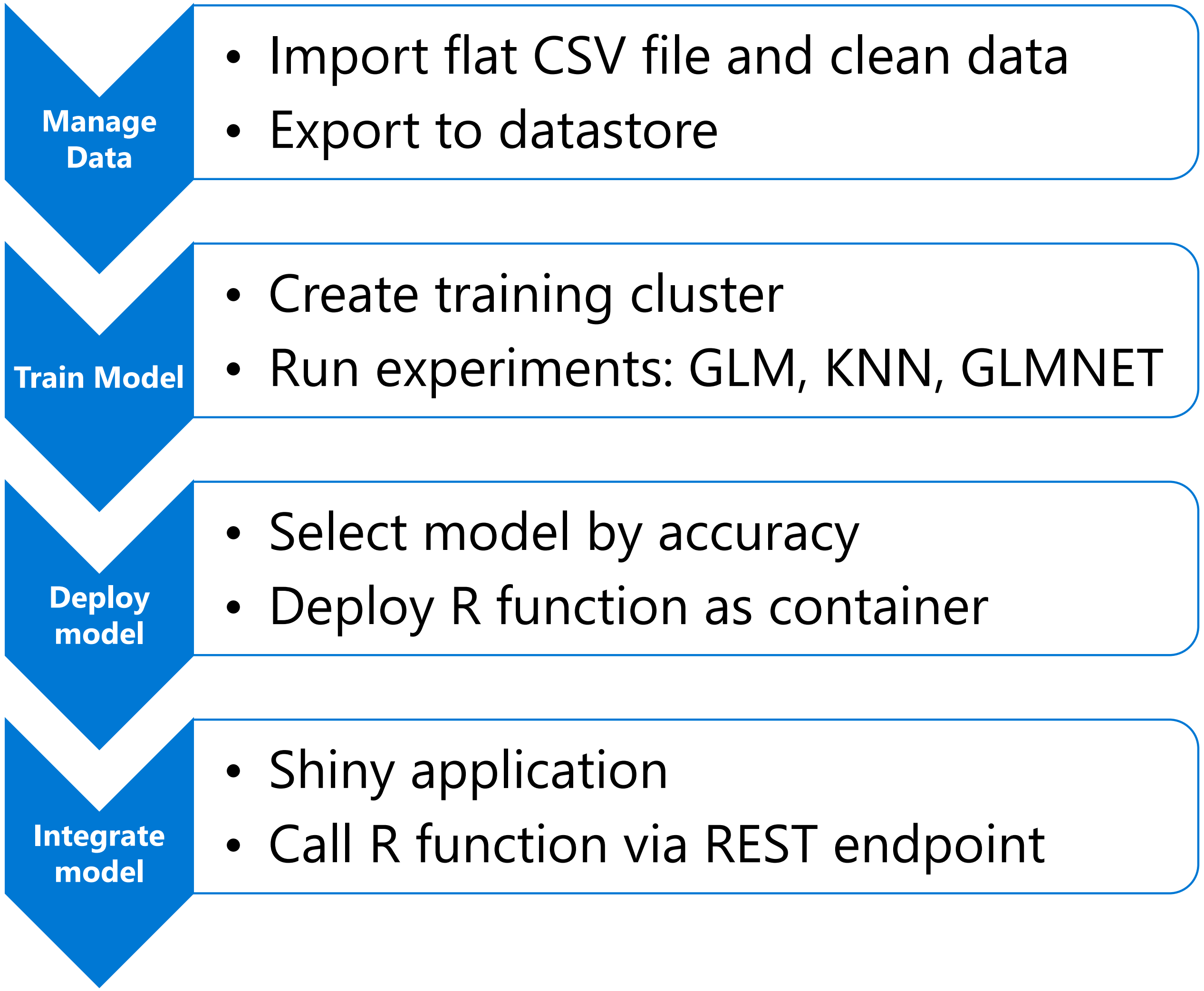
Datastores – Connections to data

Azure ML service SDK for R

Open-source R package for use with CRAN R: **azuremlsdk**

- Create Workspaces, Experiments, Compute, Models, and other artifacts with R commands
- Use any R function/package (and track requirements for deployment)
- HyperDrive support: smart hyperparameter search with parallel compute
- Publish models as web services (in Azure or your own infra)
- Trigger training / deployment pipelines from CI/CD services

