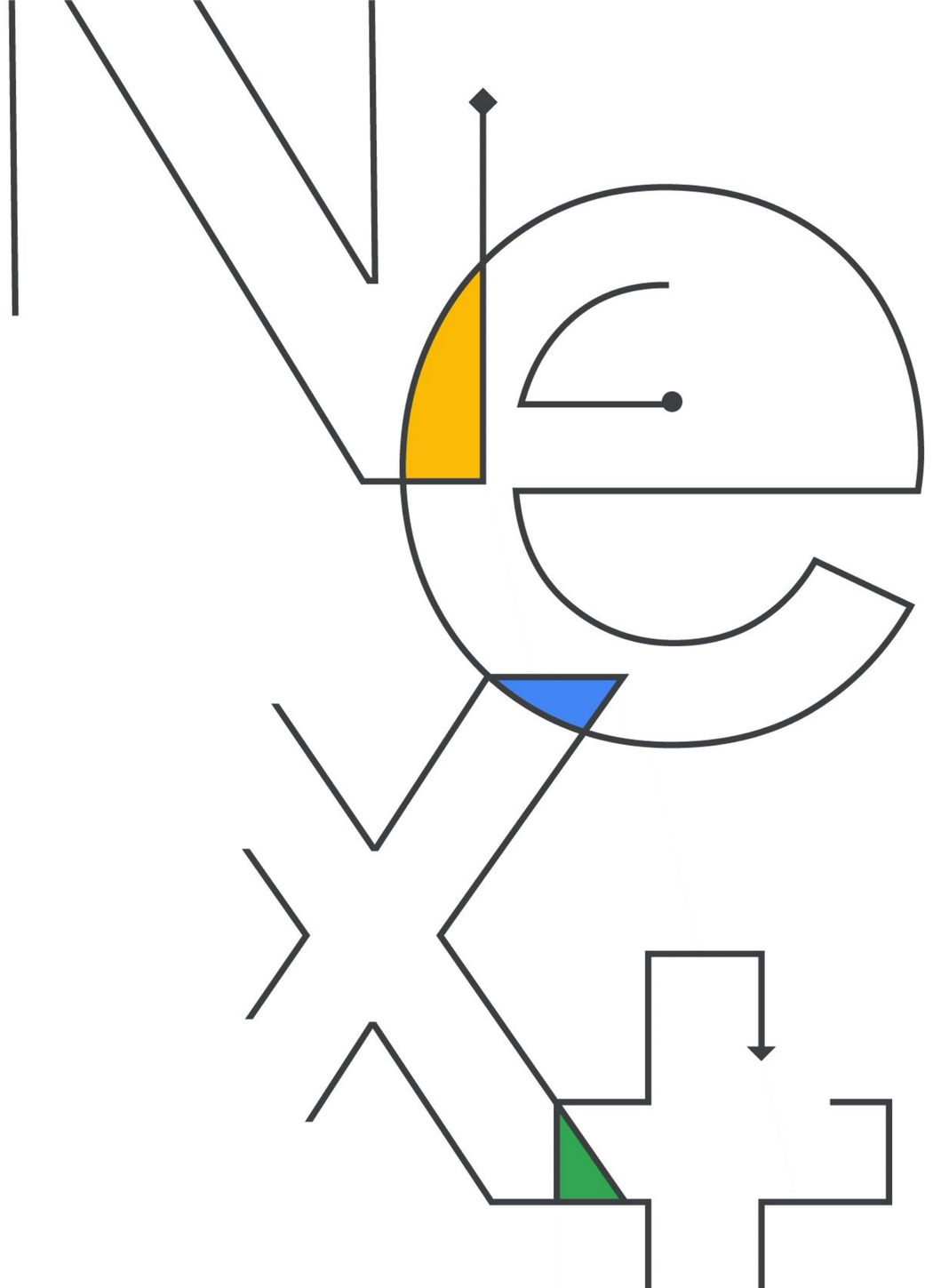
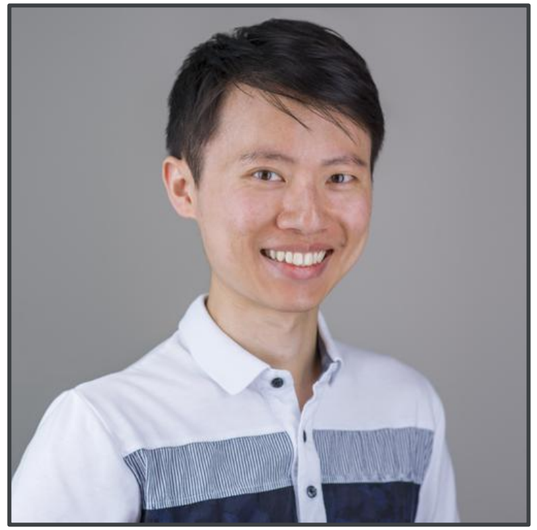
Google Cloud

