走入雲端，擁抱 AI

— 從雲端代理商，看未來的 AI 技術應用發展

余佑駿 Youjun (宏庭科技 Microfusion Technology)
@紐約Baruch College 全球校友百業分享堂
Mar. 2022

余佑駿 Youjun
Cloud Architect

littlefish0331

you-jun-yu

littlefish0331

littlefish0331@gmail.com

經歷】

2021 宏庭股份有限公司 Microfusion Technology
GCP解決方案部 雲端架構師

2019 國家高速網路與計算中心 資料科學家

2017 國立政治大學 統計所

2012 國立清華大學 數學系輔修教育學程

Linktree

履歷、專案
社群分享

競賽】

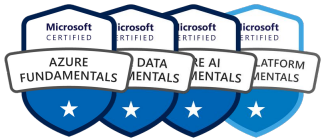
2021 (Tbrain) Tomofun 狗音辨識: 4th + 評審獎

2019 (Kaggle) LANL Earthquake Prediction: top 23%

宏庭科技
股份有限公司

我目前公司。
上雲需求歡迎

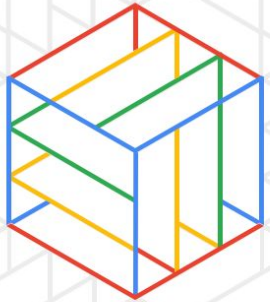
塵世中一個迷途小書僮，從資料科學走向雲端科技。
平時喜歡數據分析、機器學習、參與社群活動。
希望用科技創造更美好的世界。



可以交流
考證照經驗

AGENDA

- 什麼是雲端科技
- 人工智慧與機器學習的發展與普及
- 打造一個人工智慧的王國
- Demo: 物件辨識模型
- 擁抱 AI & ML 的下一步
- 漫漫長路，承先啟後



什麼是雲端科技

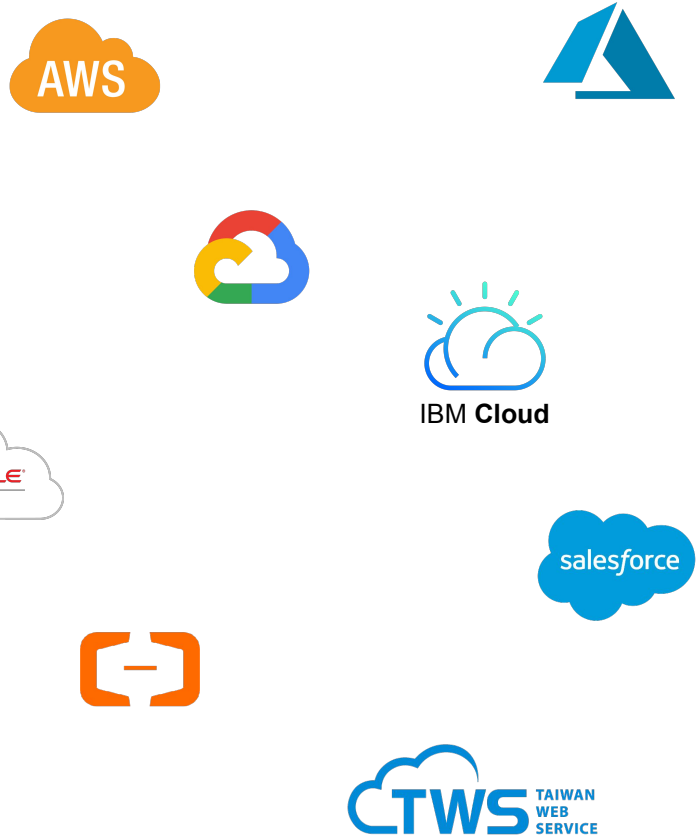
歲月靜好,
是有人替你負重前行

- 雲端原廠與雲端代理商
- 風光產業下的酸「?」苦辣
- 地盤怎麼分
- 見見大哥與經營理念

雲端原廠與雲端代理商

雲端原廠 Cloud Provider

- Amazon Web Service
- Microsoft Azure
- Alibaba
- Google Cloud Platform
- ...



以 IaaS 為基礎

提供 PaaS、SaaS 等服務

雲端原廠與雲端代理商

雲端原廠 Cloud Provider

- Amazon Web Service
- Microsoft Azure
- Alibaba
- Google Cloud Platform
- ...

以 IaaS 為基礎

提供 PaaS、SaaS 等服務

雲端代理商的功能

- 幫忙推廣產品
- 協助客戶打造客製化的雲端架構
- 「研發、生產、製造」以外的項目

MICROFUSION 宏庭科技

還是可以開發自有產品

享受最新的技術、解決問題

貼近客戶真實場景，應用落地

風光產業下的酸「?」苦辣

- 「甜」咧?
- 夾在中間

地盤怎麼分

AWS

- 2006 年推出第一個服務 (S3)
- 主打中小企業，提供解決方案

GCP

- 2008 年推出第一個服務(App Engine)
- 技術與開發能力著稱

Table 1. Worldwide IaaS Public Cloud Services Market Share, 2019-2020 (Millions of U.S. Dollars)

Company	2020 Revenue	2020 Market Share (%)	2019 Revenue	2019 Market Share (%)	2019-2020 Growth (%)
Amazon	26,201	40.8	20,365	44.6	28.7
Microsoft	12,658	19.7	7,950	17.4	59.2
Alibaba	6,117	9.5	4,004	8.8	52.8
Google	3,932	6.1	2,367	5.2	66.1
Huawei	2,672	4.2	882	1.9	202.8
Others	12,706	19.8	10,115	22.1	25.6
Total	**64,286**	**100.0**	**45,684**	**100.0**	**40.7**

Source: Gartner (June 2021)

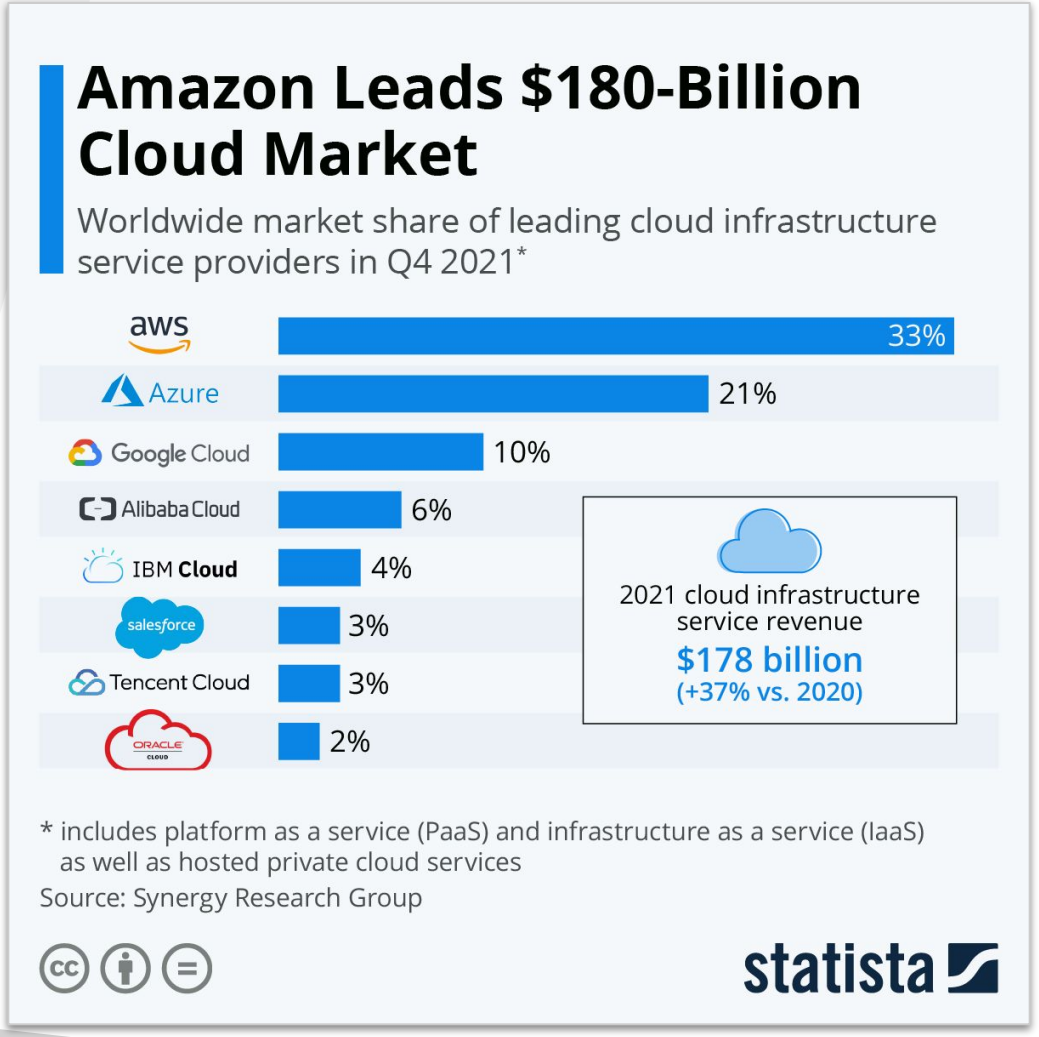
地盤怎麼分

AWS

- 2006 年推出第一個服務 (S3)
- 主打中小企業，提供解決方案

GCP

- 2008 年推出第一個服務(App Engine)
- 技術與開發能力著稱



地盤怎麼分

《讓子彈飛 Let the Bullets Fly》— 2010

地盤怎麼分

《讓子彈飛 Let the Bullets Fly》— 2010

地盤怎麼分

《讓子彈飛 Let the Bullets Fly》— 2010

地盤怎麼分

《讓子彈飛 Let the Bullets Fly》— 2010

地盤怎麼分

Google 怎麼做

- 和系統整合商合作
  (打不贏就買、欣賞就收購)
- 人工智慧是賣點
- 技術開源持續領先
  (K8s, istio, Tensorflow)
- 著眼混合雲(Hybrid Cloud)的解決方案(Anthos)

Table 1. Worldwide IaaS Public Cloud Services Market Share, 2019-2020 (Millions of U.S. Dollars)

Company	2020 Revenue	2020 Market Share (%)	2019 Revenue	2019 Market Share (%)	2019-2020 Growth (%)
Amazon	26,201	40.8	20,365	44.6	28.7
Microsoft	12,658	19.7	7,950	17.4	59.2
Alibaba	6,117	9.5	4,004	8.8	52.8
Google	3,932	6.1	2,367	5.2	66.1
Huawei	2,672	4.2	882	1.9	202.8
Others	12,706	19.8	10,115	22.1	25.6
Total	**64,286**	**100.0**	**45,684**	**100.0**	**40.7**

Source: Gartner (June 2021)

見見大哥與經營理念

以ML為切入

AWS

- 以需求出發
- 鎖定特定領域提供 Solution，解決 80% 的問題。
- 開發平台 SageMaker[1]，偏向開發者；但近期有推出 Canvas[2]

產業

ML Solutions Lab 已經成功為世界各地各行各業的客戶提供協助，包括生產製造、醫療保健和生命科學、金融服務、體育、公共部門和汽車，以建立採用機器學習技術的全新解決方案。

生產製造

ML Solutions Lab 專精於異常偵測、預測等方面的專業知識，可以協助製造公司改善其核心生產、維護、安全和品質，以及研發和供應鏈功能。例如，ML Solutions Lab 協助 Formosa Plastics 運用機器學習更準確地偵測缺陷，並將員工花費在手動檢查上的時間減少了一半。

醫療保健與生命科學

ML Solutions Lab 已使用 ML 協助醫療保健與生命科學領域的客戶降低成本並改善患者照護。例如，ML Solutions Lab 與 Cerner 合作建立解決方案，以便研究人員分析匿名患者資料，從而開發能在臨床表現之前最長 15 個月預測出充血性心力衰竭的演算法。

金融服務

ML Solutions Lab 與銀行、投資組織、保險公司和抵押公司等金融機構合作，以改善預測，讓監視系統能夠標記新的或正在出現的威脅，針對金融產品產生個人化建議，自動化文件處理，以及改善客戶使用聊天機器人和對話界面的體驗。

運動

ML Solutions Lab 擁有與 Formula 1、National Football League 和 Seattle Seahawks 等體育組織合作的豐富經驗，從而改善球迷的體驗並提高比賽品質。在 AWS 與 NFL 的合作中，ML Solutions Lab 確定並建立了一種全新的方式，讓球迷可以透過 Next Stats 參與運動賽事，並且目前正在運用 ML 來支援 NFL 的球員健康和安全計劃。

環境與社會影響

ML Solutions Lab 密切配合解決了一些全球最大的挑戰，從人口販運到飢荒，再到使用機器學習更充分地了解我們的世界。例如，ML Solutions Lab 與 Maxar 合作，透過機器學習更有效地分析衛星資料，讓組織能夠更快地採取行動以應對野火等災難，並向非洲村莊提供疫苗。

汽車

在汽車產業，ML Solutions Lab 與客戶在各領域密切合作，包括供應鏈最佳化、車載娛樂體驗和自動駕駛。這包括透過準確的道路情境感知和進階分析，來提高駕駛員和行人的安全。

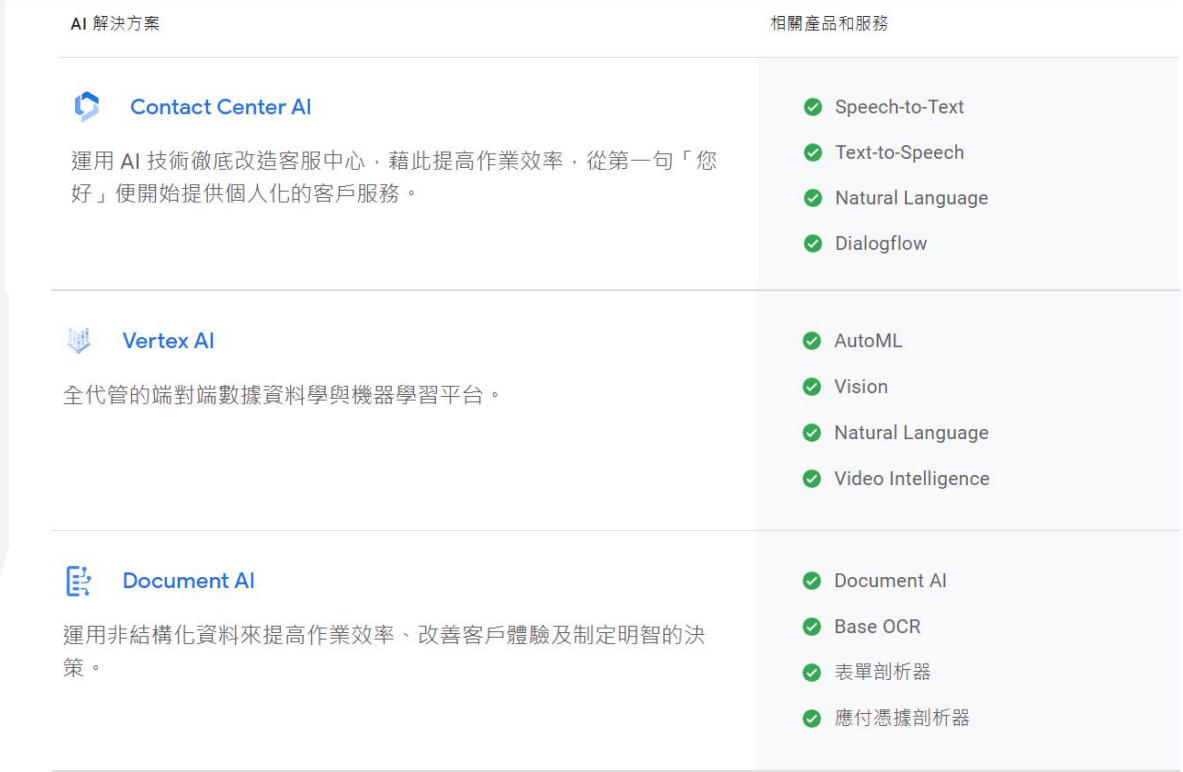
[1] Amazon SageMaker Studio – 第一個用於機器學習的DE
[2] Announcing Amazon SageMaker Canvas – a Visual, No Code Machine Learning Capability for Business Analysts | AWS News Blog

