.NET Conf

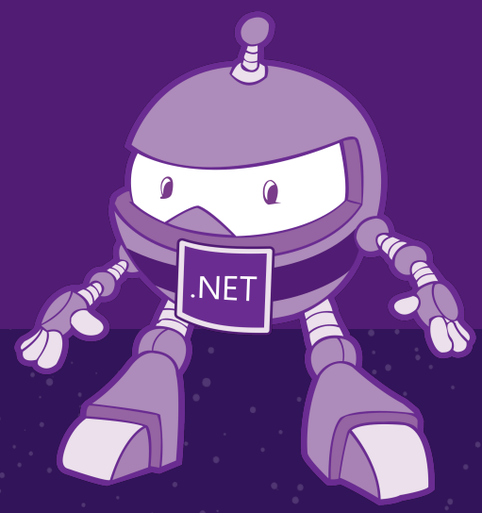
探索 .NET 新世界

Host by

STUDY4

SUMMIT SUEN

從機器學習平台的發展觀察如何打造資料科學團隊



SUMMIT SUEN
孫玉峰

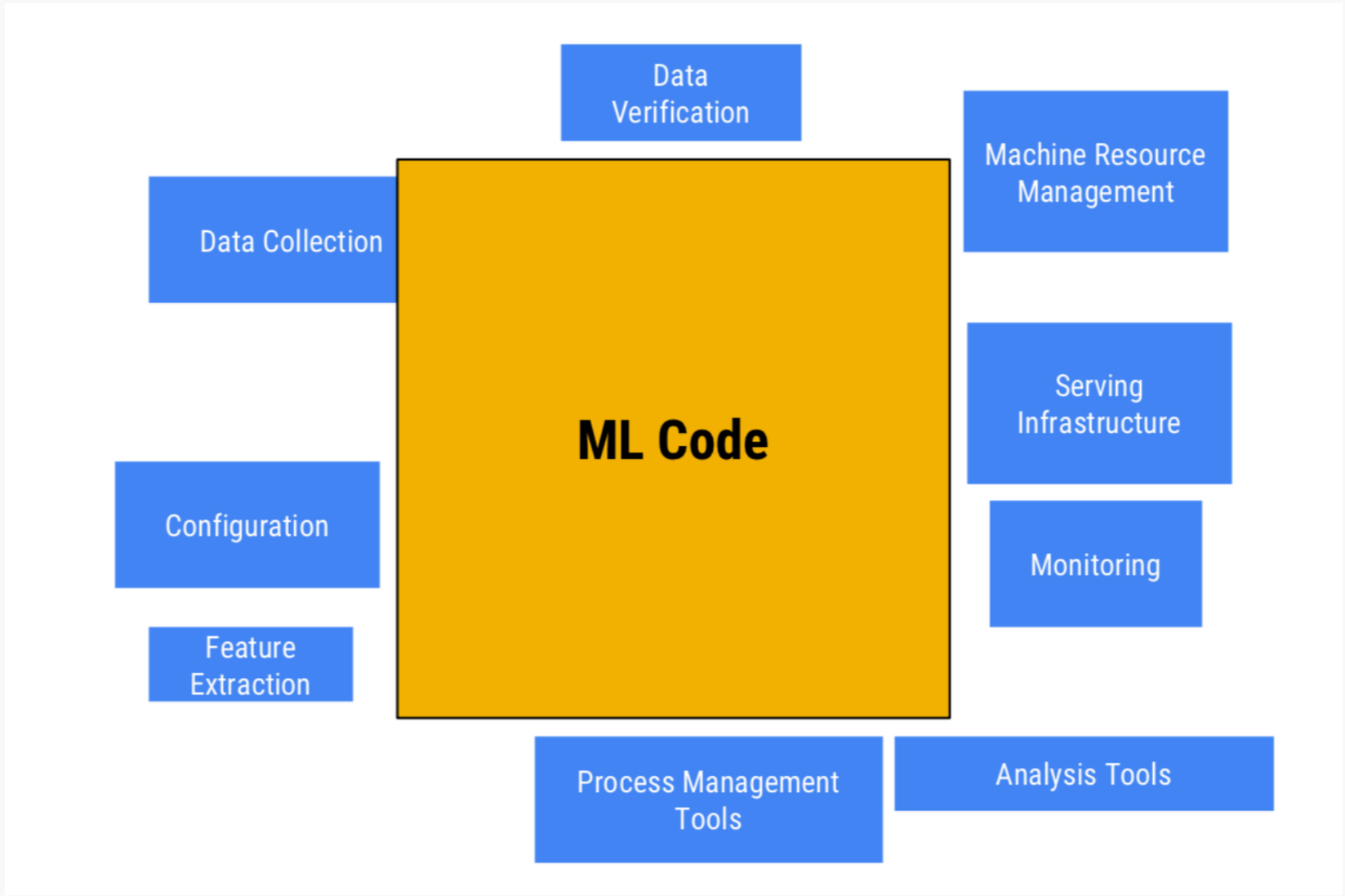
Taiwan R User Group
臺灣R語言使用者社群主持人

曾任新創公司研發總監 /
資料科學家

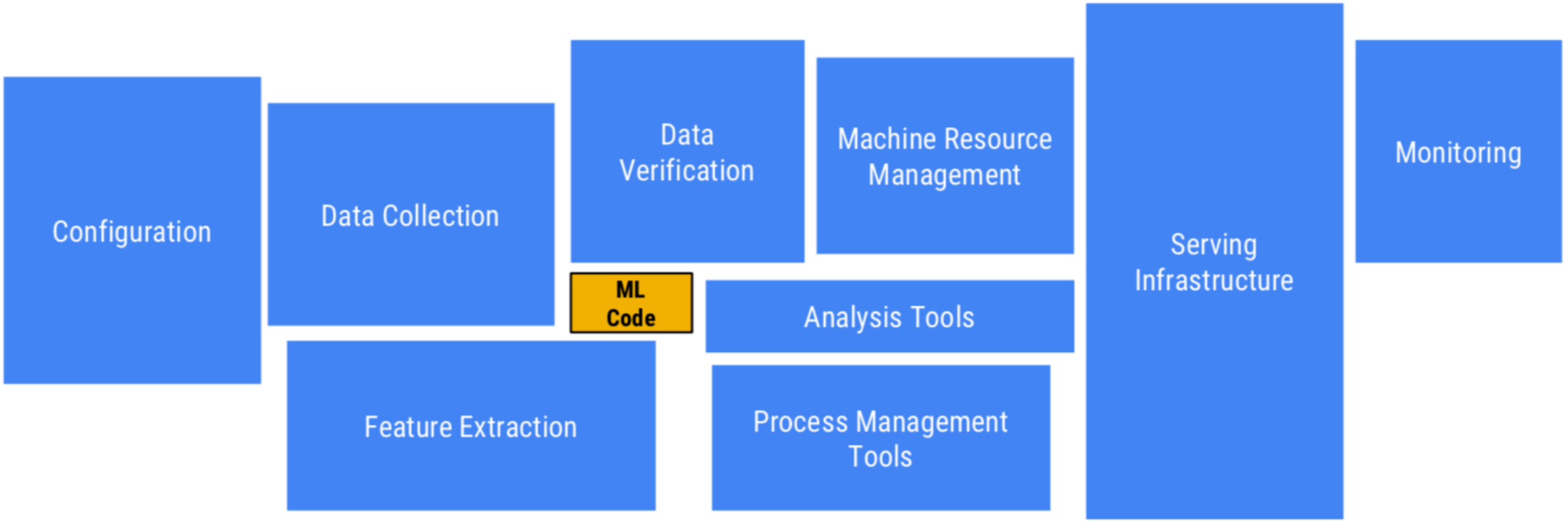
2017-2019 微軟最有價值專家
Most Valuable Professionals