Next '22

Move from raw data to ML faster with BigQuery and Vertex AI

Oct
11–13





Polong Lin

Developer Advocate,
Data Science &
Machine Learning
Google Cloud

Erwin Huizenga

Developer Advocate
Machine Learning
Google Cloud

From Data to ML

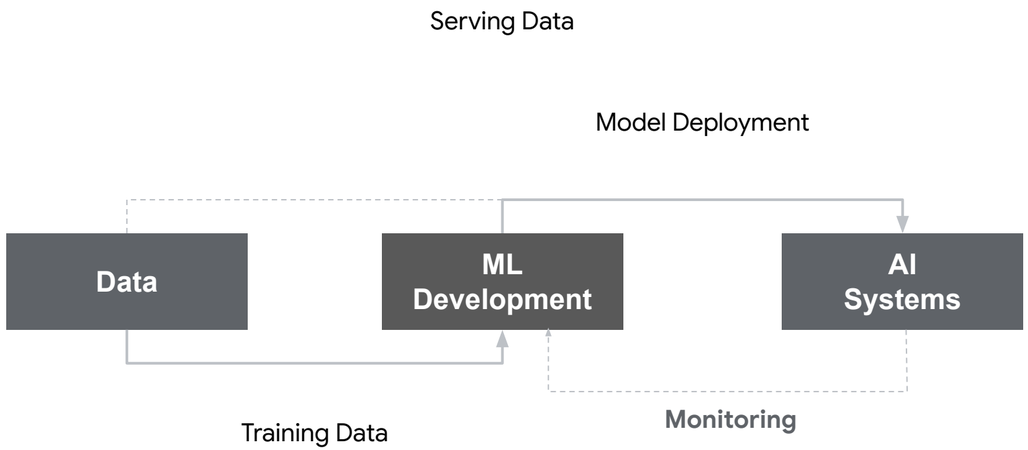
Barriers between Data & Machine Learning continue to slow innovation

Data is big and multi-format

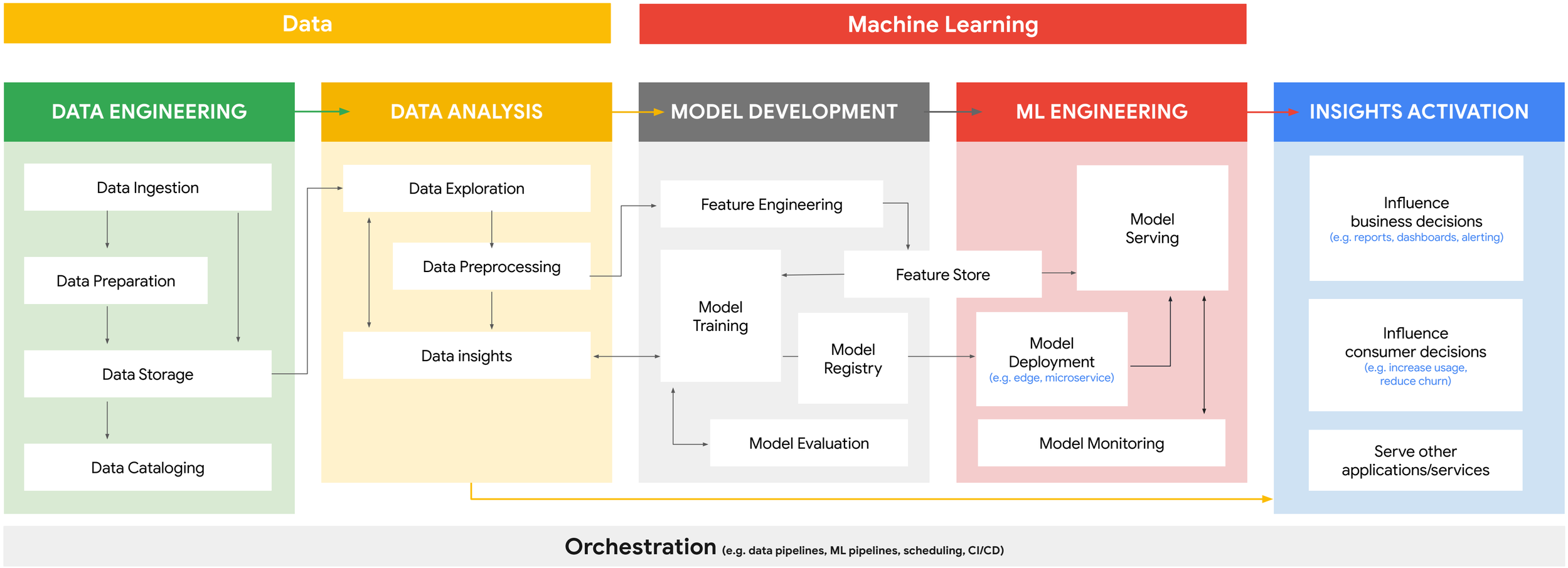
- More data silos
- More data copies
- More tools