Accident fatality prediction app



Manage Data

- Import flat CSV file and clean data
- Export to datastore

Train Model

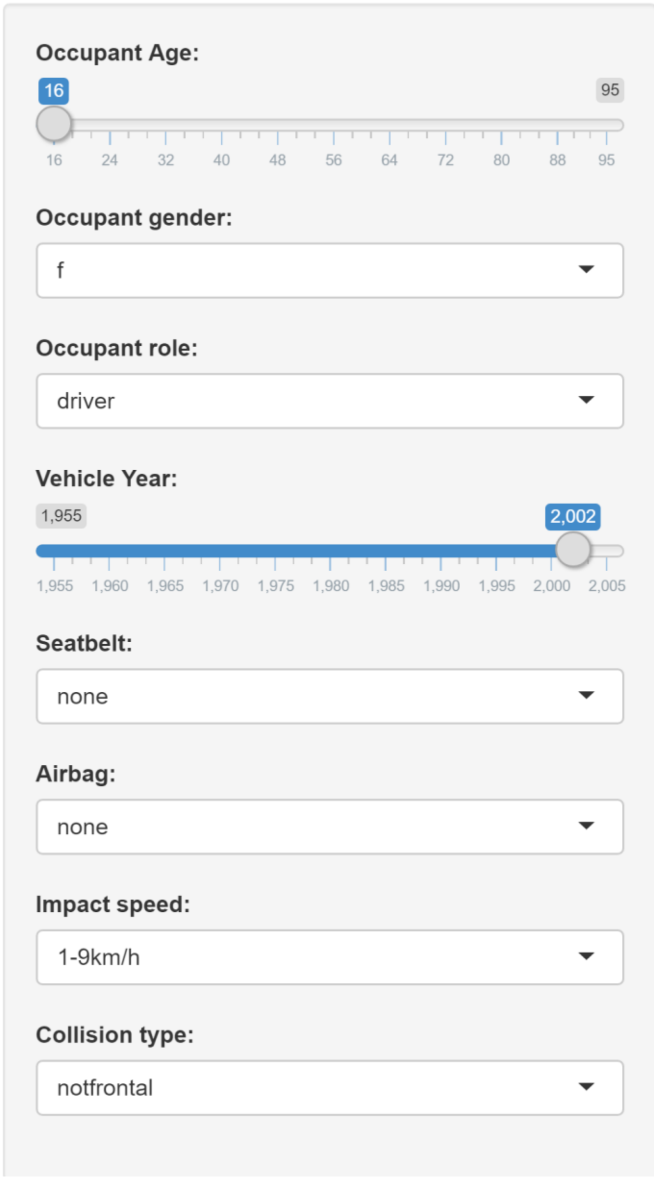
- Create training cluster
- Run experiments: GLM, KNN, GLMNET

Deploy model

- Select model by accuracy
- Deploy R function as container

Integrate model

- Shiny application
- Call R function via REST endpoint



Occupant Age: 16

Occupant gender: f

Occupant role: driver

Vehicle Year: 1,955 – 2,002

Seatbelt: none

Airbag: none

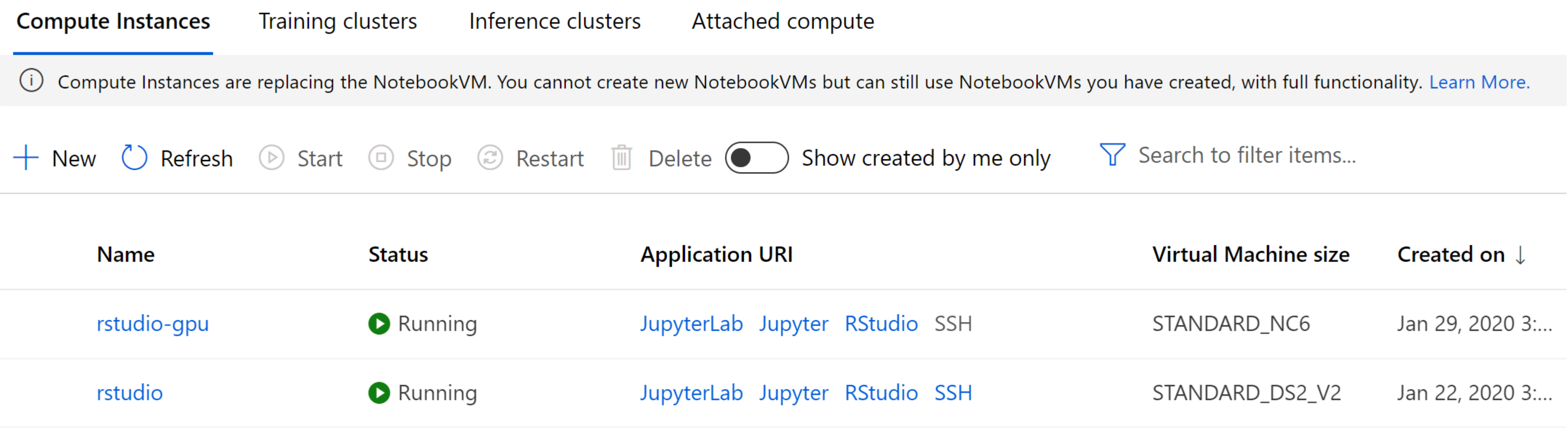
Impact speed: 1-9km/h

Collision type: notfrontal

aka.ms/mlops-r

@revodavid at #rstudioconf

Create compute instance for interactive work



Compute Instances | Training clusters | Inference clusters | Attached compute

ⓘ Compute Instances are replacing the NotebookVM. You cannot create new NotebookVMs but can still use NotebookVMs you have created, with full functionality. Learn More.

+ New | Refresh | Start | Stop | Restart | Delete | Show created by me only | Search to filter items...

Name	Status	Application URI	Virtual Machine size	Created on ↓
rstudio-gpu	Running	JupyterLab Jupyter RStudio SSH	STANDARD_NC6	Jan 29, 2020 3:...
rstudio	Running	JupyterLab Jupyter RStudio SSH	STANDARD_DS2_V2	Jan 22, 2020 3:...