資料科學家與他們的產地

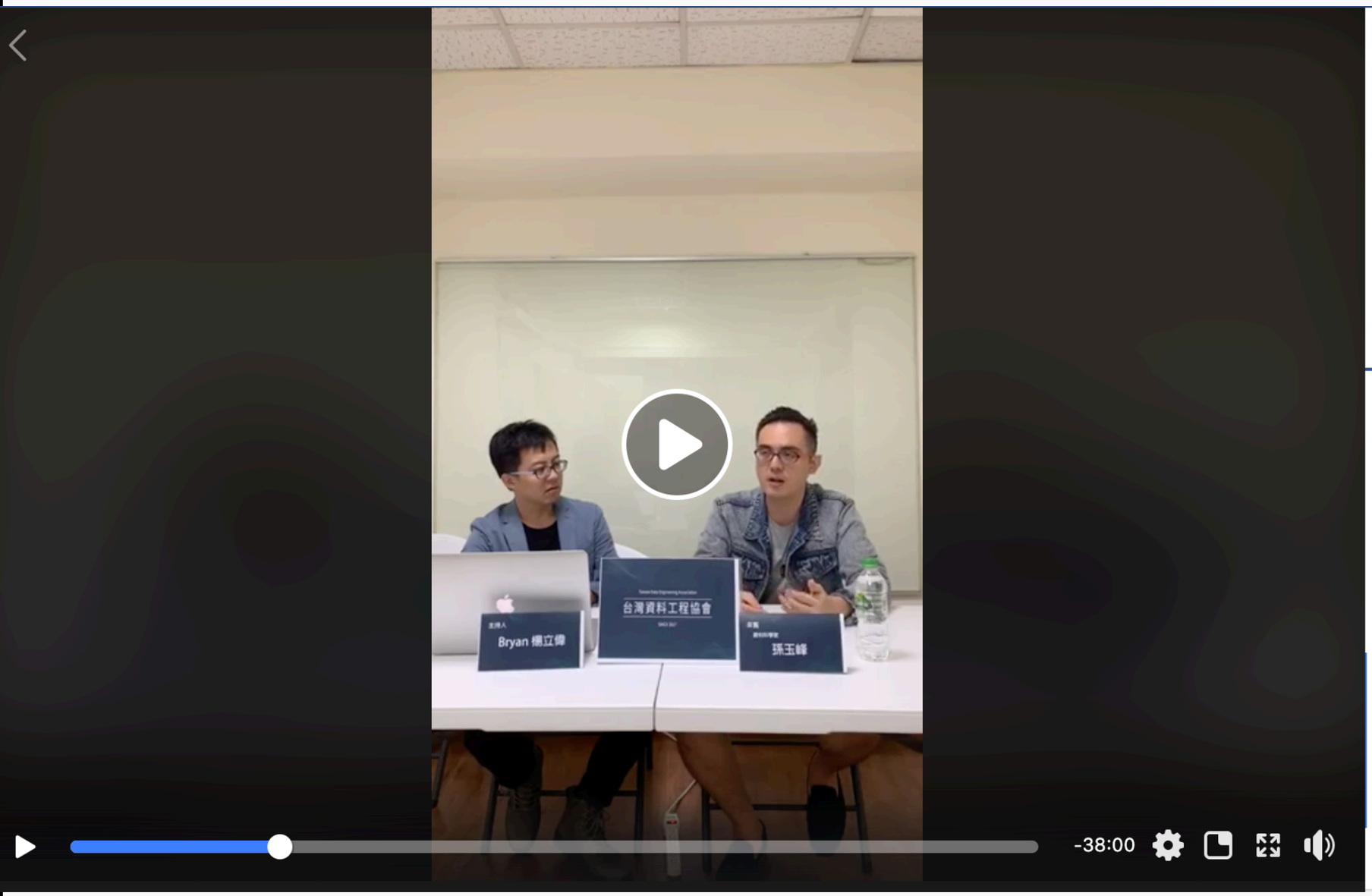
我以為的資料科學家



實際上的資料科學家


Configuration
Data Collection
Data Verification
Machine Resource Management
Serving Infrastructure
Monitoring
ML Code
Analysis Tools
Feature Extraction
Process Management Tools

