



# .NET Conf

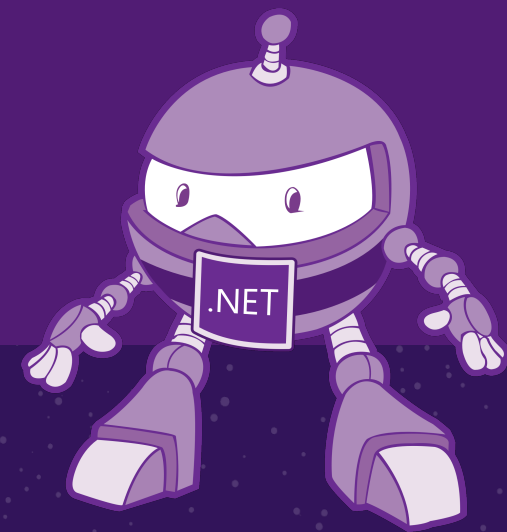
探索 .NET 新世界

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



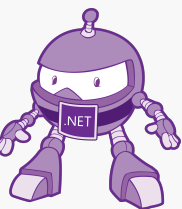
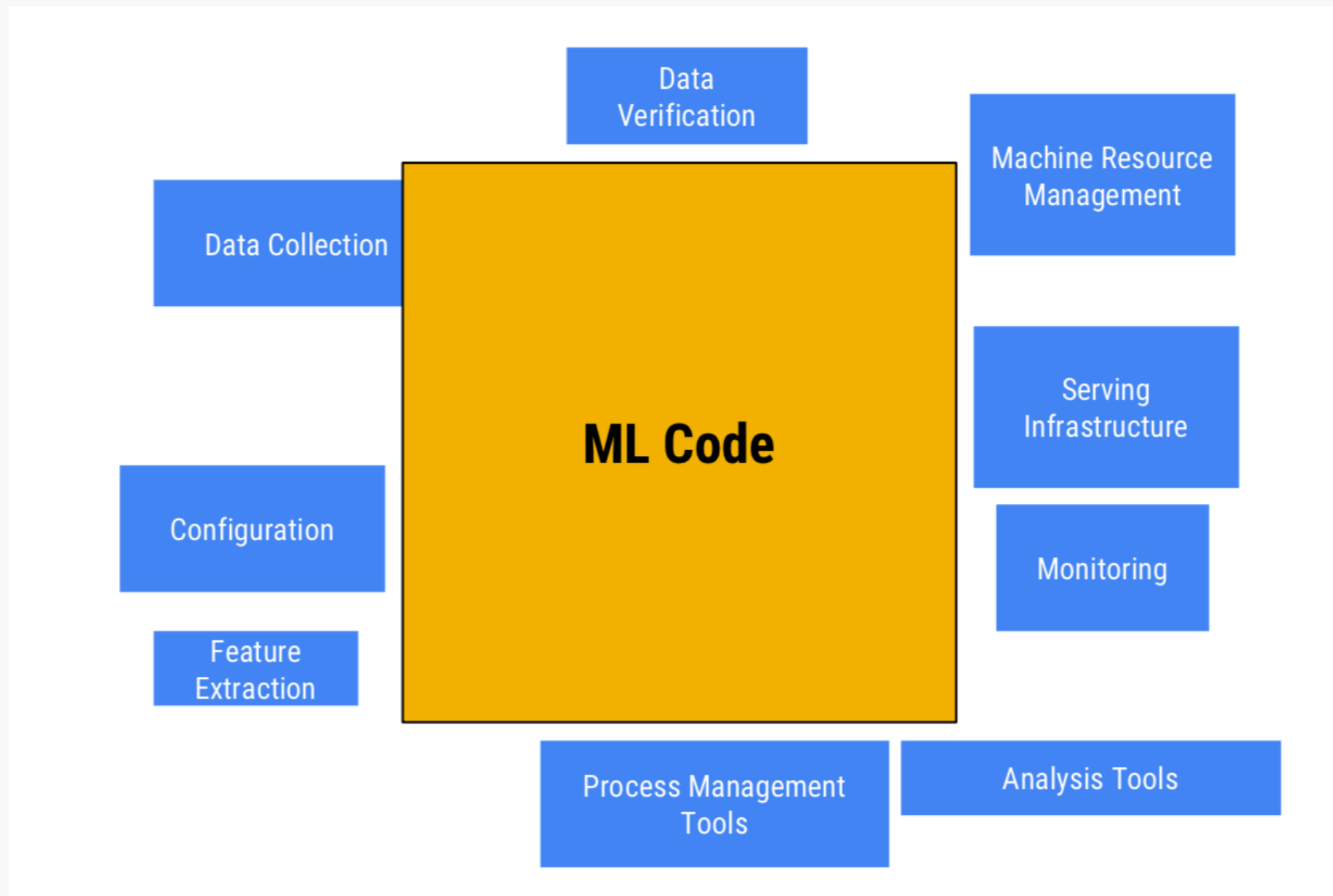