Instances are persistent, charge at usual VM rates

… or just use desktop R/Rstudio with azuremlsdk package

… or use your own attached compute server

Use Compute Instance to prep and share data

```
library(azuremlsdk)
ws <- load_workspace_from_config()
```

azuremlsdk package and auth keys pre-installed on compute instance

```
nassCDS <- read.csv("nassCDS.csv")
# Lots of cleaning code
saveRDS(accidents,
        file="accidents.Rd")
```

Normal R code for data / model experimentation

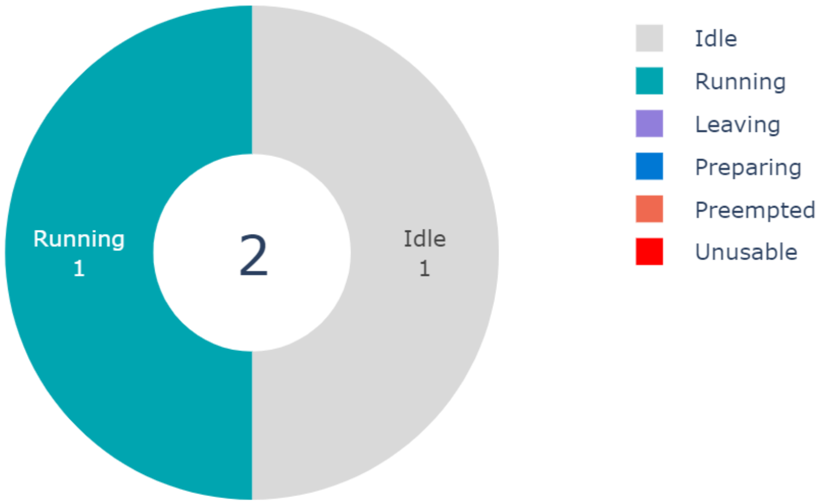
```
ds <- get_default_datastore(ws)
target_path <- "accidentdata"
upload_files_to_datastore(ds,
          list("./accidents.Rd"),
          target_path = target_path,
          overwrite = TRUE)
```

Export data to shared, persistent datastore

Can use for files other than data, too

Create 2-node training cluster

```r
ws <- load_workspace_from_config()
compute_target <- create_aml_compute(
                    workspace = ws,
                    cluster_name = "rcluster",
                    vm_size = "STANDARD_D2_V2",
                    vm_priority = "lowpriority",
                    min_nodes = 1,
                    max_nodes = 2)
```



Nodes pre-loaded with R and R packages

Scales up as jobs added to queue, scales down when unused

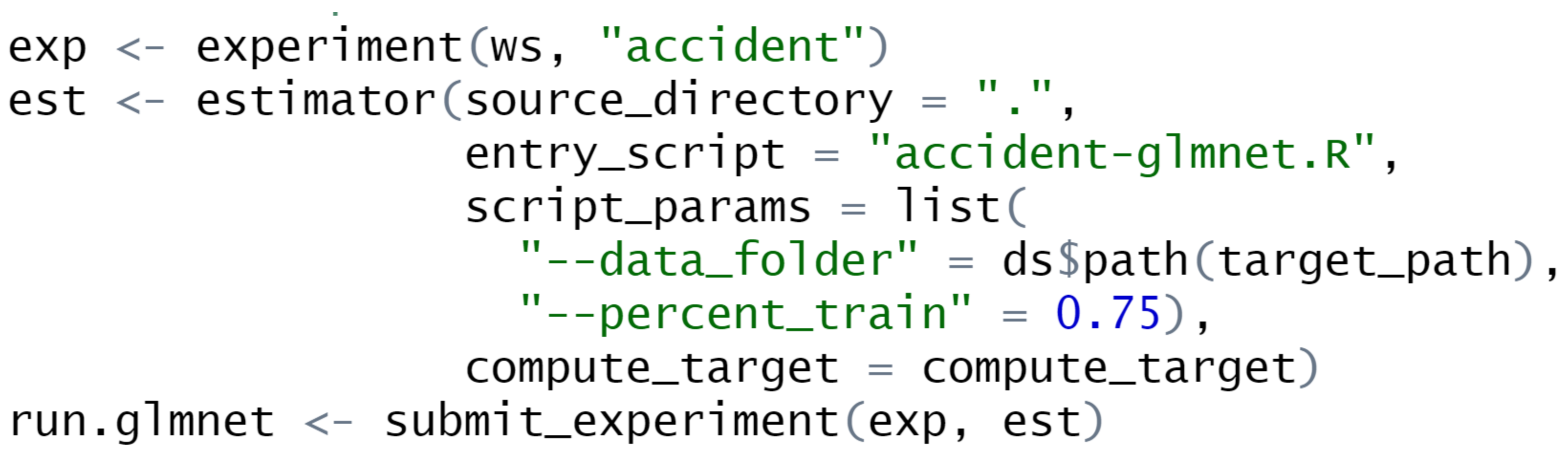
Resize as needed

Shared between workspace users (Python and R)