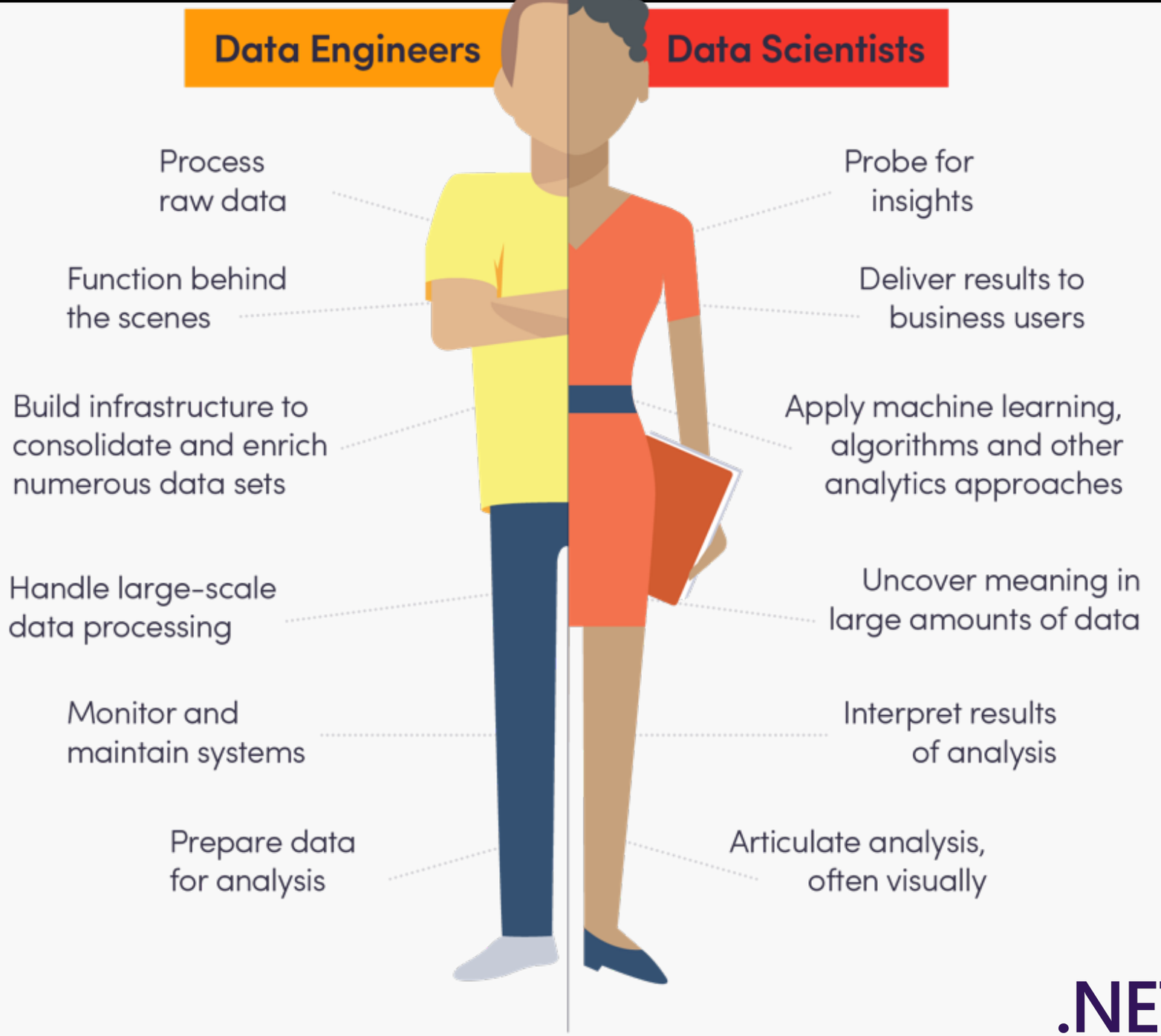
一條龍資料科學家~~你就是那條龍~~



STUDY4.TW 為學習而生

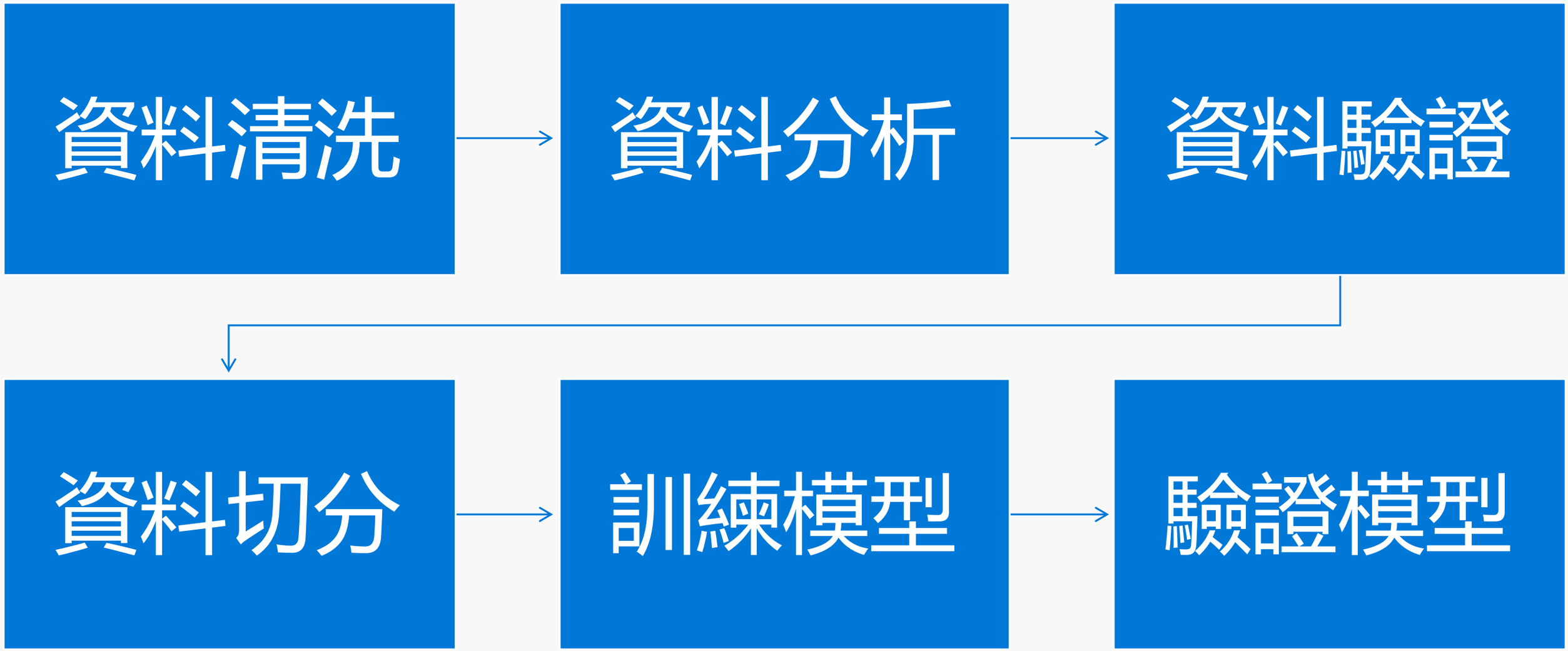
.NET Conf

資料科學家~~與他的快樂夥伴們~~



.NET Conf

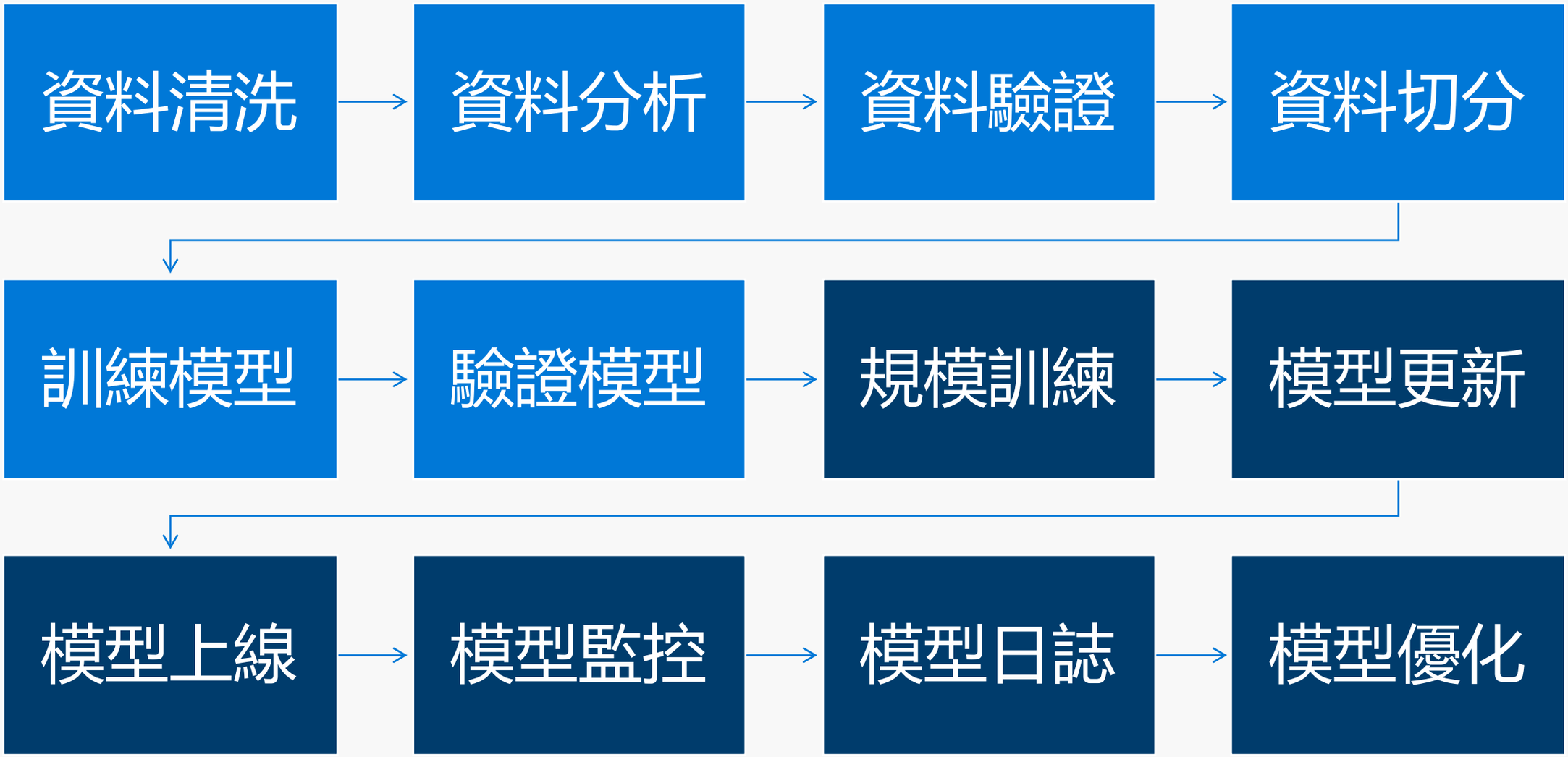
開發資料科學 PROJECT PIPELINE



STUDY4.TW
為學習而生

.NET Conf

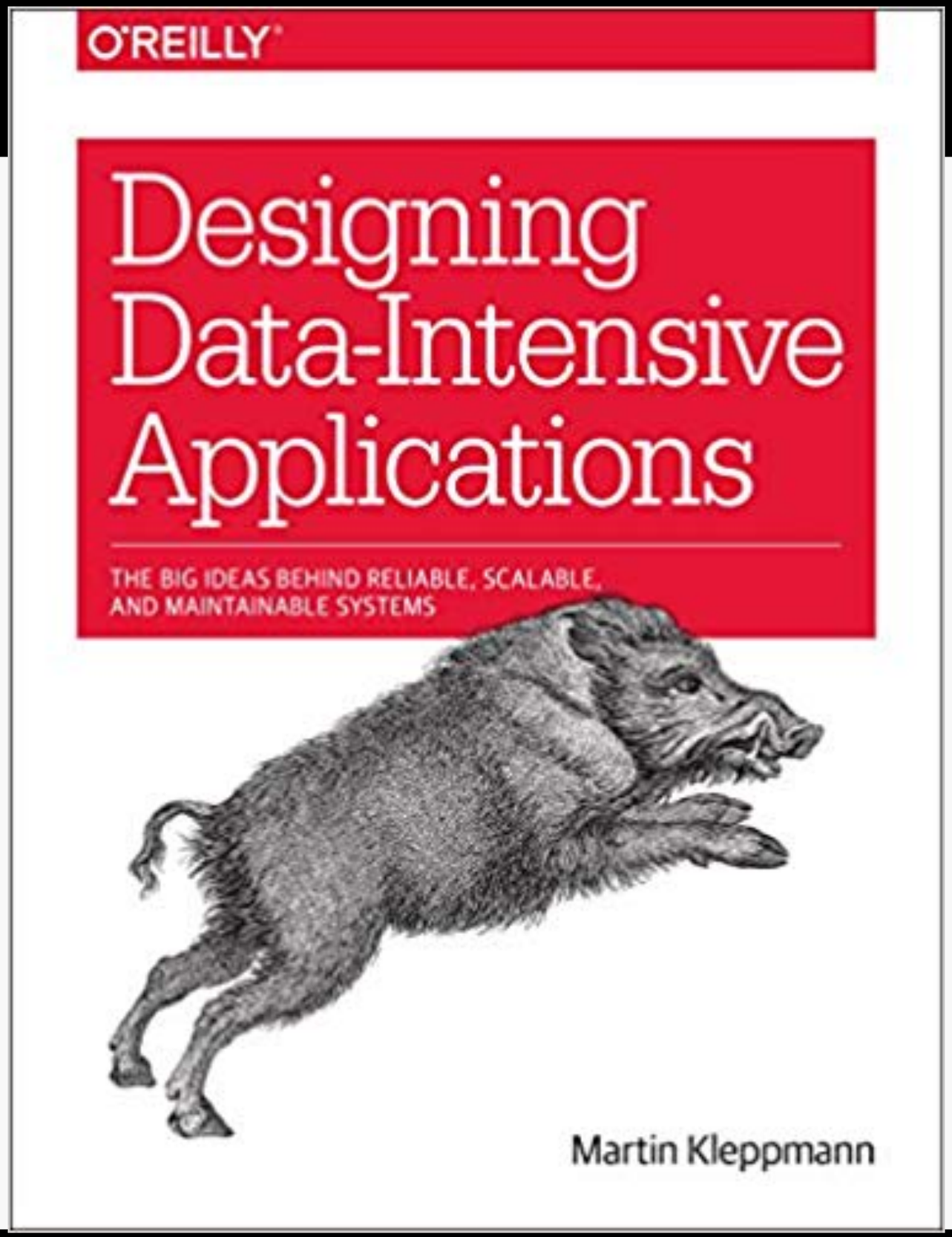
開發資料科學 PRODUCT PIPELINE



STUDY4.TW
為學習而生

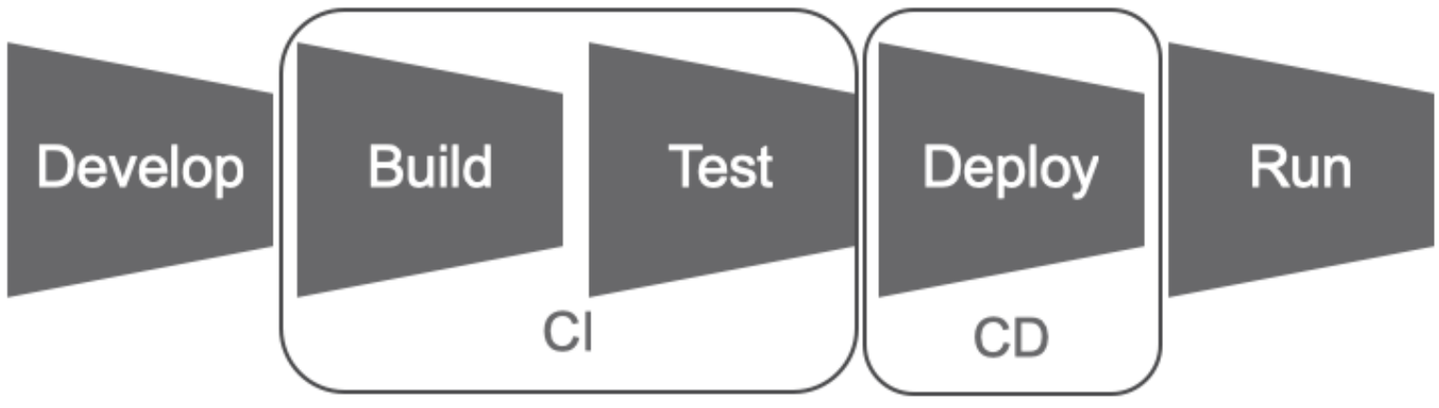
.NET Conf

延伸閱讀

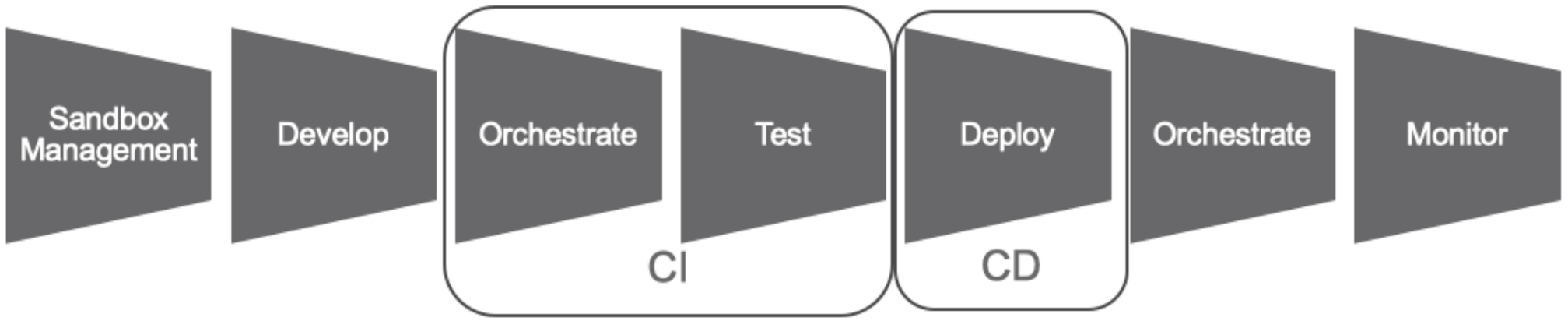


DATAOPS (AIOPS / MLOPS)

DevOps Process



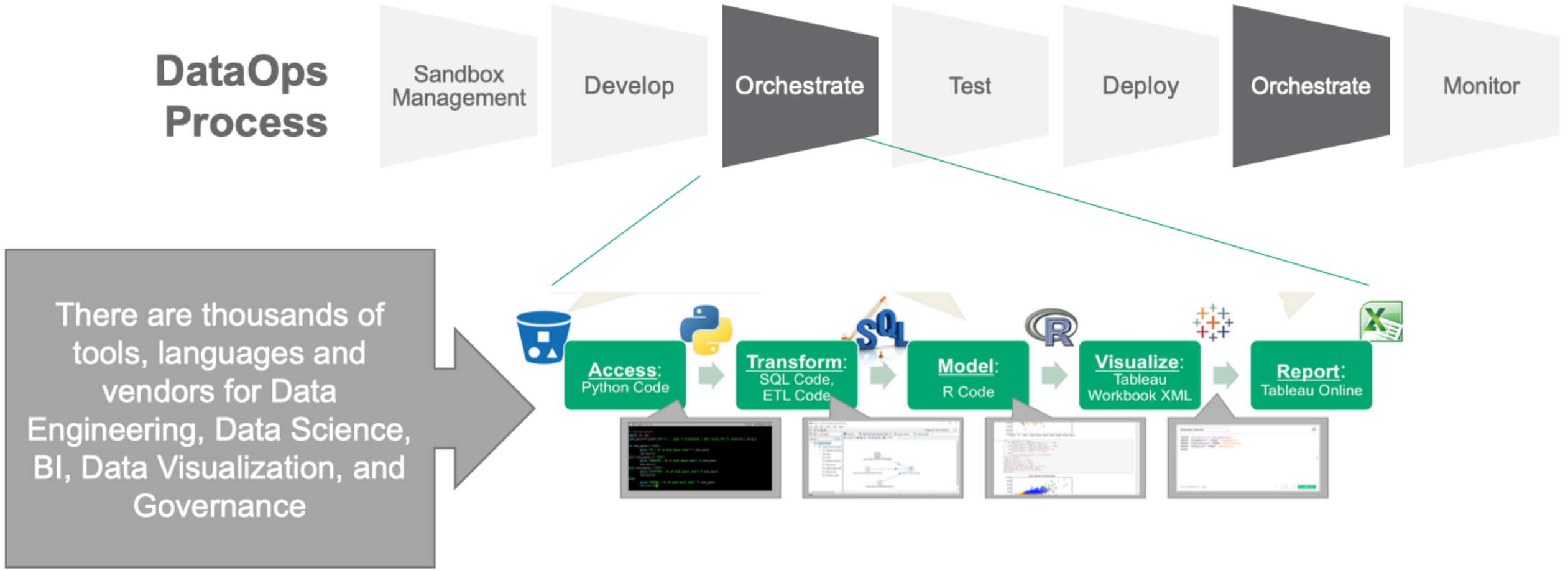
DataOps Process