# 資料科學家與他們的產地

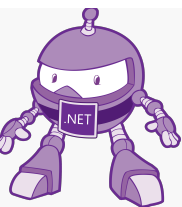
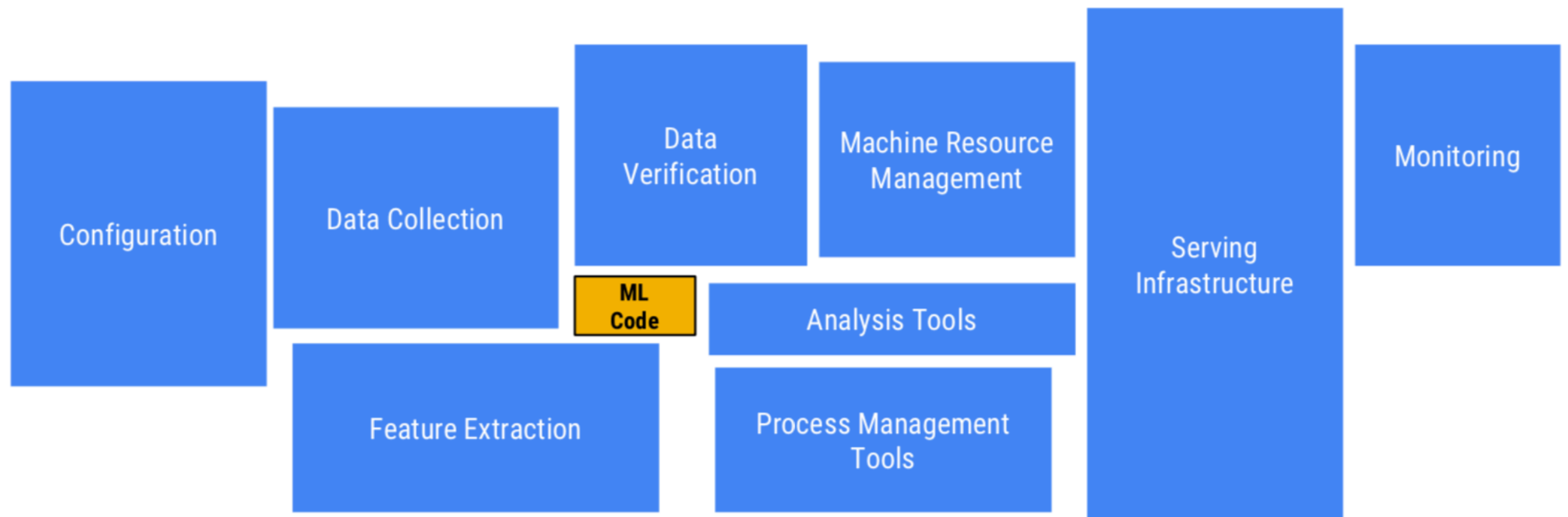




# 我以為的資料科學家



# 實際上的資料科學家



# 一條龍資料科學家你就是那條龍



Yang Li Wei was live in 臺灣資料工程協會.  
on Wednesday · 🌐

今天很高興邀請到孫玉峰大大來到現場～  
#資料工程協會  
#專家線上見

👍❤️😂 38

13 Shares 2.1K Views



Love



Comment



Share

Comments

Up Next

Most Relevant ▾



Anna Yen · 1:49 感謝玉峰大大蒞臨臺灣資料工程協會！

Like · Reply · 3d



林冠廷 · 1:37 就位

Like · Reply · 3d



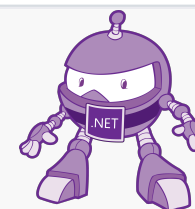
楊明翰 · 5:16



Like · Reply · 3d

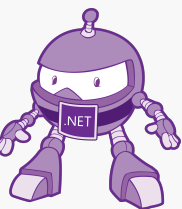
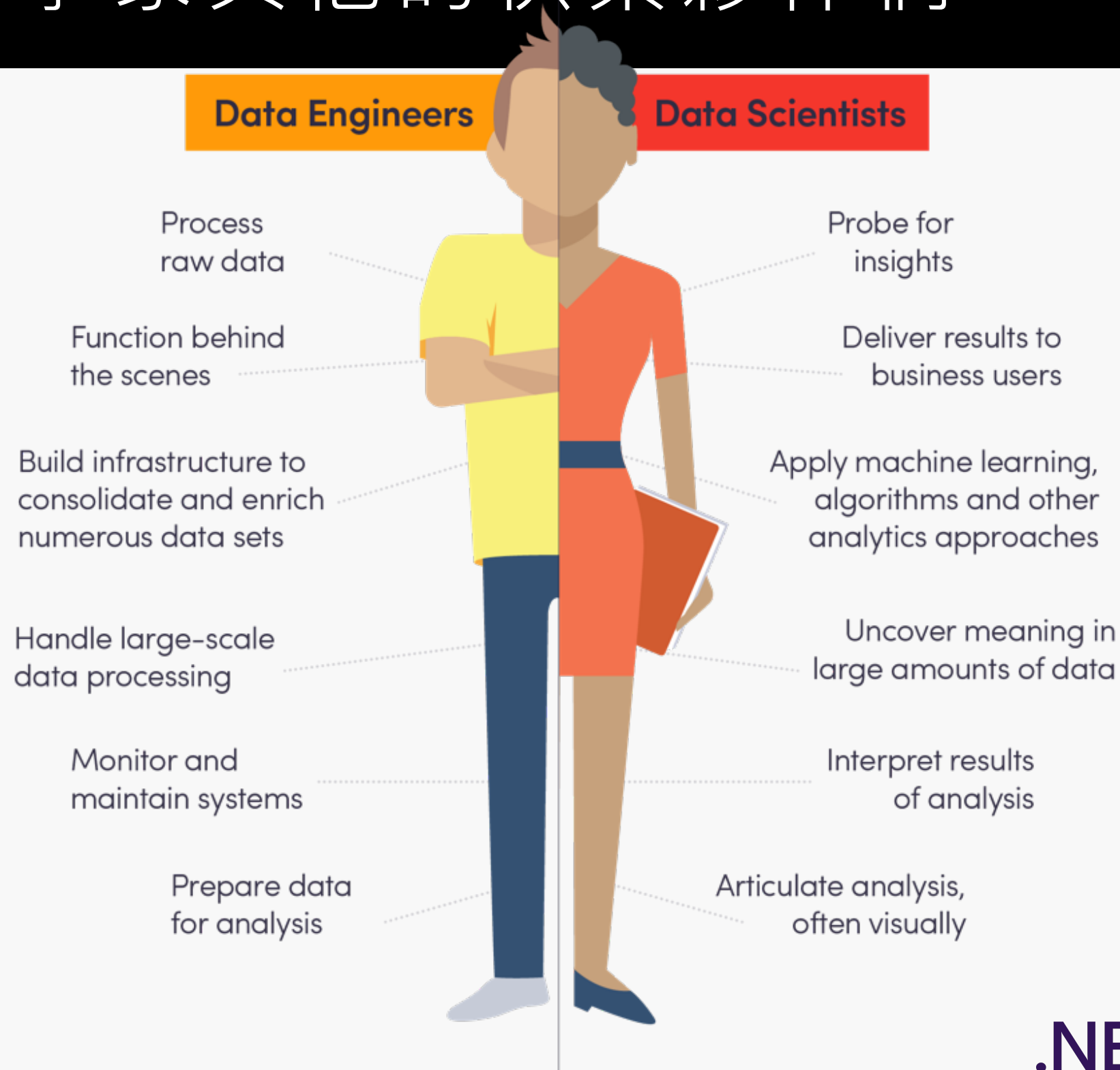