Low priority nodes for ~80% discount

GPU clusters available

Train models and choose one to deploy

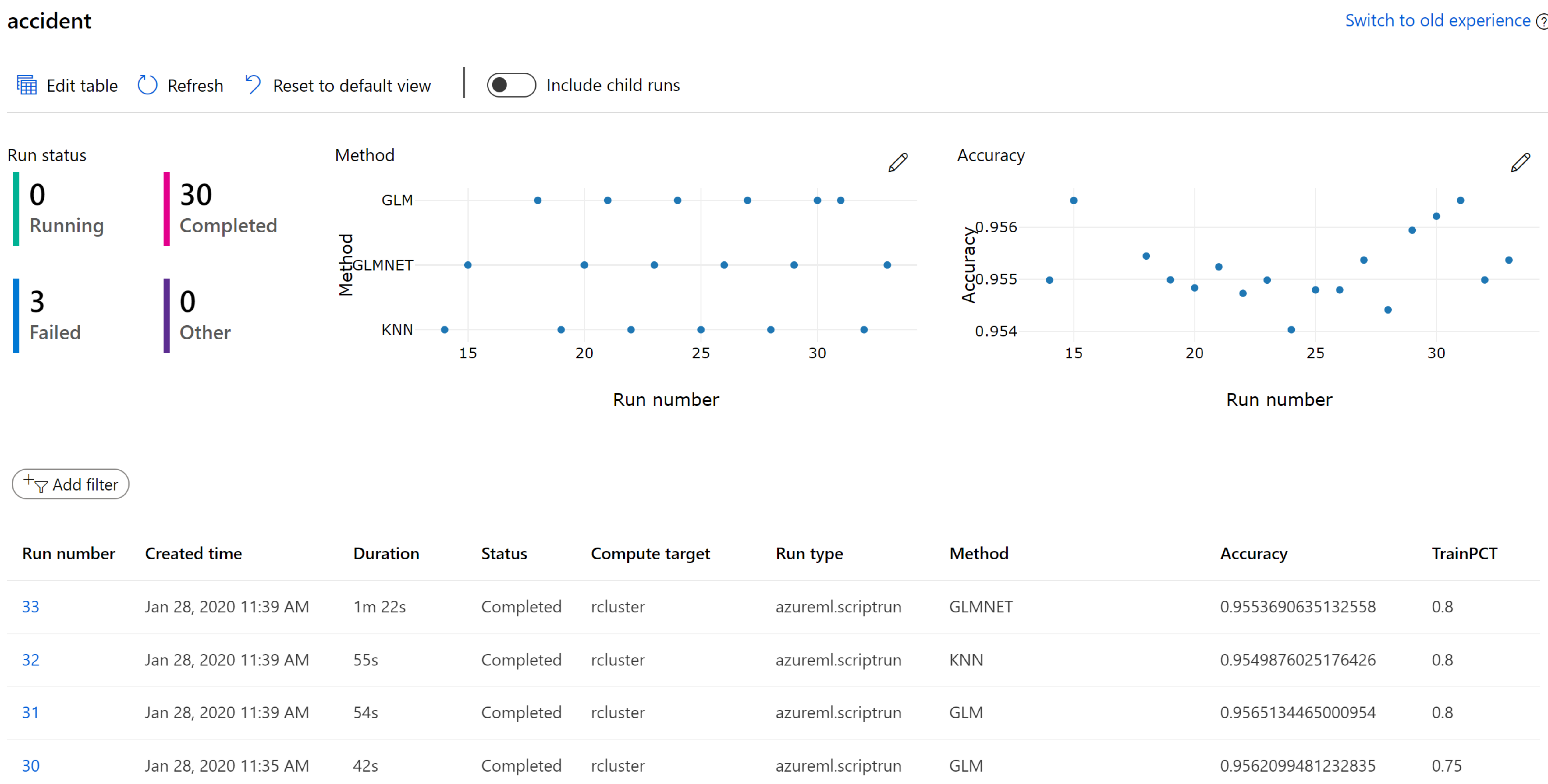


```r
exp <- experiment(ws, "accident")
est <- estimator(source_directory = ".",
                 entry_script = "accident-glmnet.R",
                 script_params = list(
                   "--data_folder" = ds$path(target_path),
                   "--percent_train" = 0.75),
                 compute_target = compute_target)
run.glmnet <- submit_experiment(exp, est)
```

Experiments tracked with source scripts and recorded metrics

Control execution with command line parameters

Most packages pre-loaded, custom packages supported



accident

Switch to old experience

Edit table · Refresh · Reset to default view · Include child runs

Run status

0 Running	30 Completed
3 Failed	0 Other

Run number	Created time	Duration	Status	Compute target	Run type	Method	Accuracy	TrainPCT
33	Jan 28, 2020 11:39 AM	1m 22s	Completed	rcluster	azureml.scriptrun	GLMNET	0.9553690635132558	0.8
32	Jan 28, 2020 11:39 AM	55s	Completed	rcluster	azureml.scriptrun	KNN	0.9549876025176426	0.8
31	Jan 28, 2020 11:39 AM	54s	Completed	rcluster	azureml.scriptrun	GLM	0.9565134465000954	0.8
30	Jan 28, 2020 11:35 AM	42s	Completed	rcluster	azureml.scriptrun	GLM	0.9562099481232835	0.75

Register model

```r
model <- register_model(
  ws,
  model_path = "outputs/model.rds",
  model_name = "accidents_model_caret",
  description = "Predict accident probability")
```

Register model with R model object

Models can be versioned

```r
r_env <- r_environment(name = "basic_env",
                       cran_packages="caret")

inference_config <- inference_config(
  entry_script = "accident_predict_caret.R",
  source_directory = ".",
  environment = r_env)
```

Packages (CRAN, GitHub or private) needed to execute.