STUDY4.TW 為學習而生

.NET Conf

DEV PHASE


DataOps Process
Sandbox Management
Develop
Orchestrate
Test
Deploy
Orchestrate
Monitor
There are thousands of tools, languages and vendors for Data Engineering, Data Science, BI, Data Visualization, and Governance
Access: Python Code
Transform: SQL Code, ETL Code
Model: R Code
Visualize: Tableau Workbook XML
Report: Tableau Online

OPS PHASE

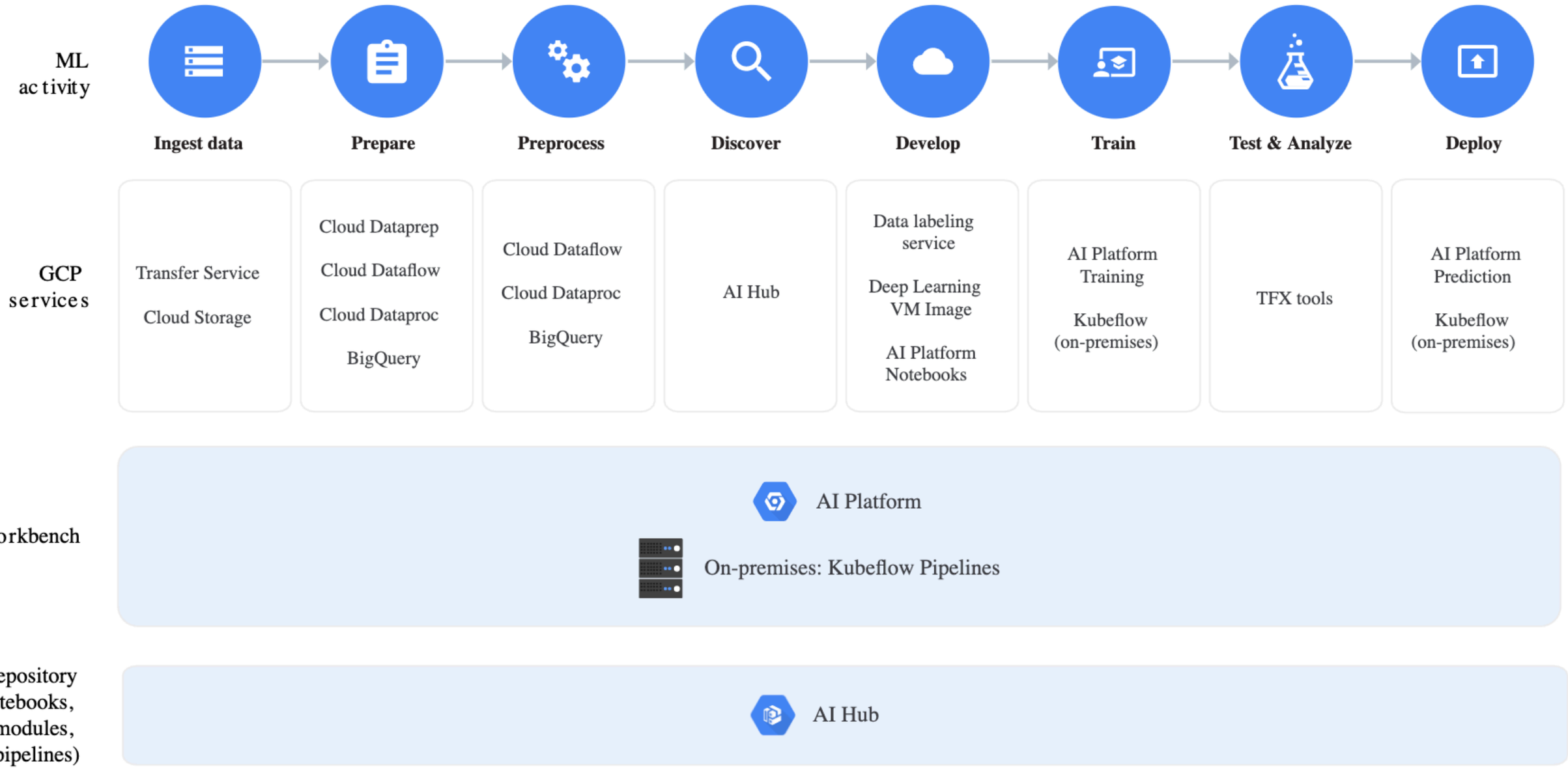
DataOps Process

Sandbox Management → Develop → **Orchestrate** → Test → Deploy → **Orchestrate** → Monitor