AI & ML systems isolated from data

- More integrations
- More security risk
- Unsuccessful ML deployments



From Data to Machine Learning



GCP users benefit from an integrated Data ⇒ Value journey

ANY USER

Data engineers
Clean, useful data

ML engineers
Integrated intelligence

Data scientists
Models that work

Developers
Intelligent apps

Data analysts
Query and analyze

Business users
Insights Everywhere

Consumers
Value

ANY DATA

- Corporate and 3rd Systems
- E-Commerce
- SaaS Applications
- IoT Devices
- Social Media
- Geospatial

Google's Data Cloud

ANY WORKLOAD

DATA ENGINEERING	DATA ANALYSIS	MODEL DEVELOPMENT	ML ENGINEERING	INSIGHTS ACTIVATION
Dataproc Dataflow	BigQuery	Vertex AI & BigQuery ML	Vertex AI	Vertex AI Looker App Engine

Cloud Composer

Vertex AI Pipelines

ANY INNOVATION

- Decision Making
- App Development
- SaaS Applications
- Exception Management
- Operational Intelligence
- Data Monetization

Google Cloud

Crawl

Start with notebooks

I chose Google Cloud from many platforms because, at the time of selection, it had multiple managed services such as Vertex AI, the managed notebooks option, and Vertex AI Training that were useful for AI development.
- SUBARU Corporation

BigLake

BigQuery Storage

Cloud Storage

Vertex AI Workbench

Seamless visual and code-based integrations with analytics and Vertex AI services

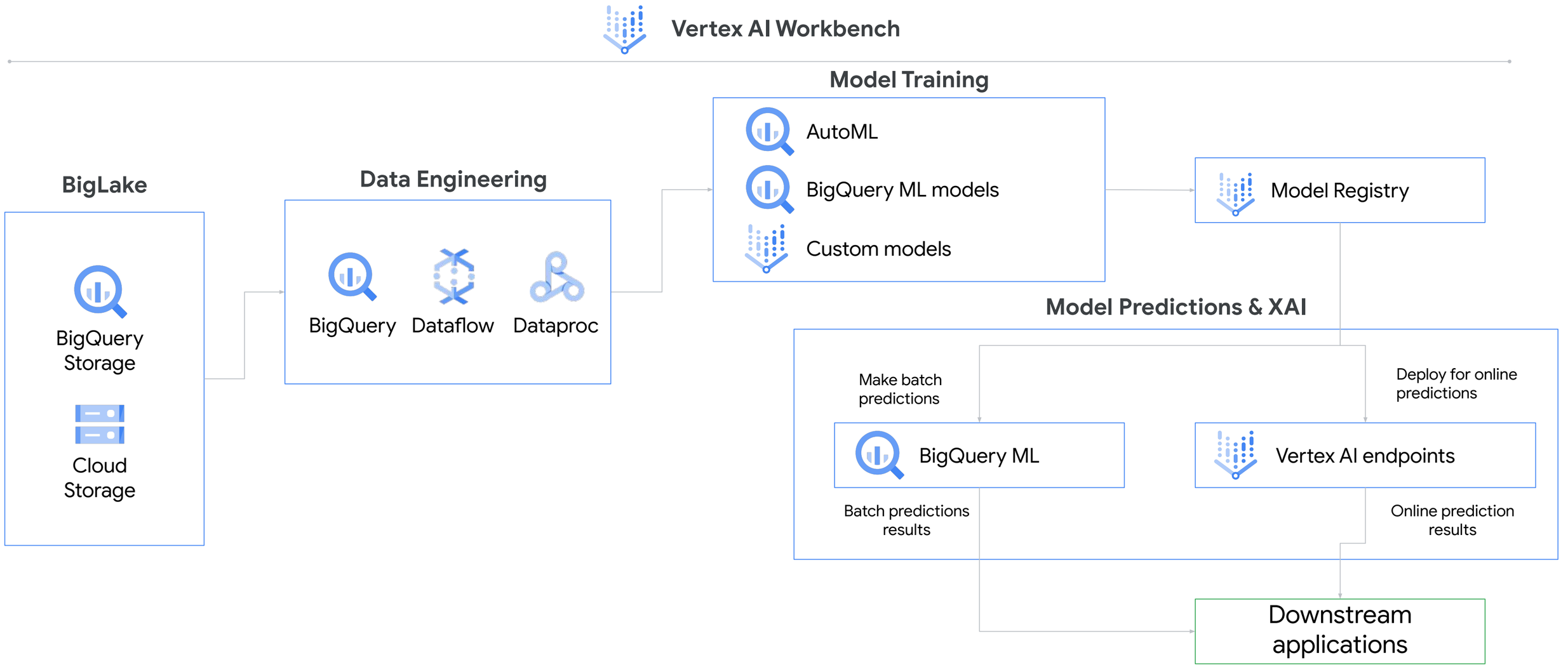
Simplified Data Access: Extensions will seamlessly connect to the entire data estate including BigQuery, Data Lake and Big-Data services

