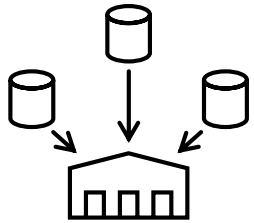
Microsoft Azure

Leverage high-scale data with R language in Azure Databricks

Krit Kamtuo
Machine Learning Engineer
Commercial Software Engineering
Microsoft
github.com/kritcs18
@kritcs18

How companies are transforming

Serving business users and end users with **intelligent** and **dynamic** applications



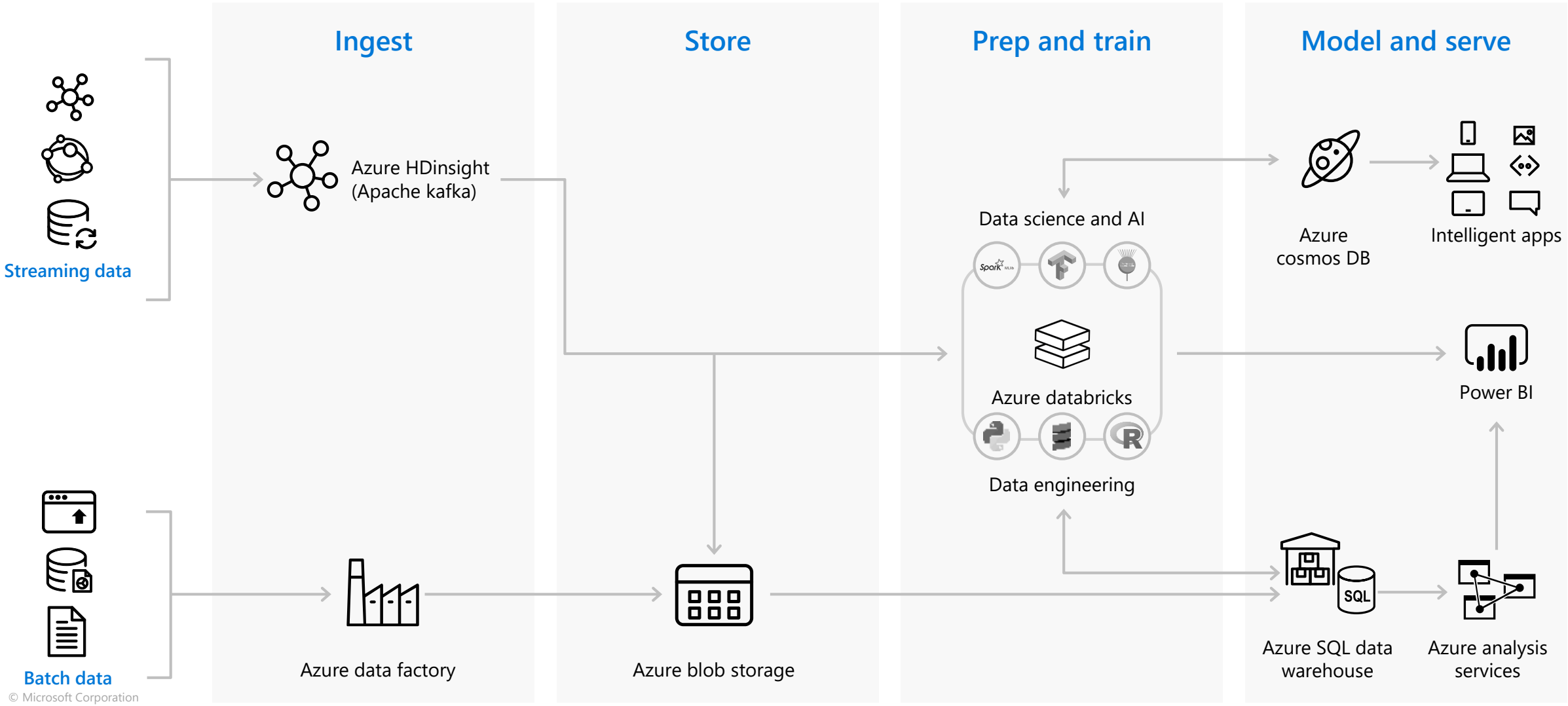
Build a unified and usable data pipeline

Train ML and DL models to derive insights

Operationalize models and distribute insights at scale

What approach should I take to overcome these challenges?