見見大哥與經營理念

以ML為切入

Google

- 以技術出發
- 完整 Solution 鎖定的行業為**客服中心**，以及以**文件處理為主的需求**
- Vertex AI 是一個完整開發平台，適用於各種角色開發、使用。



AI 解決方案	相關產品和服務
Contact Center AI 運用 AI 技術徹底改造客服中心，藉此提高作業效率，從第一句「您好」便開始提供個人化的客戶服務。	Speech-to-Text Text-to-Speech Natural Language Dialogflow
Vertex AI 全代管的端對端數據資料學與機器學習平台。	AutoML Vision Natural Language Video Intelligence
Document AI 運用非結構化資料來提高作業效率、改善客戶體驗及制定明智的決策。	Document AI Base OCR 表單剖析器 應付憑據剖析器

以GCP Solution 為例

我阿罵不只打籃球比你強,
連訓練的模型都比你準

人工智慧與機器學習的發展與普及

02

Google Cloud AI/ML Strategy

Tools & Solutions

Solution	senario or tools
ML API	for tech (e.g. develop)
AutoML Services	build GUI for ML API. (e.g. document AI, NLP, image, Vision, Table, etc.)
AI Platform	integrate AutoML Service. (ready to deprecate[1])
Vertex AI	integrate AutoML[2], ML Engine, Kubeflow, and TensorFlow Extended (TFX)

PaaS for tech and non-tech

Tools & Solutions

Solution	senario or tools
Contact Center AI (CCAI)	integrate Dialogflow, STT, TTS, NLP
Retail AI	integrate Vision Product Search, Recommendations AI, Retail Search

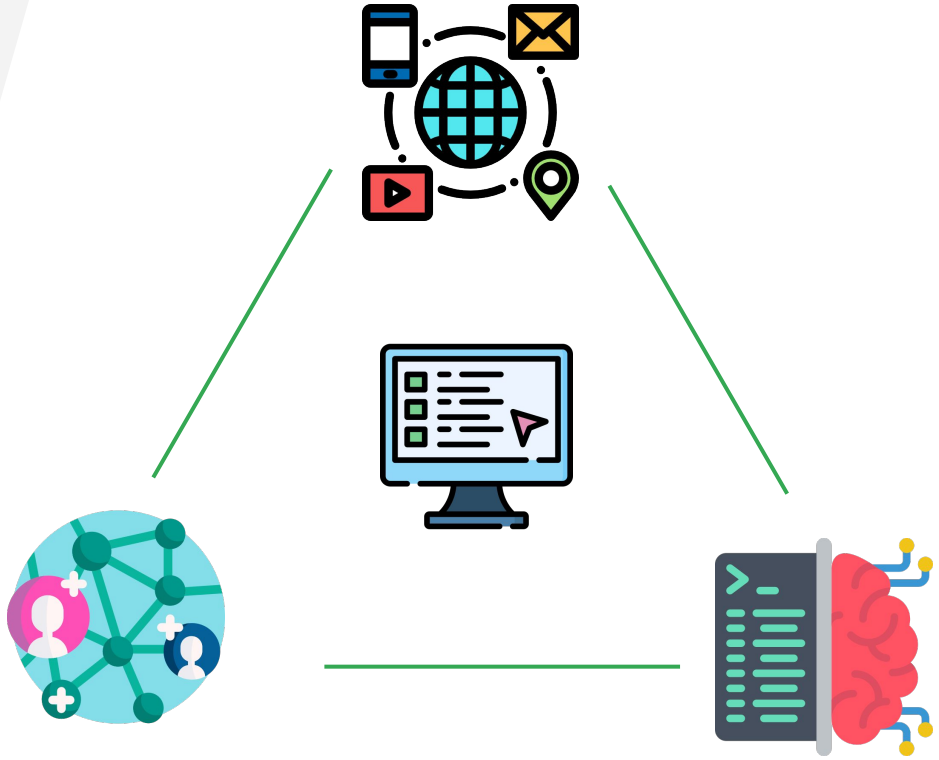
[1] Google Cloud 沒有明說，但是許多的文件都會導到Vertex AI 了。
[2] Vertex AI 的 AutoML 和 AutoML 其實是不同團隊在維護，但是東西是差不多的。

雲端上的機器學習

- 普及:一般的機器 + 用戶資料 + 便利的網路 + Google AutoML

- 優點: 突破開發限制
  - 訓練資源的極限
  - 預算有限

- 缺點: 不是萬能的
  - AutoML 應用範圍有限

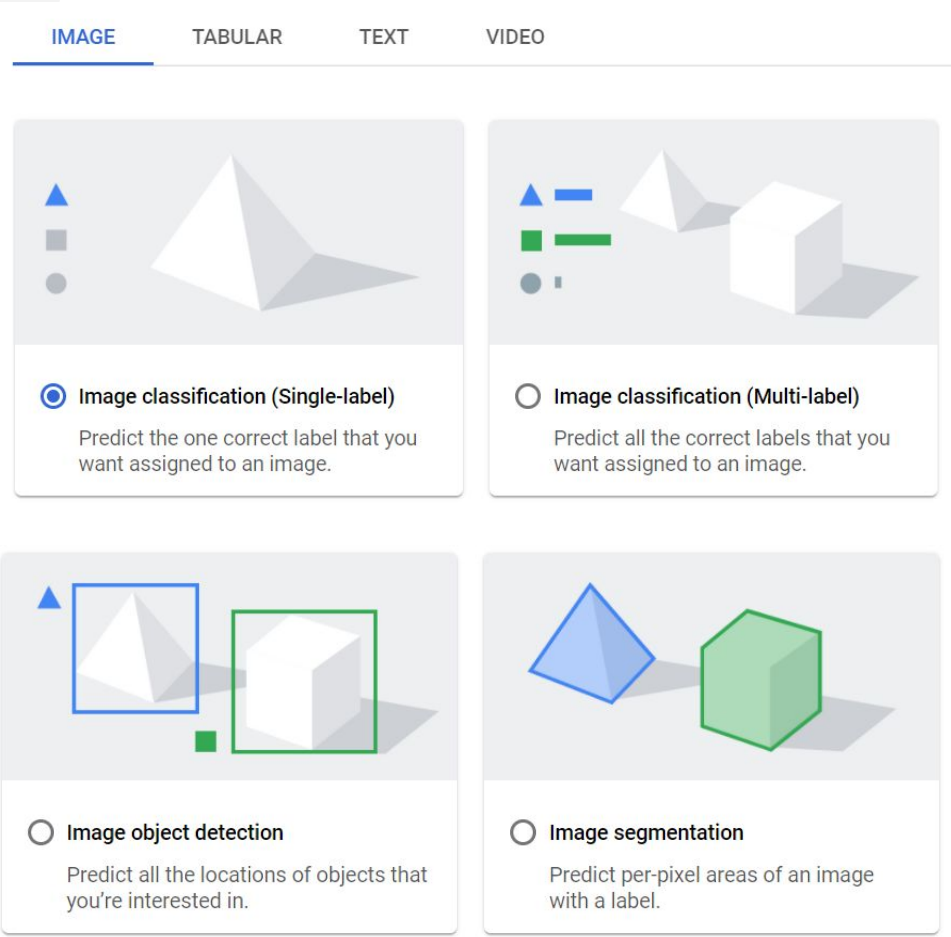


雲端上的機器學習

圖像分類(單標籤、多標籤)

圖像物件偵測 **圖像分割**

- **使用情境案例**
  - **口罩辨識**
  - **違規行為偵測**
  - **火災辨識**
  - **器材預警維修**

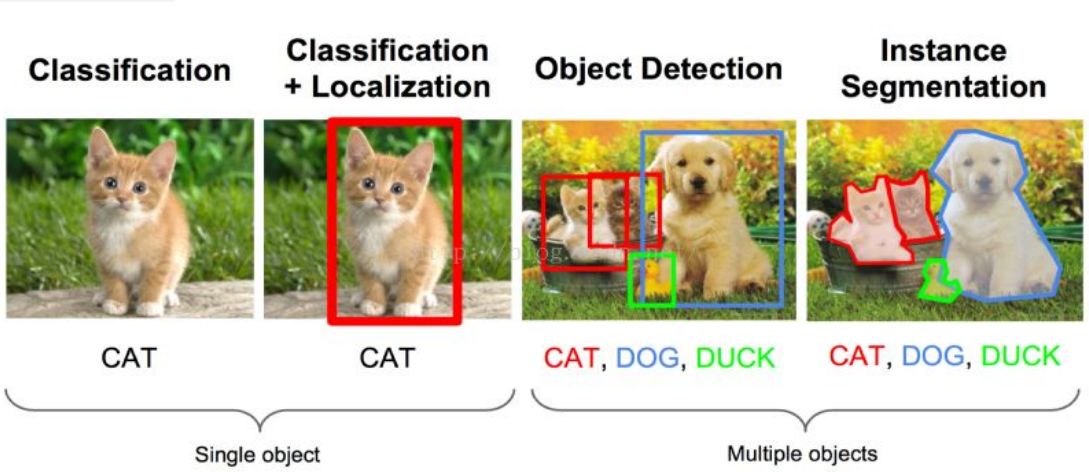


雲端上的機器學習

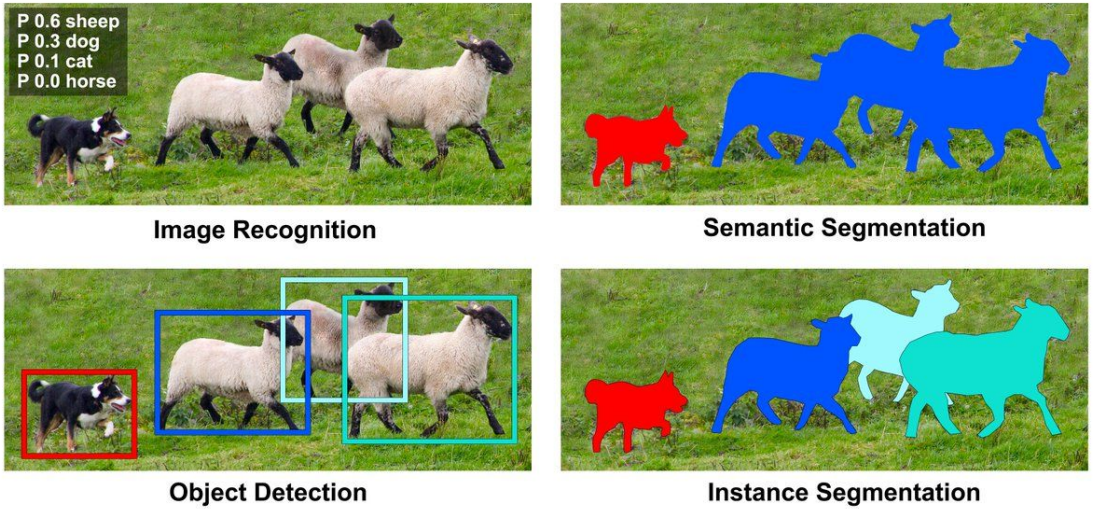
圖像分類(單標籤、多標籤)

圖像物件偵測

圖像分割



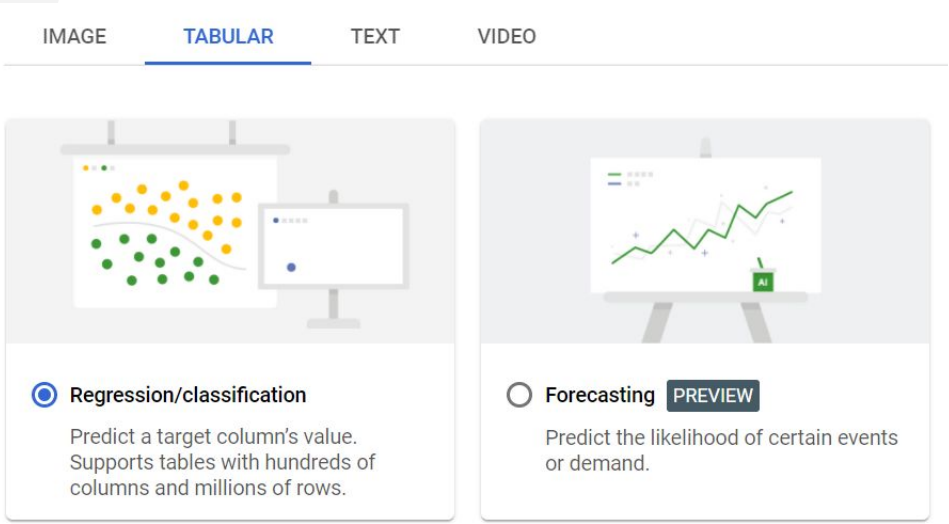
- 使用情境案例
  - 口罩辨識
  - 違規行為偵測
  - 火災辨識
  - 器材預警維修



雲端上的機器學習

回歸 / 分類　預測

- 使用情境案例
  - 庫存預測
  - 來客數量預測
  - 客戶流失
  - 垃圾郵件偵測

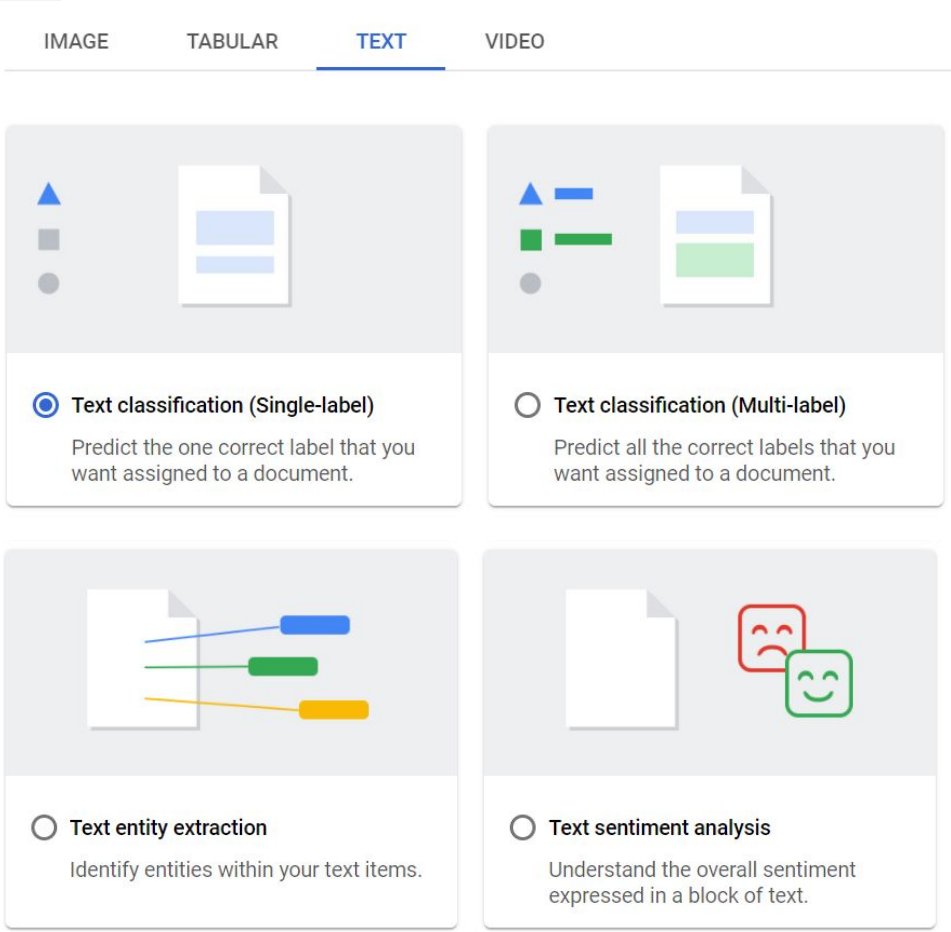


雲端上的機器學習

文章分類(單標籤、多標籤)