ErhWen Kuo · 3:20 趕快上線來看大神

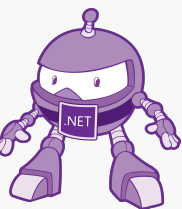




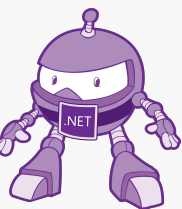
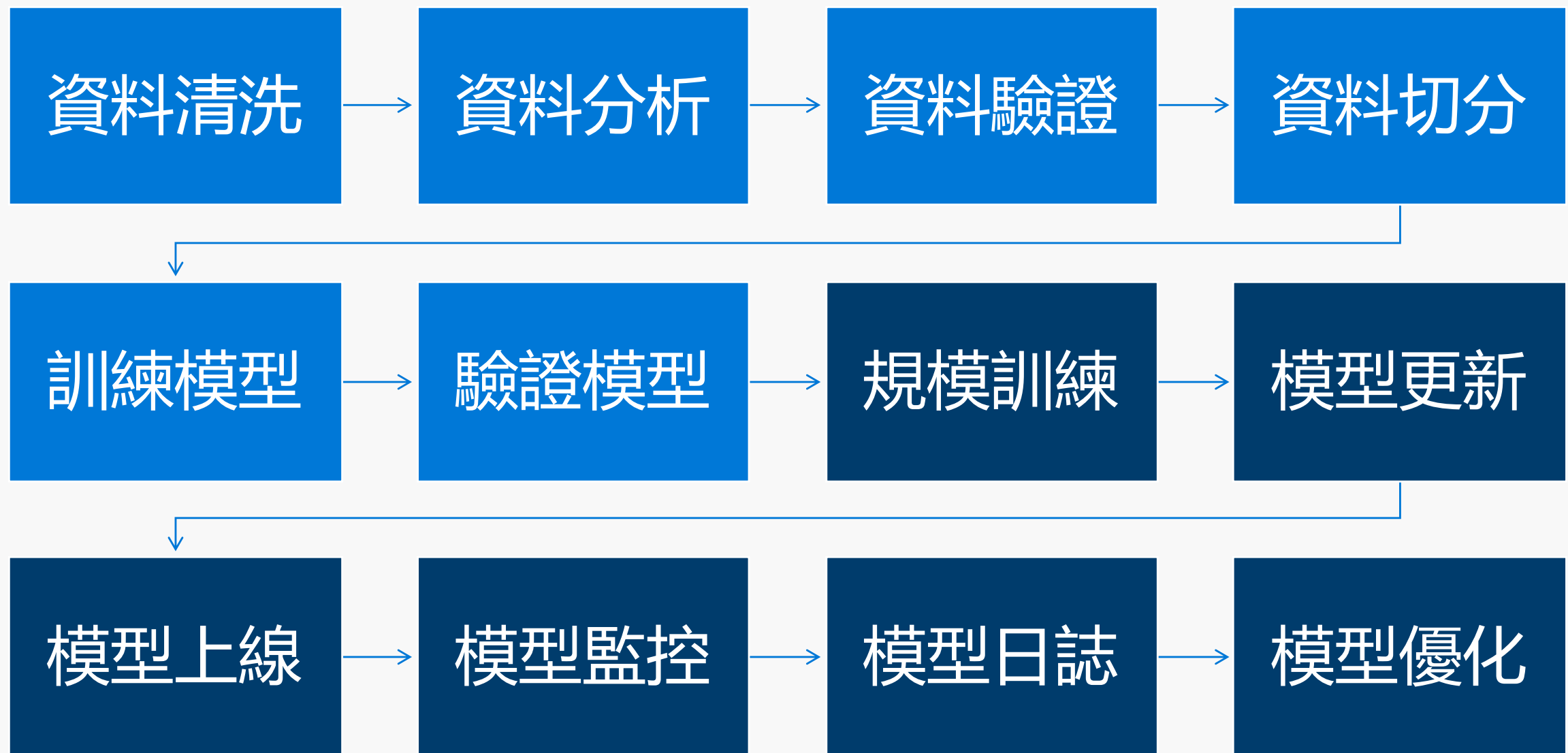
# 資料科學家與他的快樂夥伴們



