# Data science and AI applications in eCommerce

Dr. Shen Liu, Principal at Logickube

2 August 2023

logickube

# Agenda

- Introduction
- Case study 1 - Decision Engine
- Case study 2 - Vertex AI
- Case study 3 - Cloud Functions & Bigquery
- Q&A

# Introduction

# An overview of **Logickube**

We partner with leading cloud providers to provide AI and Data services to leading retailers in APAC

### Strategic advisory on AI

Build out a strategic roadmap and Proof-Of-Concepts models in collaboration with our technology and data experts to accelerate your AI journey.

### Cloud data engineering

Maximise the advantages of new technologies in the cloud by modernizing your data and ML workloads in a scalable and secure manner.

### AI engineering

Build best in class data and model pipelines to produce your own state of the art AI capabilities across retail, digital media, financial services and health.

### Personalisation & Attribution

Tailor unique interactions for each customer, across marketing channels and store experiences. Quantify personalisation benefits with robust experiment design and attribution.

Google Cloud Partner

databricks Partner Connect

aws

Azure

# Maths/Statistics is the cornerstone

Designing the right model to solve the right problem

- A systematic process to identify and define problems
- Always design the most suitable solution for each specific problem

Interpreting models and creating actionable insights

- Make black boxes transparent
- Translate maths/stats into business language

Experimental design / getting the right data

- Validate maths/stats tools in a real-world context
- Disentangle factors that jointly contribute to business success

Innovation and R&D

- Develop novel, mathematically sound methods adapting to ever-changing business needs

logickube

# Cloud computing is our go-to skillset

Leveraging cloud computing could significantly boost the performance of data and AI products

- Data **availability** and **reliability** - Data are replicated and stored in different locations, easy to backup and restore data in case of any failure
- **Big data capability** and high **efficiency** - e.g. we use Databricks to optimise performance, enabling real-time data processing, message producing and delivery
- **Cost effective**
  - Reduced cost of maintaining hardware and software
  - ‘Pay-as-you-go’: cost is only generated for what/when is used
  - Various tiers of computing power and storage classes

Cloud based solutions provide a good level of **data security**

- Advanced security features ensure data is securely stored and handled
- Data encryption in transit and at rest
- Certain protocols may be enforced to strengthen security

Our team has relevant **experience** and **qualifications**

- Certified data scientists, data engineers, machine learning engineers and solution architects across mainstream cloud platforms
- Highly experienced in building scalable cloud solutions to create end-to-end data and AI solutions

# Case study 1

## Decision Engine

# Decision Engine offering

## ML decisioning of offers

8+ basket offer constructs
10+ category offers
Hundreds of discount depths

## Superior performance

+25% inc. sales

x2 redeem rates

## Omni-channel experience

| Reach non-email marketable audience | 20X increase in return on ad spend in paid channels |
| --- | --- |

## Multi-objective decisioning

Optimisation engine that can balance multiple objectives

e.g.: costs, audience size, returns

## Improves operating efficiency

Always-on and automated pipeline free up execution resources

Marketers can focus on offer and creative designs whilst engine curates the best action

## Framework for test & learn and enabling measurement

Rapidly test out new offers through random experiments

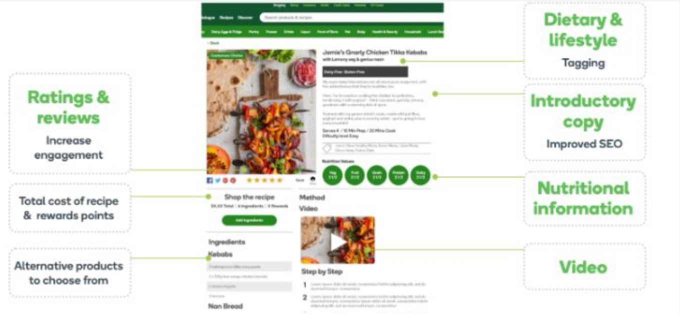
Set aside control groups for measurement purpose

# Decision Engine key features

**Responsive and personalised recipes across discovery pages**

- Home page
- Landing page
- Listing page

**Learn all metadata on a recipe, for 5K+ recipes**



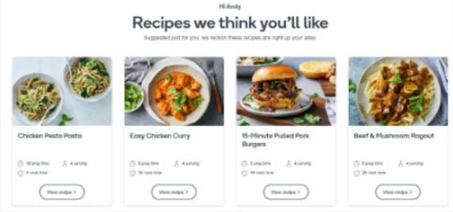

**Real time API**

- <200ms latency
- Platform agnostic

**Benefits**

- Increased engagement and discovery
- Combines offline and online events
- Responsive content
- Automate recipe tagging

**4 personalisation API**


| | |
|---|---|
| What's trending | Vegetarian/Vegan |
| Most recent | We think you'll like |

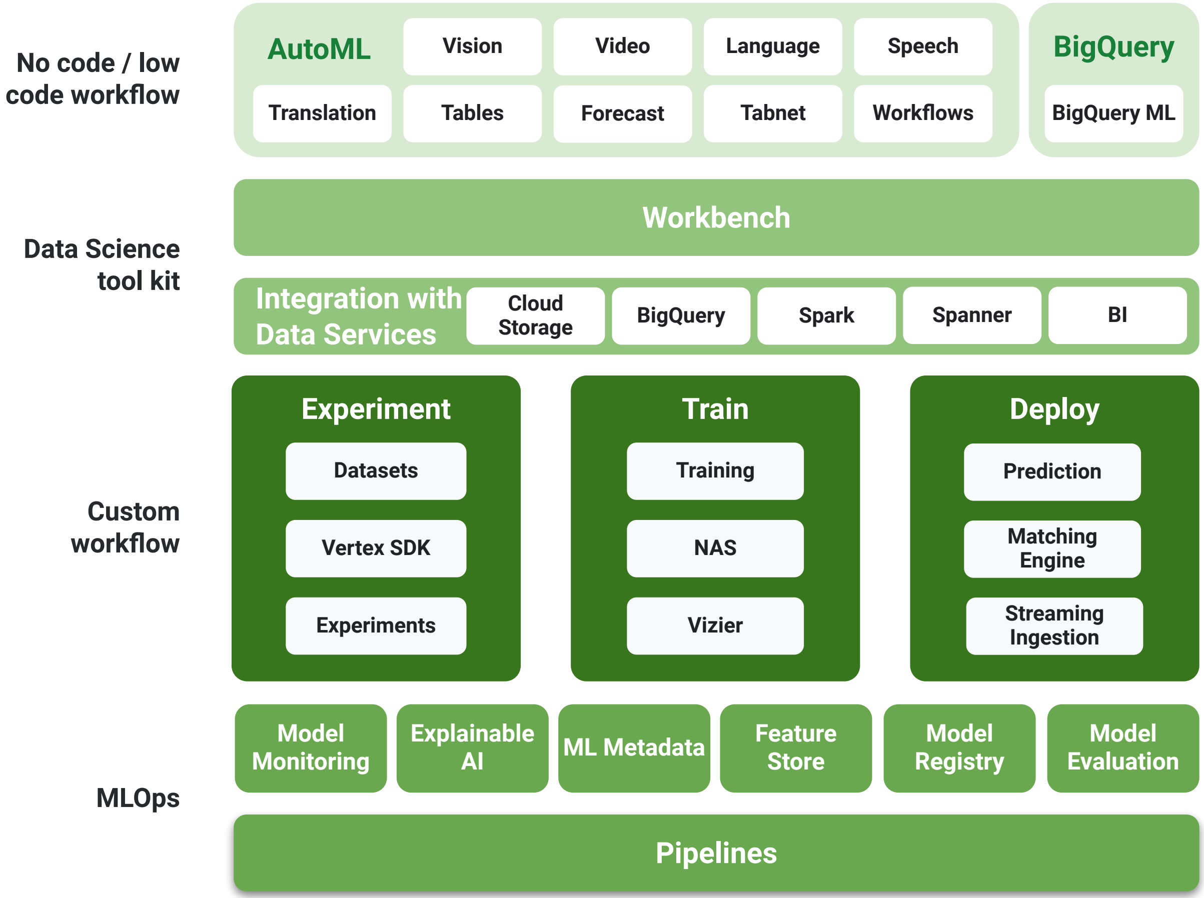
logickube

# Case study 2

## Vertex AI

# Vertex AI

- Unified development and deployment platform for data science and machine learning
- Increase productivity of data scientists and ML engineers