文字實體識別 **文字情緒分析**

- **使用情境案例**
  - **專有名詞辨識**
  - **文章分類**
  - **客服對話分析**

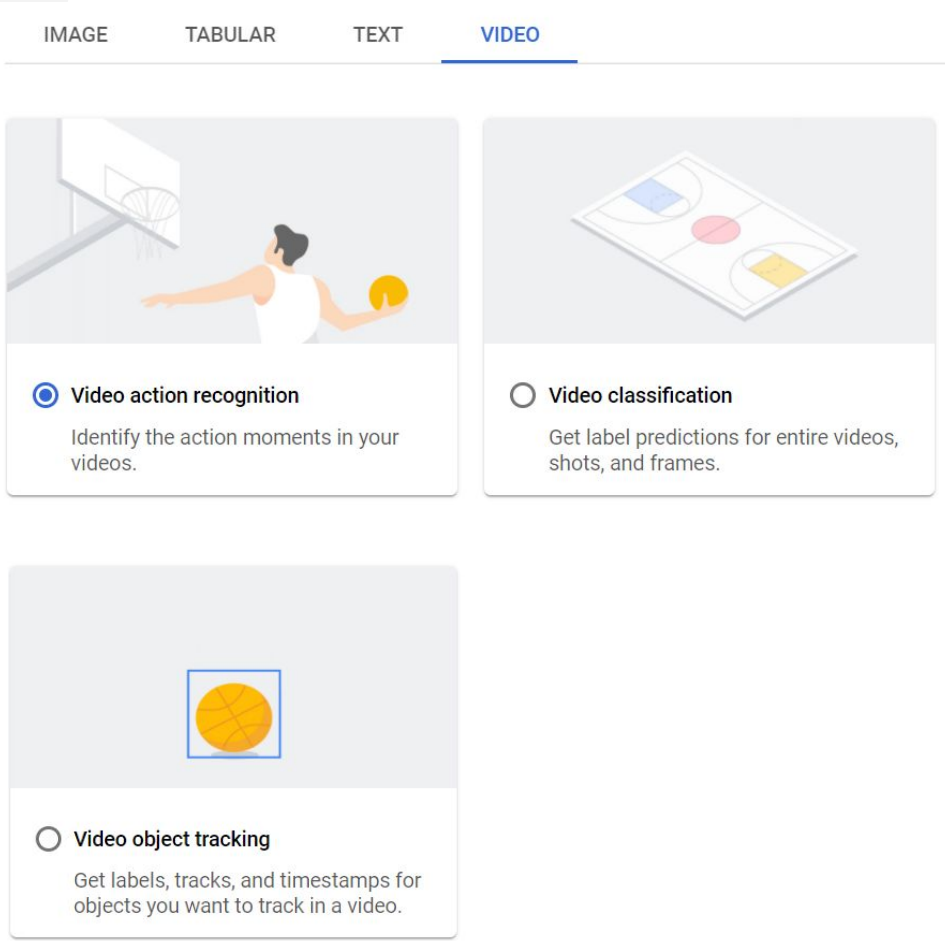


雲端上的機器學習

影像動作分類　影像分類

影像物件追蹤

- 使用情境案例
  - 運動判分輔助
  - 足球鏡頭追蹤
  - 駕駛打瞌睡判斷



雲端上的機器學習

影像動作分類　影像分類

影像物件追蹤

- 使用情境案例
  - 運動判分輔助
  - 足球鏡頭追蹤
  - 駕駛打瞌睡判斷

雲端上的機器學習

限定使用情境 + 協助完善服務脈絡

- Quotas and limits | Vertex AI | Google Cloud
- Quotas and limits | AutoML Tables
- Quotas & Limits | AutoML Vision
- Quotas & Limits | AutoML Natural Language
- Quotas and limits | AutoML Video Intelligence | Google Cloud
- Quotas & Limits | AutoML Translation Documentation | Google Cloud

打造一個
人工智慧的王國

萬事俱備，只欠東風

- ML 如何平易近人
- How About Data
- BQML
- Last Pieces?

ML 如何平易近人

Your Data + Your Models


App Engine


Cloud TPU


Cloud Dataproc


Kubernetes Engine


Compute Engine

Your Data + Google Models


AutoML Video Intelligence


Cloud AutoML


AutoML Natural Language


AutoML Translation


AutoML Tables


AutoML Vision

Google Data + Google Models


Cloud Natural Language API


Cloud Text-to-Speech


Dialogflow


Recommendations AI


Cloud Inference API


Cloud Vision API

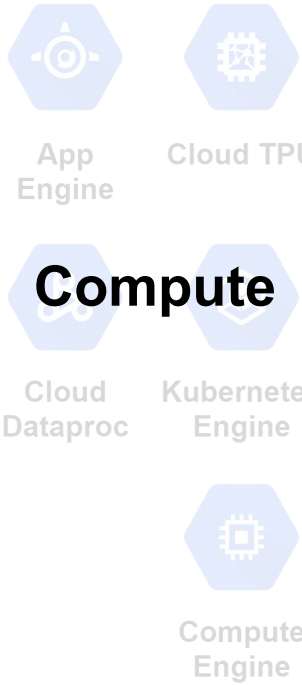

Cloud Translation API


Cloud Video Intelligence API


Cloud Speech-to-Text

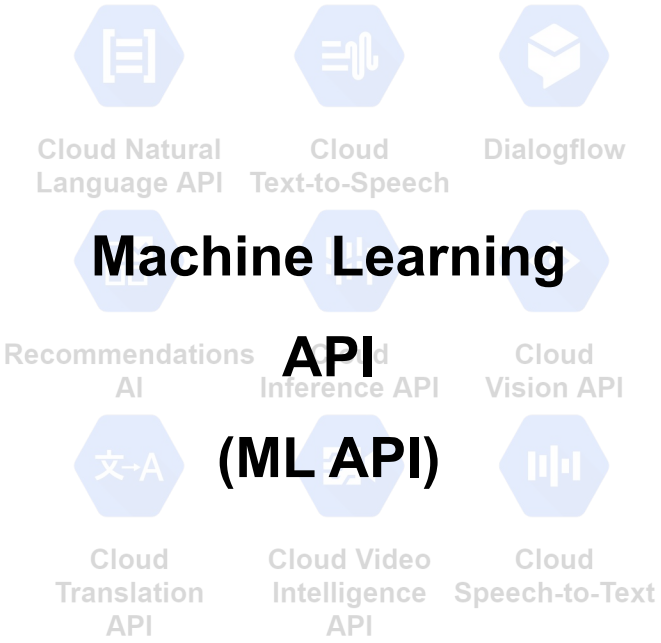
ML 如何平易近人

Your Data + Your Models



Your Data + Google Models

Google Data + Google Models

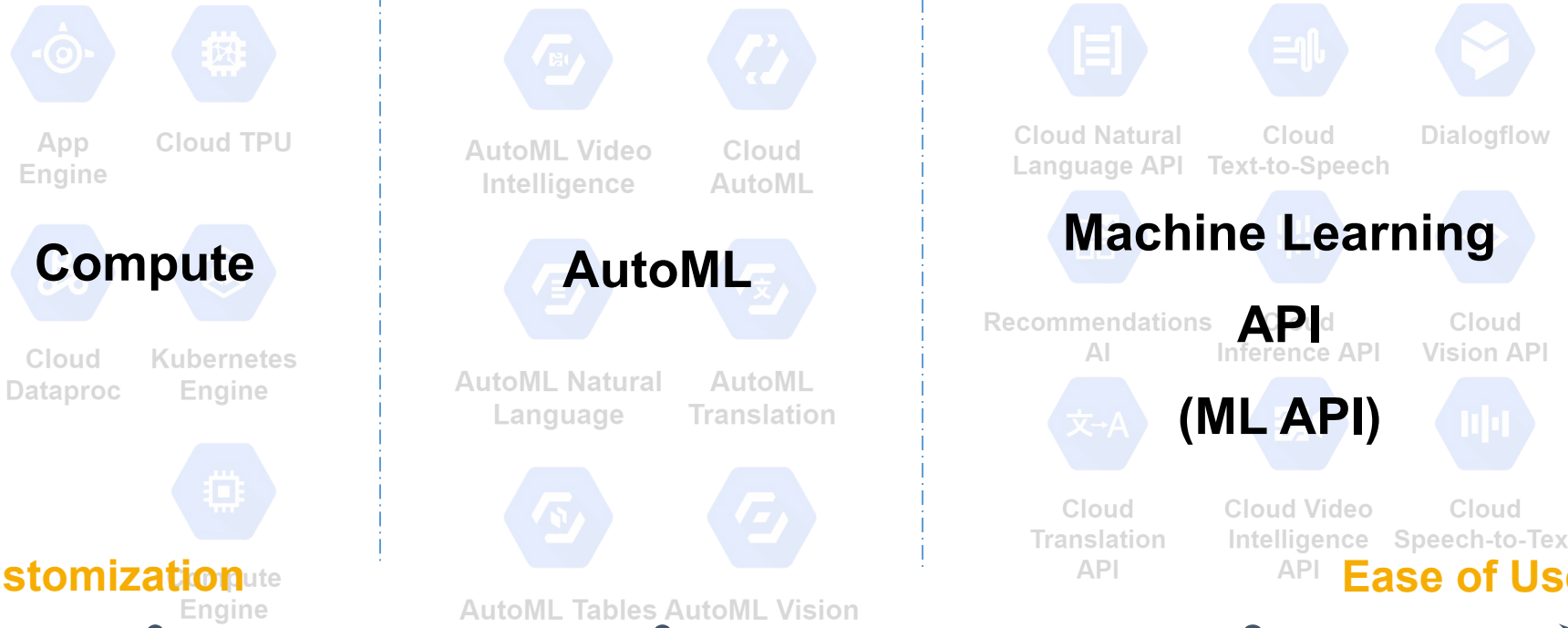


ML 如何平易近人

Your Data + **Your** Models

Your Data + **Google** Models

Google Data + **Google** Models



Customization — **Ease of Use**

Build **your** own models

Train by **Google's** state-of-the-art models

Call **Google's** perception APIs

ML 如何平易近人

Your Data + Your Models

Your Data + Google Models

Google Data + Google Models

a unified AI platform

Vertex AI

Customization

Ease of Use

Build your own models

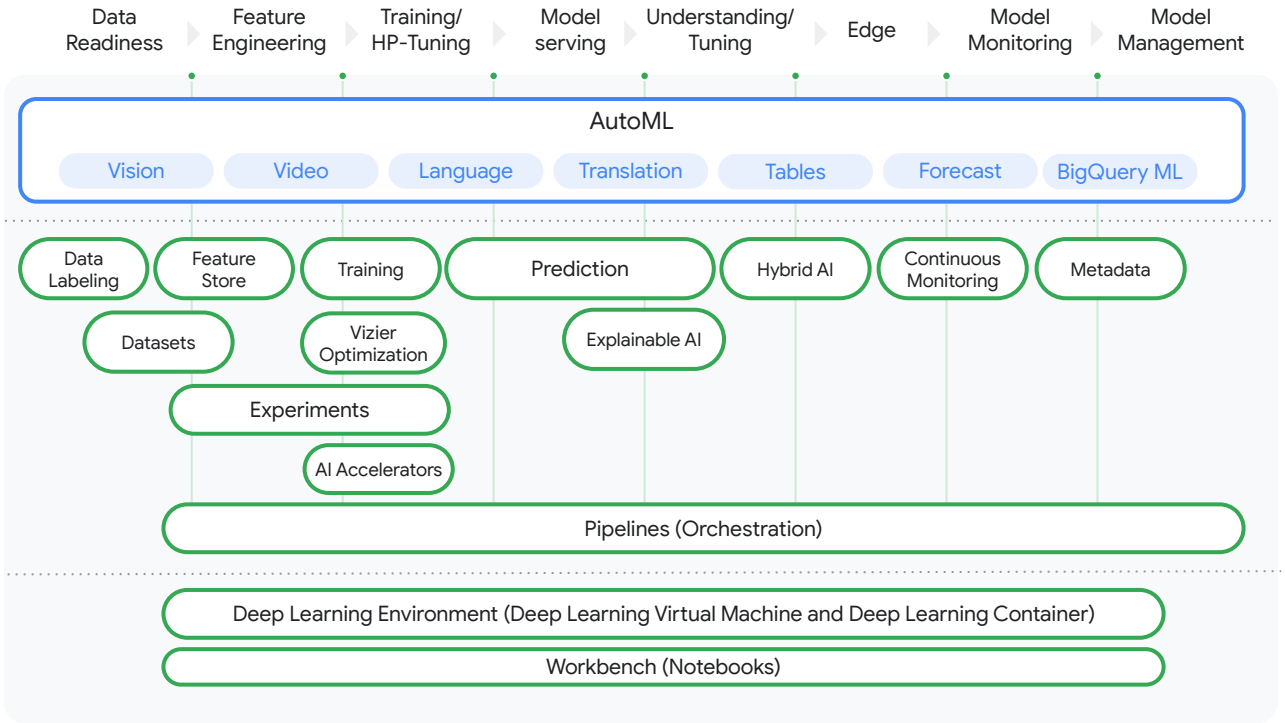
Train by Google's state-of-the-art models

Call Google's perception APIs

ML 如何平易近人

One comprehensive end-to-end platform for everything AI

Vertex AI



no-code/ low code workflow

Custom development workflow

Infrastructure services/ Add-ons

How About Data

Google Cloud is significantly simplifying big data analytics

Deliver **serverless** analytics, not infrastructure

Empower analytics across the **entire data lifecycle**

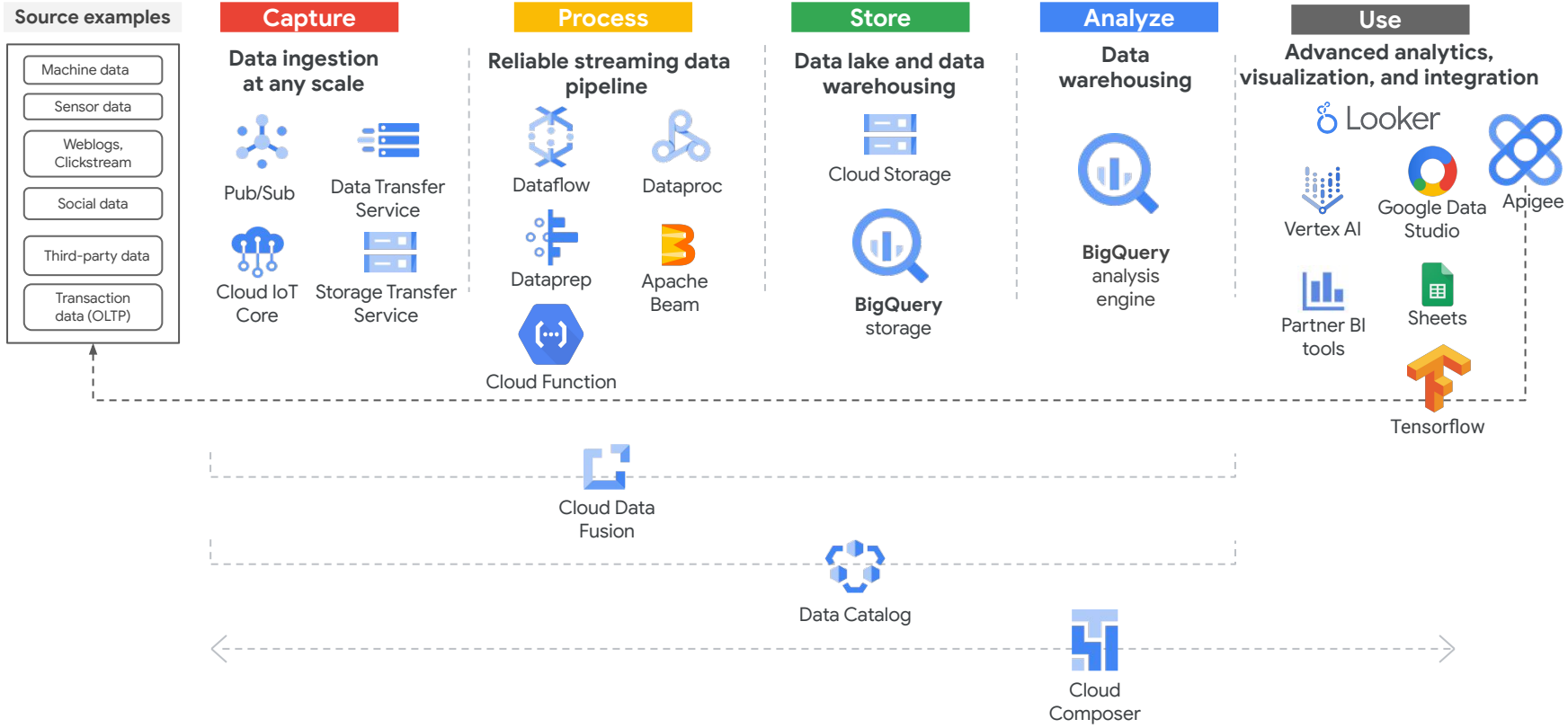
Embed ML and drive an end-to-end lifecycle