# 開發資料科學 PROJECT PIPELINE

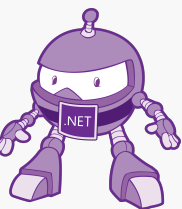


# 開發資料科學 PRODUCT PIPELINE



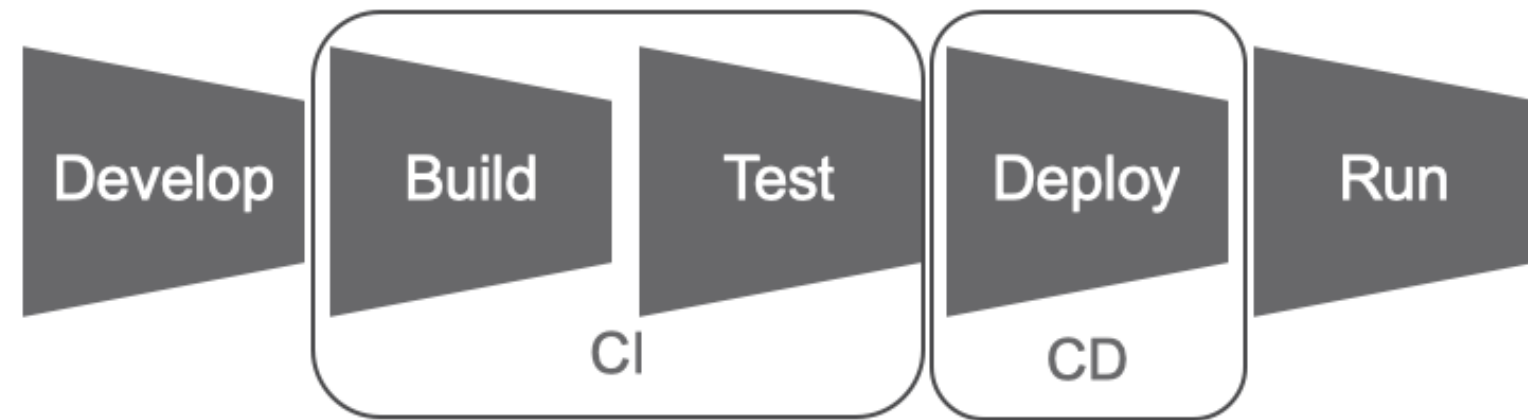


# 延伸閱讀

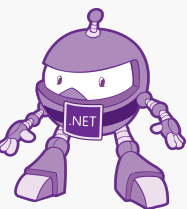
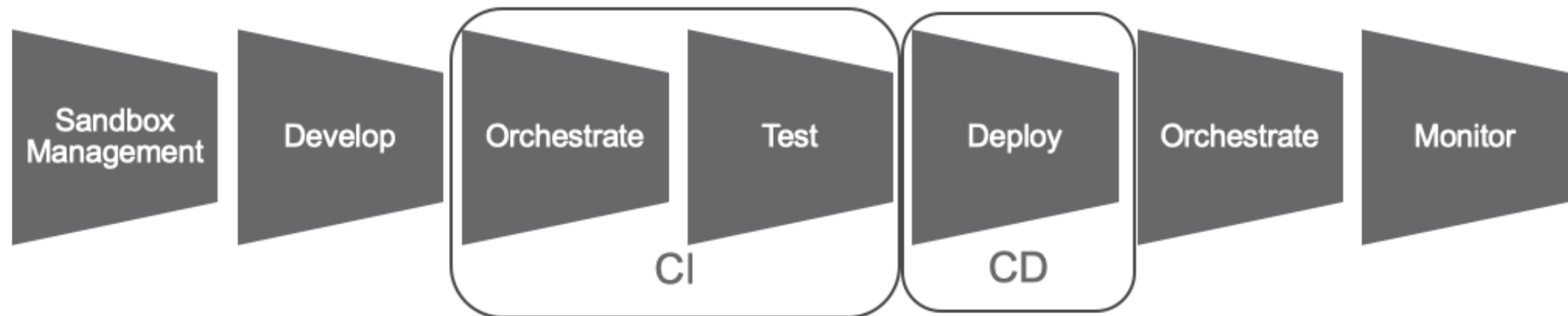


# DATAOPS (AIOPS / MLOPS)

## DevOps Process



## DataOps Process

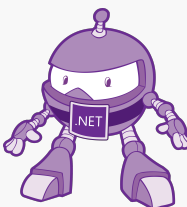


# DEV PHASE

## DataOps Process



There are thousands of tools, languages and vendors for Data Engineering, Data Science, BI, Data Visualization, and Governance





# OPS PHASE

## DataOps Process

Sandbox Management

Develop

Orchestrate

Test

Deploy

Orchestrate

Monitor

DataOps Production  
Requires Automated  
Testing, Monitoring and  
Statistical Process Control

