

Google Cloud

Partner Certification Academy



# Professional Machine Learning Engineer

pls-academy-pmle-student-slides-2-2403

The information in this presentation is classified:

## **Google confidential & proprietary**

⚠ This presentation is shared with you under NDA.

- Do **not** record or take screenshots of this presentation.
- Do **not** share or otherwise distribute the information in this presentation with anyone **inside** or **outside** of your organization.

Thank you!



Google Cloud

# Source Materials

Some of this program's content has been sourced from the following resources:

- [Google Cloud certification site](#)
- [Google Cloud documentation](#)
- [Google Cloud console](#)
- [Google Cloud courses and workshops](#)
- [Google Cloud white papers](#)
- [Google Cloud Blog](#)
- [Google Cloud YouTube channel](#)
- [Google Cloud samples](#)
- [Google codelabs](#)
- [Google Cloud partner-exclusive resources](#)

 This material is shared with you under the terms of your Google Cloud Partner **Non-Disclosure Agreement**.



## Google Cloud Skills Boost for Partners

- [Professional Machine Learning Engineer Certification](#)
- [Cloud Skills Boost for Partners Professional Machine Learning Engineer Learning Path](#)
- [Partner Learning Services Instructor-Led PMLE Curriculum](#)

## Google Cloud Partner Advantage

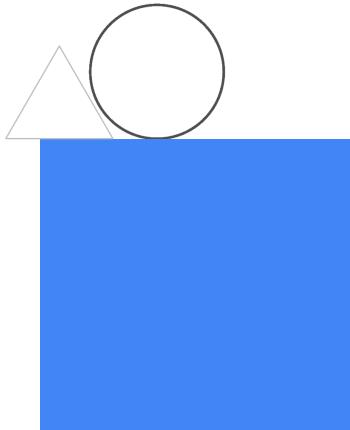
- [Best practices for implementing machine learning on Google Cloud](#)
- [Artificial Intelligence](#)
- [End-to-End MLOps Go-to-Market Kit](#)

# Session Logistics

- When you have a question, please:
  - Click the Raise hand button in Google Meet.
  - Or add your question to the Q&A section of Google Meet.
  - Please note that answers may be deferred until the end of the session.
- These slides are available in the Student Lecture section of your Qwiklabs classroom.
- The session is **not recorded**.
- Google Meet does not have persistent chat.
  - If you get disconnected, you will lose the chat history.
  - Please copy any important URLs to a local text file as they appear in the chat.

# Google Cloud Partner Learning Programs

- Partner Certification Academy
- Partner Delivery Readiness Index (DRI)
- Cloud Skills Boost for Partners
- Partner Advantage



## PARTNER CERTIFICATION ACADEMY

# Professional Machine Learning Engineer



A Professional Machine Learning Engineer builds, evaluates, productionizes, and optimizes ML models by using Google Cloud technologies and knowledge of proven models and techniques. The ML Engineer:

- handles large, complex datasets and creates repeatable, reusable code.
- considers responsible AI and fairness throughout the ML model development process, and collaborates closely with other job roles to ensure long-term success of ML-based applications.
- has strong programming skills and experience with data platforms and distributed data processing tools.
- is proficient in the areas of model architecture, data and ML pipeline creation, and metrics interpretation.
- is familiar with foundational concepts of MLOps, application development, infrastructure management, data engineering, and data governance.
- makes ML accessible and enables teams across the organization.

By training, retraining, deploying, scheduling, monitoring, and improving models, the ML Engineer designs and creates scalable, performant solutions.

**Recommended candidate:**

- Has in-depth experience setting up cloud environments for an organization
- Has experience deploying services and solutions based on business requirements

Google Cloud

## PARTNER CERTIFICATION ACADEMY

# Professional Machine Learning Engineer



A Professional Machine Learning Engineer builds, evaluates, productionizes, and optimizes ML models by using Google Cloud technologies and knowledge of proven models and techniques. The ML Engineer:

- handles large, complex datasets and creates repeatable, reusable code.
- considers responsible AI and fairness throughout the ML model development process, and collaborates closely with other job roles to ensure long-term success of ML-based applications.
- has strong programming skills and experience with data platforms and distributed data processing tools.
- is proficient in the areas of model architecture, data and ML pipeline creation, and metrics interpretation.
- is familiar with foundational concepts of MLOps, application development, infrastructure management, data engineering, and data governance.
- makes ML accessible and enables teams across the organization.

By training, retraining, deploying, scheduling, monitoring, and improving models, the ML Engineer designs and creates scalable, performant solutions.

**Recommended candidate:**

- Has in-depth experience setting up cloud environments for an organization
- Has experience deploying services and solutions based on business requirements

Google Cloud

## Learner Commitment

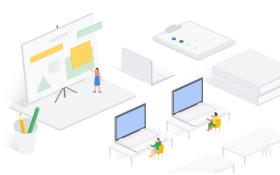
Each week, learners are to complete the learning path's course content, Cloud Skills Boost for Partner Quests/Challenge Labs and material that the mentor has recommended that will support learning.

- **Workshop Day:** Meet for the cohort's weekly 'general session'. ( $\approx 2$  hours)
- **During the week:** Complete the week's course, perform hands-on labs, review any additional material suggested material for the week. ( $\approx 8 - 16$  hours)
- **Important:** Learners must allocate time between each weekly session to study and familiarize themselves with any prerequisite knowledge they may lack. It is also recommended that learners complete the next week's course prior to the scheduled workshop.

## Path to Service Excellence



Certification



Advanced Solutions Training

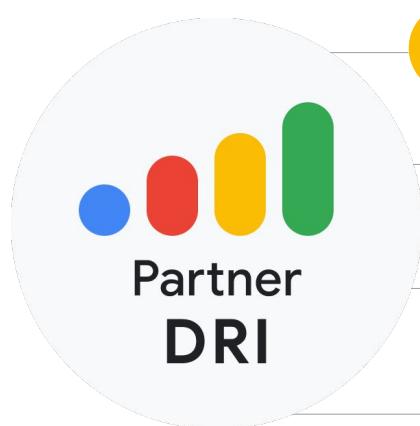


Delivery Readiness Index

Google Cloud

Certification is just one step on your professional journey. Google Cloud also offers our partners access to advanced solutions training, and a new quality-focused program called Delivery Readiness Index (DRI) to help you achieve service excellence with your customers.

## Benchmark your skills with DRI



### Assess: Partner Proficiency and Delivery Capability

Benchmark Partner individuals, project teams and practices GCP capabilities



### Analyze: Individual Partner Consultants' GCP Readiness

Showcase Partner individuals GCP knowledge, skills, and experience



### Advise: Google Assurance for Partner Delivery

Packaged offerings to bridge specific capability gaps



### Action: Tailored L&D Plan for Account Based Enablement

Personalized learning & development recommendations per individual consultant

Google Cloud

DRI helps to benchmark partner proficiency and capability at any point during the customer journey however should be used primarily as a lead measure to predict and prepare for partner delivery success.

DRI assesses and analyzes Partner Consultant GCP proficiency by creating a DRI Profile inclusive of their GCP knowledge, skills, and experience.

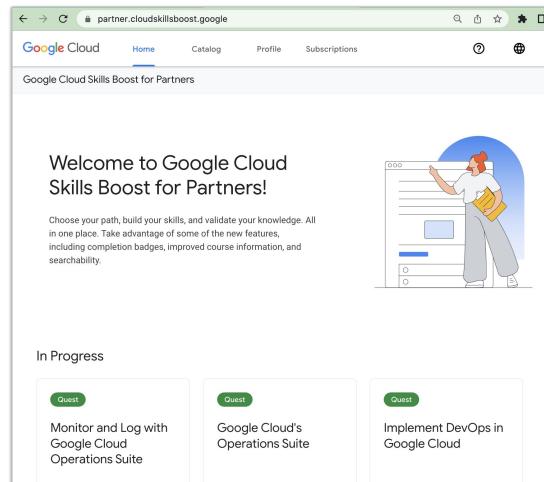
With the DRI insights, we can prescriptively advise the partner project team on the ground and bridge niche capability gaps.

DRI also takes action. For partner consultants, DRI generates a tailored L&D plan that prescribes personalized learning, training, and skill development to build GCP proficiency.

# Google Cloud Skills Boost for Partners

<https://partner.cloudskillsboost.google/>

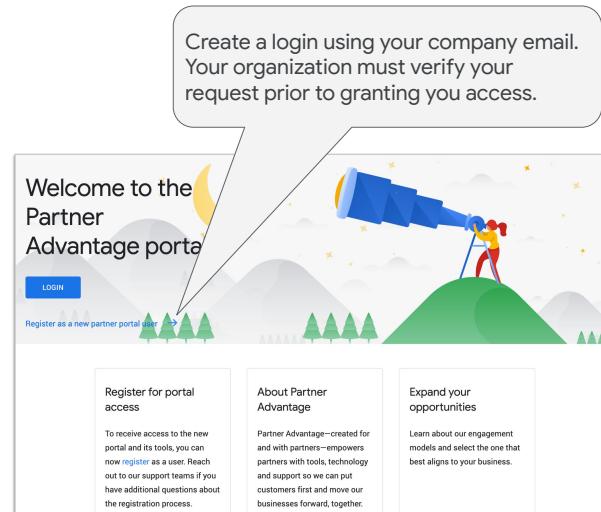
- On-demand course content
- Hands-on labs
- Skill Badges
- **FREE** to Google Cloud Partners!