Enable the **best OSS** technologies

Build for **enterprise** at any scale

How About Data

Part of our comprehensive data analytics portfolio



BQML

BigQuery ML for predictive analytics

1. **Execute** ML initiatives without moving data from BigQuery
2. **Iterate** on models in SQL in BigQuery to increase development speed
3. **Automate** common ML tasks, and hyperparameter tuning

> BigQuery is NOT just analytics

BQML

supported models and features

Classification

- Logistic regression
- DNN classifier (Tensorflow)
- XGBoost
- AutoML Tables
- Wide and Deep NNs[Preview, GA 2022]

Other Models

- k-means clustering
- Time series forcasting
- Recommendation: Matrix factorization
- Time series anomaly detection

Regression

- Linear regression
- DNN regressor
- XGBoost
- AutoML Tables
- Wide and Deep NNs[Preview, GA 2022]

Model Import/Export

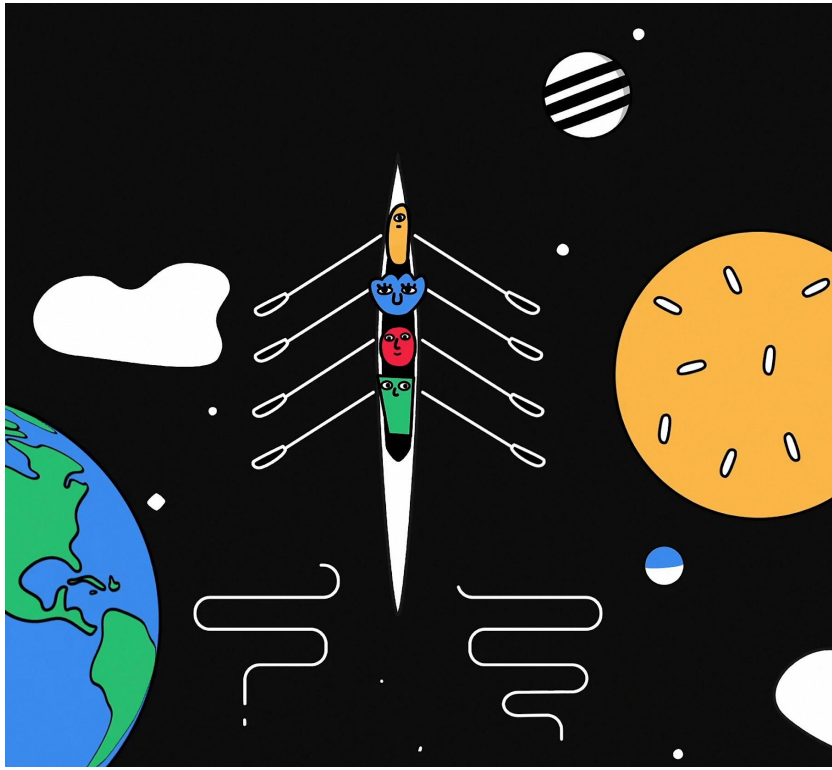
- Importing TensorFlow and XGBoost models for batch prediction
- Exporting models from BigQuery ML for online prediction
- Hyperameter tuning using Cloud AI Vizier

BQML

Behind the scenes

Through two lines of SQL

- Leverage BigQuery's processing power to build a model
- Auto-tuned learning rate
- Auto-split of data into training and test
- Null imputation
- Standardization of numeric features
- One-hot encoding of strings
- Class imbalance handling



Last Pieces → Your Data | Your Idea | What position do you stand for

Making ML accessible for all audiences

Train with	Data Analyst	ML Developer	Data Scientist	Use when	Serve with
Vertex Training & TensorFlow			✓	• Your problem doesn't match the criteria listed below for BigQuery ML or AutoML. • You're already running training on-premises or another cloud, and you need consistency across the platforms.	**Vertex Prediction**
AutoML & CloudML APIs		✓	✓	• Your problem fits into one of the types AutoML supports. Offers a point-and-click workflow. • Natural Language or Video models are served from Google Cloud. While Vision and Tables support edge / downloadable models.	
BigQuery ML	✓	✓	✓	• All your data is contained in BigQuery. • Users are most comfortable with SQL. • The set of models available in BigQuery ML matches the problem you're trying to solve.	

Demo:
物件辨識模型

04

Vertex AI

Dashboard
Datasets
Features
Labeling tasks
Notebooks
Pipelines
Training
Experiments
Models
Endpoints
Batch predictions
Metadata

Marketplace

Create dataset

Dataset name *
untitled_1632885993746
Can use up to 128 characters.

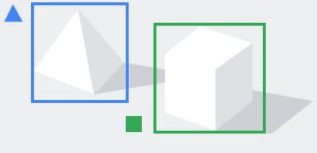
Select a data type and objective

First select the type of data your dataset will contain. Then select an objective, which is the outcome that you want to achieve with the trained model. Learn more about model types

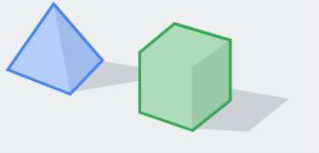
IMAGE | TABULAR | TEXT | VIDEO

- ○ **Image classification (Single-label)**
  Predict the one correct label that you want assigned to an image.

- ○ **Image classification (Multi-label)**
  Predict all the correct labels that you want assigned to an image.



- ◉ **Image object detection**
  Predict all the locations of objects that you're interested in.



- ○ **Image segmentation**
  Predict per-pixel areas of an image with a label.

Region
us-central1 (Iowa)

ADVANCED OPTIONS

You can use this dataset for other image-based objectives later by creating an annotation set. Learn more about annotation sets

CREATE CANCEL

Vertex AI

Dashboard
Datasets
Features
Labeling tasks
Notebooks
Pipelines
Training
Experiments
Models
Endpoints
Batch predictions
Metadata

Marketplace

untitled_1632885993746

untitled_1632885993746_...

CREATE LABELING TASK

IMPORT BROWSE ANALYZE

Add images to your dataset

Before you begin, read the data guide to learn how to prepare your data. Then choose an import method.

Select an import method

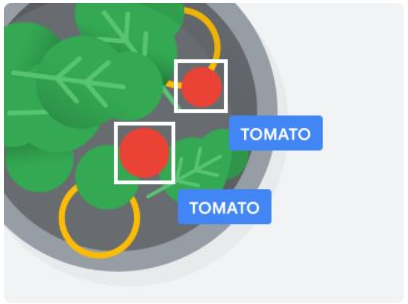
- **Upload images**: Recommended if you don't have labels yet
- **Import files**: Recommended if you already have labels. An import file is a list of Cloud Storage URIs to your images and optional data, like labels. Learn how to create an import file

- (•) Upload images from your computer
- ( ) Upload import files from your computer
- ( ) Select import files from Cloud Storage

Upload images from your computer

Add up to 500 images per upload. Images will be preprocessed and stored in Cloud Storage.

SELECT FILES



Object detection models draw bounding boxes around items of interest. For example, identifying vegetables from images of food.

Instead of creating a custom model, try Google's **Vision API** to detect generic objects, faces, and text. Learn more

訓練資料 - 上傳方法與資料限制

參考資料: [Formatting a training data CSV | AutoML Vision Object Detection | Google Cloud]

CSV formatting guidelines

To use the `importData` method, both the CSV file *and* the images it points to must be in a Google Cloud Storage bucket.

Additionally, the CSV file must also fulfill the following requirements:

- The file can have any filename, but must be in the same bucket as your image file.
- Must be UTF-8 encoded.
- Must end with a `.csv` extension.
- Has one row **for each bounding box** in the set you are uploading, ***or*** one row **for each image with no bounding box** (such as row 4 below).
- Contain **one image per line**; an image with multiple bounding boxes will be repeated on as many rows as there are bounding boxes.

  For example, rows 1 and 2 reference the same image that has 2 annotations (`car,0.1,0.1,,,0.3,0.3,,` and `bike,.7,.6,,,.8,.9,,`). Row 3 refers to an image that has only 1 annotation (`car,0.1,0.1,0.2,0.1,0.2,0.3,0.1,0.3`), while row 4 references an image with no annotations.

Four sample rows:

```
TRAIN,gs://folder/image1.png,car,0.1,0.1,,,0.3,0.3,,
TRAIN,gs://folder/image1.png,bike,.7,.6,,,.8,.9,,
UNASSIGNED,gs://folder/im2.png,car,0.1,0.1,0.2,0.1,0.2,0.3,0.1,0.3
TEST,gs://folder/im3.png,,,,,,,,,
```

標記方式與限制

參考資料: [Preparing your training data | AutoML Vision Object Detection | Google Cloud]

	Labels and bounding box requirements
Label instances for training	10 annotations (instances) minimum.
Annotation requirements	For each label you must have *at least* 10 images, each with *at least* one annotation (bounding box and the label). However, for model training purposes it's recommended you use about **1000 annotations per label**. In general, the more images per label you have the better your model will perform.
Label ratio (most common label to least common label):	The model works best when there are at most 100x more images for the most common label than for the least common label. For model performance, it is recommended that you remove very low frequency labels.
Bounding box edge length	At least **0.01 * length of a side** of an image. For example, a 1000 * 900 pixel image would require bounding boxes of at least 10 * 9 pixels.
Bounding box size (pixels)	8 pixels by 8 pixels minimum.
Bounding boxes per distinct image	500 maximum.
Bounding boxes returned from a prediction request	100 (default), 500 maximum. You can specify this value as part of the `predict` request in the `params.max_bounding_box_count` field.

Google Cloud Platform sandbox vertex

Vertex AI

- Dashboard
- Datasets
- Features
- Labeling tasks
- Notebooks
- Pipelines
- Training
- Experiments
- Models
- Endpoints
- Batch predictions
- Metadata

Marketplace

untitled_1632885993746

IMPORT BROWSE ANALYZE

All	682
Labeled	682
Unlabeled	0
Training	682
Validation	0
Test	0

Filter Filter labels +

Images

mask	654
no_mask	270

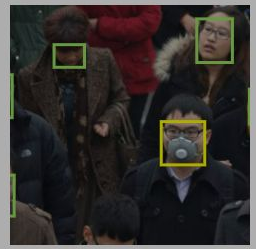
ADD NEW LABEL

Filter Filter items

Unable to import data due to e

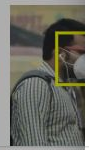
DETAILS DISMISS

Select all



no_mask (6), mask (1)

mask (1)



Items per page: 10 1 – 1

Error details

如果上傳資料有誤，都會一目了然。

Operation ID: projects/593039475026/locations/us-central1/operations/399735973493080064

Error Messages:

Error: Unable to preprocess image. for: gs://mask-ai-sean/train.csv line 2055,gs://mask-ai-sean/train.csv line 2056,gs://mask-ai-sean/train.csv line 2057

Error: Unable to preprocess image. for: gs://mask-ai-sean/train.csv line 2928,gs://mask-ai-sean/train.csv line 2929,gs://mask-ai-sean/train.csv line 2930,gs://mask-ai-sean/train.csv line 2931

Error: Unable to preprocess image. for: gs://mask-ai-sean/train.csv line 3220,gs://mask-ai-sean/train.csv line 3221,gs://mask-ai-sean/train.csv line 3222,gs://mask-ai-sean/train.csv line 3223,gs://mask-ai-sean/train.csv line 3224,gs://mask-ai-sean/train.csv line 3225

Error: Unable to preprocess image. for: gs://mask-ai-sean/train.csv line 3883,gs://mask-ai-sean/train.csv line 3884,gs://mask-ai-sean/train.csv line 3885,gs://mask-ai-sean/train.csv line 3886,gs://mask-ai-sean/train.csv line 3887

Error: Unable to preprocess image. for: gs://mask-ai-sean/train.csv line 1581

Error: Unable to preprocess image. for: gs://mask-ai-sean/train.csv line 1694,gs://mask-ai-sean/train.csv line 1695,gs://mask-ai-sean/train.csv line 1696,gs://mask-ai-sean/train.csv line 1697

Error: Unable to preprocess image. for: gs://mask-ai-sean/train.csv line 2048,gs://mask-ai-sean/train.csv line 2049,gs://mask-ai-sean/train.csv line 2050

CLOSE

模型訓練

模型的訓練主要有三種 AutoML、AutoML Edge、Custom training

- AutoML：Train by Google's state-of-the-art models
- AutoML Edge：Google 假設所訓練的模型是要部署到 edge device, 如攝影機、手機等邊緣裝置, 因此 **Google會限制模型的大小**。**訓練較快, 但是精準度較差**, 可能需要再做進一步的程式對圖像做篩選。
- Custom training：利用自己的程式碼, 做模型的訓練

Vertex AI

Dashboard
Datasets
Features
Labeling tasks
Notebooks
Pipelines
Training
Experiments
Models
Endpoints
Batch predictions
Metadata

Marketplace

untitled_1632885993746

untitled_1632885993746_...

TRAIN NEW MODEL CREATE LABELING TASK

IMPORT BROWSE ANALYZE

All	682
Labeled	682
Unlabeled	0
Training	682
Validation	0
Test	0

Filter Filter labels +

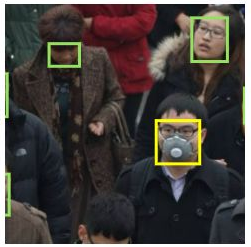
Images

mask	654
no_mask	270

ADD NEW LABEL

Filter Filter items

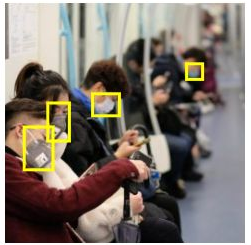
Select all

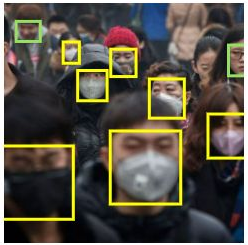

no_mask (6), mask (1)

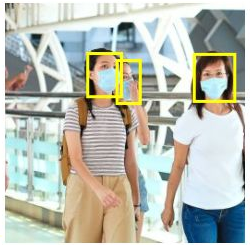
mask (1)


mask (1)


no_mask (3), mask (2)







Items per page: 10 1 – 10 of many

Vertex AI

Dashboard
Datasets
Features
Labeling tasks
Notebooks
Pipelines
Training
Experiments
Models
Endpoints
Batch predictions
Metadata

Marketplace

untitled_1632885993746

untitled_1632885993746_...