In-place Exploration: Explore data sources using a catalog. Write SQL, Spark queries from a syntax aware, auto-complete enabled Notebook cell

Data Visualization: Integrated, intelligent visualization tools will provide easy insights into data

Walk

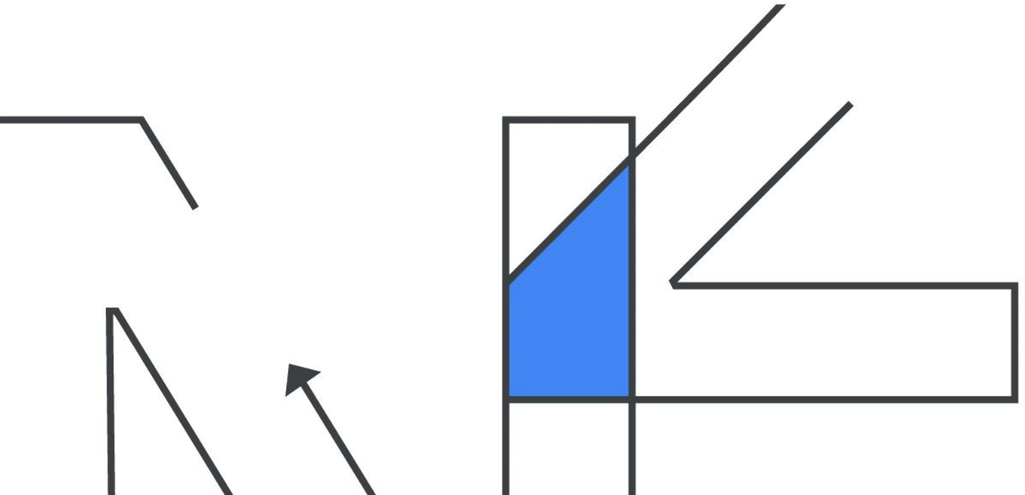
Train and Deploy your models



From BigQuery ML to Vertex AI Model Registry now GA

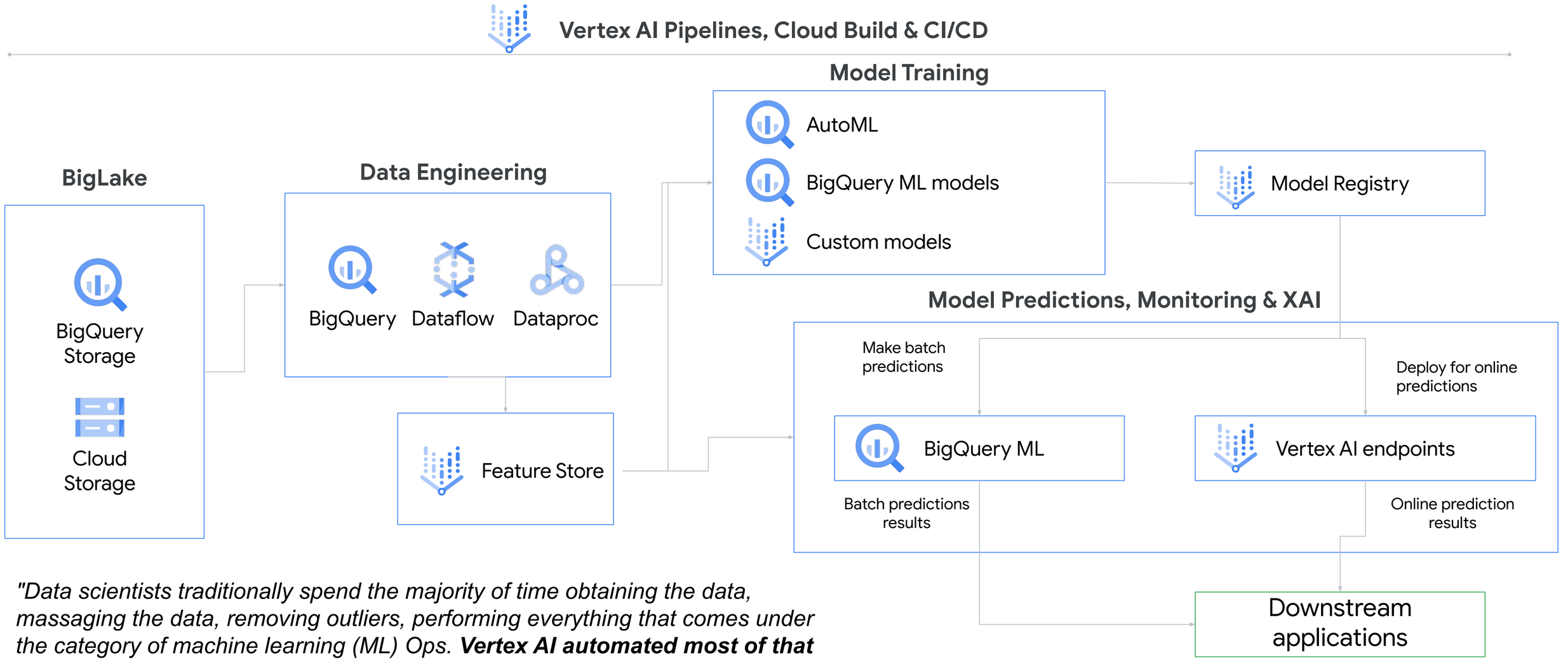
Inspect and manage all of your ML models with **Vertex AI Model Registry**