Google Cloud

# Google Cloud Partner Advantage

- Resources for Google Cloud partner organizations:
  - Recent announcements
  - Solutions/role-based training
  - Live/pre-recorded webinars on various topics
    - [Partner Advantage Live Webinars](#)
- Complements the certification self-study material presented on Google Cloud Skills Boost for Partners
- Helpful Links:
  - [Getting started on Partner Advantage](#)
  - [Join Partner Advantage](#)
  - [Get help accessing Partner Advantage](#)



<https://www.partneradvantage.googlecloud.com/>

Google Cloud

## The getting started link:

<https://support.google.com/googlecloud/topic/9198654#zippy=%22Getting+Started+%26+User+Guides%22>

Note the top section, “**Getting Started & User Guides**” and two key documents → Direct Partners to this if they need to enroll into Partner Advantage

1. Logging in to the Partner Advantage Portal - Quick Reference Guide
2. Enrolling in the Partner Advantage Program - Quick Reference Guide

## Focus from this point on:

### Some context on enrolling in PA:

Access to Partner Portal is given in 2 ways

- Partner Admin Led: Partner Administrator at Partner Company can set up users
- User Led: User can go through Self Registration
  - [https://www.partneradvantage.googlecloud.com/GCPPRM/s/partneradvantageportal/login?language=en\\_US](https://www.partneradvantage.googlecloud.com/GCPPRM/s/partneradvantageportal/login?language=en_US)
  - Or directly to the User Registration Form,  
[https://www.partneradvantage.googlecloud.com/GCPPRM/s/partnerselfregistration?language=en\\_US](https://www.partneradvantage.googlecloud.com/GCPPRM/s/partnerselfregistration?language=en_US)

### Please Note

- After a user self-registers, they receive an email that essentially states:

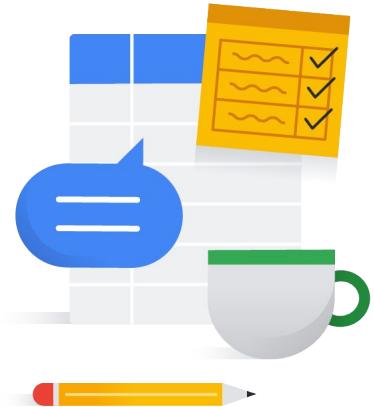
- "Hi {Partner Name}, you are one step away from joining the Google Cloud Partner Advantage Community. Please click to continue with the user registration process. See you in the cloud, The Partner Advantage Team
- Once registered, they can access limited content until their **Partner Administrator approves the user**
- Their Partner Administrator also receive an email notifying them that a member of their organization has registered themselves on their organization's Google Cloud Partner Advantage account.
  - It also states that this user has limited access to the portal
  - They are provided instructions on how to review and provision the appropriate access for the user that has registered
- Once their admin approves the user, they receive an email that states:
  - Hi {User Name}, Your Partner Administrator has updated your access to the Google Cloud Partner Advantage portal. You have been granted edit access to additional account information on the portal on behalf of your organization to help build your business. For additional access needs, please work with your Partner Administrator. See you in the cloud, The Partner Advantage Team

The net takeaway is, on the Support Page (the first link on this slide) [Google Cloud Partner Advantage Support](#), there's a section "**Issue accessing Partner Advantage Portal? Click here for troubleshooting steps**"

- The source of their issue can be related to the different items shown
- Additionally, there's a Partner Administrator / Partner Adminstrator Team at their partner organization that has to approve their access.. Until that step is completed, they will have access issues/limitation. They will need to identify who this person or team is at their organization

## Program issues or concerns?

- Problems with **accessing** Cloud Skills Boost for Partners
  - [cloud-partner-training@google.com](mailto:cloud-partner-training@google.com)
- Problems with **a lab** (locked out, etc.)
  - [support@qwiklabs.com](mailto:support@qwiklabs.com)
- Problems with accessing Partner Advantage
  - <https://support.google.com/googlecloud/topic/9198654>



Google Cloud

- Problems with accessing **Cloud Skills Boost for Partners**
  - [cloud-partner-training@google.com](mailto:cloud-partner-training@google.com)
- Problems with **a lab** (locked out, etc.)
  - [support@qwiklab.com](mailto:support@qwiklab.com)
- Problems with accessing **Partner Advantage**
  - <https://support.google.com/googlecloud/topic/9198654>

## Module 2

# Architecting low-code ML solutions

Google Cloud

# Module Agenda

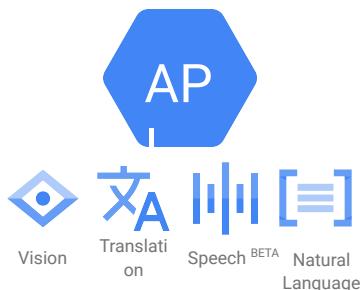
- 01** Developing ML models by using BigQuery ML
- 02** Building AI solutions by using ML APIs
- 03** Training models by using AutoML



Developing ML  
models by using  
BigQuery ML

Google Cloud

# BigQuery ML is a way to train custom models on data already in BigQuery



Pre-trained models

very common to **enrich your data** with pre-trained models, to take advantage of unstructured data



Build your own model

Developers  
Data Analysts  
ML engineers

Google Cloud

## Baselines are important; It helps to know what error metric is “reasonable” and/or “good” for the problem

A baseline helps you set a goal for a good value for the error metric. Often a simple heuristic rule can function as a good benchmark.

What's a good benchmark for the taxi fare prediction?



Google Cloud

We will begin with a baseline model to help us set a goal for a good value for the error metric. Baselines are important to know because they help us determine a reasonable or good metric for the problem.

## BigQuery is a petabyte-scale fully-managed data warehouse

1. It's serverless
2. Flexible pricing model
3. Data encryption and security
4. Geospatial data types and functions
5. Foundation for BI and AI



BigQuery

Google Cloud

BigQuery is designed to be an easy to use data warehouse where you can focus on writing SQL statements on small or large datasets without worrying about infrastructure. If you've never written SQL before I'll also provide resources and labs to get you up to speed as we go.

## BigQuery is two services in one



- 1. Fast SQL Query Engine
- 2. Managed storage for datasets

Google Cloud

It's time to get a little more technical. So how does BigQuery actually work? Well it's actually two services in one as we hinted at earlier.

It's both a fast SQL query engine and also a fully-managed storage layer for loading and storing your datasets.

Keep in mind that it's a "serverless" service, meaning that it is fully managed. Let's take a look at what Google Cloud is managing for you as part of the BigQuery service.

## BigQuery supports standard SQL queries for analysis

```
#standardSQL
SELECT
    COUNT(*) AS total_trips
FROM
    `bigquery-public-data.san_francisco_bikeshare.bikeshare_trips`
```

Row	total_trips
1	1947419

Google Cloud

In the last lesson we showed this SQL statement ran using the query engine. As any data analyst will tell you, exploring your dataset with SQL is your first step in uncovering those hidden insights.

Here we're counting the total number of trips taken on the public dataset for San Francisco Bike Share trips. Let's explore this public dataset in greater detail with a demo.

Example SF Bike Share query

[https://console.cloud.google.com/bigquery?sq=1057666841514:6e9100c9fbaf421bac62a862efbe934c&\\_ga=2.65081326.-156181847.1534452292](https://console.cloud.google.com/bigquery?sq=1057666841514:6e9100c9fbaf421bac62a862efbe934c&_ga=2.65081326.-156181847.1534452292)

## Common SQL operations include deduplication and cleansing

```
# Top 5 stations for casual bike share users
SELECT
    start_station_name,
    COUNT(*) AS total_trips
FROM
    `bigquery-public-data.san_francisco_bikeshare.bikeshare_trips`  

    WHERE c_subscription_type = 'Customer'
    AND member_birth_year IS NOT NULL

GROUP BY start_station_name
ORDER BY total_trips DESC
LIMIT 5
```

Google Cloud

As we saw in the demo, a lot of the time you aren't sure of the quality of the data in your dataset so you can use SQL to explore and filter for anomalies. Here we filtered out customer records who did not provide a birth date by setting member\_birth\_year IS NOT NULL as a filter. I'll also provide a SQL syntax reference that you can bookmark and refer to. While I have been working with SQL awhile I certainly don't have all the functions and clauses committed to memory yet and am constantly on Google searching for syntax.