R script receives model and data (JSON) to make prediction

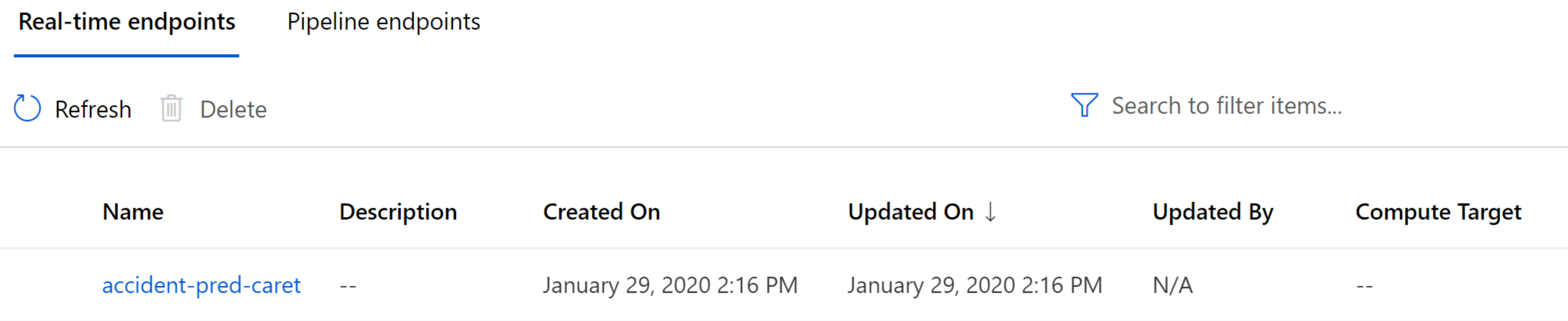
Deploy Model as service

```
aci_config <-
  aci_webservice_deployment_config(
    cpu_cores = 1, memory_gb = 0.5)

aci_service <- deploy_model(ws,
  'accident-pred-caret',
  list(model),
  inference_config,
  aci_config)

accident.endpoint <- get_webservice(
  ws, "accident-pred-caret")$scoring_uri
```

Small container instance for testing

(use Kubernetes for production)

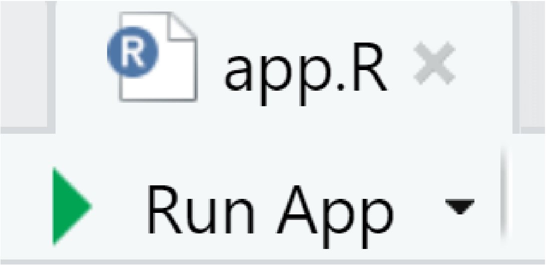
Get REST endpoint for use in app



Name	Description	Created On	Updated On ↓	Updated By	Compute Target
accident-pred-caret	--	January 29, 2020 2:16 PM	January 29, 2020 2:16 PM	N/A	--

Integrate model into Shiny app

```
library(httr)
v <- POST(accident.endpoint,
             body=input,
             encode="json")

pred <- content(v)[[1]]
```