IMPORT | BROWSE | ANALYZE

All	682
Labeled	682
Unlabeled	0
Training	682
Validation	0
Test	0

Filter Filter labels +

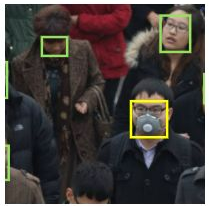
Images

mask	654
no_mask	270

ADD NEW LABEL

Filter Filter items

Select all

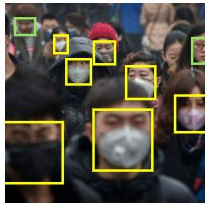

no_mask (6), mask (1)


mask (1)

mask (1)


no_mask (3), mask (2)


mask (4)


mask (10), no_mask (2)

Items per page: 10 1 – 10 of many

Training jobs and models

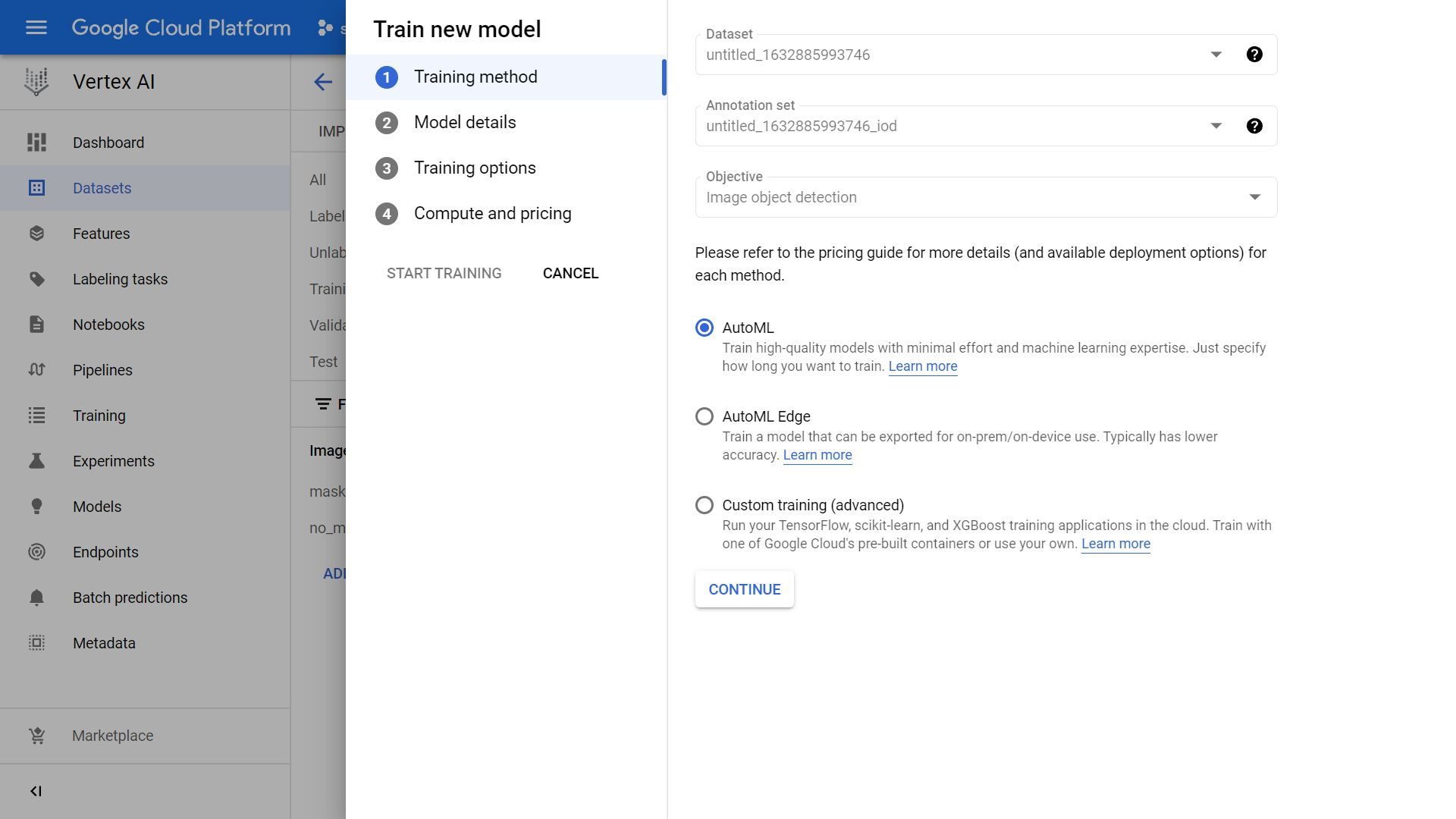
Use this dataset and annotation set to train a new machine learning model with AutoML or custom code

TRAIN NEW MODEL

Labeling tasks

If your data still needs to be labeled, create a labeling task to have others label it for you

CREATE LABELING TASK


Google Cloud Platform
Vertex AI
Dashboard
Datasets
Features
Labeling tasks
Notebooks
Pipelines
Training
Experiments
Models
Endpoints
Batch predictions
Metadata
Marketplace
Train new model
1 Training method
2 Model details
3 Training options
4 Compute and pricing
START TRAINING
CANCEL
Dataset
untitled_1632885993746
Annotation set
untitled_1632885993746_iod
Objective
Image object detection
Please refer to the pricing guide for more details (and available deployment options) for each method.
AutoML
Train high-quality models with minimal effort and machine learning expertise. Just specify how long you want to train. Learn more
AutoML Edge
Train a model that can be exported for on-prem/on-device use. Typically has lower accuracy. Learn more
Custom training (advanced)
Run your TensorFlow, scikit-learn, and XGBoost training applications in the cloud. Train with one of Google Cloud's pre-built containers or use your own. Learn more
CONTINUE

Google Cloud Platform

Vertex AI

- Dashboard
- Datasets
- Features
- Labeling tasks
- Notebooks
- Pipelines
- Training
- Experiments
- Models
- Endpoints
- Batch predictions
- Metadata

Marketplace

Train new model

1. Training method
2. Model details
3. Training options
4. Compute and pricing

START TRAINING CANCEL

Model name *
untitled_1632885993746_20219294146

Data split

(•) Randomly assigned ( ) Manual (Advanced)

Your dataset will be automatically randomized and split into training, validation, and test sets using the following ratios.

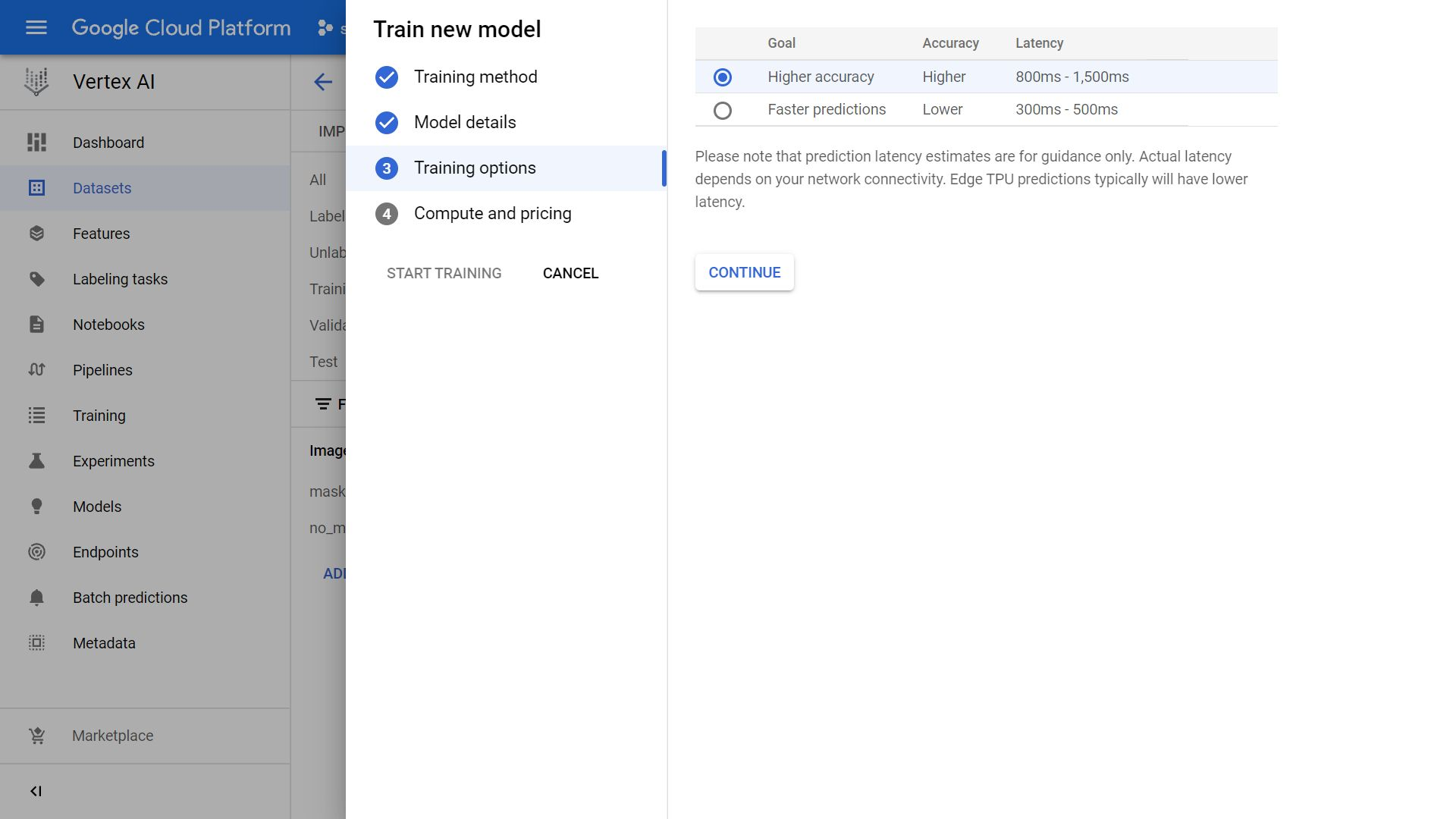
Training	Validation	Test
80 %	10 %	10 %

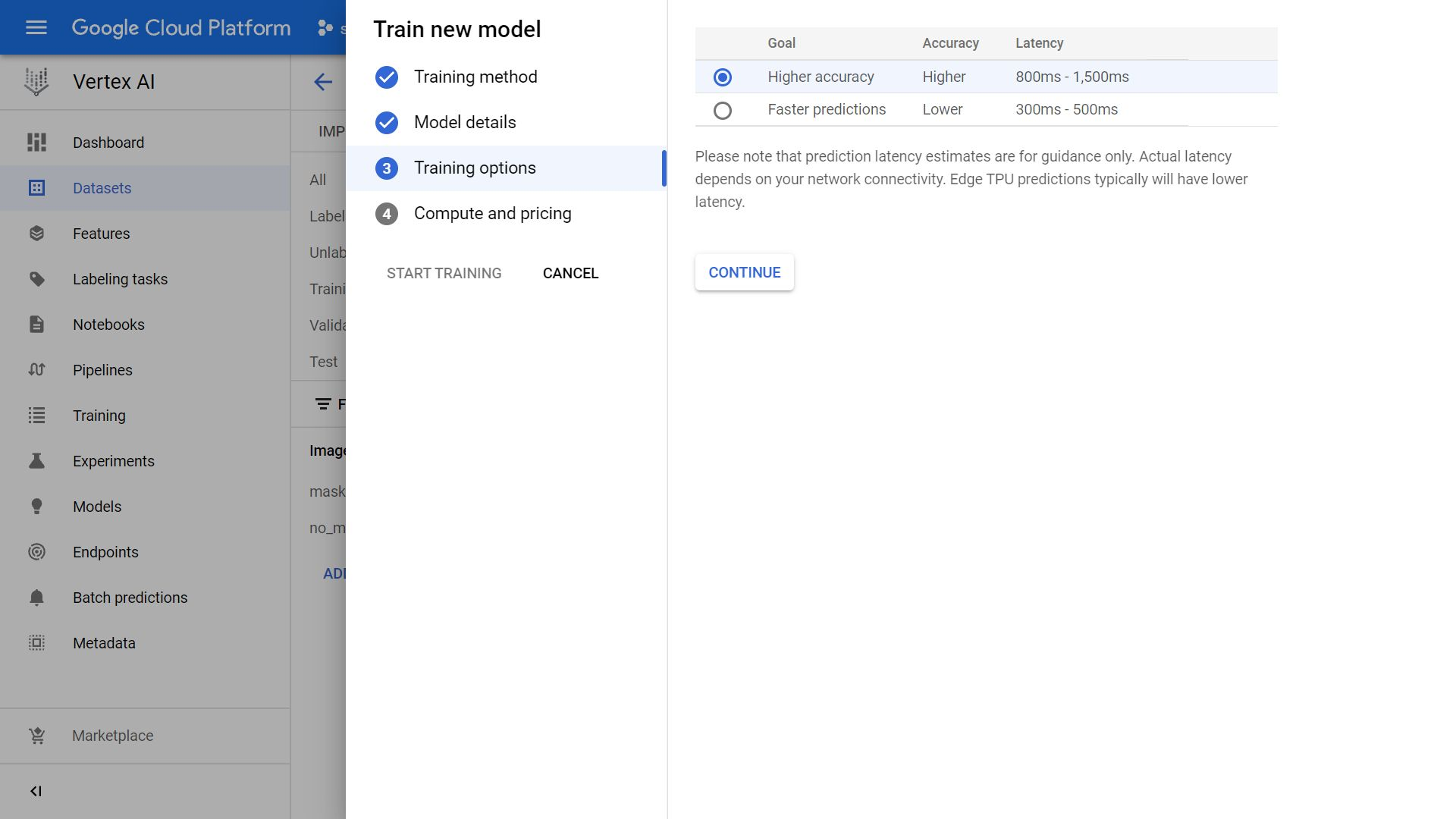
Encryption

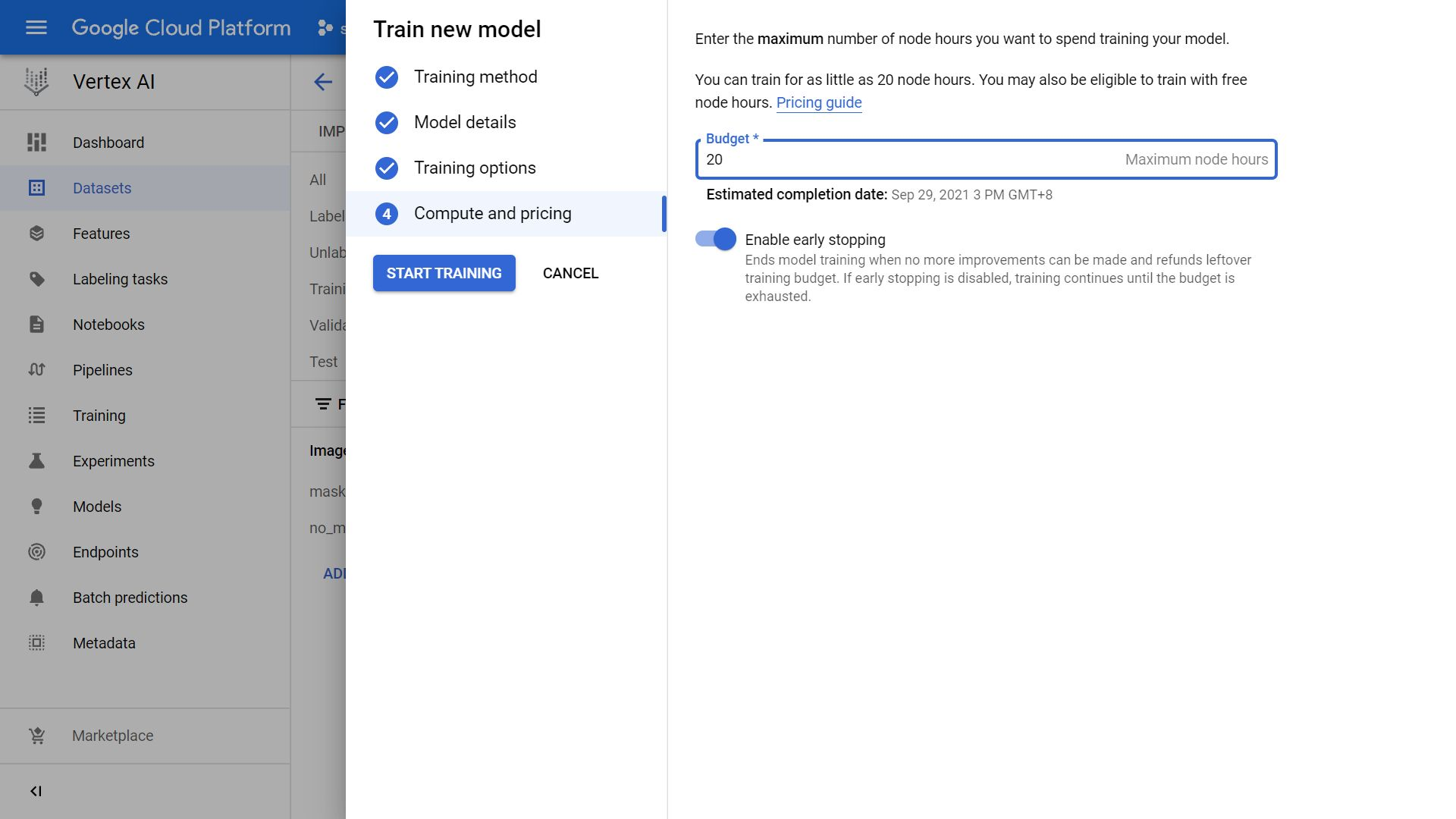
[ ] Use a customer-managed encryption key (CMEK)