<https://www.fordgobike.com/system-data>

# Cell Magic from a Jupyter Notebook

```
%%bq --project $PROJECT_ID  
SELECT  
    data_split,  
    COUNT(*) AS n_examples  
FROM bqmlga4.ml_features  
GROUP BY data_split
```

Job ID 9188a34f-ddda-4038-

bb70-bcd6df720725

successfully executed: 100%

Downloading:

100%

	data_split	n_examples
0	TEST	799
1	TRAIN	6386
2	EVAL	846

This is built-in with Vertex AI Workbench notebooks, or if calling from a non-Workbench notebook, load this functionality with:

```
%load_ext google.cloud.bigquery
```

Google Cloud

<https://cloud.google.com/python/docs/reference/bigquery/latest/magics>

## BigQuery is two services in one

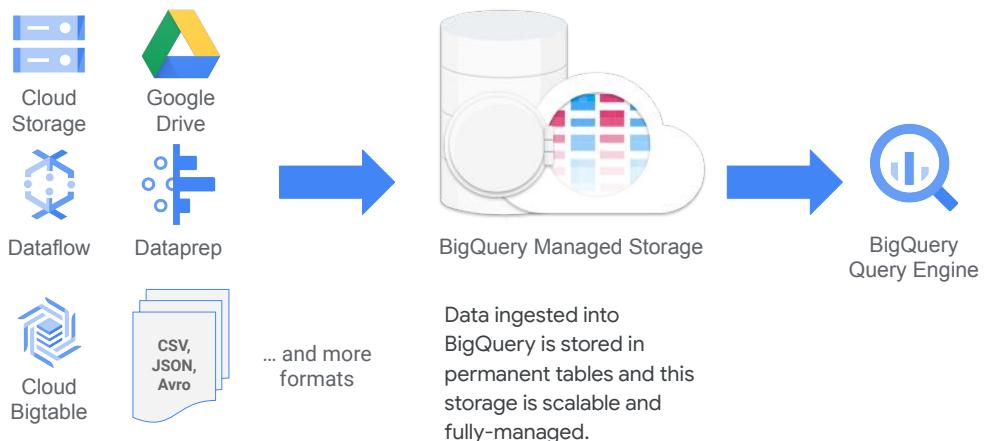


- 1. Fast SQL Query Engine
- 2. Managed storage for datasets

Google Cloud

In addition to super fast query execution times, BigQuery also manages the storage and metadata for your datasets.

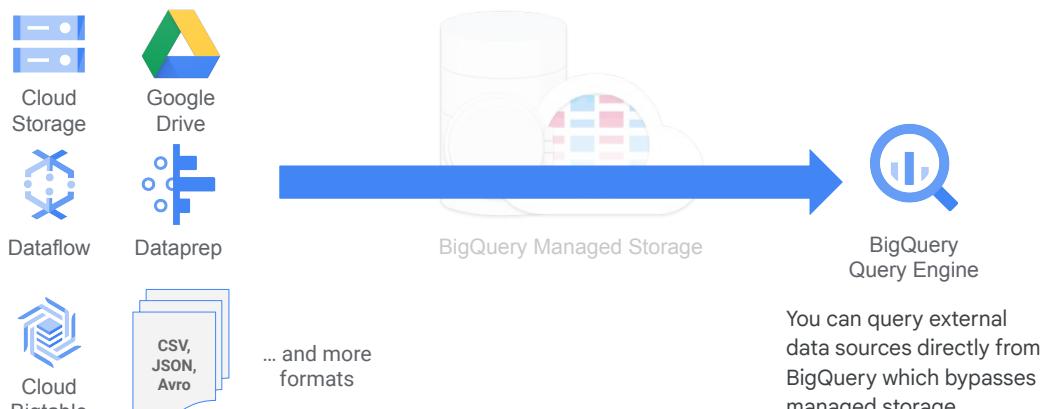
## Use native BigQuery storage for the highest performance



Google Cloud

As we mentioned earlier, BigQuery can ingest datasets from a variety of different formats. Once inside BigQuery native storage, it is fully managed by the BigQuery team here at Google and is automatically replicated, backed up, and setup to autoscale. You can even recover recently deleted dataset tables within a certain period too.

## BigQuery can query external (aka federated) data sources in Cloud Storage and Drive directly



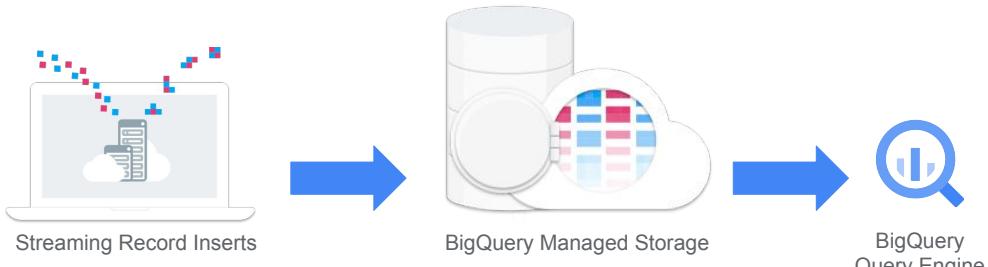
Google Cloud

You also have the option of querying external data sources directly and bypassing BigQuery managed storage. That means if you have a raw CSV file in Cloud Storage or a Google Sheet you can write a query against it without having to ingest the data into BigQuery first.

A common use case for this is having an external data source that is small and constantly changes like say a price-list for commodities that another team maintains.

But there are a few reasons why you should not do this. The first is that data consistency is not guaranteed. If the data source changes mid-flight, BigQuery doesn't guarantee that those updates were captured. If you're concerned about that, consider building a streaming data pipeline into BigQuery with Dataflow which we will cover in the next module.

## Stream records into BigQuery through the API



Streaming Record Inserts

Streaming data allows you to query data without waiting for a full batch load.

BigQuery Managed Storage

BigQuery Query Engine

Google Cloud

In addition to ingesting datasets as a batch, like uploading a CSV, you can also stream records into BigQuery using the API.

Note that there are a few quota restrictions that you should be aware of. The max row size is 1 MB and the maximum throughput is 100,000 records per second per project.

If you need higher throughput (on the order of millions of records per second) for use cases like application logging or real-time event consider using Cloud Bigtable instead.

<https://cloud.google.com/bigquery/streaming-data-into-bigquery>

## BigQuery supports streaming data ingestion directly or via Dataflow pipelines



BigQuery



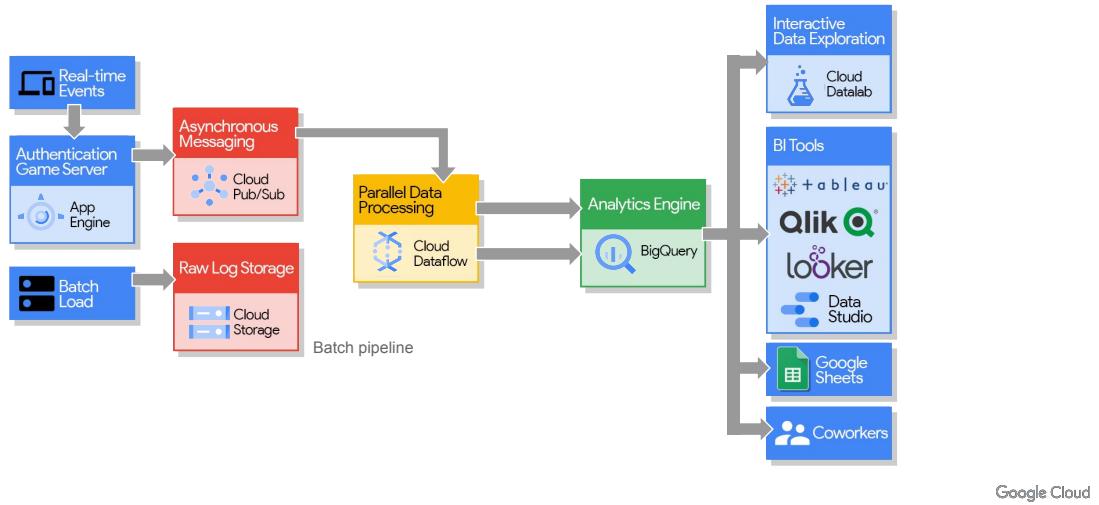
Dataflow

Google Cloud

Before you start streaming thousands of records into BigQuery using the API, consider the other options you have for streaming solutions.

If you have data that needs to be transformed or aggregated mid-flight into table and row format or even joined against other data sources mid-stream or if you want to take just a window or segment of that data you should consider using Dataflow for your streaming data pipeline. We'll cover working with that solution in the very next module.

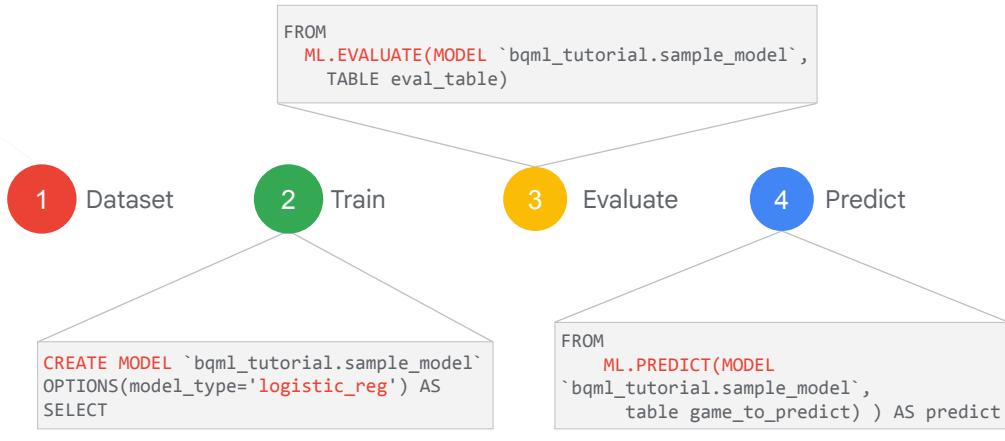
# A typical BigQuery data warehouse architecture



So what does a typical data warehouse solution architecture look like?

Take a look at the green box. BigQuery is the analytics engine that sits at the end of the data pipeline. It stores all the incoming data from the left and then allows you to do your analysis and model building.