logickube

# What Data Science & Machine Learning Engineering teams want

**Accelerate time to market**

Vertex AI

**Unified data and AI platform for all users to accelerate time to value**

**Enhance stability and reliability of ML systems**

Vertex AI

**End-to-end MLOps to efficiently and responsibly manage and govern AI**

**Improve AI team productivity**

Vertex AI

**Open and scalable AI infrastructure to flexibly and successfully deploy AI**

# Vertex AI is a platform for all users throughout the ML lifecycle

## Data analyst
Query and analyse

**Endless EDW**
BigQuery

**Self-managed data pipelines**
Cloud Data Fusion, Dataflow

**Data models, catalog**
Looker, Data Catalog

**Machine learning in SQL**
BigQuery ML

## Data engineer
Get clean, useful data

**Self-driving infra**
BigQuery, Dataflow, Cloud Composer

**Broad choice of tools/language**
Dataproc, Dataflow

**Data quality /lineage**
Vertex AI, BigQuery, Dataflow

**Real-time capabilities**
BigQuery, Dataflow

## Data scientist
Models that work

**Portable notebooks**
Managed Notebooks

**Model eval and selection**
Vertex Explainable AI, Vertex AI Experiments

**Point-and-click dev**
AutoML

**Collaboration**
Vertex AI Feature Store, Vertex AI Pipelines

## ML engineer
Models in production

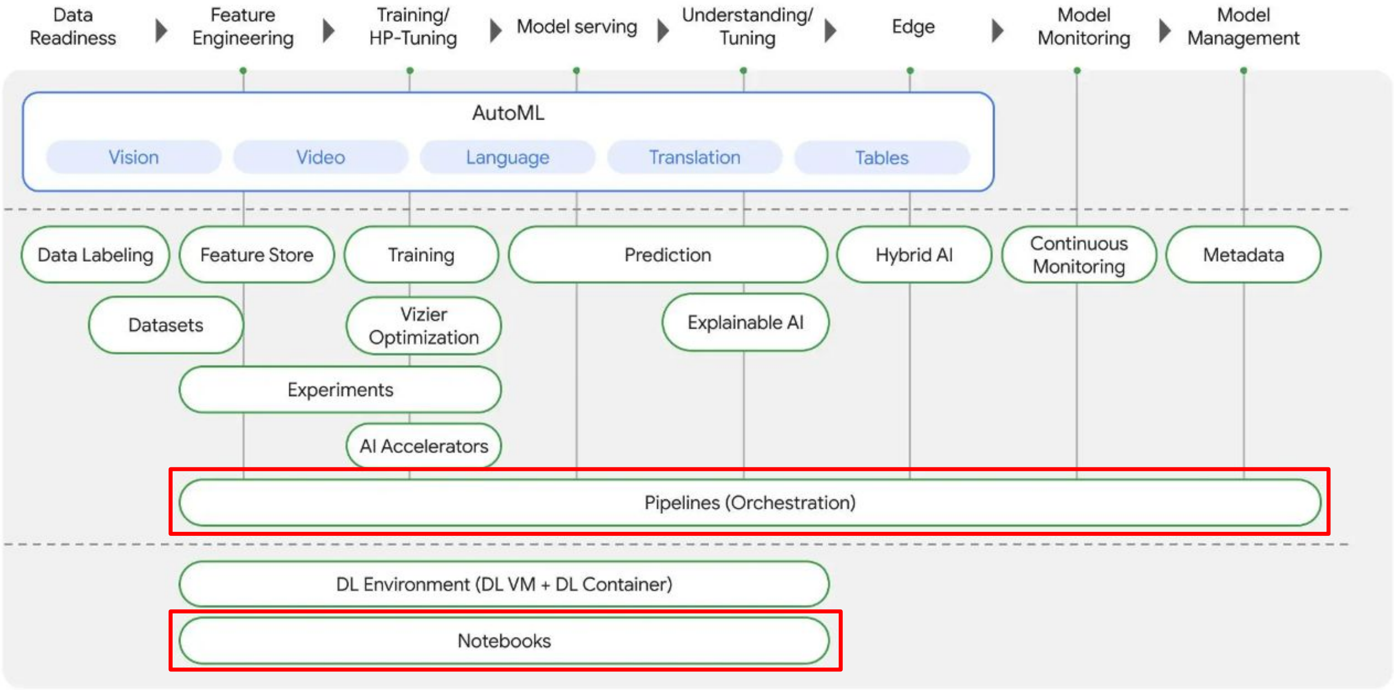
**Scalable model hosting**
Vertex AI Prediction

**ML CI/CD and orchestration**
Vertex AI Pipelines

**Provenance and lineage**
Vertex ML Metadata

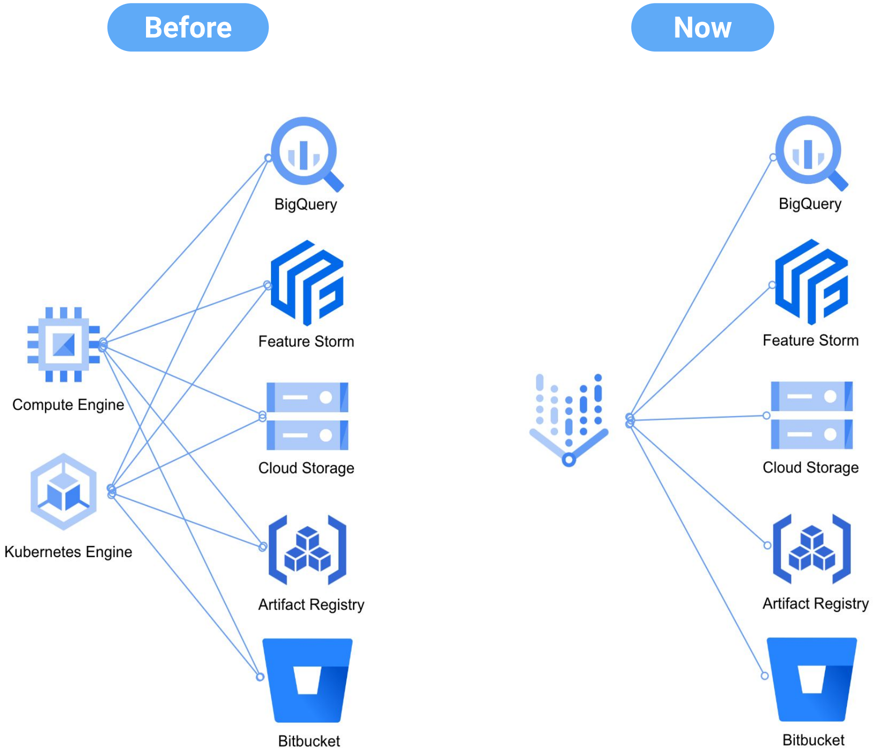
**Improvements and retraining**
Cloud Monitoring