Microsoft has a recommended reference architecture



Introducing Azure Databricks

Fast, easy, and collaborative Apache Spark™-based analytics platform

Increase productivity

Build on a secure, trusted cloud

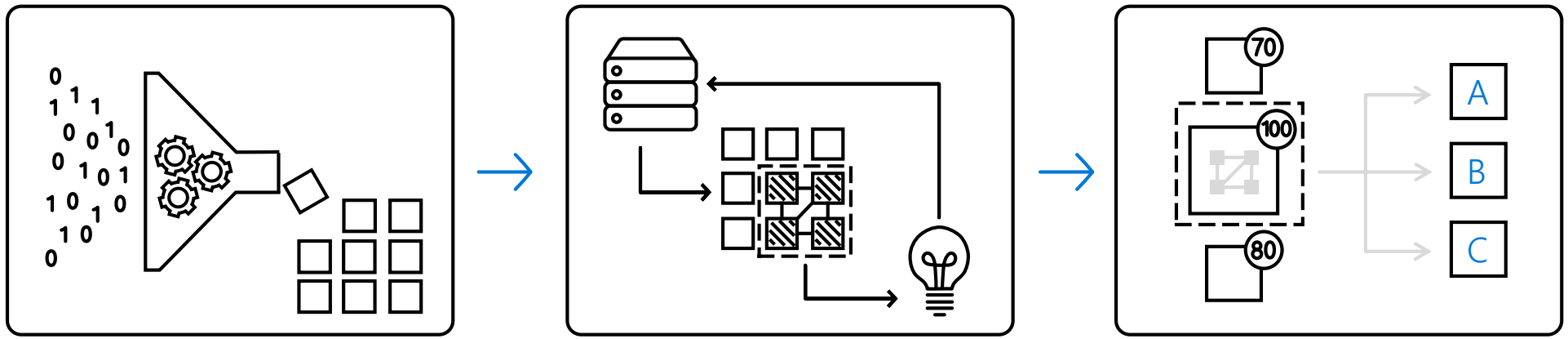
Scale without limits

Built with your needs in mind

- Role-based access controls
- Effortless autoscaling
- Live collaboration
- Enterprise-grade SLAs
- Best-in-class notebooks
- Simple job scheduling