# Working with BigQuery ML



Google Cloud

In four major steps it looks like this:

1. Write SQL query to extract training data from BigQuery
2. Create a model, specifying model type
3. Evaluate model and verify that it meets requirements
4. Predict using model on data extracted from BigQuery

BigQuery ML models can be exported to Vertex AI for online predictions.



Google Cloud

```
SELECT
    url, title
FROM
    `bigquery-public-data.hacker_news.stories`
WHERE
    LENGTH(title) > 10
    AND LENGTH(url) > 0
LIMIT 10
```

## Use SQL query to extract data

```
SELECT
  url, title
FROM
  `bigquery-public-data.hacker_news.stories`
WHERE
  LENGTH(title) > 10
  AND LENGTH(url) > 0
LIMIT 10
```

url	title
http://www.bbc.co.uk/news/business-27732743	Vodafone reveals direct government wiretaps
https://www.kickstarter.com/projects/appdocu/a...	Doc - App: The Human Story
http://www.starwebworld.com/android-jelly-bean...	Android Jelly Bean: Streaming Audio Through th...
http://www.myplanetdigital.com/digital.strateg...	Why Canadian Tech Entrepreneurs Need to Man/Wo...
http://startupslandconference.com/index.html	StartupConference June 13. - 16. 2013, HVAR Cr...
http://kopimism.org/	Kopimism Hactivism Meetup Tomorrow (Sunday) in...
http://unearthedgadget.com/xbox-live-gold-2/14...	Xbox Live Gold Membership Is It Really Worth ...
https://evertale.com	Evertale changes the way people remember
http://www.racketboy.com/retro/commodore-amiga...	Commodore Amiga: A Beginner's Guide
http://www.extremetech.com/extreme/156393-cold...	Cold fusion reactor "Independently verified"

Google Cloud

```
SELECT
  url, title
FROM
  `bigquery-public-data.hacker_news.stories`
WHERE
  LENGTH(title) > 10
  AND LENGTH(url) > 0
LIMIT 10
```

# Automated Feature Preprocessing

## One-Hot Encoding

For all non-numeric columns other than TIMESTAMP, BigQuery ML performs a one-hot encoding transformation that generates a separate feature for each unique value in the column.

## Imputing Missing Values

BQML imputes missing values [according to that field's data type](#).

Google Cloud

Note: For all non-numeric columns other than TIMESTAMP, BigQuery ML performs a one-hot encoding transformation. This transformation generates a separate feature for each unique value in the column.

# Manual Feature Preprocessing

BigQuery ML provides functions to make data preprocessing easy

Let's look at some [preprocessing functions in the documentation](#).

## General functions

Use the following function on string or numerical expressions to do data cleanup:

- [ML.IMPUTER](#)

## Categorical functions

Use the following functions on string expressions to categorize data:

- [ML.FEATURE\\_CROSS](#)
- [ML.HASH\\_BUCKETIZE](#)
- [ML.LABEL\\_ENCODER](#)
- [ML.MULTI\\_HOT\\_ENCODER](#)
- [ML.NGRAMS](#)
- [ML.ONE\\_HOT\\_ENCODER](#)

Google Cloud

Note: For all non-numeric columns other than TIMESTAMP, BigQuery ML performs a one-hot encoding transformation. This transformation generates a separate feature for each unique value in the column.

## Example:

### Use regex to get source + train on words of title

The screenshot shows the Google Cloud BigQuery interface. On the left, there is a code editor window containing a SQL query named "txtclass\_words". The query uses regular expressions to extract words from news titles and categorize them by source (github, nytimes, techcrunch). It then concatenates these words into a single array for each row. The right side of the interface shows the "Query results" pane, which displays 204 rows of data. The columns are labeled: source, words, and title. The first few rows show entries like "nytimes" followed by arrays of words such as "[the, socratic, shrink, NULL, NULL]" and "[still, stuck, in, a, climate]". At the bottom of the interface, there are buttons for "Run", "Save query", "Save view", and "More". A note says "This query will process 204".

```

-- txtclass_words
WITH extracted AS (
  SELECT source, REGEXP_REPLACE(LOWER(REGEXP_REPLACE(title, '[^a-zA-Z0-9 $.-]', '')), " ", "") AS title
  FROM (
    SELECT
      ARRAY_REVERSE(SPLIT(REGEXP_EXTRACT(url, '.*://(.*/+)/'), '.'))[OFFSET(1)] AS source,
      title
    FROM
      `bigquery-public-data.hacker_news.stories`
    WHERE
      REGEXP_CONTAINS(REGEXP_EXTRACT(url, '.*://(.*/+)/'), '.com$')
      AND LENGTH(title) > 10
  )
  , ds AS (
    SELECT ARRAY_CONCAT(SPLIT(title, " "), ['NULL', 'NULL', 'NULL', 'NULL', 'NULL']) AS words
    FROM extracted
    WHERE (source = 'github' OR source = 'nytimes' OR source = 'techcrunch')
  )
  SELECT
    source,
    words[OFFSET(0)] AS word1,
    words[OFFSET(1)] AS word2,
    words[OFFSET(2)] AS word3,
    words[OFFSET(3)] AS word4,
    words[OFFSET(4)] AS word5
  FROM ds
)

```

Google Cloud

```

WITH extracted AS (
  SELECT source, REGEXP_REPLACE(LOWER(REGEXP_REPLACE(title, '[^a-zA-Z0-9 $.-]', '')), " ", "") AS title
  FROM (
    SELECT
      ARRAY_REVERSE(SPLIT(REGEXP_EXTRACT(url, '.*://(.*/+)/'), '.'))[OFFSET(1)]
    AS source,
      title
    FROM
      `bigquery-public-data.hacker_news.stories`
    WHERE
      REGEXP_CONTAINS(REGEXP_EXTRACT(url, '.*://(.*/+)/'), '.com$')
      AND LENGTH(title) > 10
  )
  , ds AS (
    SELECT ARRAY_CONCAT(SPLIT(title, " "), ['NULL', 'NULL', 'NULL', 'NULL', 'NULL']) AS words
    FROM extracted
    WHERE (source = 'github' OR source = 'nytimes' OR source = 'techcrunch')
  )
  SELECT
    source,
    words[OFFSET(0)] AS word1,
    words[OFFSET(1)] AS word2,
    words[OFFSET(2)] AS word3,

```

```
words[OFFSET(3)] AS word4,  
words[OFFSET(4)] AS word5  
FROM ds
```

# Create a classification model

```
CREATE OR REPLACE MODEL advdata.txtclass
OPTIONS(model_type='logistic_reg', input_label_cols=['source'])
AS
```

```
WITH extracted AS (
...
),
ds AS (
SELECT ARRAY_CONCAT(SPLIT(title, " "), ['NULL', 'NULL', 'NULL',
'NULL', 'NULL']) AS words, source FROM extracted
WHERE (source = 'github' OR source = 'nytimes' OR source =
'techcrunch')
)
SELECT
source,
words[OFFSET(0)] AS word1,
...
FROM ds
```

Create a model

From these data

Google Cloud

A model feels like just another table that is being created.

Predict Visitor Purchases with a Classification Model in BigQuery ML:

[https://partner.cloudskillsboost.google/catalog\\_lab/1101](https://partner.cloudskillsboost.google/catalog_lab/1101)

## Evaluate model

```
SELECT * FROM ML.EVALUATE(MODEL advdata.txtclass)
```

precision	recall	accuracy	f1_score	log_loss	roc_auc
		0.79	0.783	0.858	0.918
Actual labels	Predicted labels	github	nytimes	techcrunch	% samples
github	github	88.8%	5.29%	5.9%	37.83%
nytimes	nytimes	6.34%	70.92%	22.74%	31.26%
techcrunch	techcrunch	5.54%	19.35%	75.11%	30.9%

Google Cloud

```
SELECT * FROM ML.EVALUATE(MODEL advdata.txtclass)
```

BigQuery ML splits the training data and reports evaluation statistics on the held-out set.

## Predict using trained model

```
SELECT * FROM ML.PREDICT(MODEL advdata.txtclass,
    SELECT 'government' AS word1, 'shutdown' AS word2, 'leaves'
AS word3, 'workers' AS word4, 'reeling' AS word5
    UNION ALL SELECT 'unlikely', 'partnership', 'in', 'house',
'gives'
    UNION ALL SELECT 'fitbit', 's', 'fitness', 'tracker', 'is'
    UNION ALL SELECT 'downloading', 'the', 'android', 'studio',
'project'
))
```

Row	predicted_source	word1	word2	word3	word4	word5
1	nytimes	government	shutdown	leaves	workers	reeling
2	nytimes	unlikely	partnership	in	house	gives
3	techcrunch	fitbit	s	fitness	tracker	is
4	techcrunch	downloading	the	android	studio	project

[Source](#)

Google Cloud

```
SELECT * FROM ML.PREDICT(MODEL advdata.txtclass,
    SELECT 'government' AS word1, 'shutdown' AS word2, 'leaves' AS word3,
'workers' AS word4, 'reeling' AS word5
    UNION ALL SELECT 'unlikely', 'partnership', 'in', 'house', 'gives'
    UNION ALL SELECT 'fitbit', 's', 'fitness', 'tracker', 'is'
    UNION ALL SELECT 'downloading', 'the', 'android', 'studio', 'project'
))
```

[https://console.cloud.google.com/bigquery?sq=663413318684:4d854a43ae93416eae\\_b349e1fc4888cb](https://console.cloud.google.com/bigquery?sq=663413318684:4d854a43ae93416eae_b349e1fc4888cb)