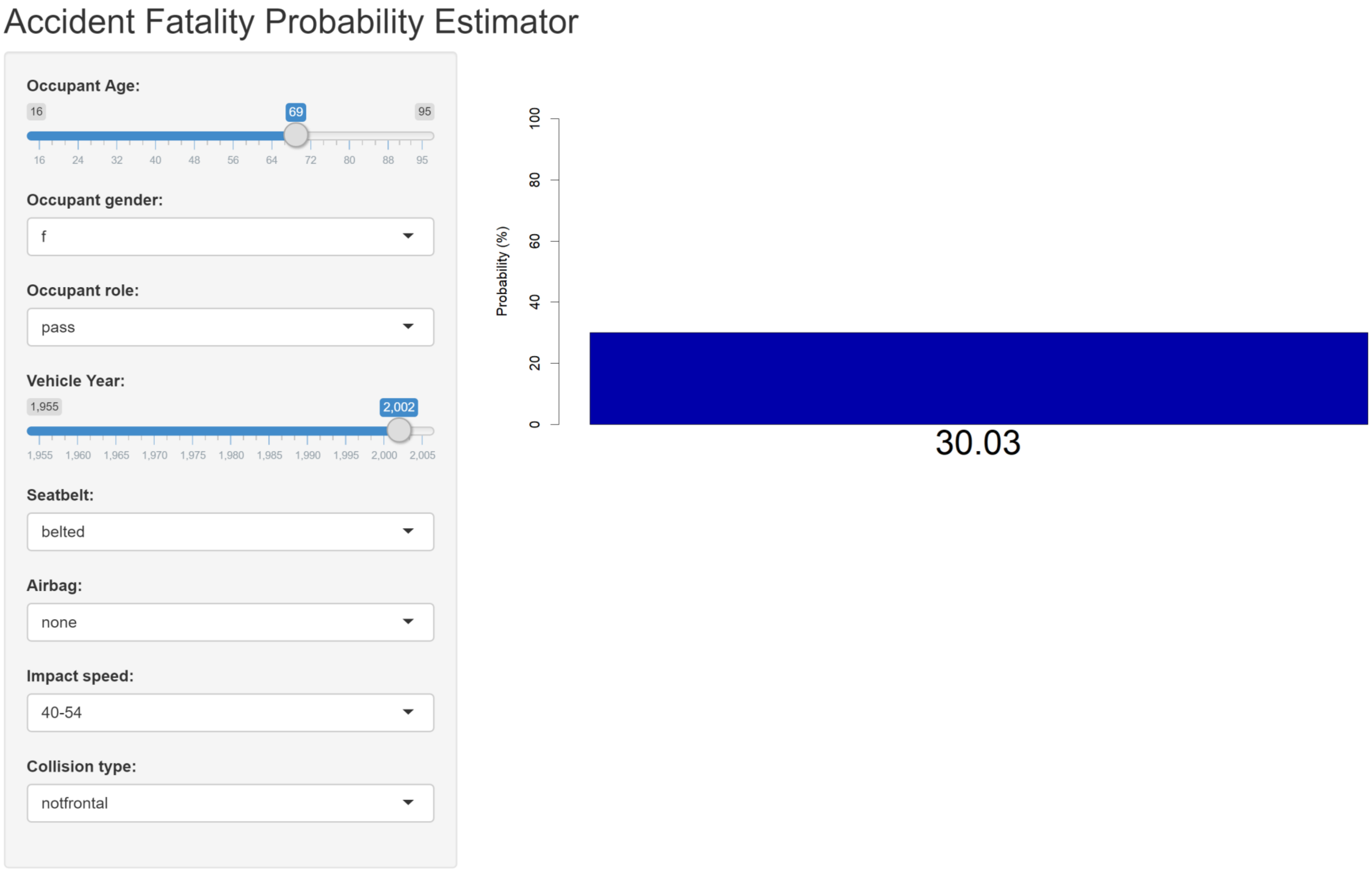


Pass Shiny input data frame to endpoint

Extract prediction from endpoint response

Deploy Shiny app to any server

… or call from any another app



Accident Fatality Probability Estimator

Occupant Age: 69

Occupant gender: f

Occupant role: pass

Vehicle Year: 2,002

Seatbelt: belted

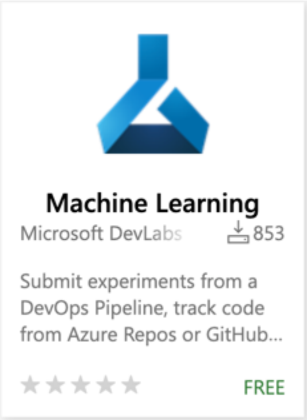
Airbag: none

Impact speed: 40-54

Collision type: notfrontal

Deliver ML apps with Pipelines

Azure ML Pipelines



Machine Learning

Microsoft DevLabs 853

Submit experiments from a DevOps Pipeline, track code from Azure Repos or GitHub...

★★★★★ FREE

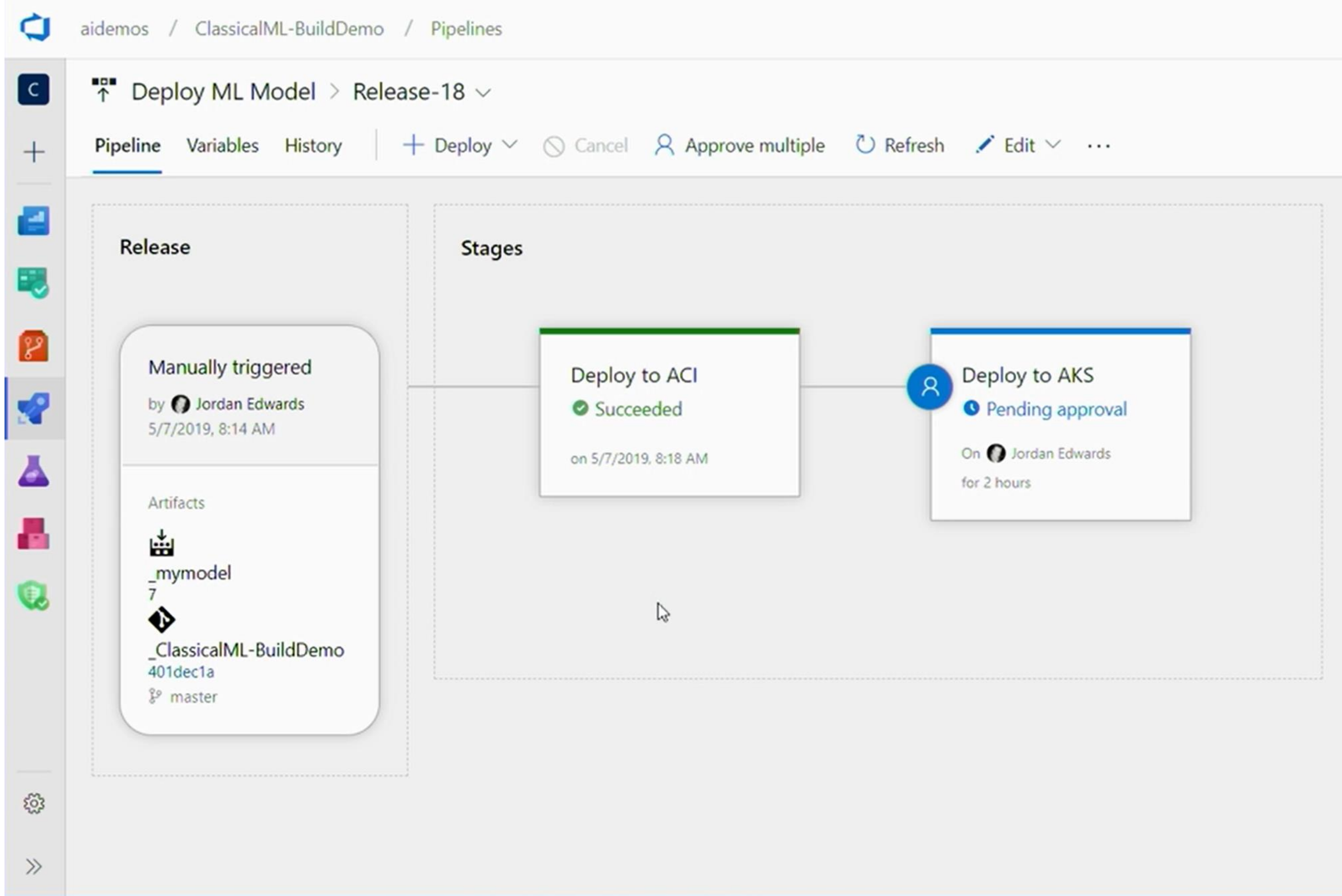
```
az extension add -n azure-cli-ml
```

```
pip install --upgrade azureml-sdk
```