Tests Timings

Passed Log Warning Failed

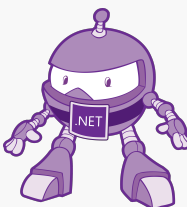
qa\_count\_unique\_bad\_ids  
Node: CheckDataInSQLServer



qa\_count\_all\_bad\_ids  
Node: CheckDataInSQLServer



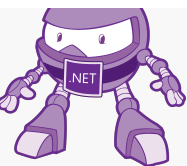
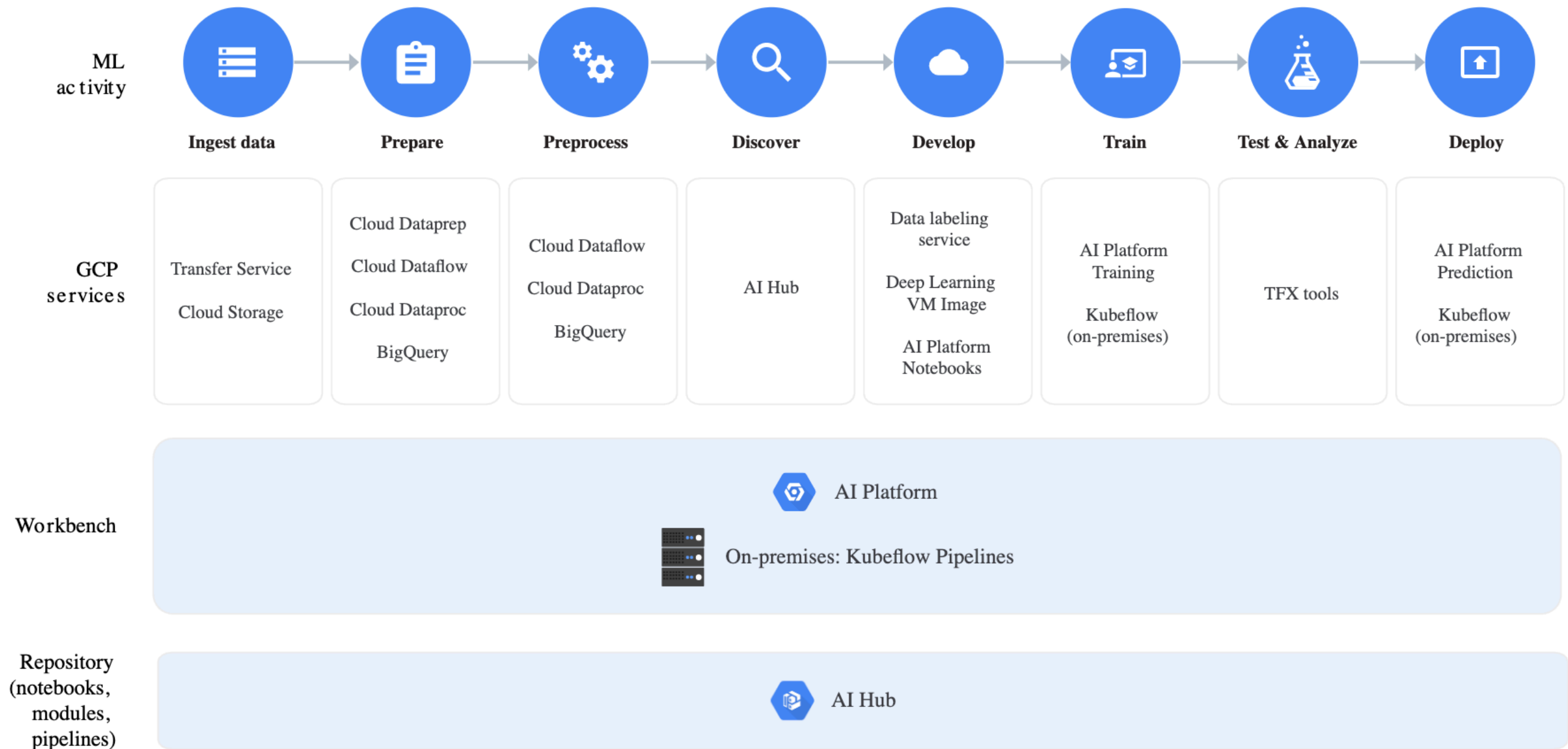
count\_raw\_order\_rows  
Node: CheckDataInSQLServer



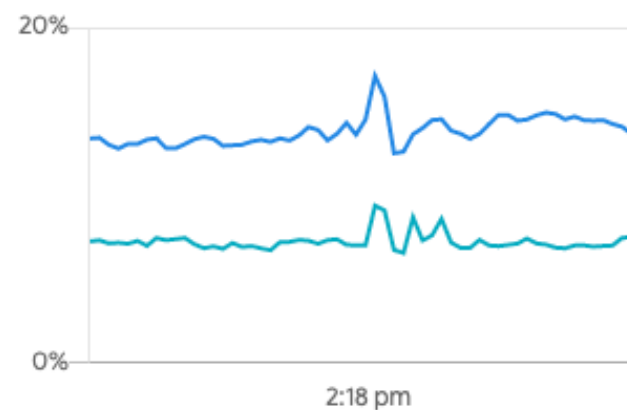
來做一點趨勢觀察



# GCP: AI PLATFORM







# AWS: SAGEMAKER



[Contact Sales](#) [Support](#) [English](#) [My Account](#)

[Sign In to the Console](#)

[Products](#) [Solutions](#) [Pricing](#) [Documentation](#) [Learn](#) [Partner Network](#) [AWS Marketplace](#) [Customer Enablement](#) [Events](#) [Explore More](#) [Q](#)

**Amazon SageMaker**

[Overview](#)

[AI/ML Services](#)

[Features](#)

[Pricing](#)

[FAQs](#)

[Developer Resources](#)

[Customers](#)

## BUILD

### Collect & prepare training data

Data labeling & pre-built notebooks for common problems

### Choose & optimize your ML algorithm

Built-in, high-performance algorithms and hundreds of ready to use algorithms in [AWS Marketplace](#)

## TRAIN

### Set up & manage environments for training

One-click training using Amazon EC2 On-Demand or Spot instances

### Train & tune model

Train once, run anywhere & model optimization

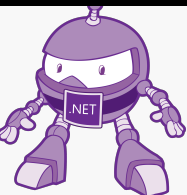
## DEPLOY

### Deploy model in production

One-click deployment

### Scale & manage the production environment

Fully managed with auto-scaling for 75% less



# AZURE: ML SERVICES

Go to your studio web experience



Build and train



Deploy and manage



Home

Author

Automated ML

Designer

Notebooks

Assets

Datasets

Experiments

Models

Endpoints

Manage

Compute

## Welcome!



Create new ▾



### Automated ML

Automatically train and tune a model using a target metric.