# Vertex AI for large enterprises



# Vertex AI for large enterprises

- Vertex AI enables seamless connections with data sources such as BigQuery and Cloud Storage
- Less infrastructure configuration
- End-to-end ML workflows in one place



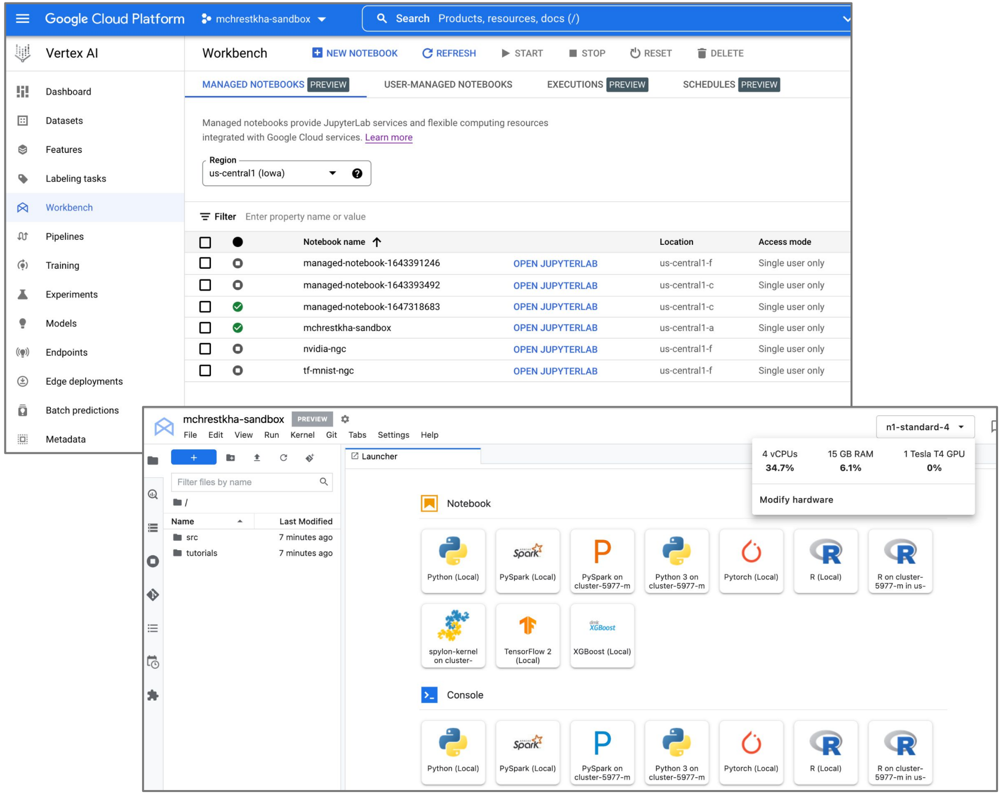

logickube

# Vertex AI **Workbench**: One-stop surface for data science

### Fully managed compute with admin control
A Jupyter-based fully managed, scalable, enterprise-ready compute infrastructure with easily enforceable policies and user management

### Fast workflow for data tasks
Seamless visual and code-based integrations with data & analytics services

### At-your-fingertips integration
Load and share notebooks alongside your AI and data tasks. Run tasks without extra code




logickube

# Benefits

**Easy data exploration and analysis** with Easy access to data in BigQuery and Cloud Storage within a Jupyter notebook

```
[2]: %%bigquery regions_by_country
     SELECT
       country_code,
       country_name,
       COUNT(DISTINCT region_code) AS num_regions
     FROM
       `bigquery-public-data.google_trends.international_top_terms`
     WHERE
       refresh_date = DATE_SUB(CURRENT_DATE, INTERVAL 1 DAY)
     GROUP BY
       country_code, country_name
     ORDER BY
       num_regions DESC;

     Query complete after 0.19s: 100%|██████████| 4/4 [00:00<00:00, 1
     Downloading: 100%|██████████| 41/41 [00:02<00:00, 16.35rows/s]

[5]: regions_by_country.info()

     <class 'pandas.core.frame.DataFrame'>
     RangeIndex: 41 entries, 0 to 40
     Data columns (total 3 columns):
      #   Column        Non-Null Count  Dtype
     ---  ------        --------------  -----
      0   country_code  41 non-null     object
      1   country_name  41 non-null     object
      2   num_regions   41 non-null     int64
     dtypes: int64(1), object(2)
     memory usage: 1.1+ KB
```

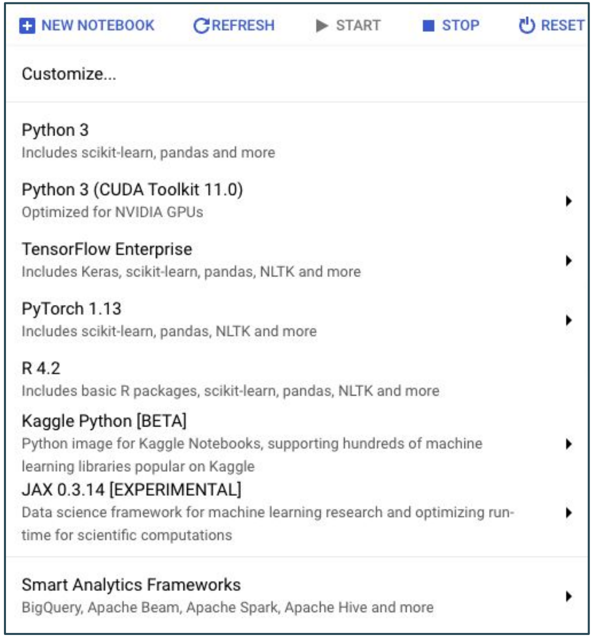
**Fast prototyping and model development** by creating a new notebook under 1 minute and connecting to other GC services within it

```
[ ]: subprocess.run(f'{sys.executable} {AKL_DIR}/exec_build_config.py', shell=True)

[ ]: subprocess.run(f'{sys.executable} {AKL_DIR}/exec_structure_create.py', shell=True)

[ ]: subprocess.run(f'{sys.executable} {AKL_DIR}/exec_preprocessor.py', shell=True)

[ ]: subprocess.run(f'{sys.executable} {AKL_DIR}/exec_feature_selection.py', shell=True)

[ ]: subprocess.run(f'{sys.executable} {AKL_DIR}/exec_combiner.py', shell=True)

[ ]: subprocess.run(f'{sys.executable} {AKL_DIR}/exec_generate_queue.py', shell=True)

[ ]: for i in range(4):
         subprocess.run(f'{sys.executable} {AKL_DIR}/exec_fit.py {i}', shell=True)

[ ]: subprocess.run(f'{sys.executable} {AKL_DIR}/exec_model_selector.py', shell=True)

[ ]: subprocess.run(f'{sys.executable} {AKL_DIR}/exec_diag_scoring.py', shell=True)

[ ]: subprocess.run(f'{sys.executable} {AKL_DIR}/exec_shap.py', shell=True)
```