- All your models (AutoML, BigQuery ML, custom) in one place for **easy reuse** and **tracking** and **model versioning**
- Add **labels** and **aliasing** to your models
- Quick and easy model deployment to Vertex AI Endpoints
- Integration through the Vertex AI Python SDK



Run

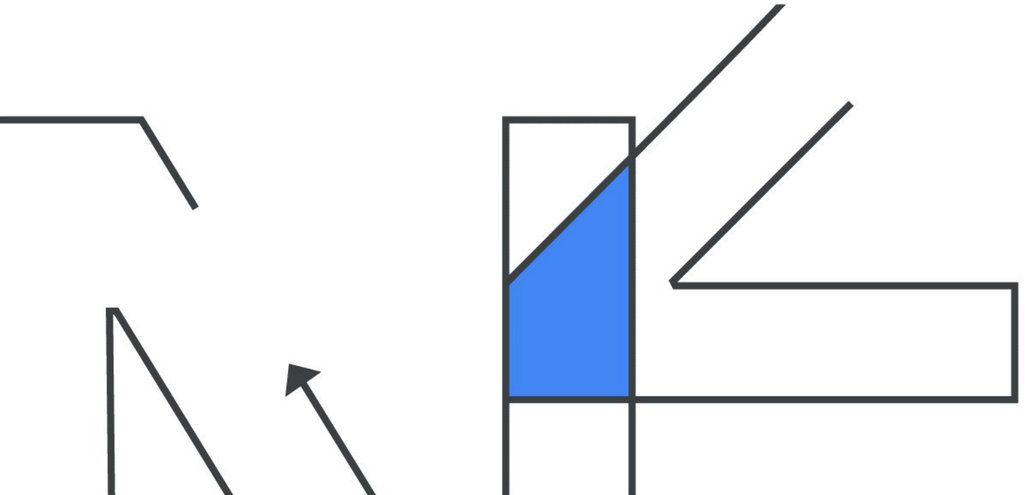
Machine Learning Operations

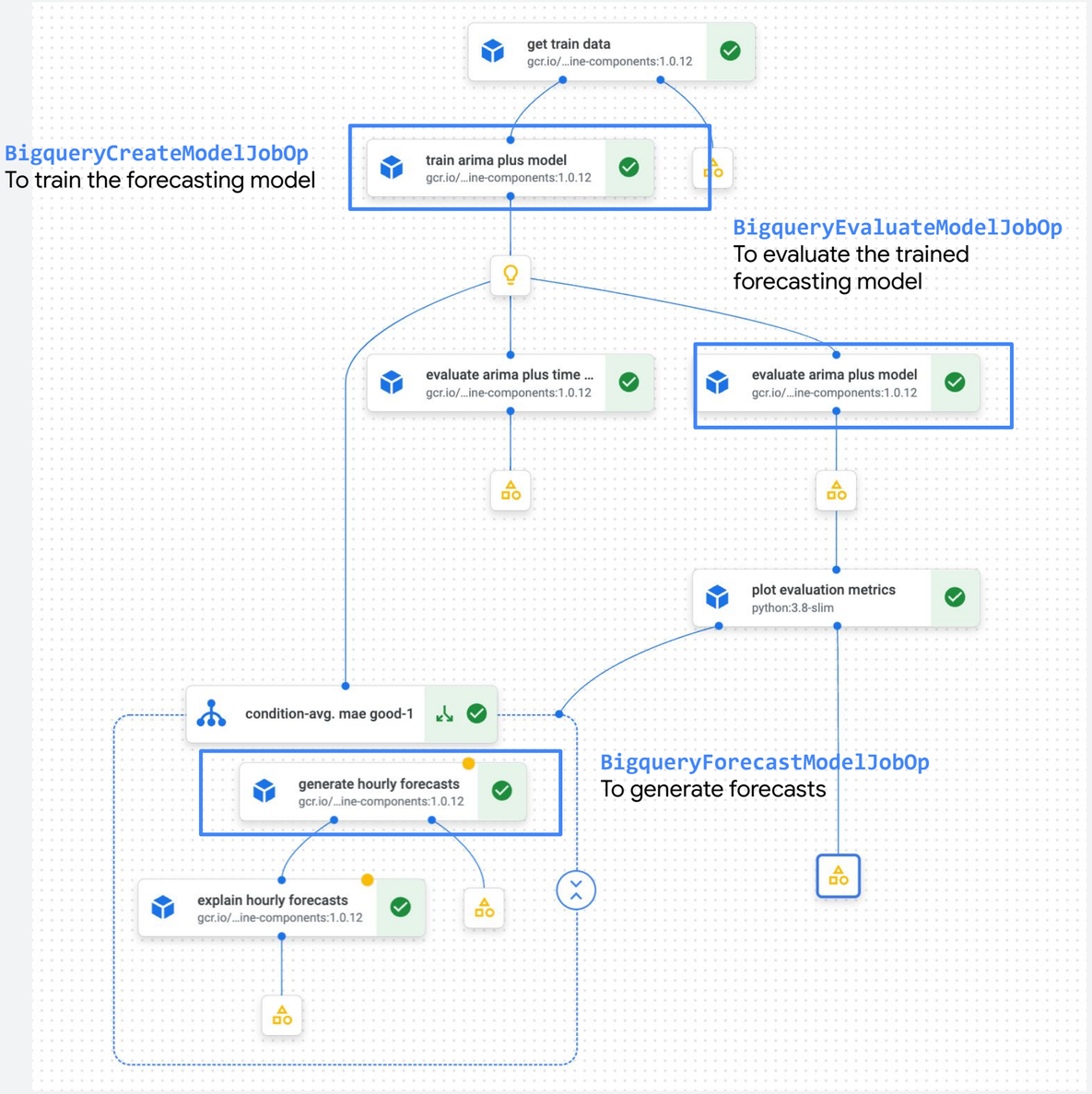


"Data scientists traditionally spend the majority of time obtaining the data, massaging the data, removing outliers, performing everything that comes under the