DataOps Production Requires Automated Testing, Monitoring and Statistical Process Control

來做一點趨勢觀察

GCP: AI PLATFORM



Kubeflow

default

Home

Pipelines

Notebook Servers

Katib

Artifact Store

GitHub

Documentation

Privacy • Usage Reporting

build version v1beta1

Dashboard | Activity

Quick shortcuts

Upload a pipeline
Pipelines

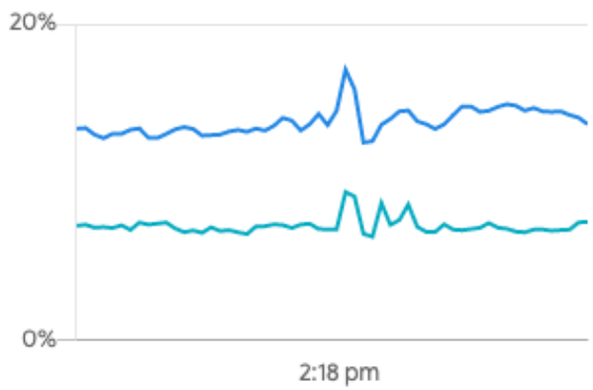
View all pipeline runs
Pipelines

Create a new Notebook server
Notebook Servers

View Katib Studies
Katib

View Metadata Artifacts
Artifact Store

Cluster CPU Utilization



Recent Notebooks

No Notebooks in namespace default

Recent Pipelines

[Sample] Basic - Exit Handler
Created 05/08/2019, 14:07:26

[Sample] Basic - Conditional execution
Created 05/08/2019, 14:07:24

[Sample] Basic - Parallel execution
Created 05/08/2019, 14:07:22

[Sample] Basic - Sequential execution
Created 05/08/2019, 14:07:21

[Sample] ML - TFX - Taxi Tip Prediction ...
Created 05/08/2019, 14:07:19

Recent Pipeline Runs

None Found

Google Cloud Platform

Stackdriver Logging
View cluster logs for the past hour

Project Overview
Manage your GCP Project

Deployment Manager
View your deployments

Kubernetes Engine
Administer your GKE clusters

Documentation

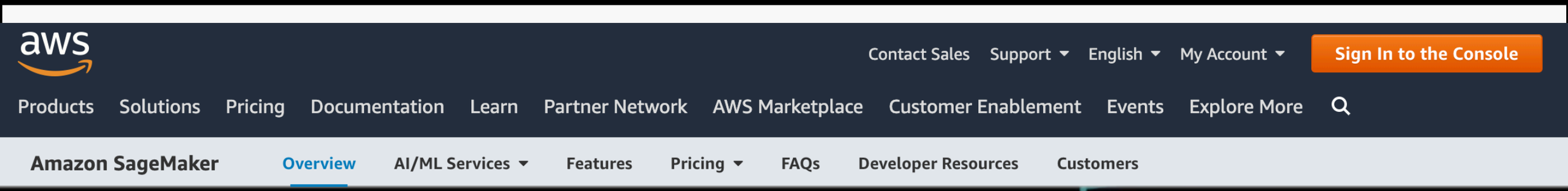
Getting Started with Kubeflow
Get your machine-learning workflow up and running on Kubeflow

MiniKF
A fast and easy way to deploy Kubeflow locally

Microk8s for Kubeflow
Quickly get Kubeflow running locally on native hypervisors

Minikube for Kubeflow
Quickly get Kubeflow running locally

AWS: SAGEMAKER



STUDY4.TW 為學習而生

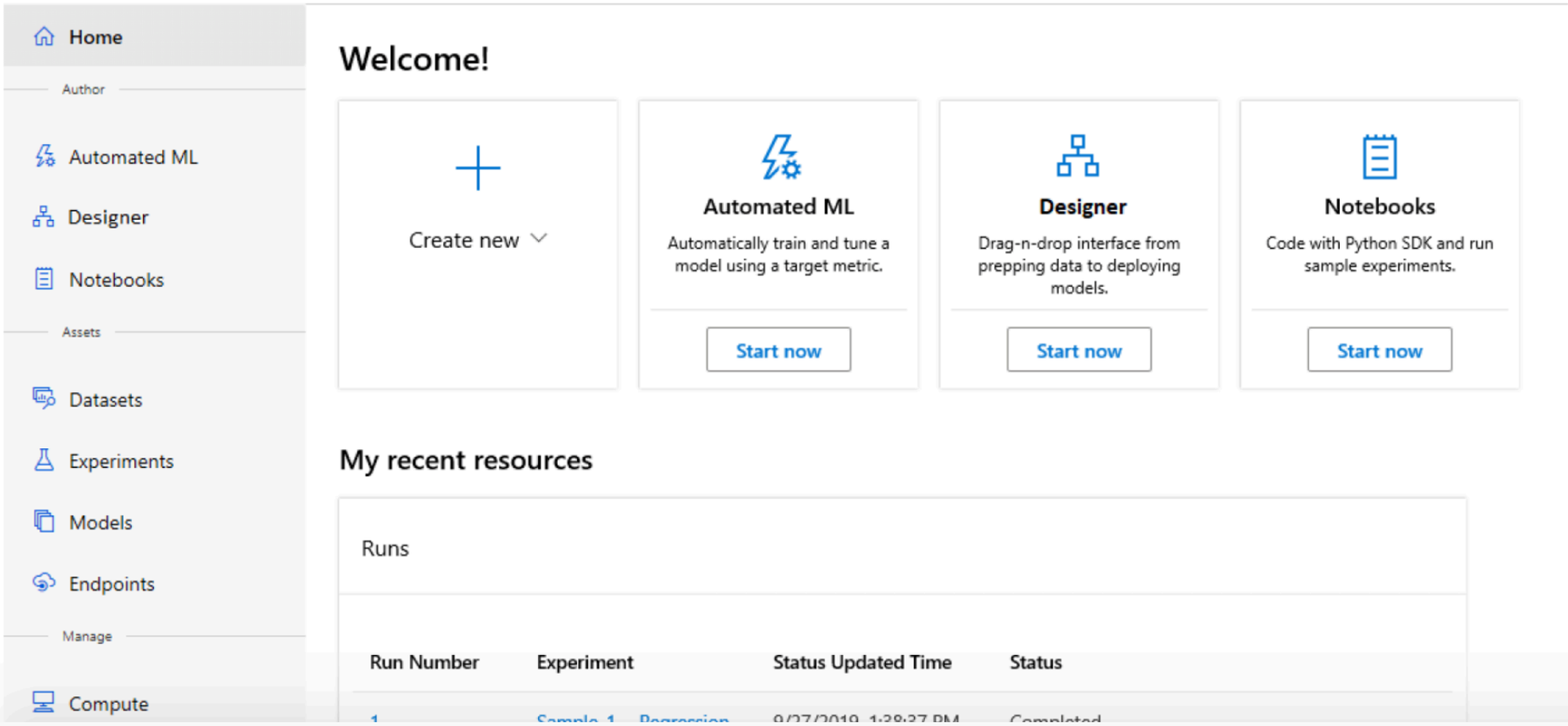
.NET Conf

AZURE: ML SERVICES

Go to your studio web experience

Build and train

Deploy and manage



You can author new models and store your compute targets, models, deployments, metrics, and run histories in the cloud.