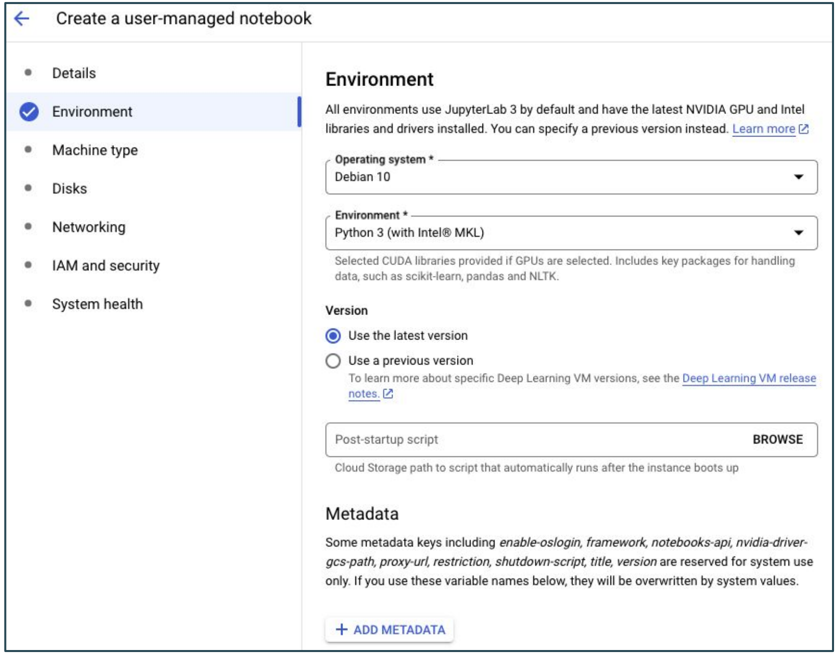
# User-Managed Notebook

User-managed notebooks are high customisable VM instances and suitable for data exploration, analysis and model development

**A preinstalled suite of ML/DL packages**




**Similar setup process to GCE**




logickube

# Managed Notebook vs User-Managed Notebook

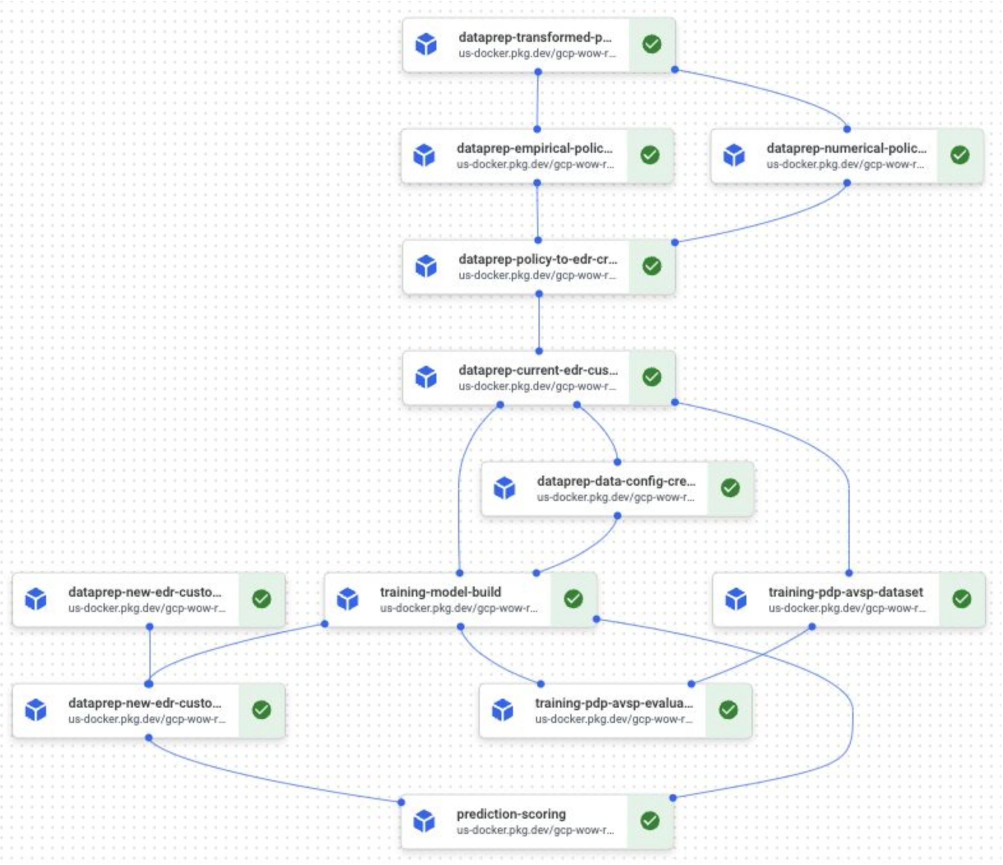
| Feature | Managed Notebook | User-Managed Notebook |
|---|---|---|
| **Flexibility** | Low | High |
| **Custom environment** | Yes | Yes |
| **Switch machine type** | Within JupyterLab | Shutdown, switch, and restart |
| **GCS navigation** | Within JupyterLab | In GCS |
| **BigQuery navigation** | Within JupyterLab | In BigQuery |
| **Scheduled runs** | Supported | Not supported |
| **Management fees** | $0.05 per vCPU per hour | $0.005 per vCPU per hour |
| **Idle shutdown** | Supported | Not supported |

# **Productionise** Models in Vertex AI

Vertex AI **Pipelines** orchestrate ML workflows serverlessly, and automate and monitor repeatable workflows such as model training and production.

Benefits:

- Serverless service
- Lower costs
- Workflow automation
- Composable and reusable pipelines
- Python function-based components




logickube

# When to use Pipelines

1. **Train/productionise models with well-defined and reusable workflows**

   When ML workflows are finalised and will be reused for multiple times, consider packaging the dependencies into a Docker image and migrate the workflows from notebook to Pipelines to save time and improve reliability.

2. **Automate model training/production**

   Manual weekly/monthly model scoring or refitting could be tedious, and schedule pipeline execution or trigger pipeline runs with Pub/Sub could be a game changer.

3. **Scalable model production**

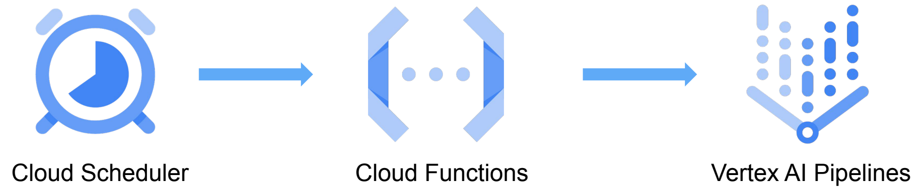
   Built on top of Kubernetes, Vertex AI Pipelines are serverless and scalable. Users are able to specify different level of resources for different steps and design parallel processing to boost speed.

# Automate Model Training/Production

Model training/production can be automated by scheduling or triggering pipeline runs.