Start now



### Designer

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

Start now



### Notebooks

Code with Python SDK and run sample experiments.

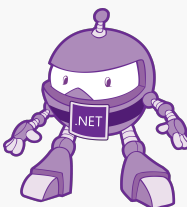
Start now

## My recent resources

Runs

Run Number	Experiment	Status Updated Time	Status
1	Sample 1 Regression	9/27/2019 1:29:27 PM	Completed

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
<a href="#">Mytestexperiment</a>	<a href="#">VotingEnsemble</a>		31m 54.10s	Completed	Jun 13, 2019 3:46 PM	<a href="#">Key:Value</a> <a href="#">Key:Value</a>
<a href="#">Mytestrun</a>	<a href="#">MinMaxScaler, RandomForest</a>		31m 54.10s	Completed	Jun 13, 2019 3:46 PM	<a href="#">Key:Value</a>
<a href="#">Testexperiment</a>	<a href="#">MinMaxScaler, ExtremeRandomTrees</a>		31m 54.10s	Completed	Jun 13, 2019 3:46 PM	<a href="#">Key:Value</a>
<a href="#">MyAutoML Run</a>	<a href="#">RobustScaler, ExtremeRandomTrees</a>		31m 54.10s	Completed	Jun 13, 2019 3:46 PM	<a href="#">Key:Value</a>
<a href="#">My AutoML run 2</a>	<a href="#">StandardScalerWrapper, RandomForest</a>		31m 54.10s	Completed	Jun 13, 2019 3:46 PM	<a href="#">Key:Value</a>
<a href="#">My AutoML run 3</a>	<a href="#">MinMaxScaler, ExtremeRandomTrees</a>		31m 54.10s	Completed	Jun 13, 2019 3:46 PM	<a href="#">Key:Value</a>

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



Authoring

Pipeline

Titanic-real time inference



Auto save on



Navigator

## Set up Real-Time Endpoint

☒ Deploy new real-time endpoint ☐ Replace existing real-time endpoint

Real-time endpoint name \*

titanic-real-time-inference

Endpoint description (optional)

Compute

Existing compute target(s)

Refresh

Compute name	Node count	Region	Status ↓
DefaultAKS	3	eastus	Succeeded

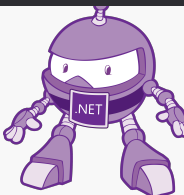
Import Data



Score Model



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

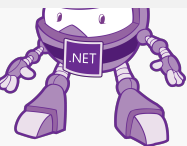
Istio service mesh (optional)