[https://console.cloud.google.com/bigquery?sq=663413318684:4d854a43ae93416eae\\_b349e1fc4888cb](https://console.cloud.google.com/bigquery?sq=663413318684:4d854a43ae93416eae_b349e1fc4888cb)

## Linear regression in SQL

```
CREATE OR REPLACE MODEL ch09edu.bicycle_model
OPTIONS(input_label_cols=['duration'],
        model_type='linear_reg')
AS

SELECT
    duration
    , start_station_name
    , CAST(EXTRACT(dayofweek from start_date) AS STRING)
        as dayofweek
    , CAST(EXTRACT(hour from start_date) AS STRING)
        as hourofday
FROM
    `bigquery-public-data.london_bicycles.cycle_hire`
```

Google Cloud

Here is a simple forecasting model to predict the duration of a bike ride in london. We are using Linear Regression as our model type and our label is 'duration' as you see here.

```
CREATE OR REPLACE MODEL ch09edu.bicycle_model
OPTIONS(input_label_cols=['duration'],
        model_type='linear_reg')
AS

SELECT
    duration
    , start_station_name
    , CAST(EXTRACT(dayofweek from start_date) AS STRING)
        as dayofweek
    , CAST(EXTRACT(hour from start_date) AS STRING)
        as hourofday
FROM
    `bigquery-public-data.london_bicycles.cycle_hire`
```

## DNN regression in SQL

```
CREATE OR REPLACE MODEL ch09edu.bicycle_model
OPTIONS(input_label_cols=['duration'],
        model_type='dnn_regressor', hidden_units=[32, 4])
AS

SELECT
    duration
    , start_station_name
    , CAST(EXTRACT(dayofweek from start_date) AS STRING)
        as dayofweek
    , CAST(EXTRACT(hour from start_date) AS STRING)
        as hourofday
FROM
    `bigquery-public-data.london_bicycles.cycle_hire`
```

Google Cloud

And here is the DNN\_regressor that you will be creating as part of your lab. Note the model type and hidden units parameters in model options.

```
CREATE OR REPLACE MODEL ch09edu.bicycle_model
OPTIONS(input_label_cols=['duration'],
        model_type='dnn_regressor', hidden_units=[32, 4])
AS

SELECT
    duration
    , start_station_name
    , CAST(EXTRACT(dayofweek from start_date) AS STRING)
        as dayofweek
    , CAST(EXTRACT(hour from start_date) AS STRING)
        as hourofday
FROM
    `bigquery-public-data.london_bicycles.cycle_hire`
```

## Example: Linear Classifier (Logistic Regression)

```
create or replace model models.will_buy_banana_example
options(model_type='logistic_reg', input_label_cols=['banana']) AS

with purchases_data AS (
select
    CAST(user_id AS string) customer_id,
    CAST(zip_code AS string) home_zipcode,
    ['dawn', 'morning', 'afternoon', 'night'][OFFSET(CAST
(TRUNC(order_hour_of_day/6) AS INT64))] AS time_of_day,
    (SELECT 24852 IN (SELECT product_id FROM UNNEST(order_lines))) AS
banana
FROM operations.orders_with_lines
JOIN operations.customers_loyalty
ON user_id = id
)
select * from purchases_data
```

Google Cloud

Show them running this model.

## Example: DNN Classifier

```
create or replace model models.will_buy_banana_example
options(model_type='dnn_classifier', hidden_units=[64, 8],
input_label_cols=['banana']) AS

with purchases_data AS (
select
  CAST(user_id AS string) customer_id,
  CAST(zip_code AS string) home_zipcode,
  ['dawn', 'morning', 'afternoon', 'night'][OFFSET(CAST
(TRUNC(order_hour_of_day/6) AS INT64))] AS time_of_day,
  (SELECT 24852 IN (SELECT product_id FROM UNNEST(order_lines))) AS
banana
FROM operations.orders_with_lines
JOIN operations.customers_loyalty
ON user_id = id
....
```

Google Cloud

Show them running this model.

There are two types of DNN models available as alpha in Qwiklabs (this will work in Qwiklabs but not your own accounts yet as of Sept 2019). The dnn\_classifier and the dnn\_regressor as we will see shortly.

# Hyperparameter tuning in BigQuery ML

BigQuery ML [models have tunable hyperparameters](#) that are unique to each model type.

Google recommends using at least 10 trials for one hyperparameter, so the total number of trials should be at least:

`10 * num_hyperparameters`

Model type	Hyperparameter objectives	Hyperparameter	Valid range	Default range	Scale type
LINEAR_REG	MEAN_ABSOLUTE_ERROR	L1_REG	(0, ∞]	(0, 10]	LOG
	MEAN_SQUARED_ERROR	L2_REG	(0, ∞]	(0, 10]	LOG
	MEAN_SQUARED_LOG_ERROR				
	MEDIAN_ABSOLUTE_ERROR				
	R2_SCORE (default)				
	EXPLAINED_VARIANCE				
LOGISTIC_REG	PRECISION	L1_REG	(0, ∞]	(0, 10]	LOG
	RECALL	L2_REG	(0, ∞]	(0, 10]	LOG
	ACCURACY				
	F1_SCORE				
	LOG_LOSS				
	ROC_AUC (default)				

Google Cloud

Note: For all non-numeric columns other than TIMESTAMP, BigQuery ML performs a one-hot encoding transformation. This transformation generates a separate feature for each unique value in the column.



## Building AI solutions by using ML APIs

Google Cloud

# Translation API

- Translate text to other languages
- Google provides [a few options for translation](#) APIs or models

	Translation API Basic	Translation API Advanced	AutoML Translation
Content types	Text, HTML	Text, HTML, DOCx, PPTx, XLSx, PDF	Text
Primary actions	Text translation	Text translation	Model training, management and text translations
Language detection	✓	✓	
Glossary		✓	

Google Cloud

The [Translation API](#) can help you translate text.

## Speech-to-Text API

- Accurately convert speech into text

## Text-to-Speech API

- Create audio of text being read by a human-sounding voice

Google Cloud

[Speech-to-Text](#)

[Text-to-Speech Basics](#)

# Natural Language API

- Analyze Sentiment
- Analyze Entities
- Analyze Syntax
- Analyze Entity Sentiment
- Classify Content
- Moderate Text

```
{  
  "entities": [  
    {  
      "name": "Lawrence of Arabia",  
      "type": "WORK_OF_ART",  
      "metadata": {  
        "mid": "/m/0bx0l",  
        "wikipedia_url": "http://en.wikipedia.org/wiki/Lawrence_of_Arabia_(1962_film)"  
      },  
      "salience": 0.75222147,  
      "mentions": [  
        {  
          "text": {  
            "content": "Lawrence of Arabia",  
            "beginOffset": 1  
          },  
          "type": "PROPER"  
        },  
        {  
          "text": {  
            "content": "the movie",  
            "beginOffset": 0  
          },  
          "type": "COMMON"  
        }  
      ]  
    }  
  ]  
}
```

Google Cloud

The [Natural Language API](#) can help you accomplish these tasks.

## Healthcare Natural Language API

Extracts:

- Medical concepts, such as medications, procedures, and medical conditions
- Functional features, such as temporal relationships, subjects, and certainty assessments
- Relations, such as side effects and medication dosage

```
{  
  "entityMentions": [  
    {  
      "mentionId": "1",  
      "type": "MEDICINE",  
      "text": {  
        "content": "Insulin regimen human"  
      },  
      "linkedEntities": [  
        {  
          "entityId": "UMLS/C3537244"  
        },  
        {  
          "entityId": "UMLS/C3537244"  
        }  
      ]  
    }  
  ]  
}
```

Google Cloud

GCP also offers a [Healthcare Natural Language API](#) which extracts healthcare information from medical text.

# Document AI

- Pre-process documents with image quality detection and deskewing
- Extract text and layout information from document files
- Identify key-value pairs in structured forms
- Split and classify documents by type
- Extract and normalize entities
- Label and review documents
- Store, search, organize, govern, and analyze documents and metadata

The screenshot shows the 'Form Parser analysis' interface with the 'KEY-VALUE PAIR' tab selected. It displays the following key-value pairs:

- First Name: Sally
- Last Name: Walker
- Address: 12 Berry Lane
- Zip: 01882
- State: MA
- City: Somerville
- Email: sally.walker@gmail.com
- Occupation: Software Engineer
- Gender: F
- Marital Status: Single
- Emergency Contact Name: Sam Walker
- Emergency Contact Phone: (0123) 4567890
- Emergency Contact: Sam Walker

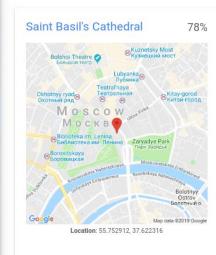
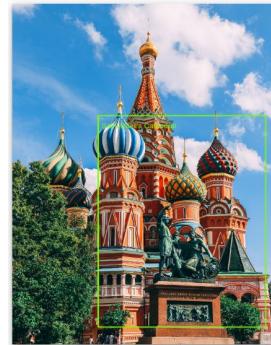
Do you currently take any medication? (If yes, please describe.)  
[Please take 25mg daily for attention.]