AZURE: ML SERVICES

Go to your studio web experience

Build and train

Deploy and manage

Home > Auto ML

Automated Machine Learning

Automatically train custom machine learning models with minimum effort and machine learning expertise. Learn more

+ New automated ML run

Recent automated ML runs View all runs →

Experiment	Best Model	Author	Duration	Status	Creation date	Tags
Mytestexperiment	VotingEnsemble		31m 54.10s	Completed	Jun 13, 2019 3:46 PM	Key:Value Key:Value
Mytestrun	MinMaxScaler, RandomForest		31m 54.10s	Completed	Jun 13, 2019 3:46 PM	Key:Value
Testexperiment	MinMaxScaler, ExtremeRandomTrees		31m 54.10s	Completed	Jun 13, 2019 3:46 PM	Key:Value
MyAutoML Run	RobustScaler, ExtremeRandomTrees		31m 54.10s	Completed	Jun 13, 2019 3:46 PM	Key:Value
My AutoML run 2	StandardScalerWrapper, RandomForest		31m 54.10s	Completed	Jun 13, 2019 3:46 PM	Key:Value
My AutoML run 3	MinMaxScaler, ExtremeRandomTrees		31m 54.10s	Completed	Jun 13, 2019 3:46 PM	Key:Value

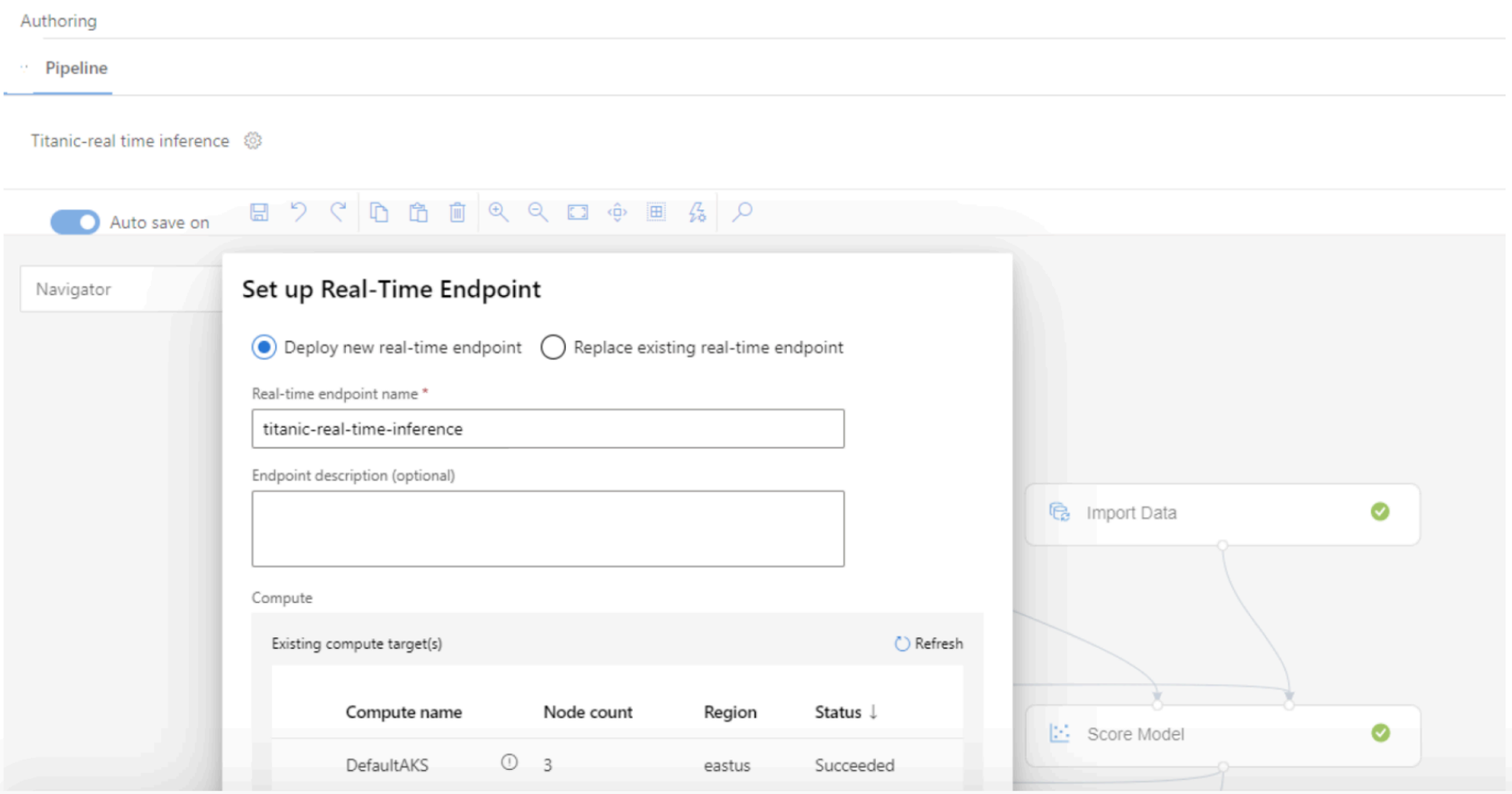
Use automated machine learning to identify algorithms and hyperparameters and track experiments in the cloud. You can also author models using notebooks or the drag and drop designer.

AZURE: ML SERVICES

Go to your studio web experience

Build and train

Deploy and manage



Deploy your machine learning model to the cloud or the edge, monitor performance, and retrain it as needed.

SELDON

SELDON DEPLOY

UI, collaboration, control, audit

Multi-arm bandits | Outlier detection | Explanation | Bias detection

SELDON CORE

runtime ML graph engine

microservices

Istio service mesh (optional)

kubernetes

deploy anywhere

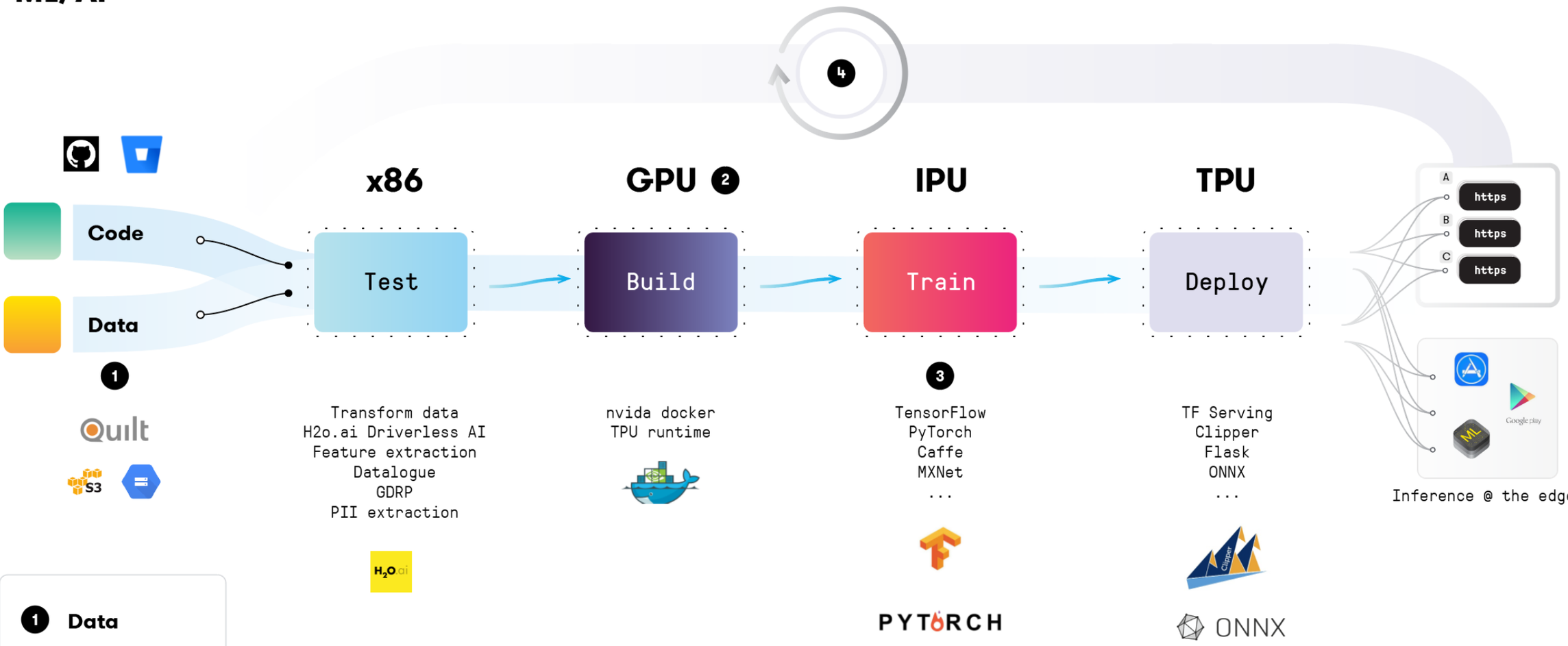
cloud services or on-prem

STUDY4.TW 為學習而生

.NET Conf

PAPERSPACE

ML/AI



1. **Data**
2. **Accelerators**
3. **Train step**
4. **Refitting**

INFUSEAI

PrimeHub Features

Cluster Computing

- Rapid construction of research environments
- Expansion to hundreds of nodes
- Container orchestration with Kubernetes
- Supports to on-premises and cloud installations

One-click Research Environment

- Develop interactively with Jupyter
- Support various deep learning frameworks
- Visualize training progress