kubernetes

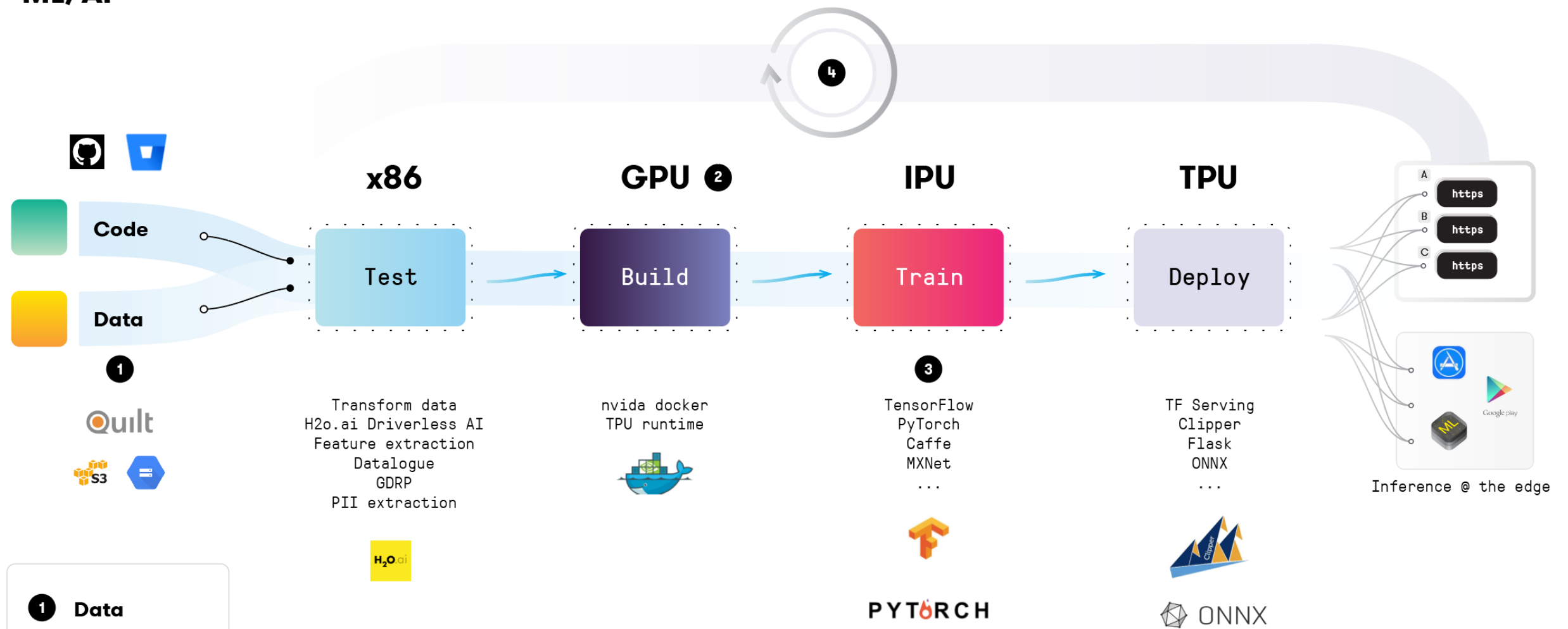
deploy anywhere

cloud services or on-prem

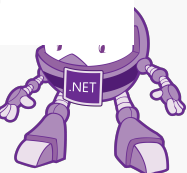


# PAPERSPACE

ML/AI




© Paperspace 2018



# INFUSEAI



[Home](#) · [PrimeHub](#) · [Documentation](#) · [GitHub](#) · [Blog](#) · [Jobs](#) · [About](#) 

## 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



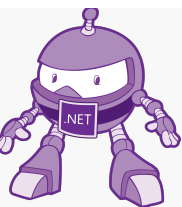
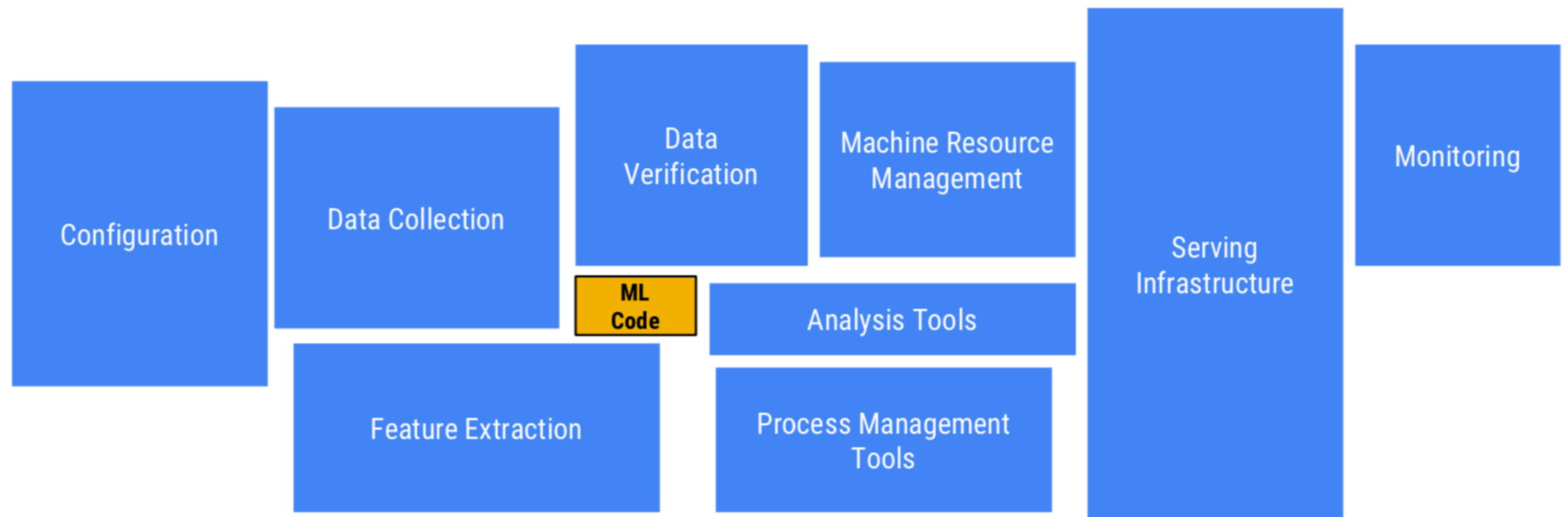
### Management of resource quota and privileges

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





# DATA SCIENCE "SILO"





# DATA ANALYST

Go to your studio web experience



Build and train



Deploy and manage



Home

Author

Automated ML

Designer

Notebooks

Assets

Datasets

Experiments

Models

Endpoints

Manage

Compute

## Welcome!



Create new ▾



### Automated ML

Automatically train and tune a model using a target metric.

Start now



### Designer

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

Start now



### Notebooks

Code with Python SDK and run sample experiments.

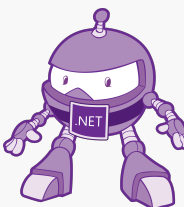
Start now

## My recent resources

Runs

Run Number	Experiment	Status Updated Time	Status
1	Sample 1 Regression	9/27/2019 1:29:27 PM	Completed

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



[Home](#) > Auto ML







## Automated Machine Learning

Automatically train custom machine learning models with minimum effort and machine learning expertise. [Learn more](#)

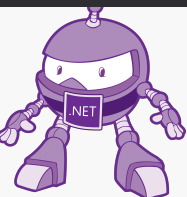
[+](#) New automated ML run

### Recent automated ML runs

[View all runs](#) →

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

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



Authoring

Pipeline

Titanic-real time inference



Auto save on



Navigator

## Set up Real-Time Endpoint

☒ Deploy new real-time endpoint ☐ Replace existing real-time endpoint

Real-time endpoint name \*

titanic-real-time-inference

Endpoint description (optional)

Compute

Existing compute target(s)