SHOW LESS

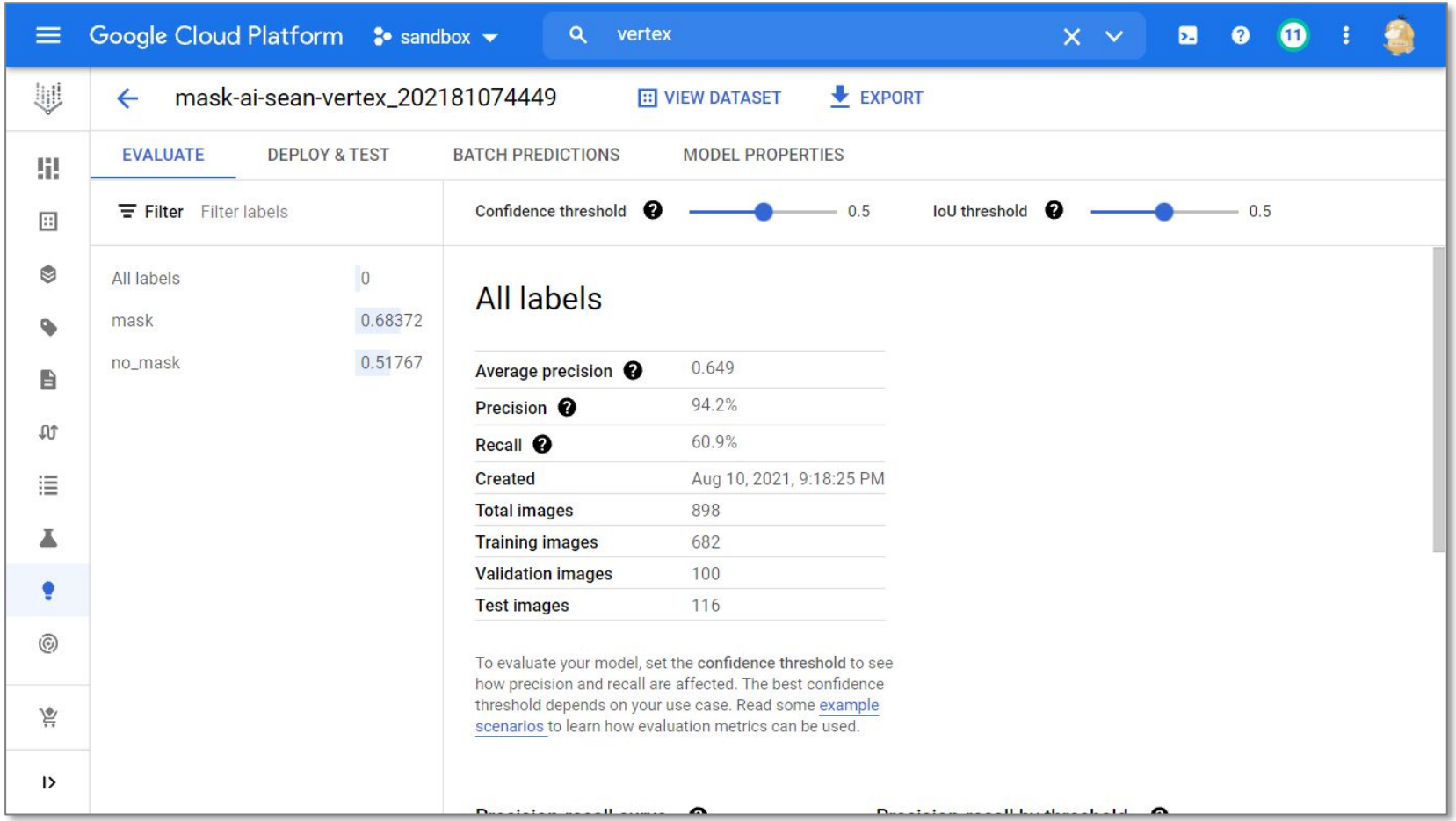
CONTINUE


Google Cloud Platform
Vertex AI
Dashboard
Datasets
Features
Labeling tasks
Notebooks
Pipelines
Training
Experiments
Models
Endpoints
Batch predictions
Metadata
Marketplace
Train new model
Training method
Model details
3
Training options
4
Compute and pricing
START TRAINING
CANCEL
Goal
Accuracy
Latency
Higher accuracy
Higher
800ms - 1,500ms
Faster predictions
Lower
300ms - 500ms
Please note that prediction latency estimates are for guidance only. Actual latency depends on your network connectivity. Edge TPU predictions typically will have lower latency.
CONTINUE

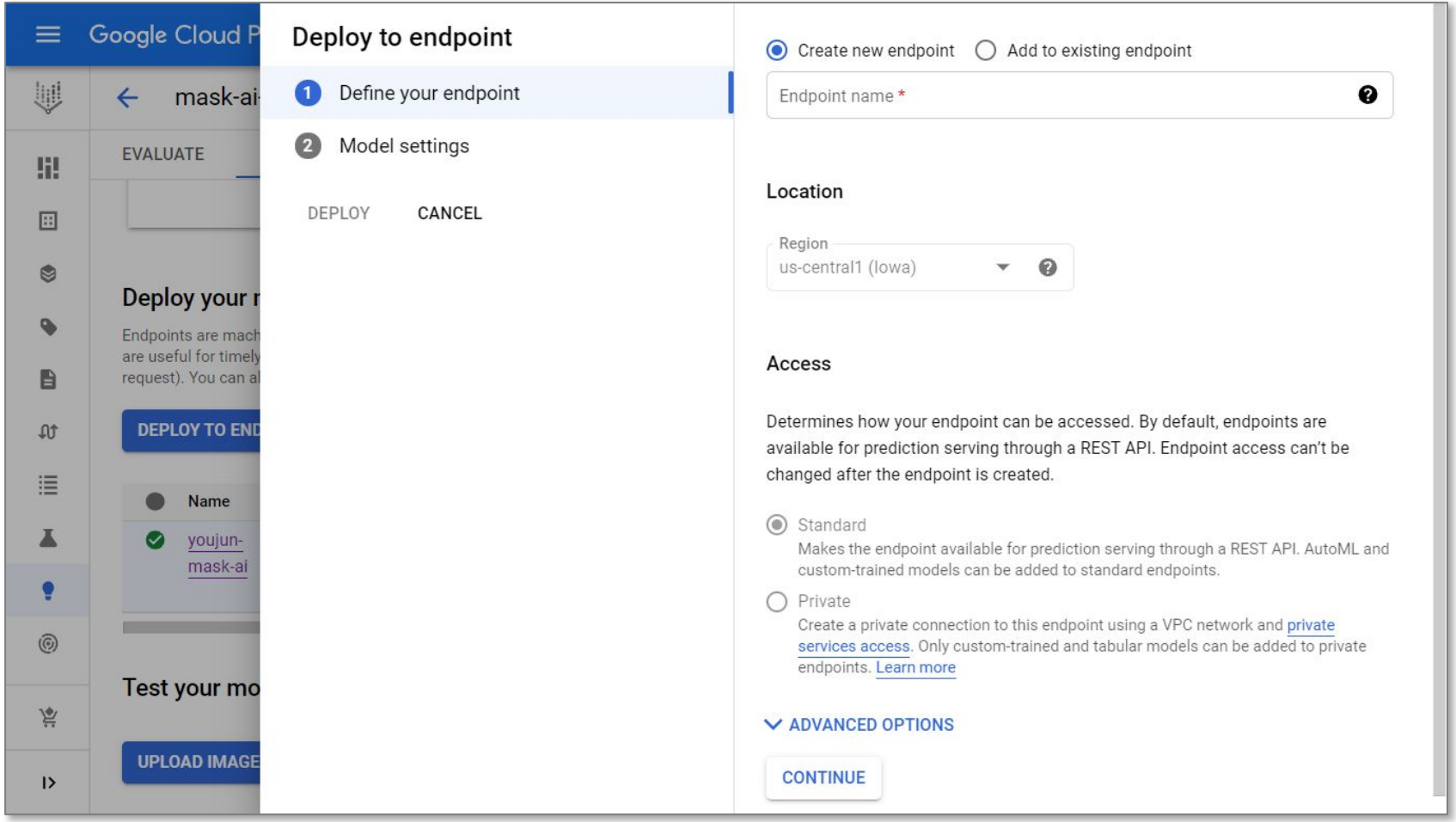

Google Cloud Platform
Vertex AI
Dashboard
Datasets
Features
Labeling tasks
Notebooks
Pipelines
Training
Experiments
Models
Endpoints
Batch predictions
Metadata
Marketplace
Train new model
Training method
Model details
3
Training options
4
Compute and pricing
START TRAINING
CANCEL
Goal
Accuracy
Latency
Higher accuracy
Higher
800ms - 1,500ms
Faster predictions
Lower
300ms - 500ms
Please note that prediction latency estimates are for guidance only. Actual latency depends on your network connectivity. Edge TPU predictions typically will have lower latency.
CONTINUE


Google Cloud Platform
Vertex AI
Dashboard
Datasets
Features
Labeling tasks
Notebooks
Pipelines
Training
Experiments
Models
Endpoints
Batch predictions
Metadata
Marketplace
Train new model
Training method
Model details
Training options
4
Compute and pricing
START TRAINING
CANCEL
Enter the maximum number of node hours you want to spend training your model.
You can train for as little as 20 node hours. You may also be eligible to train with free node hours. Pricing guide
Budget *
20
Maximum node hours
Estimated completion date: Sep 29, 2021 3 PM GMT+8
Enable early stopping
Ends model training when no more improvements can be made and refunds leftover training budget. If early stopping is disabled, training continues until the budget is exhausted.

模型評估

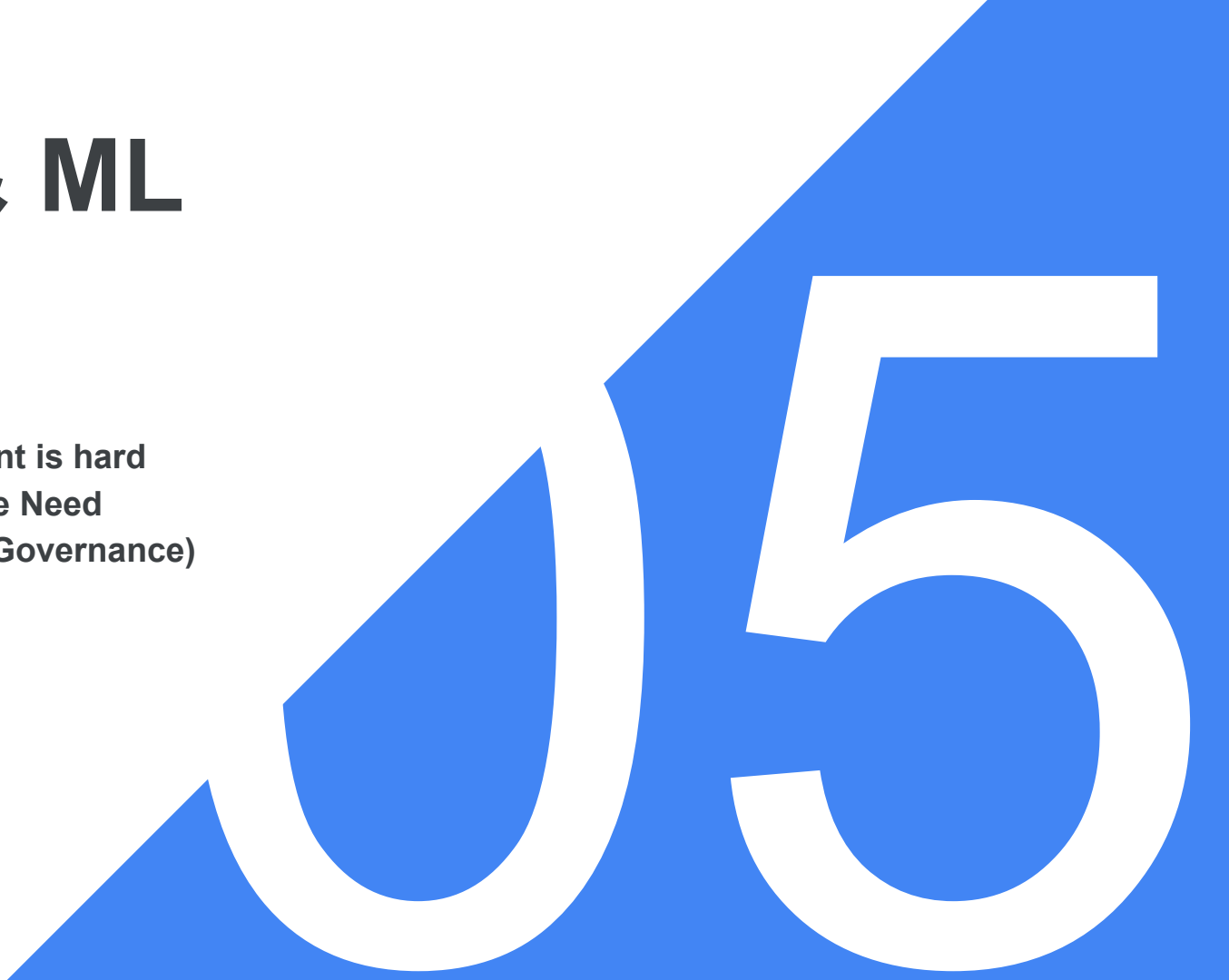

Google Cloud Platform
sandbox
vertex
11
mask-ai-sean-vertex_202181074449
VIEW DATASET
EXPORT
EVALUATE
DEPLOY & TEST
BATCH PREDICTIONS
MODEL PROPERTIES
Filter
Filter labels
Confidence threshold
0.5
IoU threshold
0.5
All labels 0
mask 0.68372
no_mask 0.51767
All labels
Average precision 0.649
Precision 94.2%
Recall 60.9%
Created Aug 10, 2021, 9:18:25 PM
Total images 898
Training images 682
Validation images 100
Test images 116
To evaluate your model, set the confidence threshold to see how precision and recall are affected. The best confidence threshold depends on your use case. Read some example scenarios to learn how evaluation metrics can be used.

模型部署

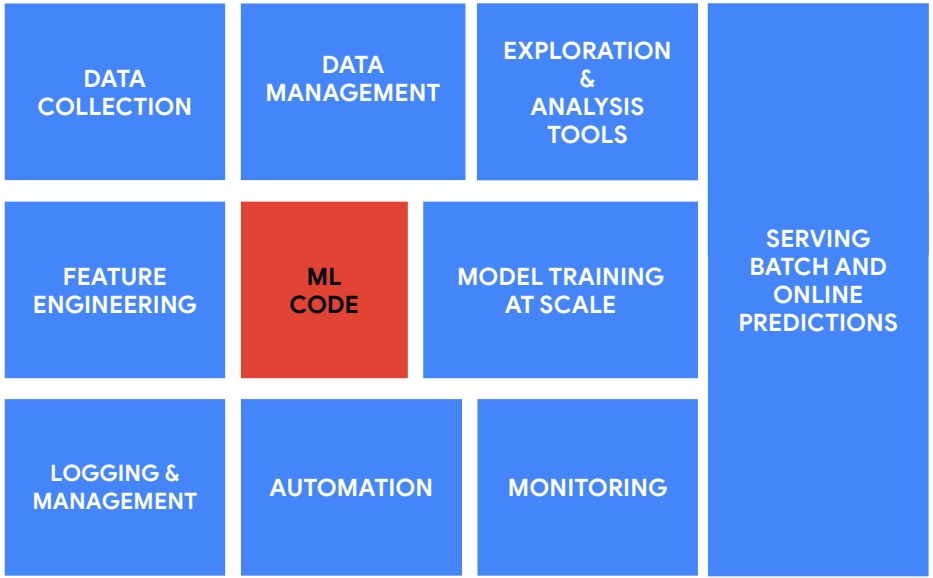

Google Cloud P
mask-ai
EVALUATE
Deploy your
Endpoints are mach
are useful for timely
request). You can al
DEPLOY TO END
Name
youjun-
mask-ai
Test your mo
UPLOAD IMAGE
Deploy to endpoint
1 Define your endpoint
2 Model settings
DEPLOY
CANCEL
Create new endpoint
Add to existing endpoint
Endpoint name *
Location
Region
us-central1 (Iowa)
Access
Determines how your endpoint can be accessed. By default, endpoints are available for prediction serving through a REST API. Endpoint access can't be changed after the endpoint is created.
Standard
Makes the endpoint available for prediction serving through a REST API. AutoML and custom-trained models can be added to standard endpoints.
Private
Create a private connection to this endpoint using a VPC network and private services access. Only custom-trained and tabular models can be added to private endpoints. Learn more
ADVANCED OPTIONS
CONTINUE

擁抱 AI & ML 的下一步

- Why Model Deployment is hard
- MLOps - Capability We Need
- 資料膨脹與治理 (Data Governance)



Why is ML deployment/operation difficult?