To schedule pipeline runs, **Cloud Scheduler** and **Cloud Functions** are also needed other than Vertex AI

- Configure Cloud Scheduler to send a JSON string to Cloud Functions on your pre-defined schedule
- Cloud Functions that you build will parse the JSON string and submit pipeline runs using ingested parameters
- Pipeline runs

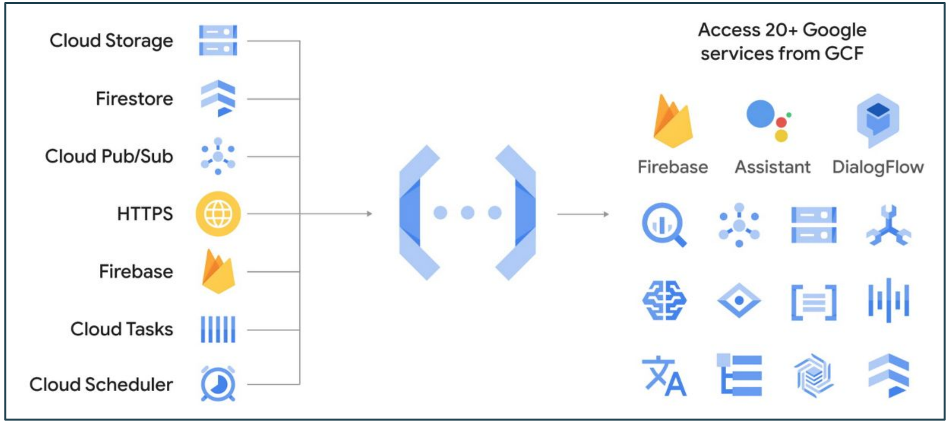


# Case study 3

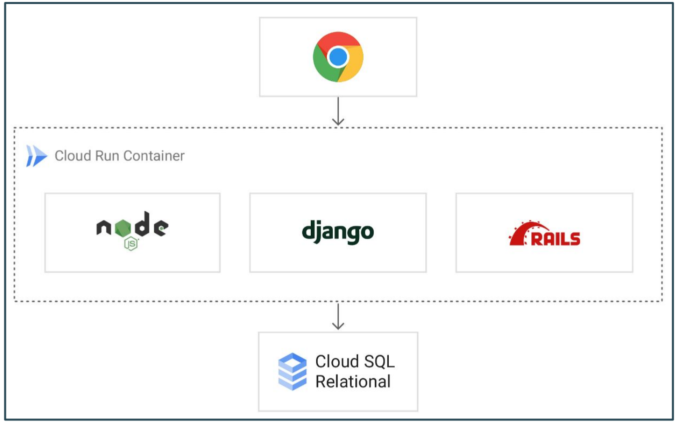
## Cloud Function & Bigquery

# A Multi-agent Orchestration Problem

- Multiple teams in the company collaborate for a common business (orchestration);
- They have different preferences in data transfer methods (Gmail, RDMS, Google Drive);
- They use different technical tools (GCP, Azure, AWS);
- They own different domain knowledge (BI, DA, DS);




**connecting different platforms**




**running long-term services**

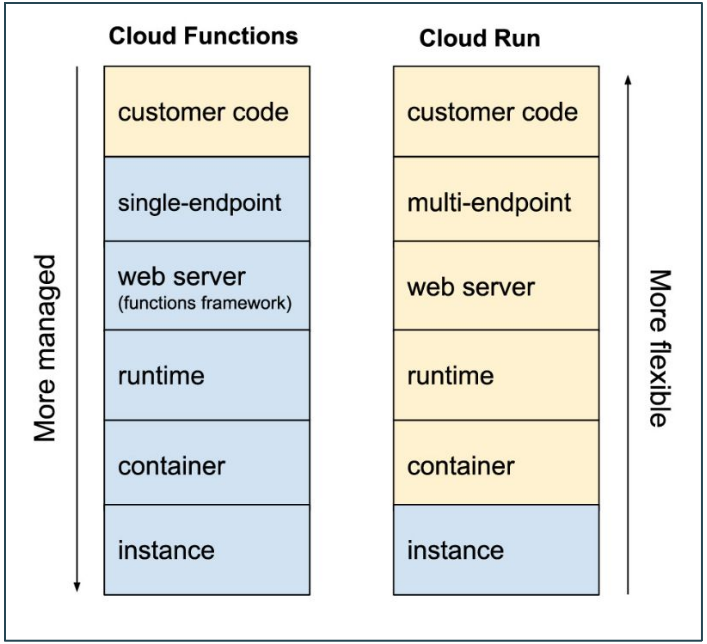
# Solution to the Orchestration Problem

## Serverless Architectures

- Faster solutions to market at lower cost
- Decreased management overhead than traditional approaches

## Cloud Function v.s. Cloud Run

- Cloud Function
  - Transforming data and loading it into BigQuery
  - Creating data summary once a BigQuery table gets updated
  - Use ML APIs to analyze data added to a database or storage bucket
- Cloud Run
  - Any web-based workload
  - REST APIs for mobile apps or games
  - Internal custom backoffice apps




## Google Cloud Function

- Function-as-a-service (FaaS) in Google Cloud;
- Serverless architectures with pay-as-you-go convenience;
- Connection or extension to services with complex applications;
- Remedy to reconcile orchestration problems;

# Main Features of Google Cloud Function (GEN 2)

## Increased compute with granular controls

- Instance concurrency (up to 1000 requests/instance)
- Fast rollbacks (version control)
- 6x longer request processing (max. 60 minutes)
- 4x larger instances (max 16GB RAM + 4 vCPUs)
- Pre-warmed instances (fast configuration)
- Support multiple programming languages
- Extensibility and portability (to Cloud run)

## Empowering Business Intelligence

- Inclusive to contributors from different backgrounds;
- Enable non-SQL functionalities;
- Combine complex operations in one go;
- Seamless data/messages digestion + broadcast;


Python

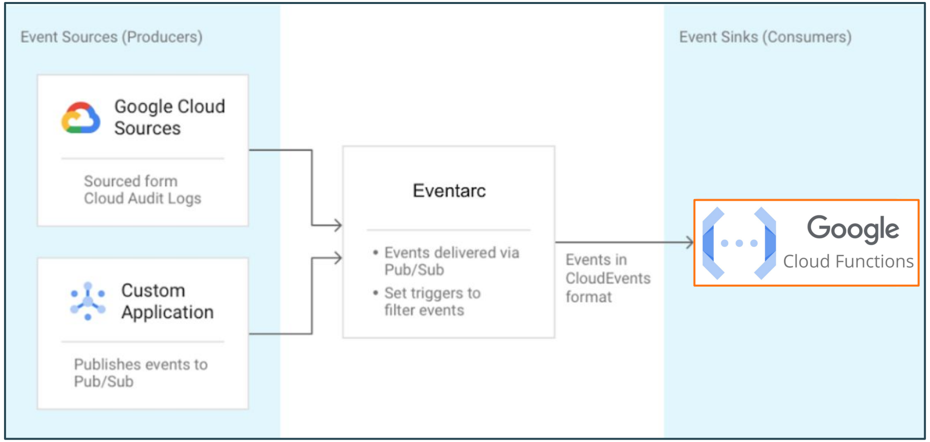
| Runtime | Operating System | Runtime ID |
|---|---|---|
| Python 3.11 (recommended) | Ubuntu 22.04 | python311 |
| Python 3.10 | Ubuntu 22.04 | python310 |
| Python 3.9 | Ubuntu 18.04 | python39 |
| Python 3.8 | Ubuntu 18.04 | python38 |
| Python 3.7 | Ubuntu 18.04 | python37 |