category of machine learning (ML) Ops. ***Vertex AI automated most of that burden, making everything simple for us."*** *- Apna*

BigQuery ML operators for Vertex AI Pipelines now GA

Over 20+ BigQuery ML operators that brings BigQuery ML to Vertex AI Pipelines. This will make it easier to operationalize your BigQuery ML models using the Vertex AI capabilities.

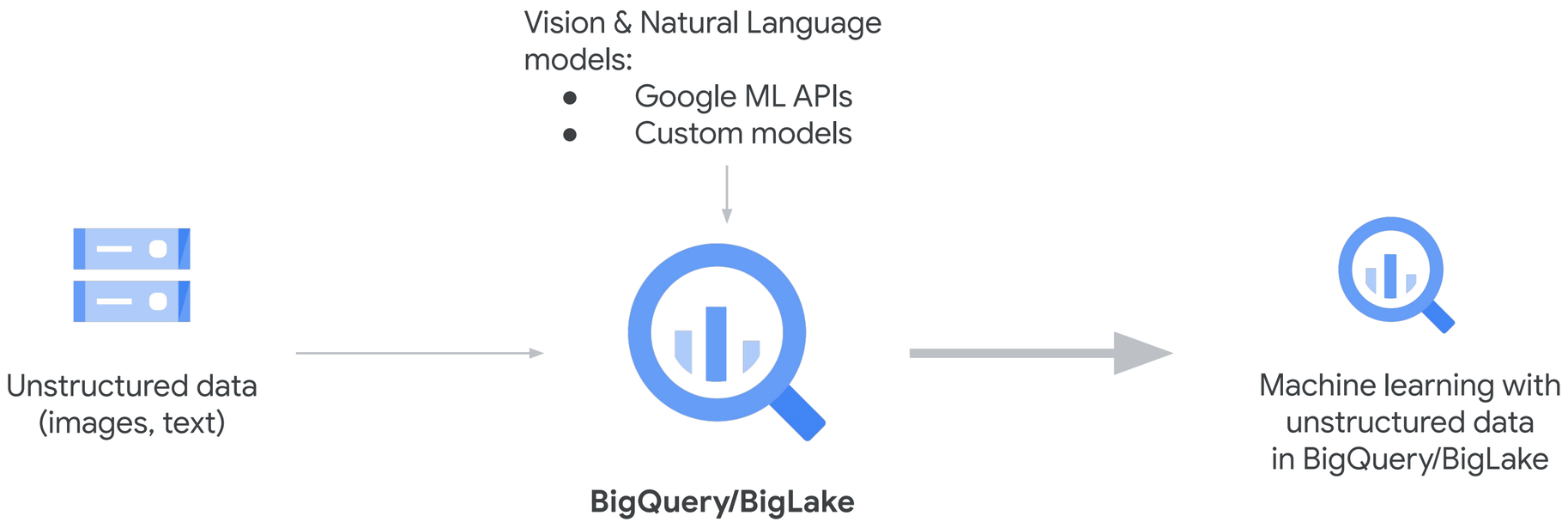




Private preview

Tap into the value of unstructured data with BigQuery ML

"Where unstructured data meets machine learning with BigQuery/BigLake"