Google Cloud

## Document AI documentation overview

# Cloud Vision API

- Text detection
- Document text detection (dense text / handwriting)
- Landmark detection
- Logo detection
- Label detection
- Image properties
- Object localization
- Crop hint detection
- Web entities and pages
- Explicit content detection (SafeSearch)
- Face detection



Google Cloud

[Cloud Vision API](#)

## Retail API

- Recommendations AI to support personalized omnichannel recommendations
- Retail Search to power high quality product search results that are customizable for your retail business needs

### Welcome to Retail

Infuse your digital properties with Google-quality recommendations and search results that enhance user engagement, deliver personalized experiences, and drive improved conversion. [Learn more](#)

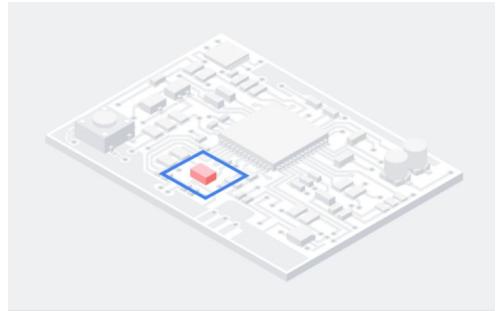
Retail API	On
Recommendations AI	On
Retail Search (enables search and browse)	On

Google Cloud

### Features of Retail API

## Visual Inspection AI

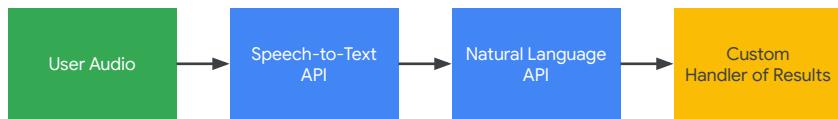
- Detect assembly inspection issues
- Detect cosmetic inspection anomalies



Google Cloud

[Visual Inspection AI](#)

## Often APIs may be used as part of a pipeline

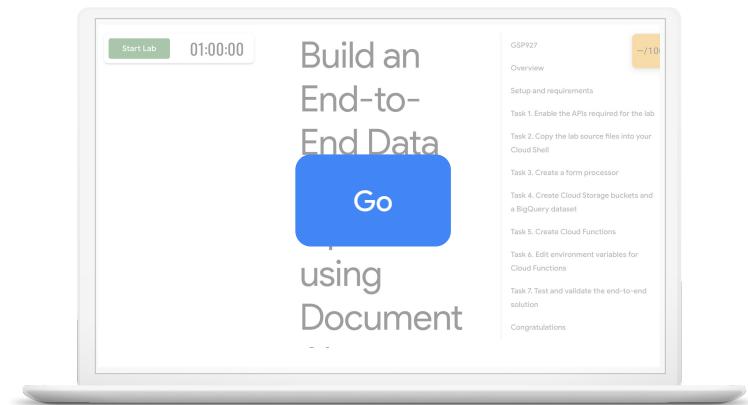


Google Cloud

[Visual Inspection AI](#)

**Recommended Lab**

**Build an  
End-to-End Data  
Capture Pipeline  
using Document AI**



Google Cloud

[Build an End-to-End Data Capture Pipeline using Document AI](#)

03



## Training models by using AutoML

Google Cloud

# Introducing AutoML

An algorithm to create Machine Learning Models



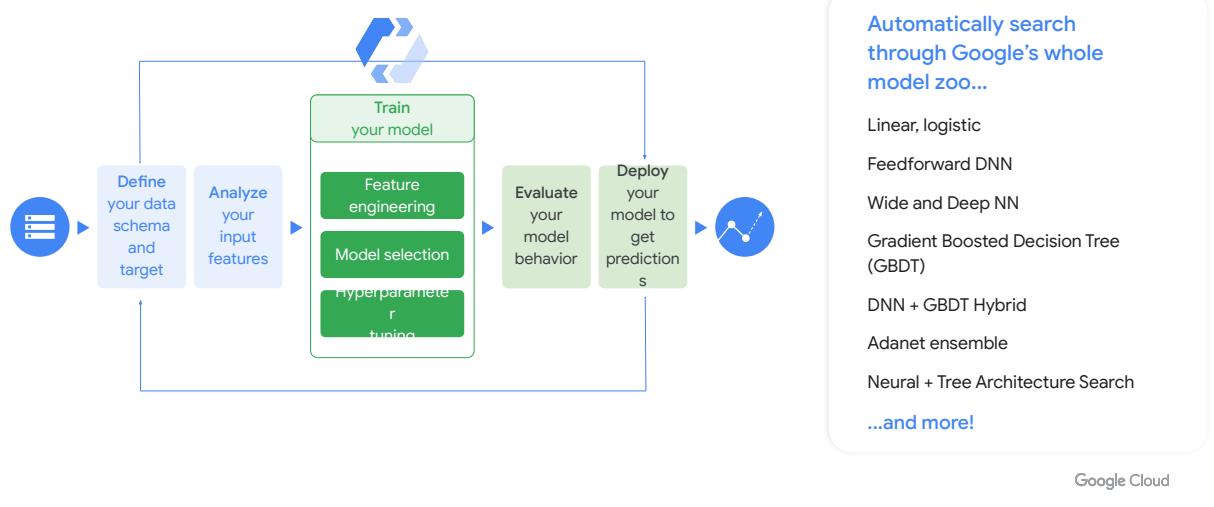
Allows quick creation of ML models.

**Production Ready:** Scales well with Google Cloud large scale computation resource

**Human Data Labeling:** For customers with no labeled training images, our in-house human labelers are available to review your custom instructions and label your images accordingly for model training.

Google Cloud

# Introducing AutoML: What does it do?



Welcome to **How Google does Machine Learning**.

Course:

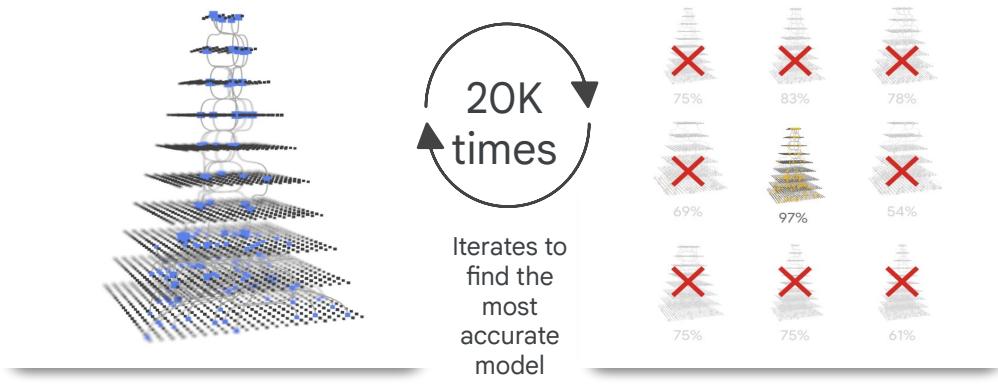
[https://partner.cloudskillsboost.google/course\\_templates/237](https://partner.cloudskillsboost.google/course_templates/237)

LAB: [Using an Image DataSet to Train an AutoML Model](#)

# # #

Qwiklab: [Identify Damaged Car Parts with AutoML](#)

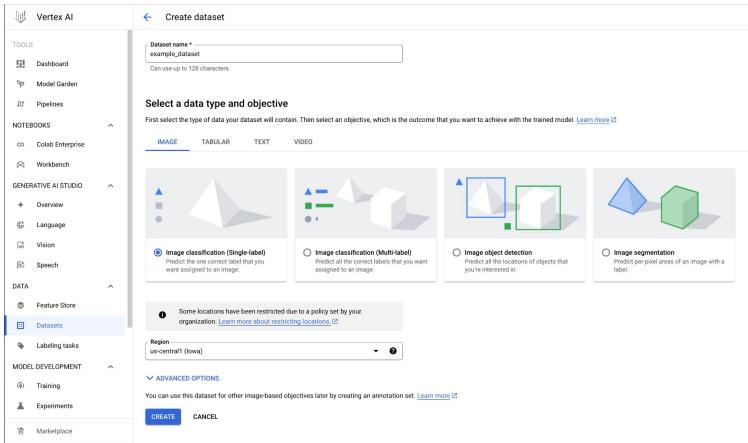
## AutoML is built with Neural Architecture Search



Google Cloud

Behind-the-scenes, AutoML is powered by the latest ML research. While your model trains, the AutoML platform actually trains and evaluates multiple models and compares them against each other. This NASNet approach or Neural Architecture Search produces an ensemble of ML models and chooses the best one.

# Prepare for AutoML training by creating a dataset & selecting a task

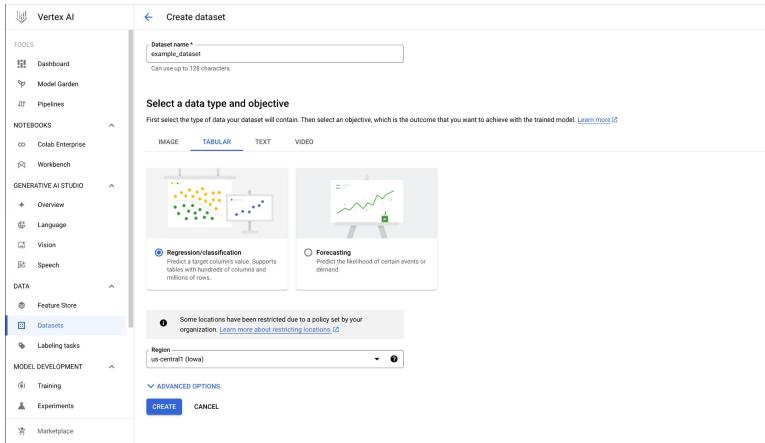


Google Cloud

Prepare for an AutoML training process by creating a dataset and selecting your task.