Seamlessly integrated with the Azure Portfolio

Prep and train



Collect and prepare data

- Azure data factory
- Azure databricks

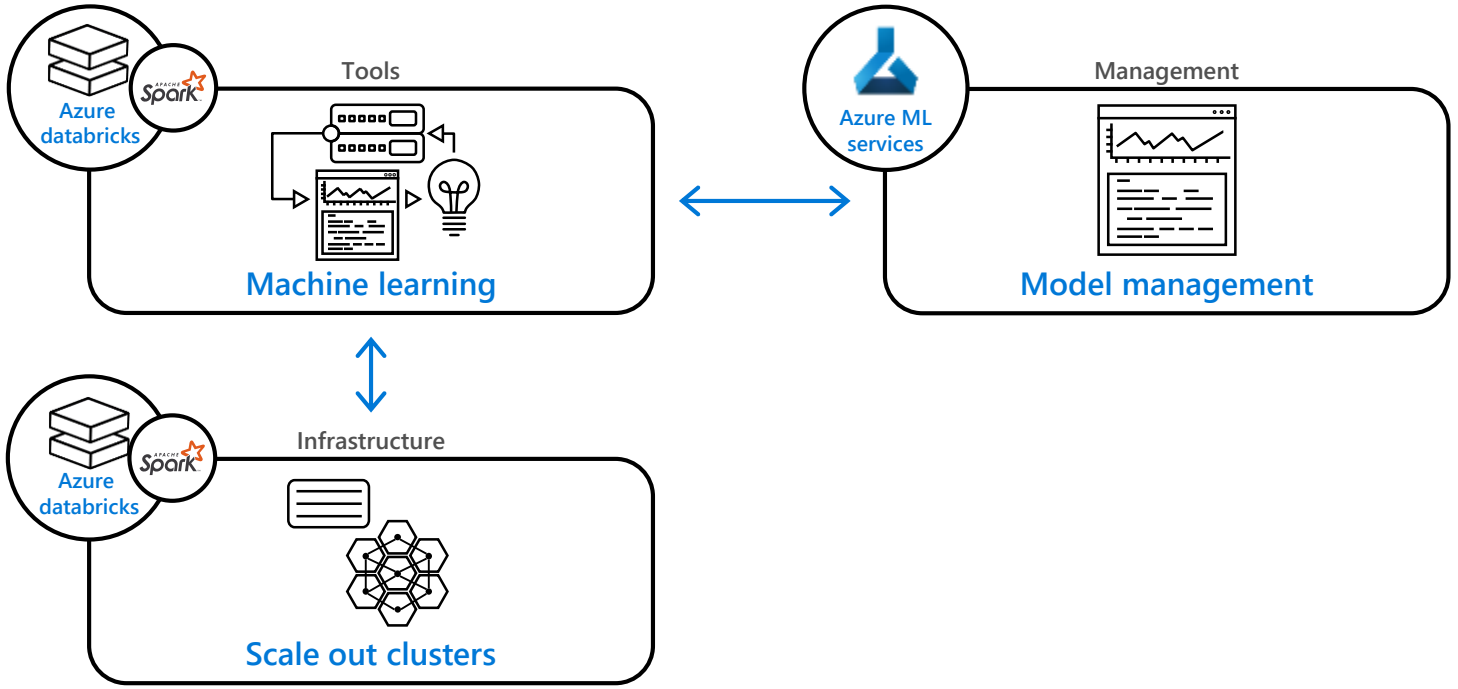
Train and evaluate model

- Azure databricks

Operationalize and manage

- Azure ML services
- Azure databricks

Train and evaluate Machine Learning models



Simplify model development

Collaborate in interactive workspaces
Access a library of battle-tested models
Automate job execution

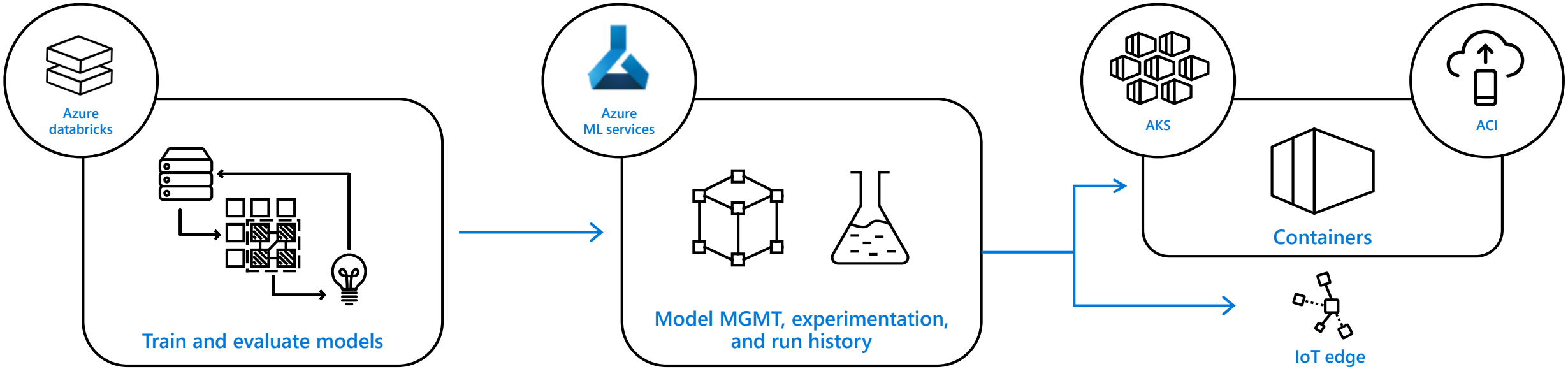
Scale compute resources to meet your needs