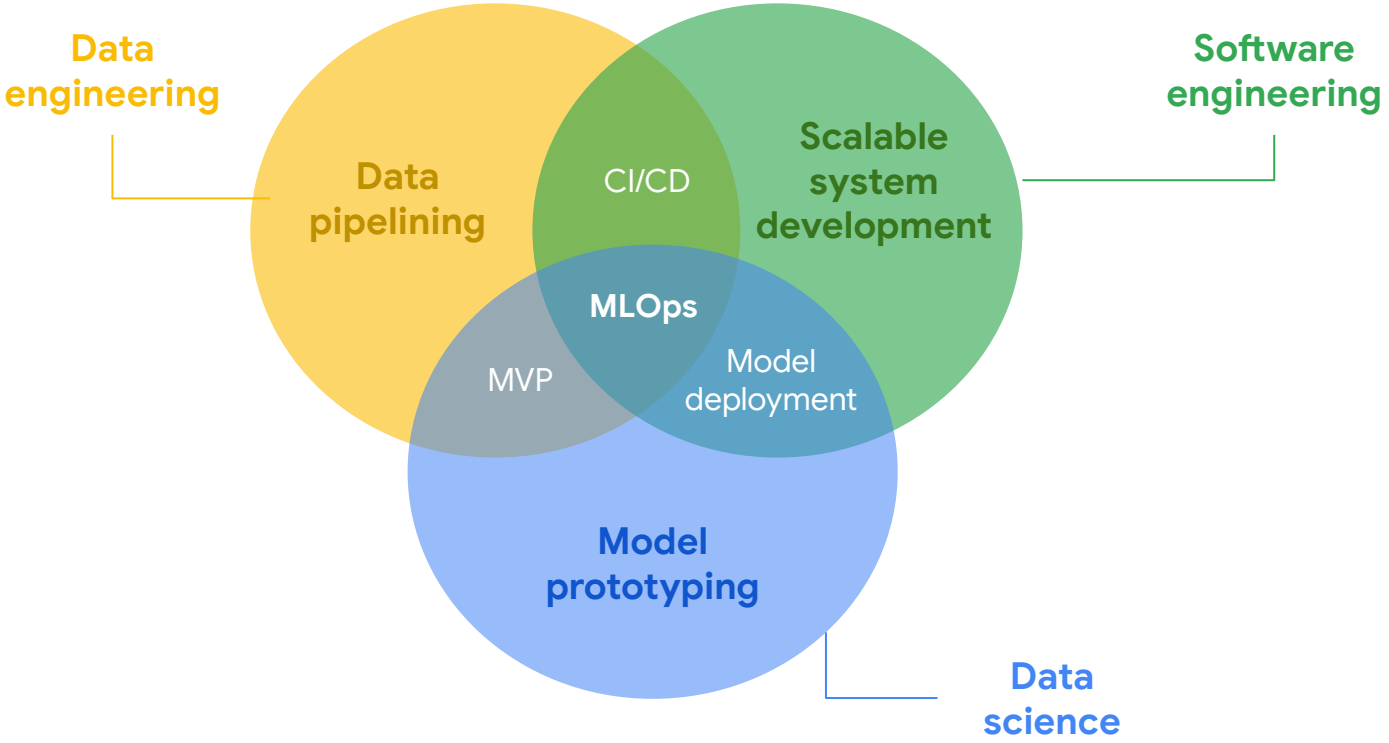
In an IT system

Behavior of system is defined by **Code**

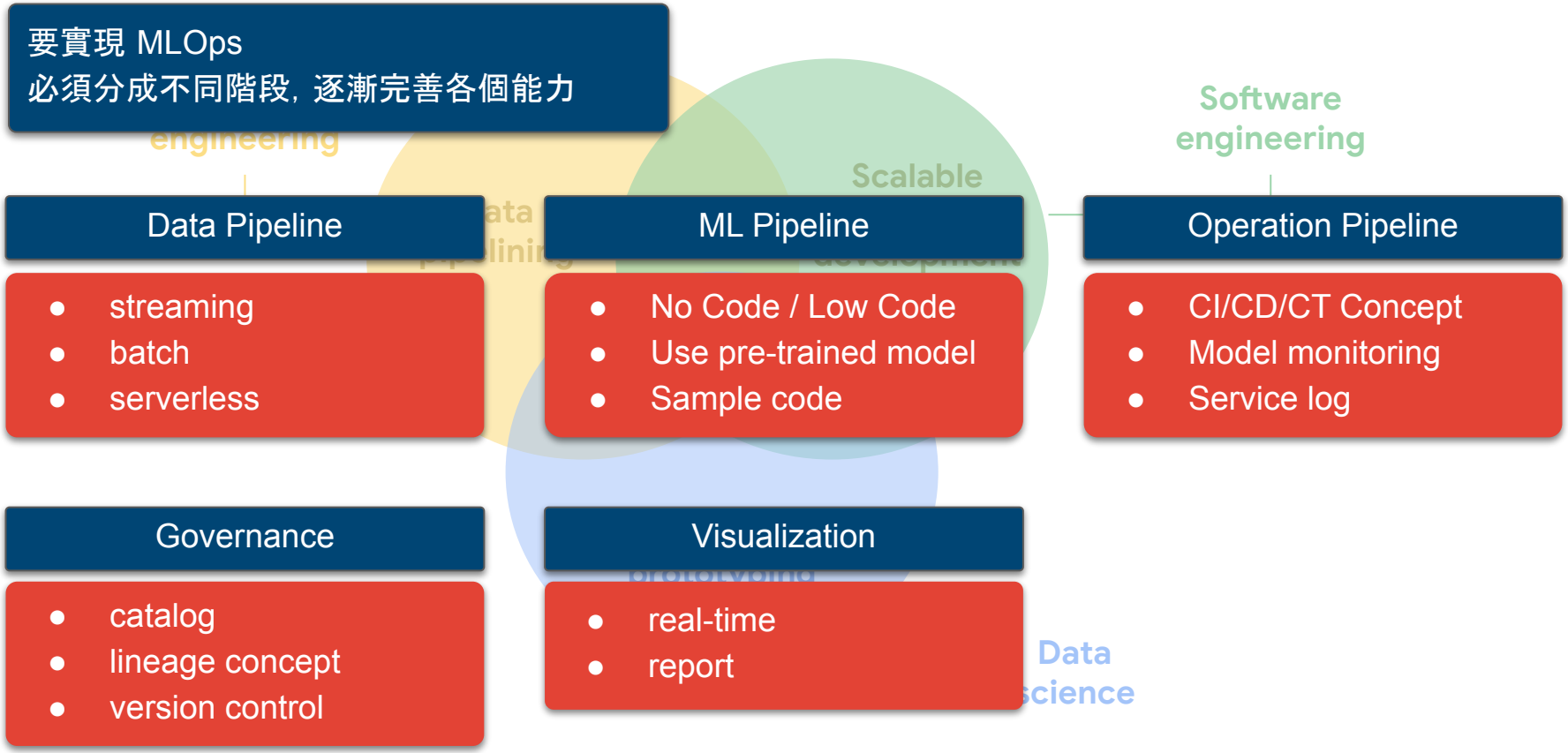
In an AI/ML System

Behavior of system is defined by **Data**

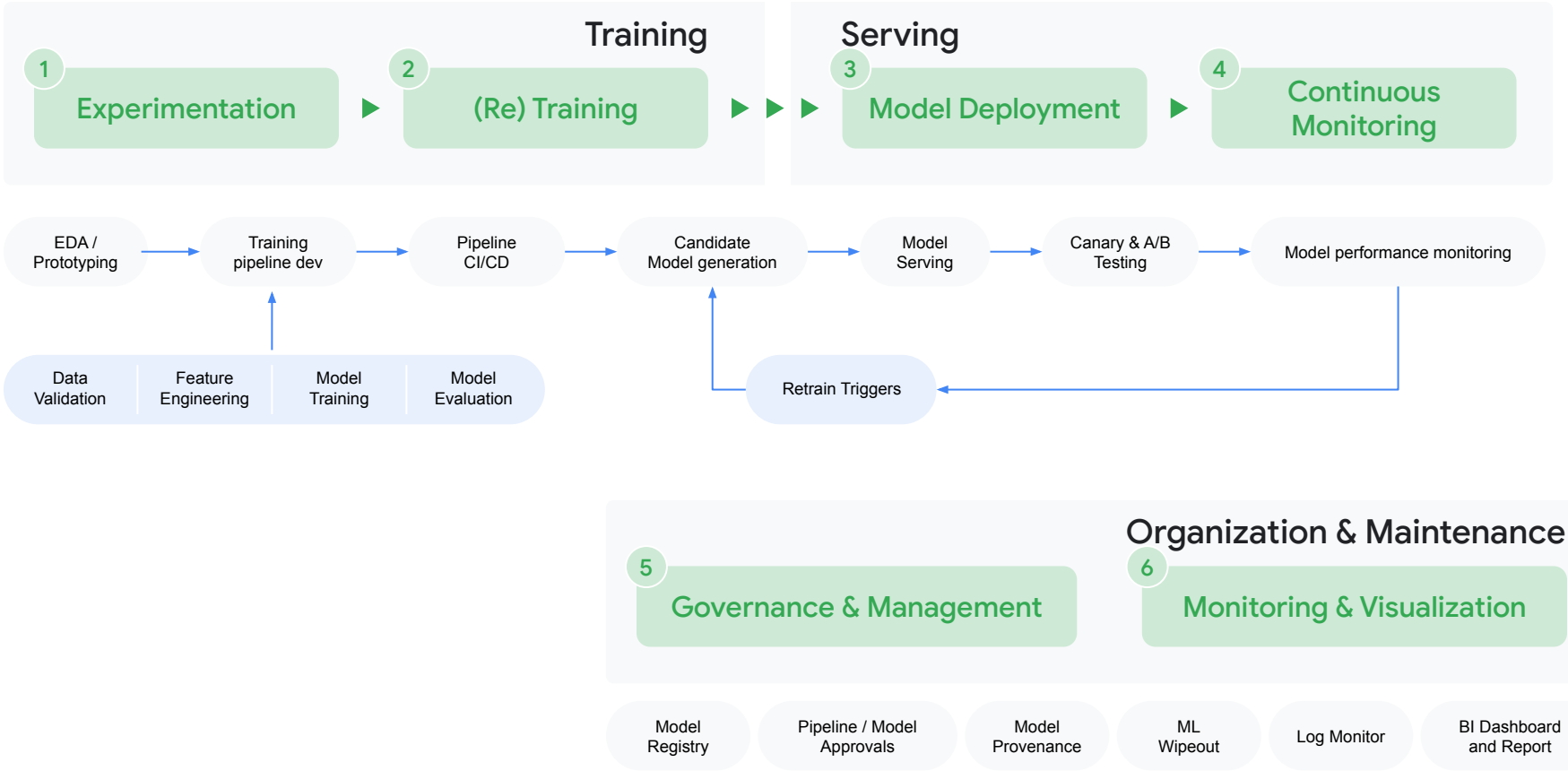
MLOps - Capability We Need

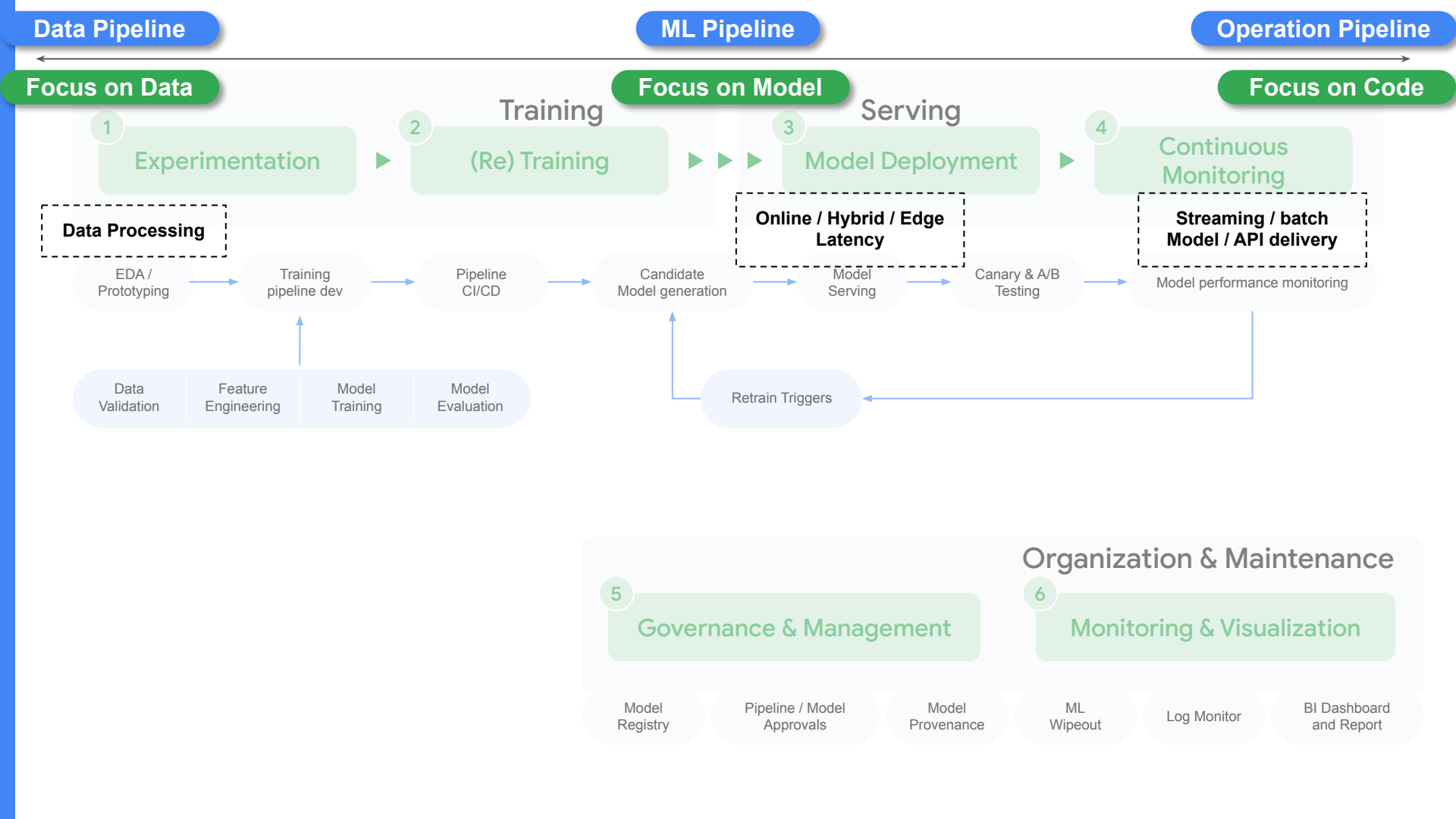


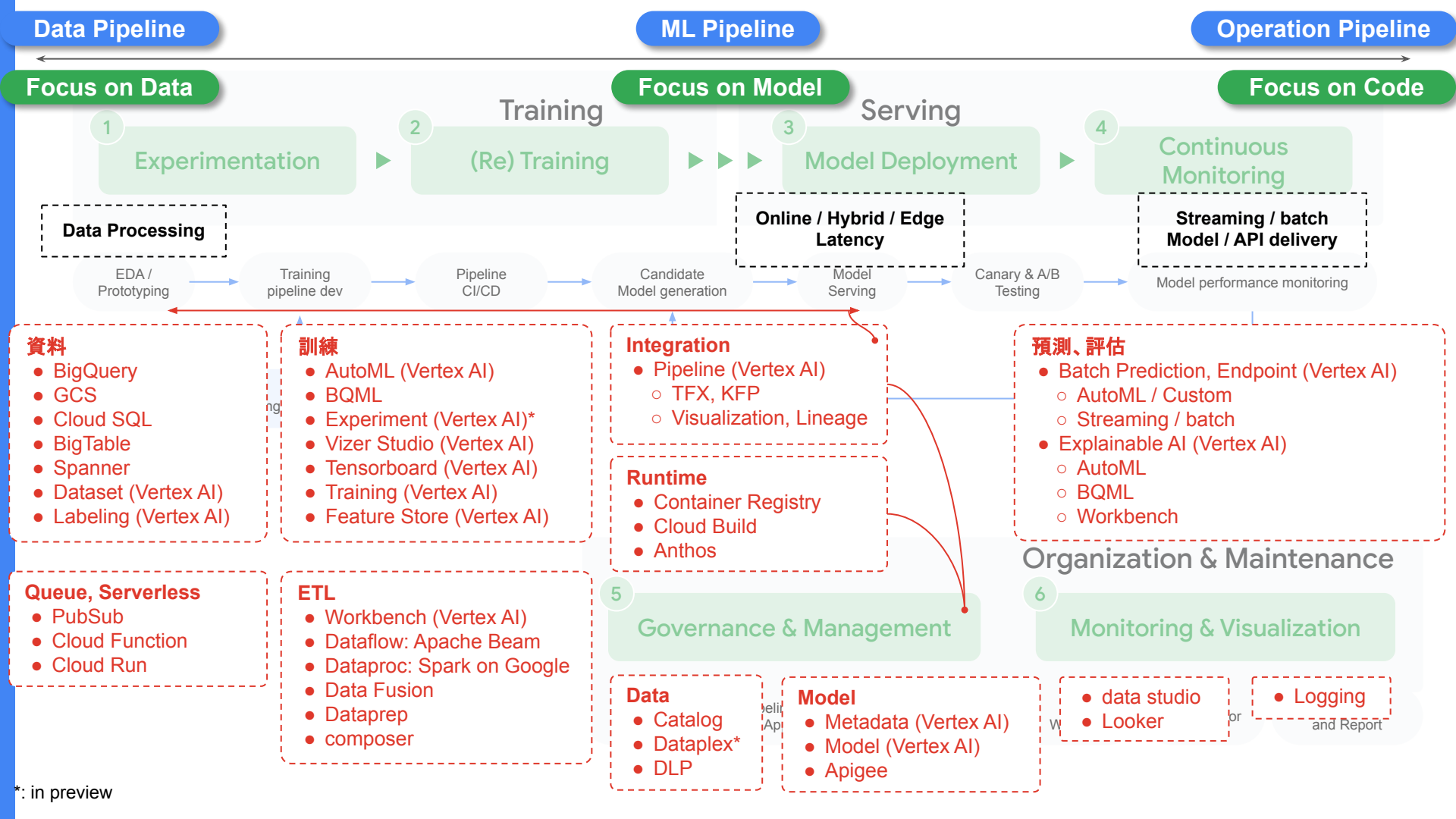
MLOps - Step by Step



Streamline and scale with MLOps



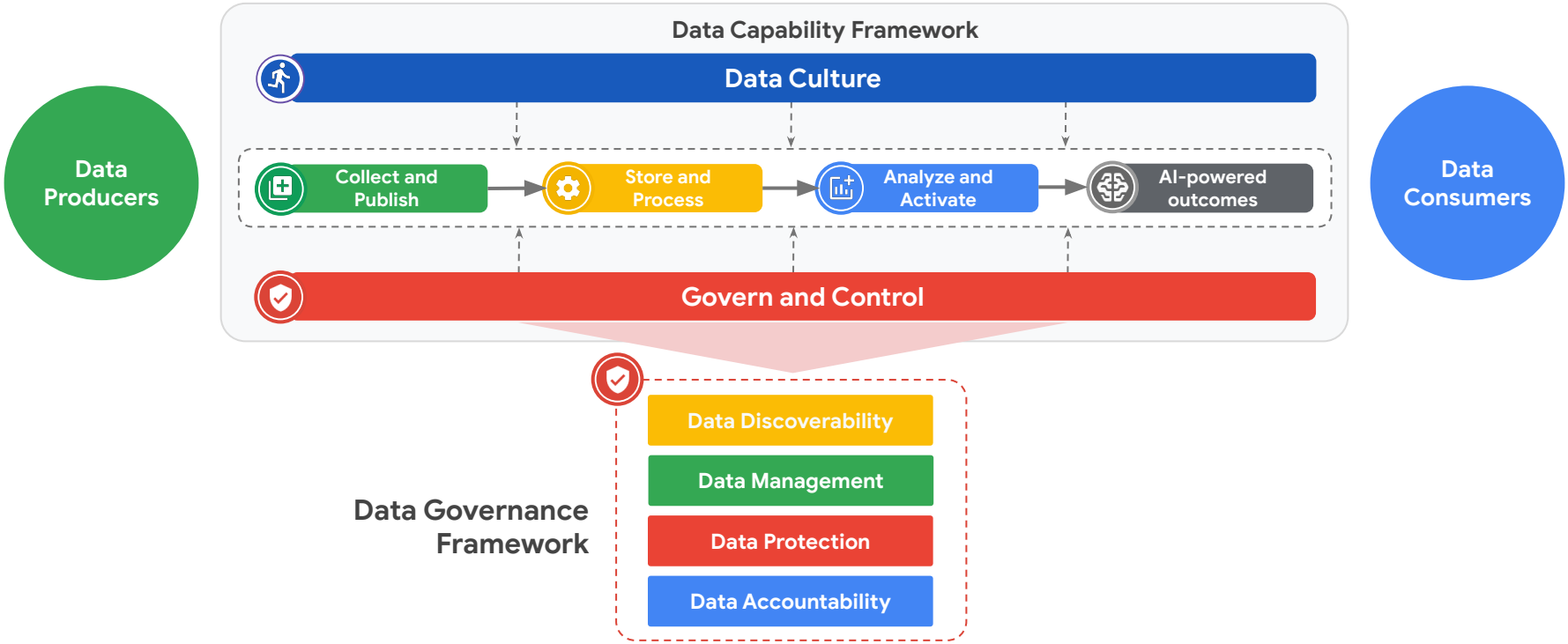

Data Pipeline
ML Pipeline
Operation Pipeline
Focus on Data
Focus on Model
Focus on Code
Training
Serving
1
Experimentation
2
(Re) Training
3
Model Deployment
4
Continuous Monitoring
Data Processing
Online / Hybrid / Edge Latency
Streaming / batch Model / API delivery
EDA / Prototyping
Training pipeline dev
Pipeline CI/CD
Candidate Model generation
Model Serving
Canary & A/B Testing
Model performance monitoring
Data Validation
Feature Engineering
Model Training
Model Evaluation
Retrain Triggers
Organization & Maintenance
5
Governance & Management
6
Monitoring & Visualization
Model Registry
Pipeline / Model Approvals
Model Provenance
ML Wipeout
Log Monitor
BI Dashboard and Report


Data Pipeline
ML Pipeline
Operation Pipeline
Focus on Data
Focus on Model
Focus on Code
Training
Serving
1
Experimentation
2
(Re) Training
3
Model Deployment
4
Continuous Monitoring
Data Processing
Online / Hybrid / Edge Latency
Streaming / batch Model / API delivery
EDA / Prototyping
Training pipeline dev
Pipeline CI/CD
Candidate Model generation
Model Serving
Canary & A/B Testing