BigQuery ML with unstructured data

Private preview

Hypothetical customer

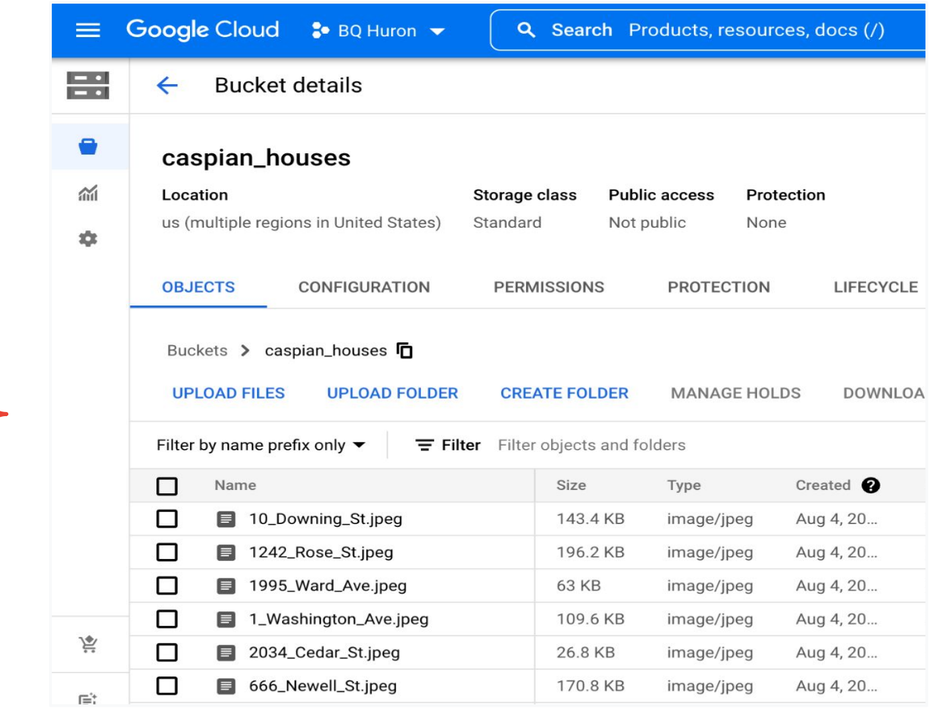
A vacation rental company wants to predict the click through rate (CTR) of the rental homes on their website

Challenges:

- The **current prediction model only looks at structured data**, like home location, square feet, etc
- **Thousands of images are not being used** that could drive insights for the business
- If they want to use images, they have to manage and process images on their own **separately**

Image files represented in an Object Table in BigQuery

Private preview




```
CREATE EXTERNAL TABLE dataset.houses
WITH CONNECTION us.demo_lake
OPTIONS (uris=['gs://caspian_houses/*'],
         object_metadata='DIRECTORY')

SELECT * FROM dataset.houses LIMIT 10
```

uri	generation	content_type	size
gs://caspian_houses/10_Downing_St.jpeg	1659659941032822	image/jpeg	146843
gs://caspian_houses/1242_Rose_St.jpeg	1659659941932150	image/jpeg	200905
gs://caspian_houses/1995_Ward_Ave.jpeg	1659659942122696	image/jpeg	64551
gs://caspian_houses/1_Washington_Ave.jpeg	1659659940828040	image/jpeg	112249
gs://caspian_houses/2034_Cedar_St.jpeg	1659659942231821	image/jpeg	27476
gs://caspian_houses/666_Newell_St.jpeg	1659659941129893	image/jpeg	174938
gs://caspian_houses/823_University_Ave.jpeg	1659659941526406	image/jpeg	78473
gs://caspian_houses/892_Fulton_Rd.jpeg	1659659941628295	image/jpeg	62013

BigQuery ML to extract latent representations of house images

Import the imagenet model into BigQuery

```
CREATE MODEL dataset.imagefv
OPTIONS (model_type='TENSORFLOW',
         color_space='RGB',
         model_path='gs://imagenet/*')
```