Azure Pipelines
in
Azure DevOps

CI/CD for apps with Azure Pipelines

Azure ML extension triggers builds/releases on model registration in Azure ML Services



Manually triggered

by Jordan Edwards

5/7/2019, 8:14 AM

_mymodel

7

_ClassicalML-BuildDemo

401dec1a

master

Deploy to ACI

Succeeded

on 5/7/2019, 8:18 AM

Deploy to AKS

Pending approval

On Jordan Edwards

for 2 hours

More details:

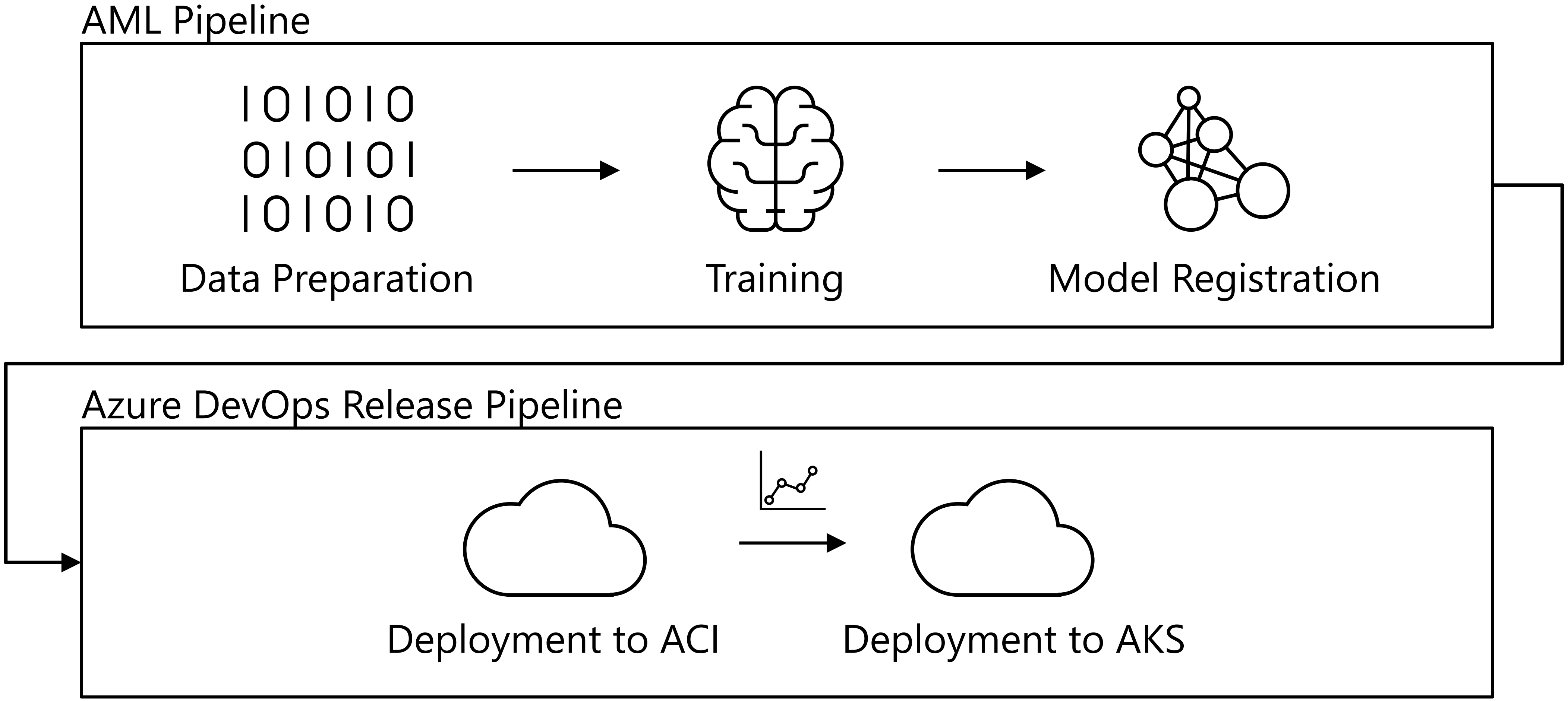
Machine learning operations: Applying DevOps to data science