You can select an image task...

# Prepare for AutoML training by creating a dataset & selecting a task



Google Cloud

tabular data task...

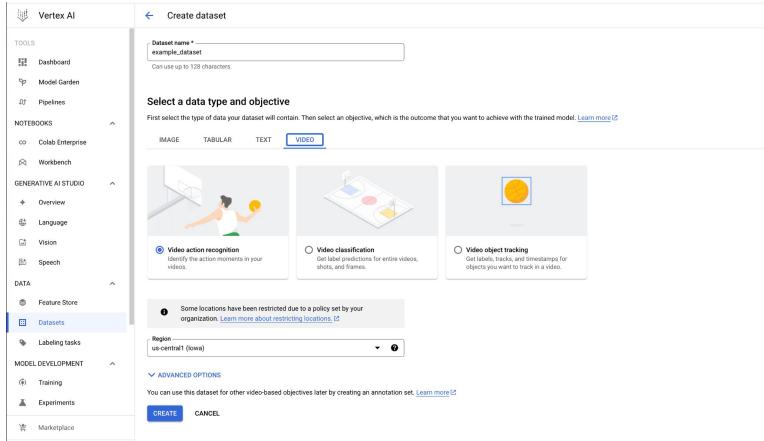
# Prepare for AutoML training by creating a dataset & selecting a task

The screenshot shows the Vertex AI interface for creating a dataset. On the left, there's a sidebar with sections like 'TOOLS', 'NOTEBOOKS', 'GENERATIVE AI STUDIO', 'DATA', and 'MODEL DEVELOPMENT'. The main area is titled 'Create dataset' and has a 'Dataset name' input field with 'example\_dataset' typed in. Below it, under 'Select a data type and objective', the 'TEXT' tab is selected. It lists four options: 'Text classification (Single-label)', 'Text classification (Multi-label)', 'Text entity extraction', and 'Text sentiment analysis'. The first option is selected. A note below says 'Some locations have been restricted due to a policy set by your organization.' At the bottom, there are 'CREATE' and 'CANCEL' buttons.

Google Cloud

text task...

# Prepare for AutoML training by creating a dataset & selecting a task



The screenshot shows the 'Create dataset' page in the Vertex AI interface. On the left, there's a sidebar with sections like 'TOOLS' (Dashboard, Model Garden, Pipelines), 'NOTEBOOKS' (Colab Enterprise, Workbench), 'GENERATIVE AI STUDIO' (Overview, Language, Vision, Speech), 'DATA' (Feature Store, Datasets, Labeling tasks), 'MODEL DEVELOPMENT' (Training, Experiments), and 'Marketplace'. The main area has a header 'Create dataset' with a back arrow. It starts with a 'Dataset name' input field containing 'example\_dataset' with a note 'Can use up to 128 characters'. Below it is a section titled 'Select a data type and objective' with tabs for IMAGE, TABULAR, TEXT, and VIDEO (which is selected). Under 'VIDEO', there are three options: 'Video action recognition' (selected, described as identifying action moments in videos), 'Video classification' (described as getting label predictions for entire videos, shots, and frames), and 'Video object tracking' (described as getting labels, tracks, and timestamps for objects you want to track in a video). A note below says 'Some locations have been restricted due to a policy set by your organization.' with a link to learn more. A 'Region' dropdown is set to 'us-central1 (Iowa)'. At the bottom are 'CREATE' and 'CANCEL' buttons.

Google Cloud

or video task.

# Upload data or an import file pointing to your data in Cloud Storage

**TOOLS**

- Dashboard
- Model Garden
- Pipelines

**NOTEBOOKS**

- Colab Enterprise
- Workbench

**GENERATIVE AI STUDIO**

- Overview
- Language
- Vision
- Speech

**DATA**

- Feature Store
- Datasets
- Labeling tasks

**MODEL DEVELOPMENT**

- Training
- Experiments
- Marketplace

**IMPORT**    **BROWSE**    **ANALYZE**

Add images to your dataset

Before you begin, review the data guide to make sure your data is formatted correctly and optimized for the best results. Supported image file formats: JPEGs, PNGs, GIF, BMP, TIFF, WebP, ICO.

[VIEW DATA GUIDE](#)

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 URLs for your images and optional data, like labels. [Learn how to create an import file](#)
- 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](#)

**Summary**

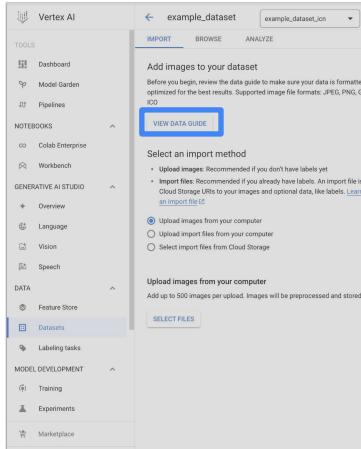
CUMULUS    CIRRUS    STRATUS

Google Cloud

You'll be asked to add data to your dataset.

Typically your data will be stored in Cloud Storage. When it's time to train, data will be loaded in batches.

# The Data Guide can provide some minimum recommendations on dataset size



The screenshot shows the Vertex AI Data Studio interface. On the left, there's a sidebar with sections like 'TOOLS', 'NOTEBOOKS', 'GENERATIVE AI STUDIO', 'DATA', 'MODEL DEVELOPMENT', and 'Marketplace'. The main area is titled 'example\_dataset' and has tabs for 'IMPORT', 'BROWSE', and 'ANALYZE'. A 'VIEW DATA GUIDE' button is highlighted with a blue box. Below it, there's a section for 'Add images to your dataset' with instructions and options for 'Upload images from your computer' or 'Upload import files from your computer'. There's also a 'SELECT FILES' button. To the right, there's a 'Data preparation guide' with sections for 'Data requirements' (requiring 10 images per label, supported formats JPEG, PNG, GIF, BMP, TIFF, WebP, ICO, max size 1.5 MB), 'Recommended' (few labels: at least 100 images per label; many labels: at least 200 images per label), and 'NEXT' and 'DONE' buttons.

Google Cloud

# Vertex AI Documentation shows how to construct an import file

The screenshot shows the Vertex AI Documentation interface with the 'Guides' tab selected. On the left, there's a sidebar with navigation links for various Vertex AI services like Video, Model, and ML development tools. The main content area is titled 'Input files' and provides instructions for preparing data in JSON Lines or CSV format. It includes a code editor with JSONL and CSV examples, and a 'Field notes' section with detailed explanations of the fields.

```

{
  "imageGcsUri": "gs://bucket/filename.ext",
  "classificationAnnotation": [
    {
      "displayName": "LABEL",
      "annotationResourceLabels": {
        "aiplatform.googleapis.com/annotation_set_name": "displayName",
        "env": "prod"
      }
    }
  ],
  "dataItemResourceLabels": {
    "aiplatform.googleapis.com/ml_use": "training/test/validation"
  }
}

```

**Field notes:**

- **imageGcsUri** - The only required field.
- **annotationResourceLabels** - Can contain any number of key-value string pairs. The only system-reserved key-value pair is the following:
  - `aiplatform.googleapis.com/annotation_set_name": "value"`
- Where `value` is one of the display names of the existing annotation sets in the dataset.
- **dataItemResourceLabels** - Can contain any number of key-value string pairs. The only system-reserved key-value pair is the following which specifies the machine learning use set of the data item:

Google Cloud

You can prepare a CSV or JSONL file, and the documentation shows you how to structure your files to point towards your resources in Cloud Storage and designate each example for training, test, or validation.

# After uploading data, you can initiate an AutoML Training Job.

**Train new model**

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

Dataset \* example\_dataset (0 images)

Annotation set \* example\_dataset\_icn

Objective Image classification (Single-label)

Please refer to the pricing guide for more details (and available deployment options) for each method.

Model training method

**AutoML**  
Train high-quality models with minimal effort and machine learning expertise. Just specify how long you want to train. [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](#)

Choose where to use the model

**Cloud**  
Deploy to an endpoint for online predictions or use for batch predictions.

**Edge**  
Export for on-prem and on-device use. Typically has lower accuracy.

**CONTINUE**

Google Cloud

To train a model, simply select the training method, input the model details, input the number of nodes for compute and pricing and start training.

Recall there are several training methods available:

AutoML, which allows you to train models with minimum ML expertise. Within AutoML you can select:

Cloud which trains models to be run on GCP, either using online or batch predictions, or

Edge which trains a model that can be exported for on-prem/on-device use.

Or instead of AutoML, you can select a custom training (advanced) training job, where you can train models in custom containers you've prepared, or if it's a TensorFlow, scikit-learn, or XGBoost model you can use containers prepared by Google.

# Vertex AI Training can override your labels to adjust your data split.

Train new model

**Training method**  
Creates a new model group and assigns the trained model as version 1

**Train new version**  
Trains model as a version of an existing model

**2 Model details**

Name \*

Description

**3 Training options**

**4 Compute and pricing**

START TRAINING CANCEL

Data split

Random  Manual (Advanced)

Your dataset will be automatically randomized and split into training, validation, and test sets using the following ratios. [Learn more](#)

Training  % Validation  % Test  %

Training: 80%  
 Validation: 10%  
 Test: 10%

Encryption

Google-managed encryption key

Google Cloud

With most training types you can see how your data will be split, or allow Vertex AI to randomly split it.