Use ML.PREDICT to extract the latent representations of images

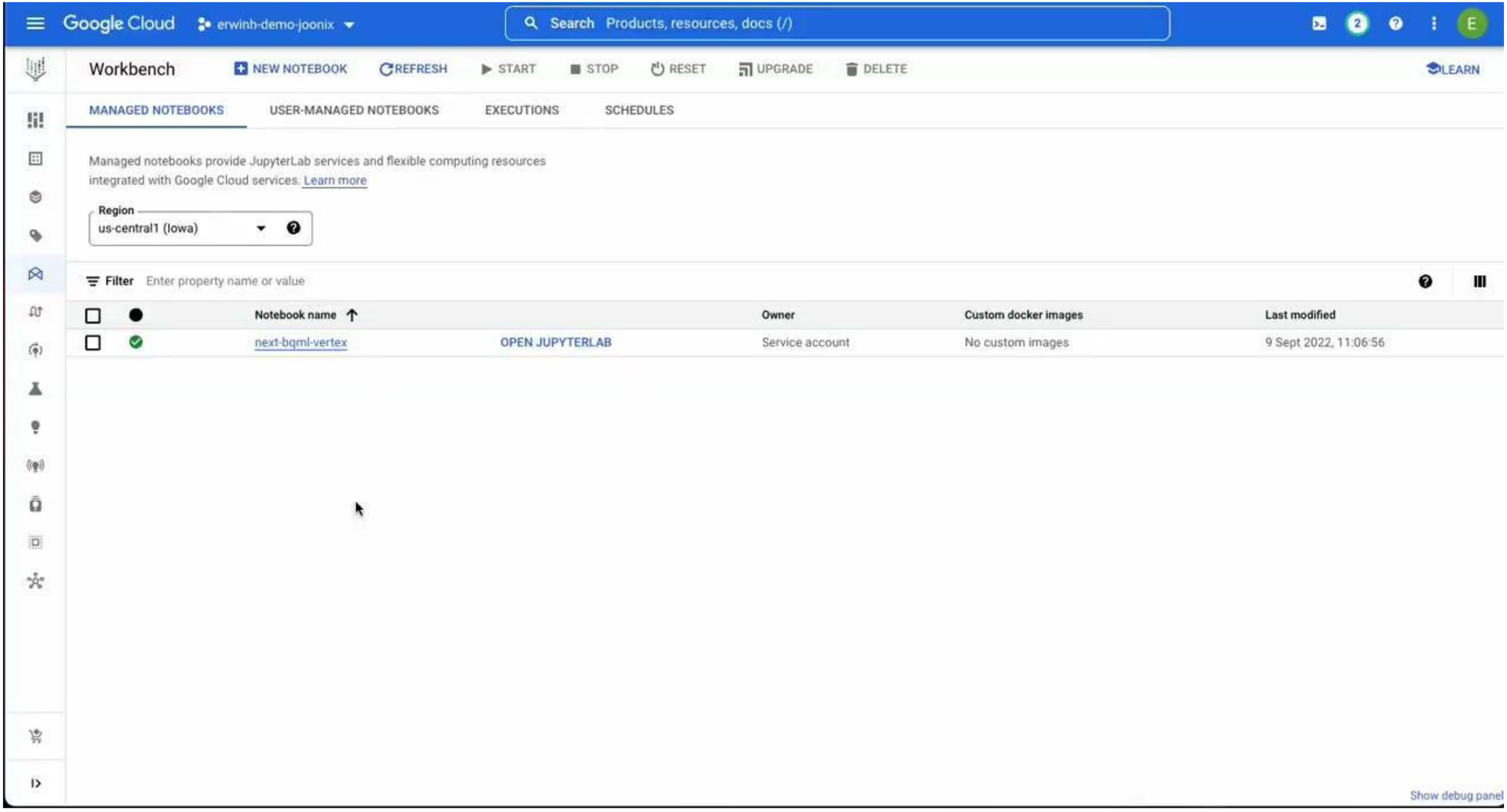
```
SELECT uri, feature_vec FROM
  ML.PREDICT(
    MODEL dataset.imagefv,
    TABLE dataset.houses)
```

uri	feature_vec
gs://caspian_houses/10_Downing_St.jpeg	["-0.6858665347099304","0.03352648764848709","1.7552795410
gs://caspian_houses/823_University_Ave.jpeg	["-0.4768984317779541","0.8592596650123596","1.6156368255
gs://caspian_houses/892_Fulton_Rd.jpeg	["0.36786502599716187","1.6773165464401245","0.6930949687
gs://caspian_houses/666_Newell_St.jpeg	["0.020243331789970398","0.24682791531085968","1.6812396049
gs://caspian_houses/2034_Cedar_St.jpeg	["-1.0804210901260376","1.1235047578811646","1.5802049
gs://caspian_houses/1242_Rose_St.jpeg	["-0.2593004107475281","-0.5539816617965698","2.4647274
gs://caspian_houses/1_Washington_Ave.jpeg	["-0.9279747605323792","-1.5976200103759766","2.9774522781
gs://caspian_houses/1995_Ward_Ave.jpeg	["-0.33510828018188477","0.9801657199859619","0.814768850

The results are then used to join the customer-house click record data as the model training dataset

Raw data to ML faster with BigQuery and Vertex AI

Demo


Google Cloud
erwinh-demo-joonix
Search Products, resources, docs (/)
Workbench
NEW NOTEBOOK
REFRESH
START
STOP
RESET
UPGRADE
DELETE
LEARN
MANAGED NOTEBOOKS
USER-MANAGED NOTEBOOKS
EXECUTIONS
SCHEDULES
Managed notebooks provide JupyterLab services and flexible computing resources integrated with Google Cloud services. Learn more
Region
us-central1 (Iowa)
Filter
Enter property name or value
Notebook name
Owner
Custom docker images
Last modified
next-bqml-vertex
OPEN JUPYTERLAB
Service account
No custom images
9 Sept 2022, 11:06:56
Show debug panel

Want to learn more?

goo.gle/data-to-ai-workshop

Thank you

Google Cloud

Next '22