aka.ms/AIML50repo

aka.ms/mlops-r

@revodavid at #rstudioconf

Complete Pipeline



AML Pipeline

Data Preparation

Training

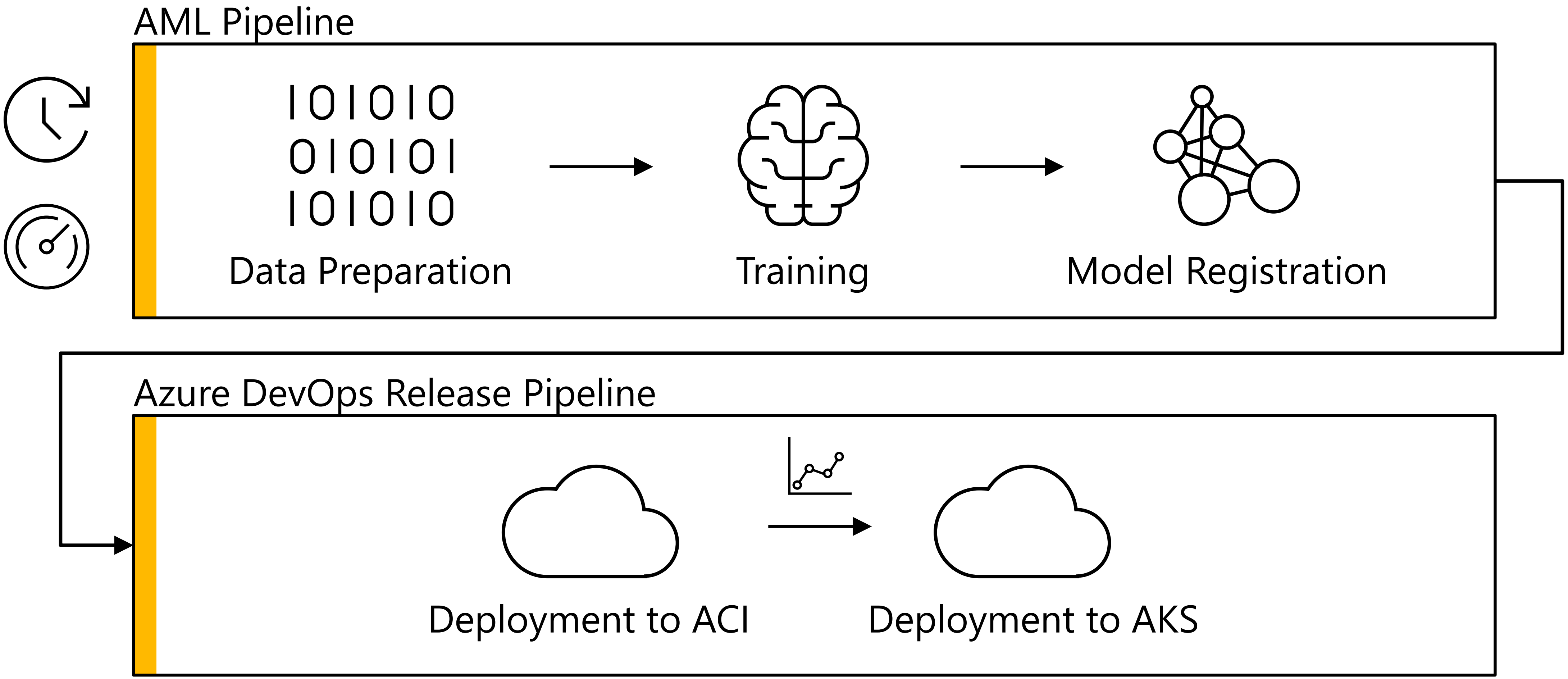
Model Registration

Azure DevOps Release Pipeline

Deployment to ACI

Deployment to AKS

Retraining


AML Pipeline
101010
010101
101010
Data Preparation
Training
Model Registration
Azure DevOps Release Pipeline
Deployment to ACI
Deployment to AKS

Azure Pipelines

➔ https://azure.com/pipelines

Free **unlimited** build minutes for public projects

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

Azure ML Service

➔ https://azure.com/ml

Free worskpaces, experiments, model registry

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

Thank you!

David Smith
Cloud Advocate, Microsoft
@revodavid

Slides and links:
aka.ms/mlops-r