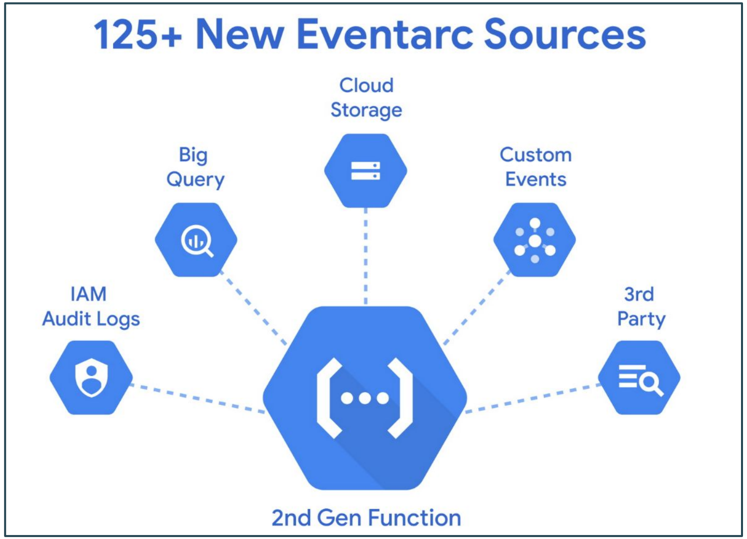
**Node.js, Go, Java, Ruby, PHP, .NET Core**

# Main Features of Google Cloud Function (GEN 2)

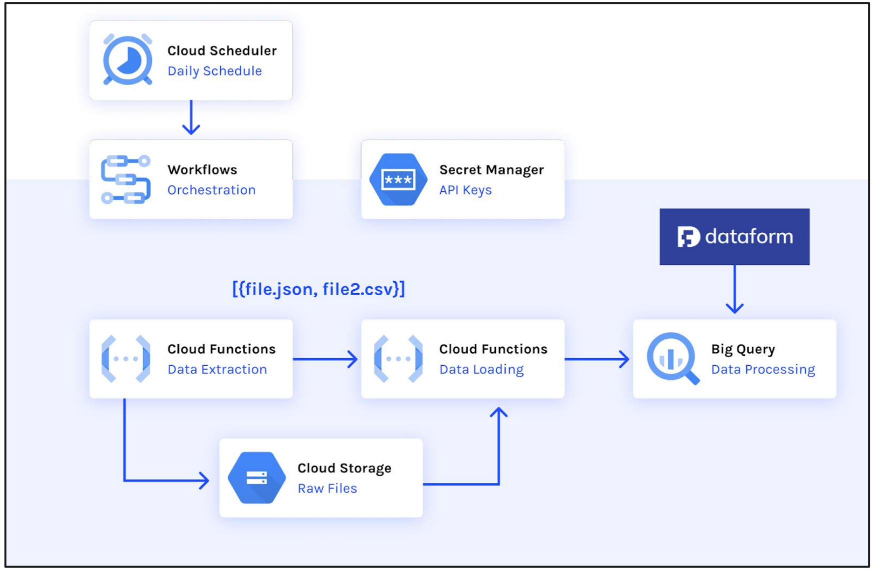
## Lots more event sources with the Eventarc

- 125+ Event sources (BigQuery, GCS, API Keys)
- Standards-based Event schema for consistent developer experience
- Customer-Managed Encryption Keys (CMEK) support

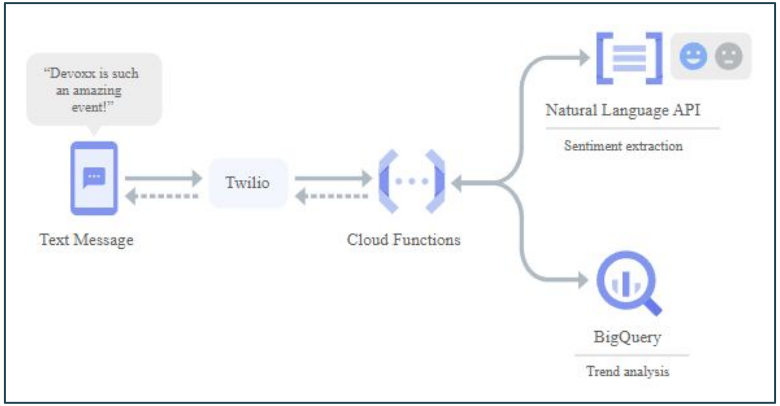


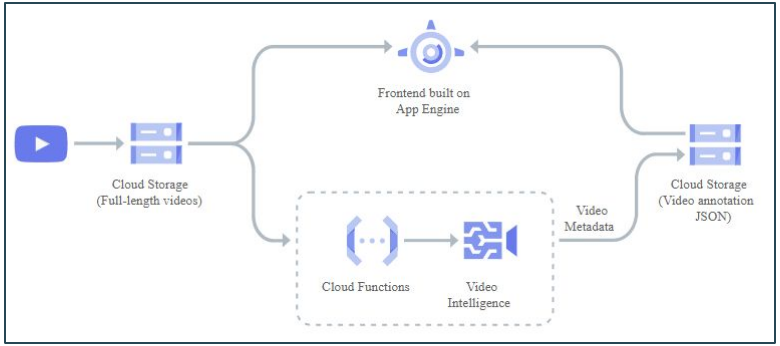
# Orchestration Connected by Cloud Functions