Refresh

Compute name	Node count	Region	Status ↓
DefaultAKS	3	eastus	Succeeded

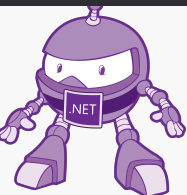
Import Data



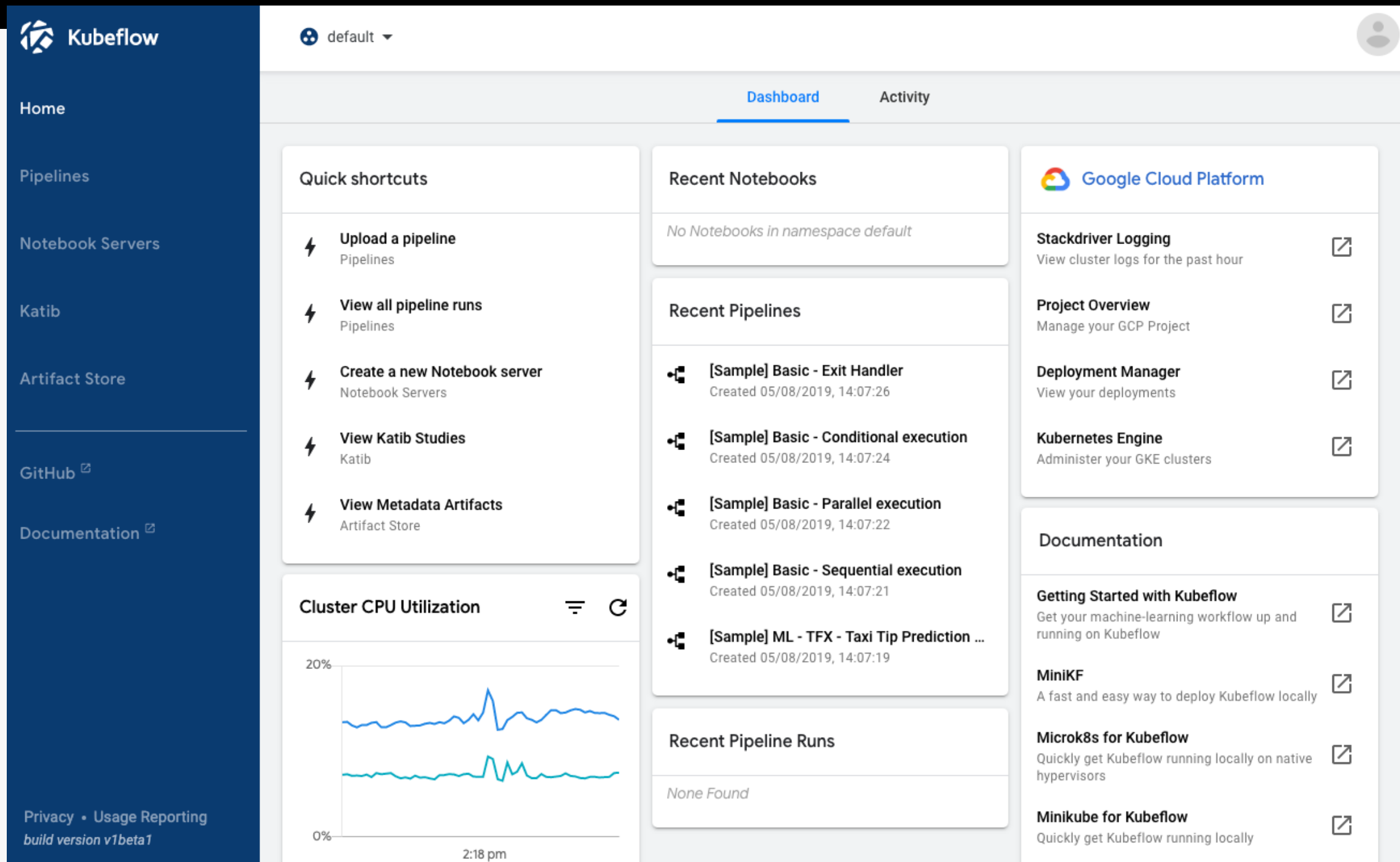
Score Model



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



# DATA SCIENTIST



The image shows the Kubeflow Dashboard interface. On the left is a dark blue sidebar with navigation links: Home, Pipelines, Notebook Servers, Katib, Artifact Store, GitHub, and Documentation. The main content area has a top bar with 'default' namespace and tabs for 'Dashboard' and 'Activity'. The dashboard is divided into several sections: 'Quick shortcuts' with links to upload pipelines, view pipeline runs, create notebook servers, view Katib studies, and view metadata artifacts; 'Recent Notebooks' showing 'No Notebooks in namespace default'; 'Recent Pipelines' listing sample pipelines like 'Exit Handler', 'Conditional execution', 'Parallel execution', 'Sequential execution', and 'ML - TFX - Taxi Tip Prediction'; 'Cluster CPU Utilization' with a line graph showing usage over time; 'Recent Pipeline Runs' showing 'None Found'; and a 'Google Cloud Platform' section with links to Stackdriver Logging, Project Overview, Deployment Manager, and Kubernetes Engine. A 'Documentation' section at the bottom right lists links for getting started, Minikf, Microk8s, and Minikube.

**Kubeflow**

default

**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

**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

**Cluster CPU Utilization**

20%  
0%  
2:18 pm

**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

Privacy • Usage Reporting  
build version v1beta1

# QUESTIONS?





# 特別感謝



R-Ladies Taipei



多奇·數位創意



## 以及各位參與活動的你們



STUDY4.TW  
為 學 習 而 生

# .NET Conf