Enterprise-class account management

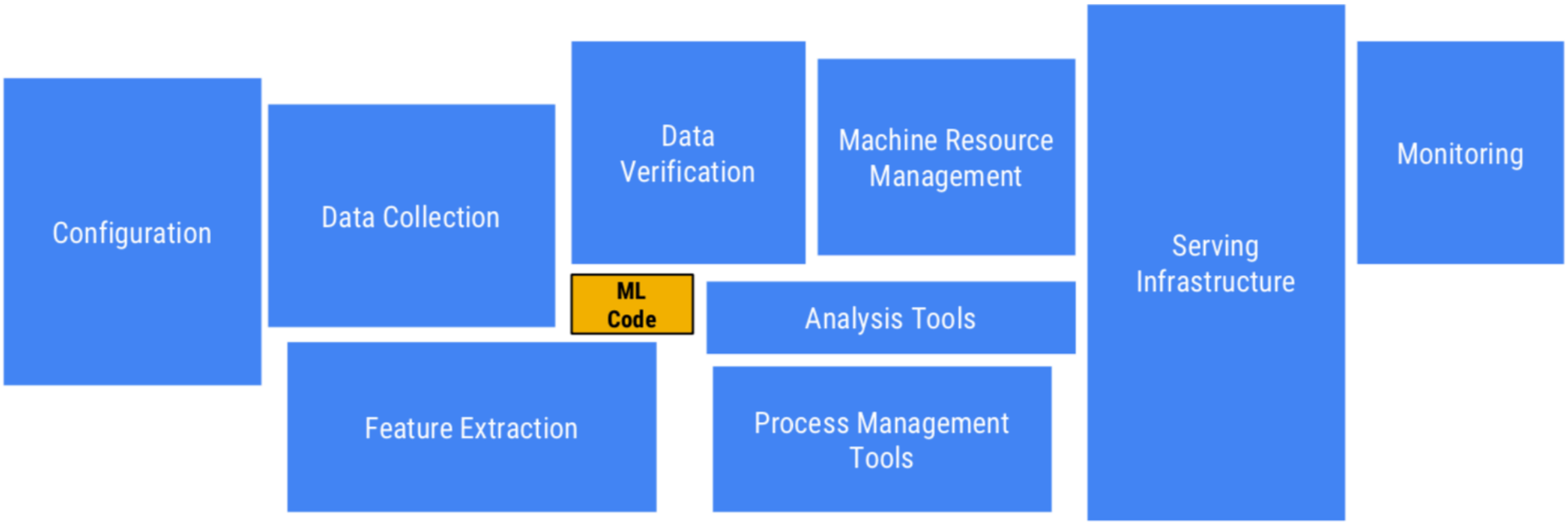
- 2FA user account protection
- Support to Single Sign-On (SSO)
- Tools for internal auditing

Management of resource quota and privileges

- Personal and shared group folders
- Fine-grained quota allocation for members and groups

資料科學團隊分工

DATA SCIENCE "SILO"

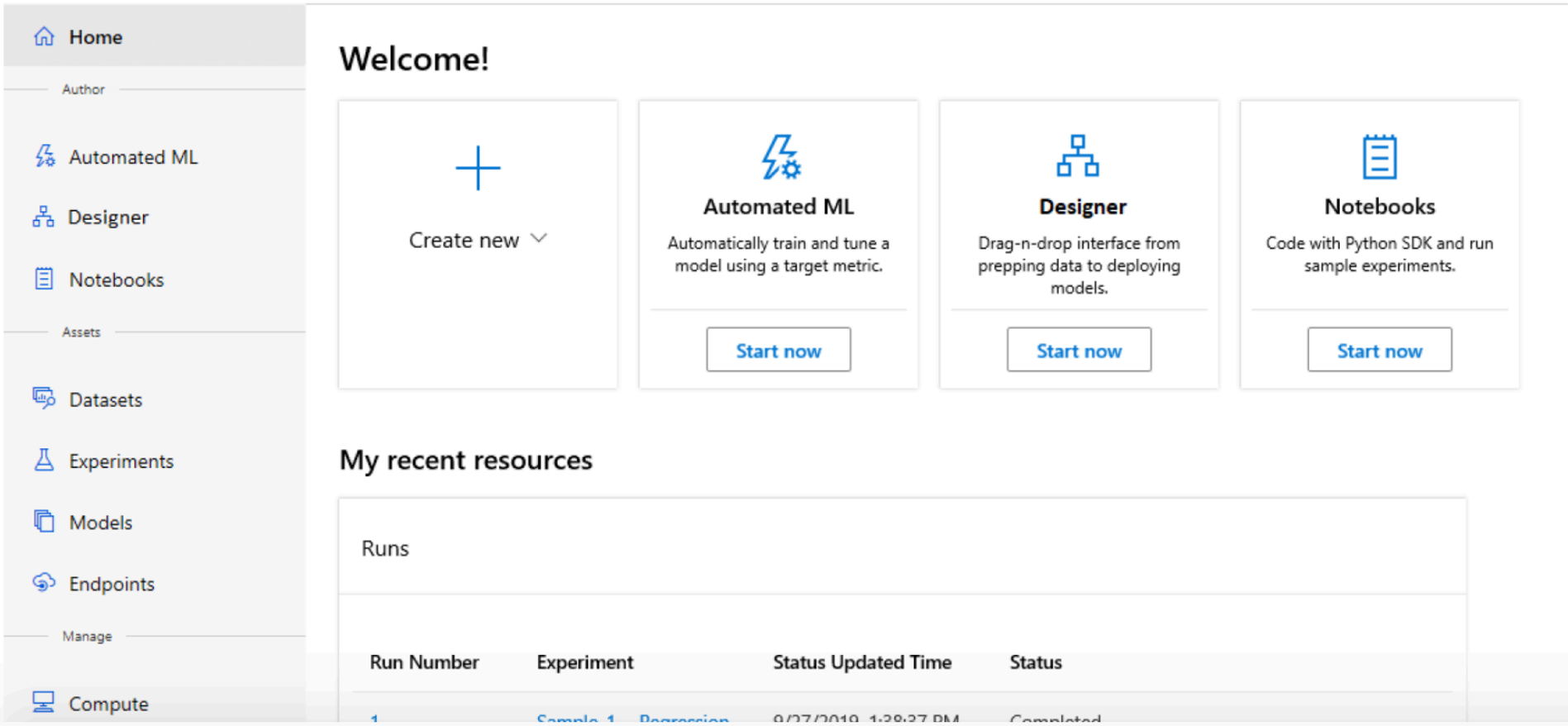


DATA ANALYST

Go to your studio web experience

Build and train

Deploy and manage



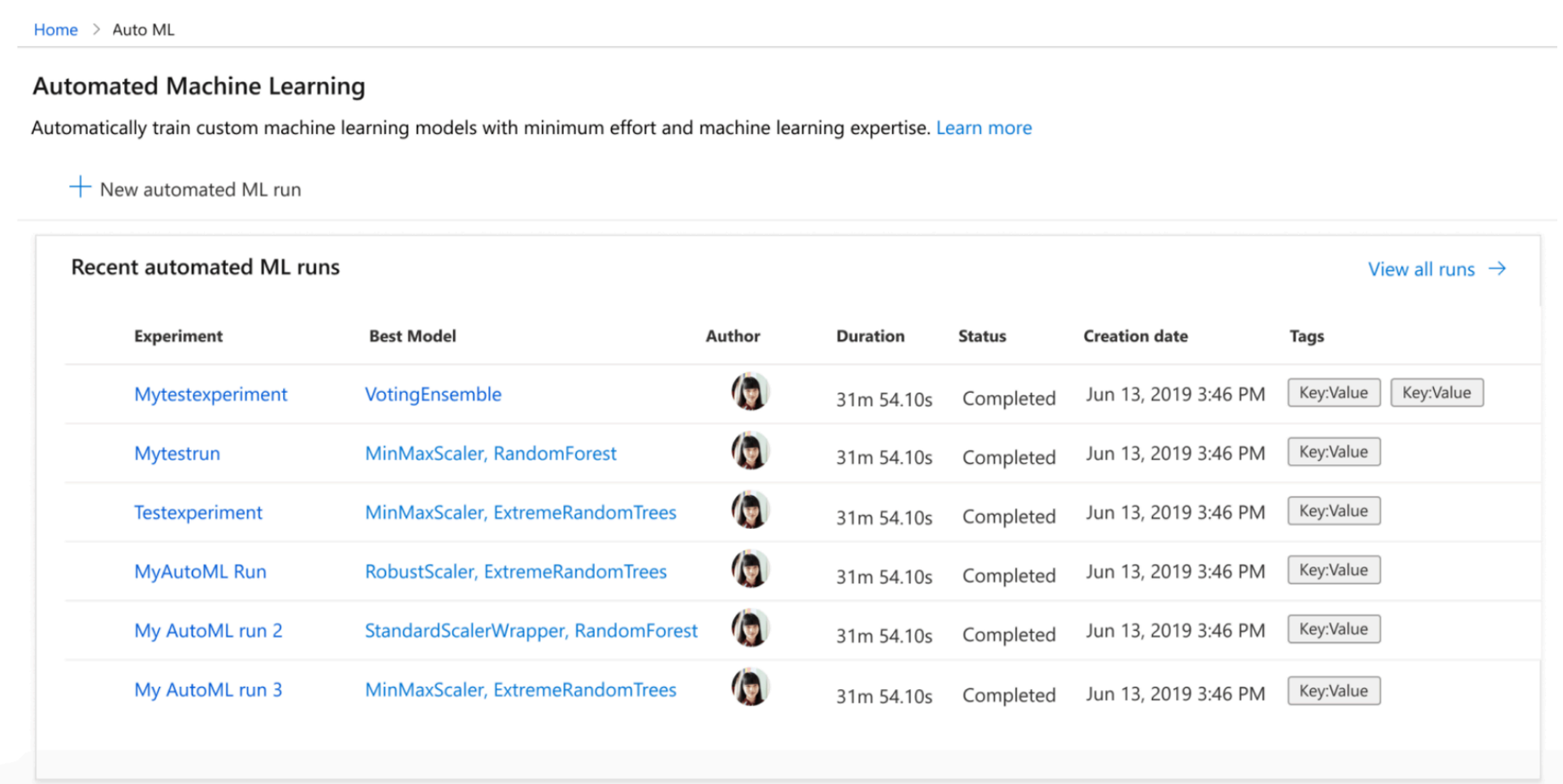
You can author new models and store your compute targets, models, deployments, metrics, and run histories in the cloud.