Easily scale up or scale out
Autoscale on a serverless infrastructure
Leverage commodity hardware

Quickly determine the right model for your data

Determine the best algorithm
Tune hyperparameters to optimize models
Rapidly prototype in agile environments

Operationalize and manage models with ease



Bring models to life quickly

Build and deploy models in minutes
Iterate quickly on serverless infrastructure
Easily change environments

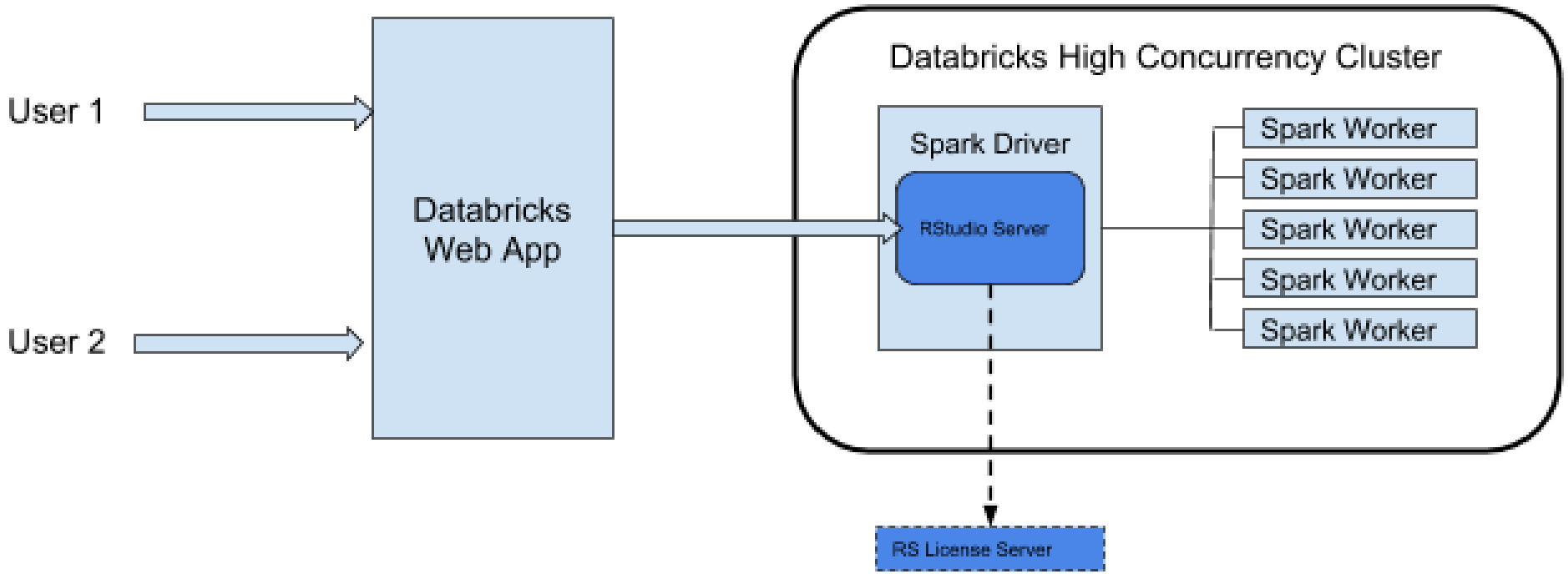
Proactively manage model performance

Identify and promote your best models
Capture model telemetry
Retrain models with APIs

Deploy models closer to your data

Deploy models anywhere
Scale out to containers
Infuse intelligence into the IoT edge

RStudio on Azure Databricks



Get Started

Try Azure:
https://azure.microsoft.com/en-us/free/

Azure Databricks docs:
https://docs.microsoft.com/en-us/azure/azure-databricks/

SparkR on Databrick guide:
https://docs.azuredatabricks.net/spark/latest/sparkr/index.html

R developer's guide to Azure:
https://docs.microsoft.com/en-us/azure/machine-learning/r-developers-guide