You'll often want to take more control of the split based on your dataset.

For example in time series, you'll want older examples in your training dataset to predict newer examples split across your validation and test sets.

Or for GPS-based records, you may want to bucketize GPS coordinates and split based on that.

## If using a custom data split, confirm you have sufficient examples of each class in each data split



Google Cloud

With most training types you can see how your data will be split, or allow Vertex AI to randomly split it.

You'll often want to take more control of the split based on your dataset.

For example in time series, you'll want older examples in your training dataset to predict newer examples split across your validation and test sets.

Or for GPS-based records, you may want to bucketize GPS coordinates and split based on that.

## Some model types (here AutoML on an image dataset) allow you to enable Explainability features.

**Train new model**

- Training method
- Model details
- Explainability (optional)
- Training options
- Compute and pricing

**START TRAINING    CANCEL**

Generate explainable bitmaps for each image in the test set

**Integrated gradients**

When you request an explanation using Integrated gradients, you'll receive a bitmap with an image overlay showing which pixels contributed to the prediction. Integrated gradients is a pixel-based attribution method that highlights important areas in the image regardless of contrast, making this method ideal for non-natural images such as X-rays.

  
Example

Visualization type \* — Outlines

Color map \* — Pink/Green

Clip below — 70   Clip above — 99.9

Overlay type — None

Number of integral steps — 50

Google Cloud

For some model types, like AutoML Edge models on images, you can turn on explainability options to see what parts of an image impacted its predicted class.

## AutoML Edge models also allow you to make accuracy / latency tradeoffs.

The screenshot shows a user interface for training a new model. On the left, a sidebar lists steps: 1. Training method (checked), 2. Model details (checked), 3. Explainability (optional), 4. Training options, and 5. Compute and pricing. Below the sidebar are 'START TRAINING' and 'CANCEL' buttons. On the right, there's a table comparing three goals:

Goal	Package size	Accuracy	Latency on iPhone X
Higher accuracy	5.6 MB	Higher	34ms
Best trade-off	3.1 MB	Medium	23ms
Faster predictions	557 KB	Lower	8ms

A note below the table states: "Please note that prediction latency estimates are for guidance only. Actual latency depends on your network connectivity." Another note says: "Models are based on state-of-the-art research at Google. Your model will be available as quantized TF Lite, TensorFlow, and Core ML packages." A "CONTINUE" button is at the bottom right.

Google Cloud

AutoML Edge models also allow you to select model size for speed / accuracy tradeoffs.

# Custom training will allow you control of Hyperparameter tuning.

The screenshot shows the 'Train new model' wizard with six steps:

- 1 Training method
- 2 Model details
- 3 Training container
- 4 Hyperparameters (optional) [Selected]
- 5 Compute and pricing
- 6 Prediction container (optional)

Fields in the 'Hyperparameters (optional)' step:

- Dataset: No managed dataset
- Annotation set: -
- Objective: Custom

Text below the fields: Please refer to the pricing guide for more details (and available deployment options) for each method.

Info box: AutoML options are only available when you train with a managed dataset.

Model training method section:

- 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](#)

Buttons at the bottom: START TRAINING CANCEL

Google Cloud

Notice if selecting a custom training job, you can also define Hyperparameter tuning options through this flow.

## You can set a Budget in node hours.

**Train new model**

- Training method
- Model details
- Explainability (optional)
- Training options
- Compute and pricing

**Budget \***  Maximum node hours

Estimated completion: Aug 31, 2023 9 PM GMT-4

Enable early stopping

Stop 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.

**START TRAINING**   **CANCEL**

Google Cloud

Finally you can define your training budget by number of node hours and enable early stopping so your model stops if training is no longer improving your validation metrics.

[Vertex AI Pricing](#)

**Vertex AI will provide appropriate metrics to use to evaluate your model depending on the model type (classification, regression, etc).**

**Confusion matrix**

This table shows how often the model classified each label correctly (in blue), and which labels were most often confused for that label (in gray).

True label	Predicted label						
	BARBUNYA	BOMBAY	CALI	DERMASON	HOROZ	SEKER	SIRA
BARBUNYA	94%	—	5%	—	—	1%	1%
BOMBAY	—	100%	—	—	—	—	—
CALI	2%	—	96%	—	1%	1%	1%
DERMASON	—	—	—	94%	—	1%	6%
HOROZ	—	—	—	1%	96%	—	3%
SEKER	0%	—	—	1%	—	96%	3%
SIRA	1%	—	—	9%	—	1%	90%

Google Cloud

After training & evaluating your model, deploy it to an endpoint for online predictions and/or use it in batch prediction pipelines.

The screenshot shows the Google Cloud AutoML interface for a project named "automl-beans1623431305". The "DEPLOY AND TEST" tab is selected. A section titled "Deploy your model" explains what endpoints are and how they work. Below this is a "DEPLOY TO ENDPOINT" button. A table lists existing endpoints, with one row selected: "automl-beans1623431305\_endpoint". The table columns are Name, ID, Models, Region, Monitoring, and Most recent monitoring job. The "Monitoring" column for the selected endpoint shows "Disabled".

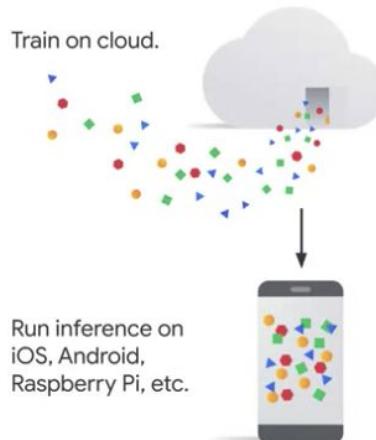
Name	ID	Models	Region	Monitoring	Most recent monitoring job
automl-beans1623431305_endpoint	3328494376161640448	1	us-central1	Disabled	–

Google Cloud

If time for them to try:

Qwiklab: [Identify Damaged Car Parts with AutoML](#)

## You can also train models to run on Edge devices



Google Cloud

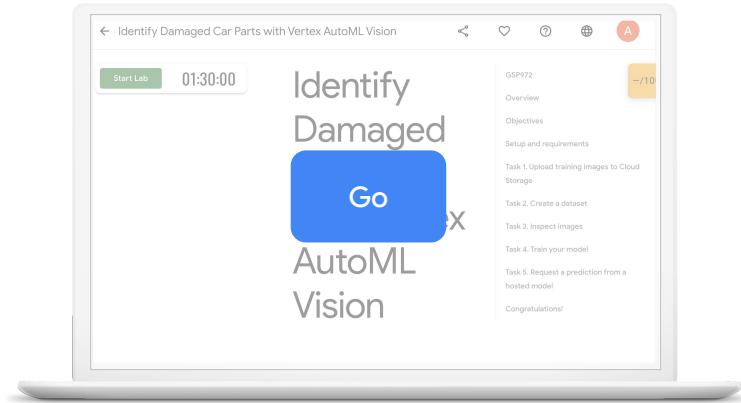
[Source - Tensorflow on Google cloud] -  
[https://partner.cloudskillsboost.google/course\\_templates/12](https://partner.cloudskillsboost.google/course_templates/12)

Portability between devices enables a lot of power and flexibility. For example, here's a common pattern. You train a TensorFlow model on the cloud, on lots, and lots, and lots of powerful hardware. Then you take the train model and put it on a device out on the edge, perhaps a mobile phone, or even an embedded chip. Then you can do predictions with the model right on the device itself offline. Have you had a chance to use the Google Translate app on an Android phone? The app can work completely offline because the train translation model is stored on the phone, and it's available for offline translation. Now I know what you're thinking, due to the limitations of processing power on your phones, the Edge model tends to be a bit smaller. Which means they're generally less powerful than what's on the cloud. However, the fact that TensorFlow allows for models to run on the Edge, it means a much faster response during predictions. So TensorFlow is this portable, powerful, production ready software, to do numeric computing. It's particularly popular for machine learning, it's the number one repository for machine learning on GitHub. Why is that? Well, it's popular among deep learning researchers because of the community around it, in the ability to extend it to do some pretty cool new things. It's popular among machine learning engineers because the ability to productionalize models, to do things at scale.

# Survey time!

**Recommended Lab**

# Identify Damaged Car Parts with Vertex AutoML Vision



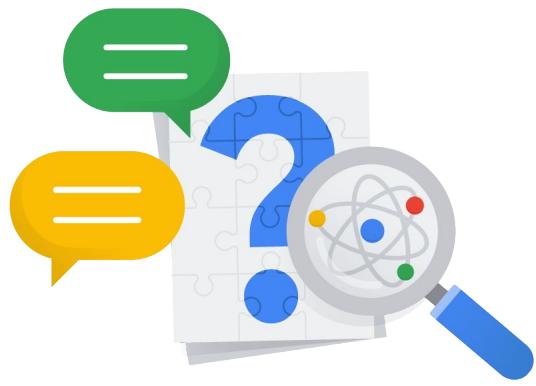
From the Course

[Build and Deploy Machine Learning Solutions on Vertex AI](#)

Google Cloud

[Identify Damaged Car Parts with Vertex AutoML Vision](#)

## Questions and answers



Google Cloud

## Thank you for attending this training!

We love your feedback! Please take a minute to complete the survey and help us improve our courses.



Google Cloud