**Simple ETL Workflow**



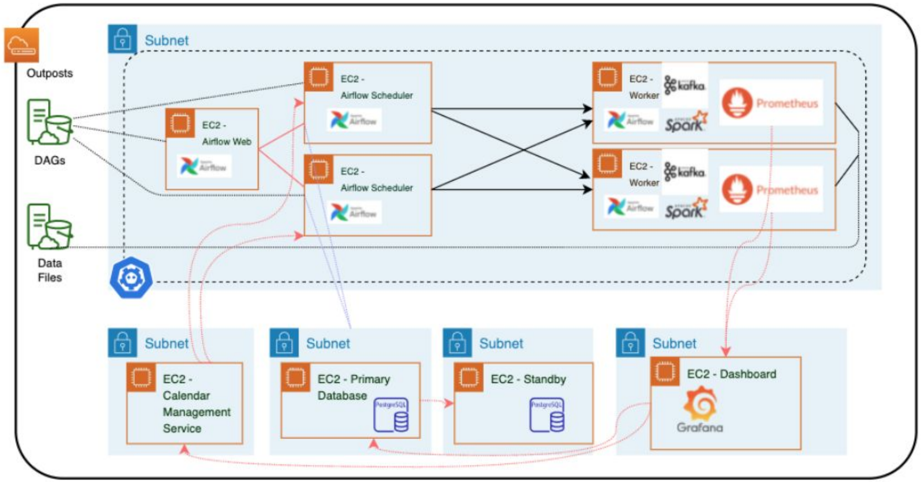

**Real-time Text Messages Recognition and Logging**

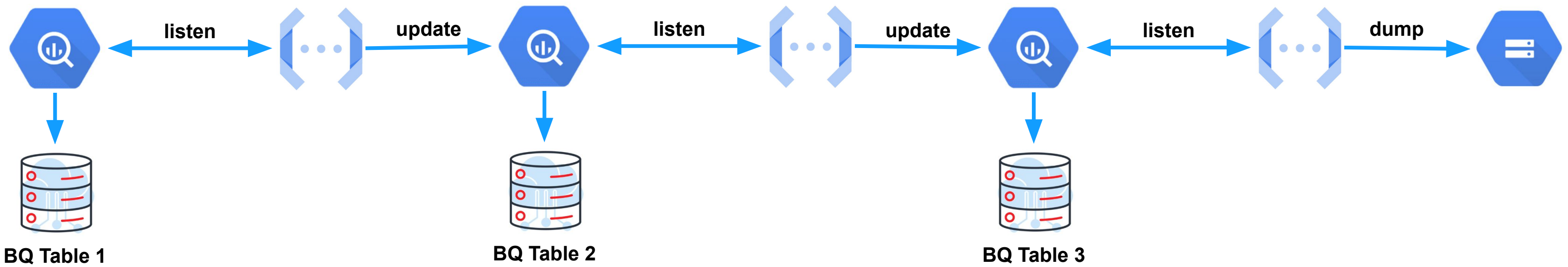



**Transfer Video Objects to GCS**

logickube

# Comparison to Conventional Orchestration Pipeline




logickube

# **BigQuery ML** Overview

BigQuery ML

```
#standardSQL
CREATE MODEL `bqml_tutorial.sample_model`
OPTIONS(model_type='logistic_reg') AS
SELECT
  IF(totals.transactions IS NULL, 0, 1) AS label,
  IFNULL(device.operatingSystem, "") AS os,
  device.isMobile AS is_mobile,
  IFNULL(geoNetwork.country, "") AS country,
  IFNULL(totals.pageviews, 0) AS pageviews
FROM
  `bigquery-public-data.google_analytics_sample.ga_sessions_*`
WHERE
  _TABLE_SUFFIX BETWEEN '20160801' AND '20170630'
```

BigQuery ML is a part of enterprise BigQuery that allows you to create and execute ML models using Google SQL queries.

# Why Use BQML?

### Easy adaptation

- Develop ML models using the language you are comfortable with
- No need to learn Python or Java and ML frameworks such as TensorFlow or PyTorch

### Increased development speed

- No need to move data in/out of BQ throughout the entire ML lifecycle.
- Bring ML to data, not the other way around.
- No need to wait for limited resources of data science team

### No more time wasted on setup

- BigQuery is serverless so no need to provision VMs for model training
- Ready to develop - no extra setup required such as installing frameworks and other dependencies

# BQML - Supported Models

**Internally trained**

| | |
|---|---|
| **Regression** | • Linear regression |
| **Classification** | • Logistic regression |
| **Others** | • K-means clustering<br>• Matrix factorisation<br>• PCA<br>• Time series forecasting |

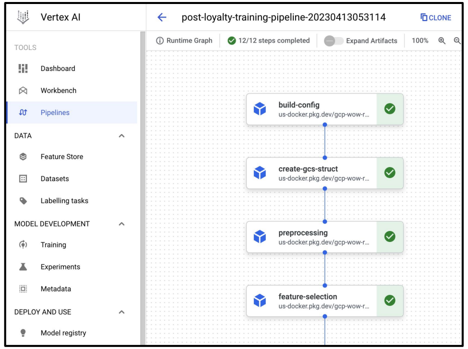
**Externally trained (Vertex AI)**

| | |
|---|---|
| **Regression** | • DNN<br>• Wide & Deep Networks<br>• Boosted Tree<br>• Random forest<br>• AutoML Tables |
| **Classification** | • DNN<br>• Wide & Deep Networks<br>• Boosted Tree<br>• Random forest<br>• AutoML Tables |
| **Others** | • Autoencoder |

# BQML in Google ML Landscape

**Out of box** → **DIY**


```sql
CREATE OR REPLACE MODEL `bqml.penguins_model`
OPTIONS (model_type='linear_reg',
         input_label_cols=['body_mass_g']) AS
SELECT * FROM `public-data.ml_datasets.penguins`
WHERE body_mass_g IS NOT NULL
```




### Pre-trained APIs & solutions

Cloud Vision API
Speech-to-Text API
...

### Custom AI with BQML and AutoML

No-code/low-code approach

### End-to-end AI with core tools

Vertex AI and TensorFlow give data scientists strong control to build and deploy models

logickube

# Import and Export Models in BQML

You can import the following models trained outside BQML and use them to perform prediction within BQ:

- Open Neural Network Exchange (ONNX) format
- TensorFlow Saved Model format
- TensorFlow Lite format
- XGBoost Booster format

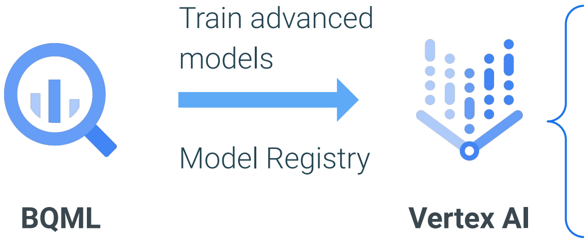

You can export most models trained in BQML in the following formats and use them in other environment:

- TensorFlow Saved Model format
- XGBoost Booster format

# Integrate BQML in Vertex AI

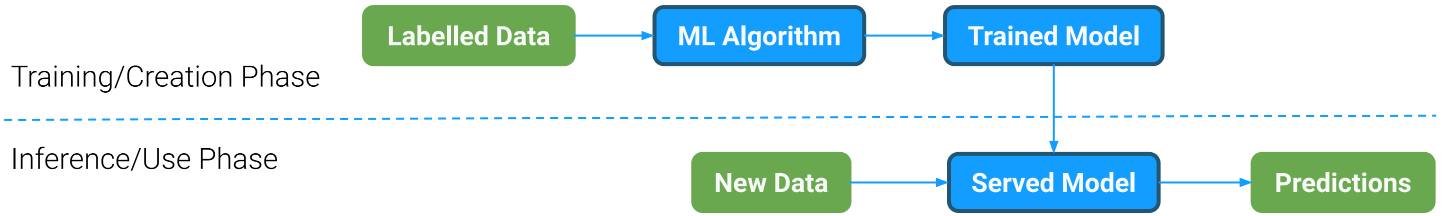
Advanced BQML models are usually trained in Vertex AI,  which is Google's unified ML platform.

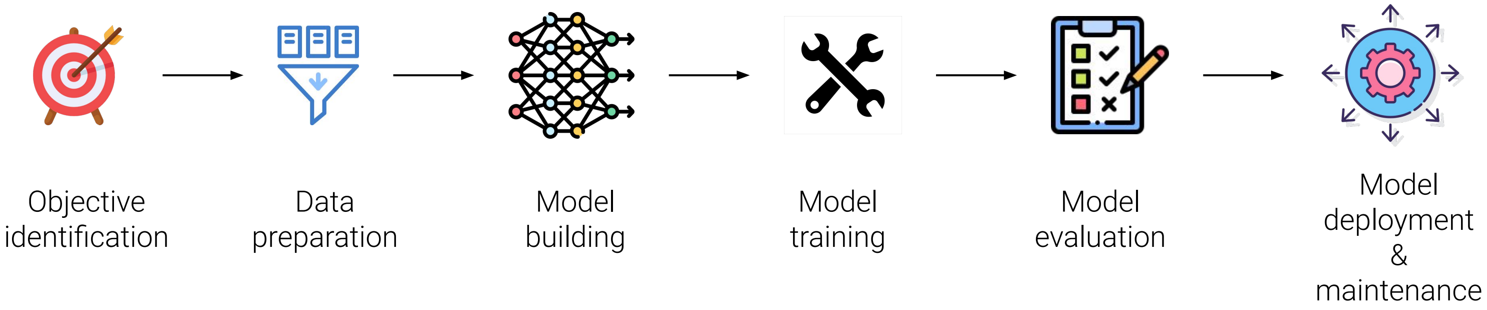
Integrating BQML in Vertex AI gives you online model serving capabilities and allows you to manage BQML models just like any other ML models via Model Registry.