ML ENGINEER/OPS

Go to your studio web experience

Build and train

Deploy and manage



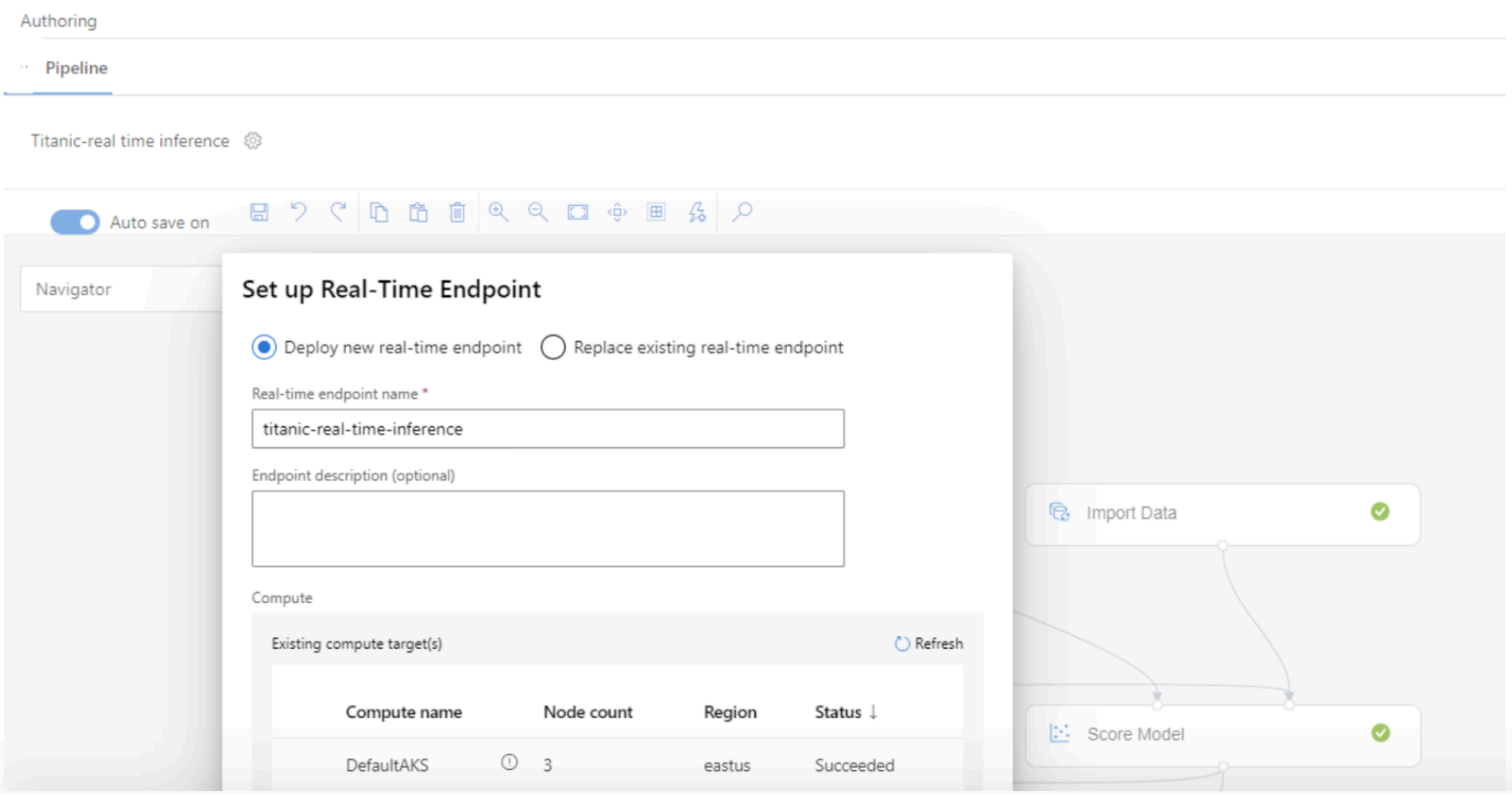
Use automated machine learning to identify algorithms and hyperparameters and track experiments in the cloud. You can also author models using notebooks or the drag and drop designer.

AI ARCHITECT/OPS

Go to your studio web experience

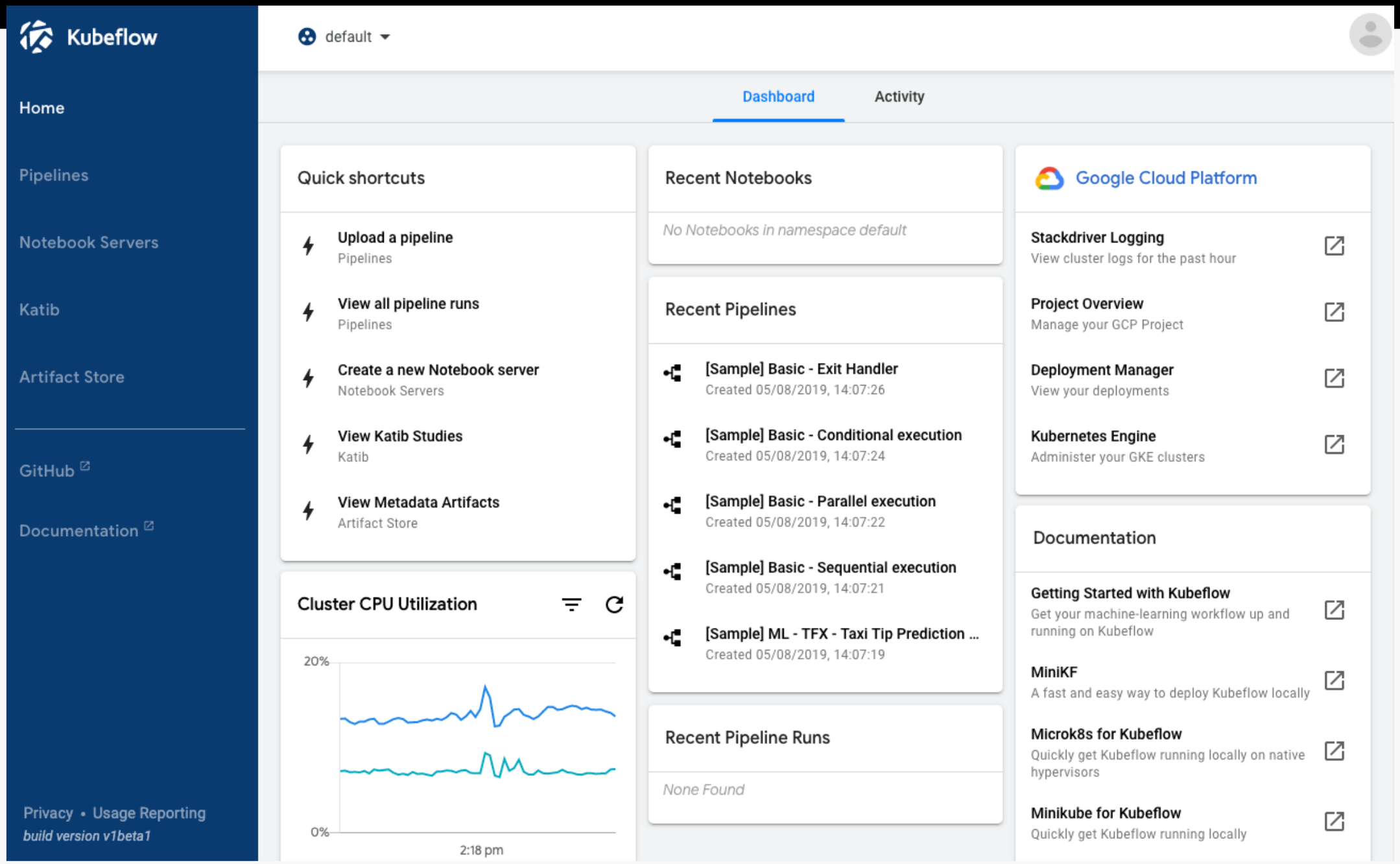
Build and train

Deploy and manage



Deploy your machine learning model to the cloud or the edge, monitor performance, and retrain it as needed.

DATA SCIENTIST



QUESTIONS?

特別感謝

Microsoft MVP Most Valuable Professional

SEA 台灣軟體工程學會 Taiwan

R-Ladies Taipei

TAIWAN

91APP

多奇·數位創意

Wishing-Soft KKTIX Geek & Nerd Studios HackMD

以及各位參與活動的你們


STUDY4.TW 為學習而生

.NET Conf