Model performance monitoring
資料
BigQuery
GCS
Cloud SQL
BigTable
Spanner
Dataset (Vertex AI)
Labeling (Vertex AI)
訓練
AutoML (Vertex AI)
BQML
Experiment (Vertex AI)*
Vizer Studio (Vertex AI)
Tensorboard (Vertex AI)
Training (Vertex AI)
Feature Store (Vertex AI)
Integration
Pipeline (Vertex AI)
TFX, KFP
Visualization, Lineage
Runtime
Container Registry
Cloud Build
Anthos
預測、評估
Batch Prediction, Endpoint (Vertex AI)
AutoML / Custom
Streaming / batch
Explainable AI (Vertex AI)
AutoML
BQML
Workbench
Queue, Serverless
PubSub
Cloud Function
Cloud Run
ETL
Workbench (Vertex AI)
Dataflow: Apache Beam
Dataproc: Spark on Google
Data Fusion
Dataprep
composer
Organization & Maintenance
5
Governance & Management
6
Monitoring & Visualization
Data
Catalog
Dataplex*
DLP
Model
Metadata (Vertex AI)
Model (Vertex AI)
Apigee
data studio
Looker
Logging
and Report

*: in preview

資料膨脹與治理

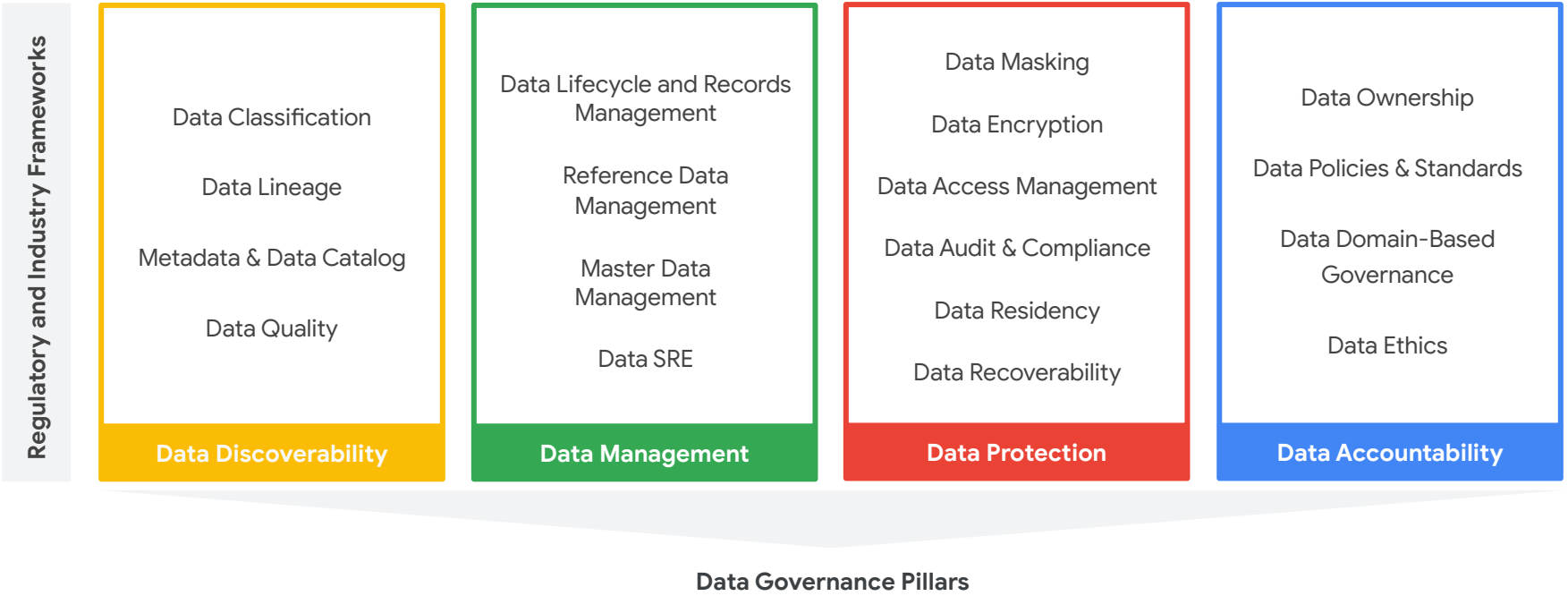
數據驅動是從數據的生產方到數據的消費者。
如果有一套**數據治理的框架**將整個旅程，結構化、系統化，那就能讓 資料應用加倍放大



資料膨脹與治理

At its most basic level, data governance is the practice of enhancing an organization's data such that it is **discoverable, understood, protected, and trusted.**

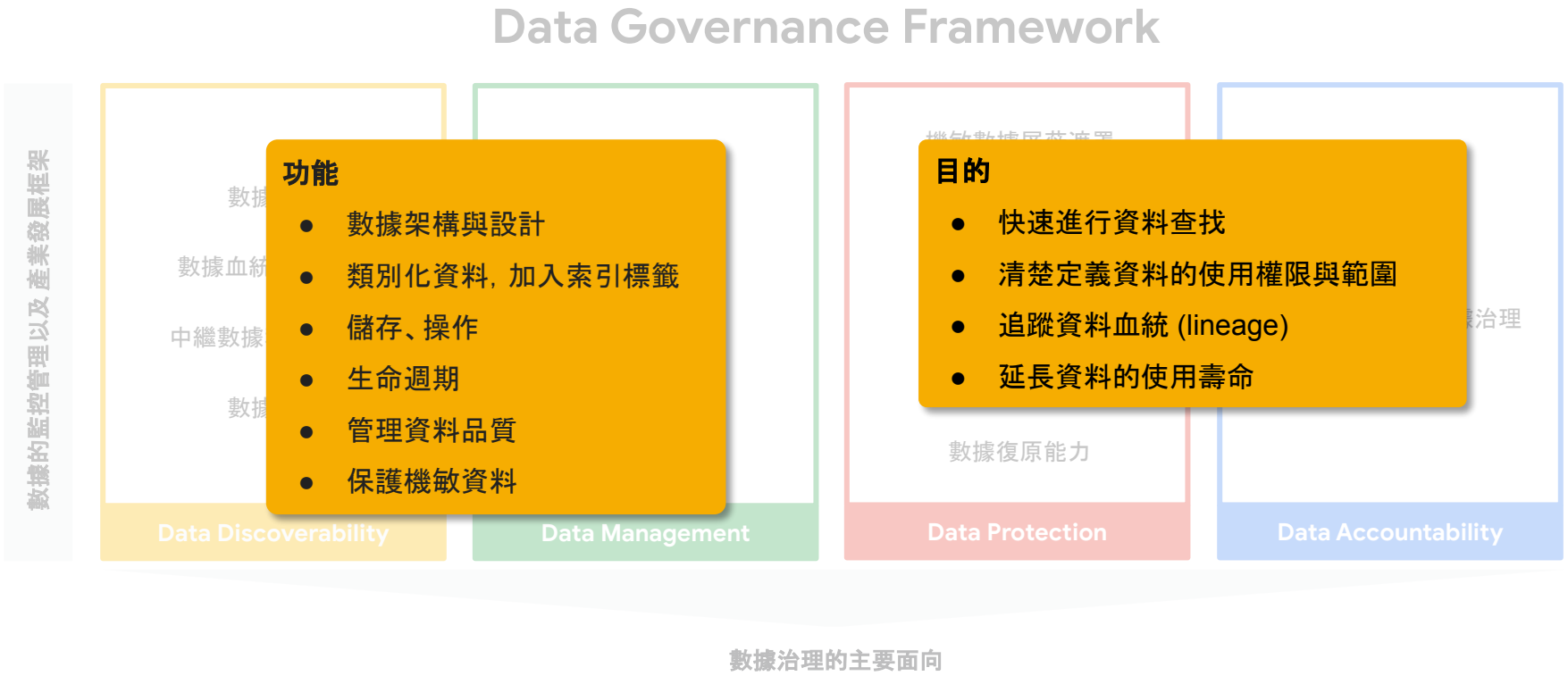
Data Governance Framework



資料膨脹與治理

你終究要做數據治理，何不一開始就用？

「數據治理」是增強整個組織在數據驅動上的實踐，使其可被 **發現、理解、管理、保護和信任** 。



漫漫長路，
承先啟後

06

全民 AI 後...

以數據治理為導向，思考如何

更有效地管理數據，

更快速地模型落地，

更創新的整合應用，

更精準的分工開發，

發揮資料科學的價值

Thank you

特別感謝「紐約 Baruch College 全球校友百業分享堂」

Linktree

履歷、專案
社群分享

MICROFUSION

我目前公司。
上雲需求歡迎

宏庭科技
股份有限公司