- All models managed in one place
- Model versioning and comparison
- Model monitoring
- Online prediction
- Batch prediction

# Typical machine learning workflows



# Create a BQML model using CREATE MODEL

Label

| Row | species | island | culmen_length | culmen_depth | flipper_length | body_mass_g | sex |
|---|---|---|---|---|---|---|---|
| 1 | Adelie Penguin (Pygoscelis ade... | Dream | 36.6 | 18.4 | 184.0 | 3475.0 | FEMALE |
| 2 | Adelie Penguin (Pygoscelis ade... | Dream | 39.8 | 19.1 | 184.0 | 4650.0 | MALE |
| 3 | Adelie Penguin (Pygoscelis ade... | Dream | 40.9 | 18.9 | 184.0 | 3900.0 | MALE |
| 4 | Chinstrap penguin (Pygoscelis ... | Dream | 46.5 | 17.9 | 192.0 | 3500.0 | FEMALE |
| 5 | Adelie Penguin (Pygoscelis ade... | Dream | 37.3 | 16.8 | 192.0 | 3000.0 | FEMALE |
| 6 | Adelie Penguin (Pygoscelis ade... | Dream | 43.2 | 18.5 | 192.0 | 4100.0 | MALE |
| 7 | Chinstrap penguin (Pygoscelis ... | Dream | 46.9 | 16.6 | 192.0 | 2700.0 | FEMALE |
| 8 | Chinstrap penguin (Pygoscelis ... | Dream | 50.5 | 18.4 | 200.0 | 3400.0 | FEMALE |
| 9 | Chinstrap penguin (Pygoscelis ... | Dream | 49.5 | 19.0 | 200.0 | 3800.0 | MALE |
| 10 | Adelie Penguin (Pygoscelis ade... | Dream | 40.2 | 20.1 | 200.0 | 3975.0 | MALE |

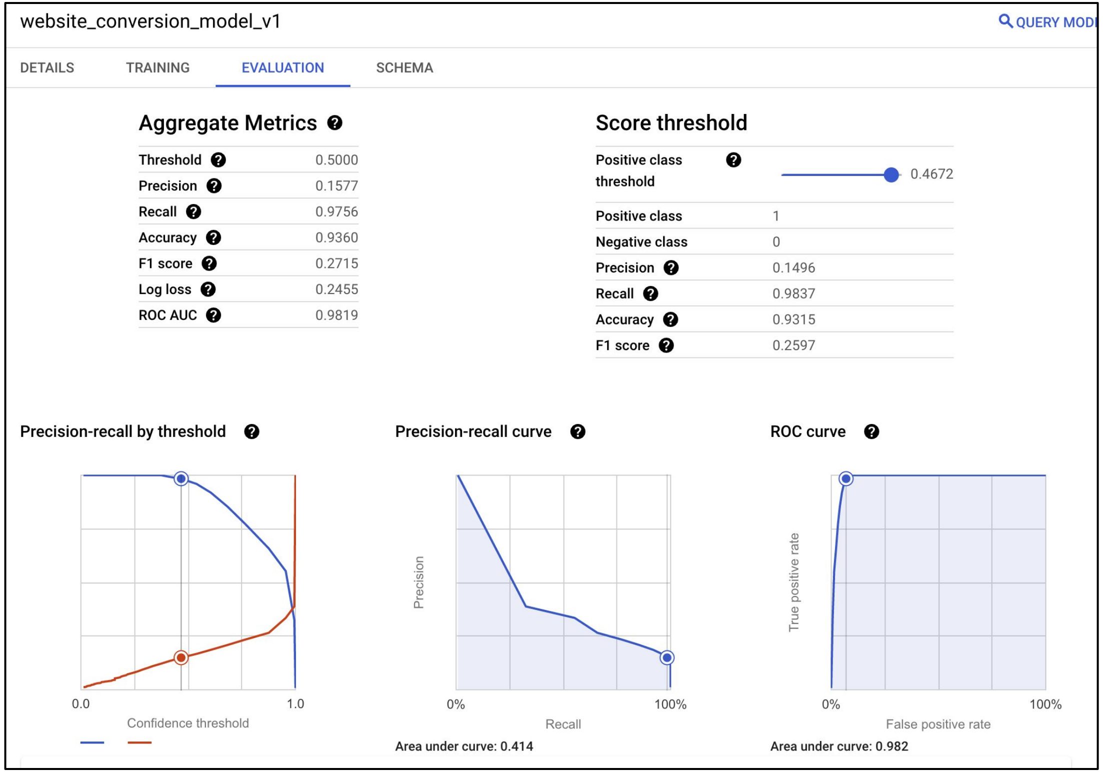
bqml_demo RUN SAVE SHAR

```
#standardSQL
CREATE OR REPLACE MODEL `bqml_demo.penguins_model`
OPTIONS
  (model_type='linear_reg',
  input_label_cols=['body_mass_g']) AS
SELECT
  *
FROM
  `bigquery-public-data.ml_datasets.penguins`
WHERE
  body_mass_g IS NOT NULL
```

# Evaluate a BQML model

Evaluation is often automatically done during model creation in BQML, to early stop the model training process to avoid **overfitting**.

The validation set is used in this process, so it is also known as **validation**.




logickube

# Use a BQML model using ML.PREDICT

Use your trained model to make predictions on new data, e.g., in model production.

model_predict_website_conversion RUN SAVE

```sql
SELECT
  *
FROM
  ML.PREDICT( MODEL `bi-workshop-2023-92764.bqml_demo.website_conver
    (
    SELECT
      IFNULL(device.operatingSystem, "") AS os,
      device.isMobile AS is_mobile,
      IFNULL(geoNetwork.country, "") AS country,
      IFNULL(totals.pageviews, 0) AS pageviews,
      IFNULL(totals.timeOnSite, 0) AS time_on_site
    FROM
      `bigquery-public-data.google_analytics_sample.ga_sessions_*`
    WHERE
      _TABLE_SUFFIX BETWEEN '20170701'
      AND '20170801' ))
```

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION GRAPH PREVIEW

| Row | predicted_label | predict...label | predict... prob | os | is_mobile | country | pageviews | time_on_site |
|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 0.020687784796... | BlackBerry | true | Indonesia | 2 | 20 |
| | | 0 | 0.979312215203... | | | | | |
| 2 | 0 | 1 | 0.017014568063... | BlackBerry | true | Indonesia | 1 | 0 |
| | | 0 | 0.982985431936... | | | | | |
| 3 | 0 | 1 | 0.015617142330... | Samsung | true | India | 1 | 0 |
| | | 0 | 0.984382857669... | | | | | |
| 4 | 0 | 1 | 0.014443974952... | Windows Phone | true | India | 1 | 0 |
| | | 0 | 0.985556025047... | | | | | |
| 5 | 0 | 1 | 0.014605231870... | Windows Phone | true | Romania | 1 | 0 |
| | | 0 | 0.985394768129... | | | | | |

logickube

# Q&A

logickube

# Contact

Shen Liu

s.liu@logickube.com

www.logickube.com

https://www.linkedin.com/company/logickube

logickube