

Google Cloud Platform CMPE260 Search products and resources

Compute Engine VM instances CREATE INSTANCE IMPORT VM REFRESH START / RESUME STOP OPERATIONS HELP ASSISTANT SHOW INFO PANEL LEARN

Virtual machines VM instances INSTANCE SCHEDULE

VM instances are highly configurable virtual machines for running workloads on Google infrastructure. Learn more

Filter Enter property name or value

Status	Name	Zone	Recommendations	In use by	Internal IP	External IP	Connect
OFF	tensorflow-2-3-20211209-130458	us-central1-a			10.128.0.4 (nic0)	None	SSH
ON	tensorflow-2-7-20211208-091022	us-central1-a			10.128.0.3 (nic0)	34.133.252.113	SSH

Related actions

View billing report Monitor VMs Explore VM logs Set up firewall rules Patch management

View and manage your Compute Engine billing View outlier VMs across metrics like CPU and network View, search, analyze, and download VM instance logs Control traffic to and from a VM instance Schedule patch updates and view patch compliance on VM instances

DISMIS

Google Cloud Platform CMPE260 Search products and resources

Vertex AI Dashboard

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

Get started with Vertex AI

Vertex AI empowers machine learning developers, data scientists, and data engineers to take their projects from ideation to deployment, quickly and cost-effectively. Learn more

Try an interactive tutorial to learn how to train, evaluate, and deploy a Vertex AI AutoML or custom-trained model

VIEW TUTORIALS

Region us-central1 (Iowa)

Recent datasets

- iowa\_daily 37 minutes ago
- + CREATE DATASET

Recent models

- beans-model-pipeline 3 hours ago
- beans-model-pipeline 3 hours ago
- + TRAIN NEW MODEL

Get predictions

After you train a model, you can use it to get predictions, either online as an endpoint or through batch requests

+ CREATE BATCH PREDICTION

Google Cloud Platform CMPE260 Search products and resources

Container Registry Repositories

Images Settings

Transition to Artifact Registry

Artifact Registry is the recommended service for managing container images. Container Registry is still supported but will only receive critical security fixes. Learn more about options to transition to Artifact Registry.

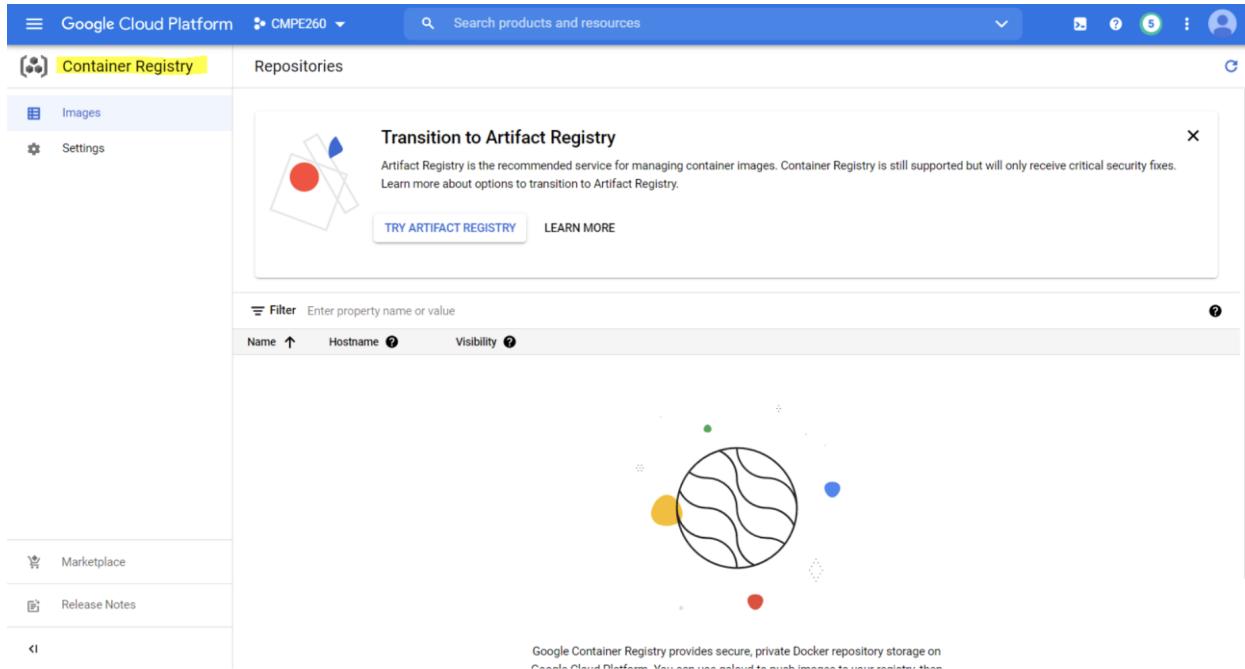
TRY ARTIFACT REGISTRY LEARN MORE

Filter Enter property name or value

Name ↑ Hostname ? Visibility ?

Marketplace Release Notes

Google Container Registry provides secure, private Docker repository storage on Google Cloud Platform. You can now switch to Artifact Registry to manage your images.



Google Cloud Platform CMPE260 Search products and resources CUSTOMIZE

RECOMMENDATIONS

Financial Services Healthcare Life Sciences Dataprep

ARTIFICIAL INTELLIGENCE

Vertex AI AI Platform Data Labeling Document AI Natural Language Recommendation... Retail Speech-to-text Tables Talent Solution

Vertex AI Workbench

Compute Engine CPU (%) 100% 80% 60% 40% 20% 0% 11:45 12 PM 12:15 12:30 instance/cpu/utilization: 0.04%

Go to Compute Engine

RPC APIs Requests (requests/sec) 1.0 0.8 0.6 0.4 ▲ No data is available for the selected time frame.

Google Cloud Platform status

Google Cloud SQL Global: Issues with Cloud SQL for MySQL instance migration to 5.7 when source databases have gtid\_mode set to ON. Began at 2021-12-07 (12:44:21)

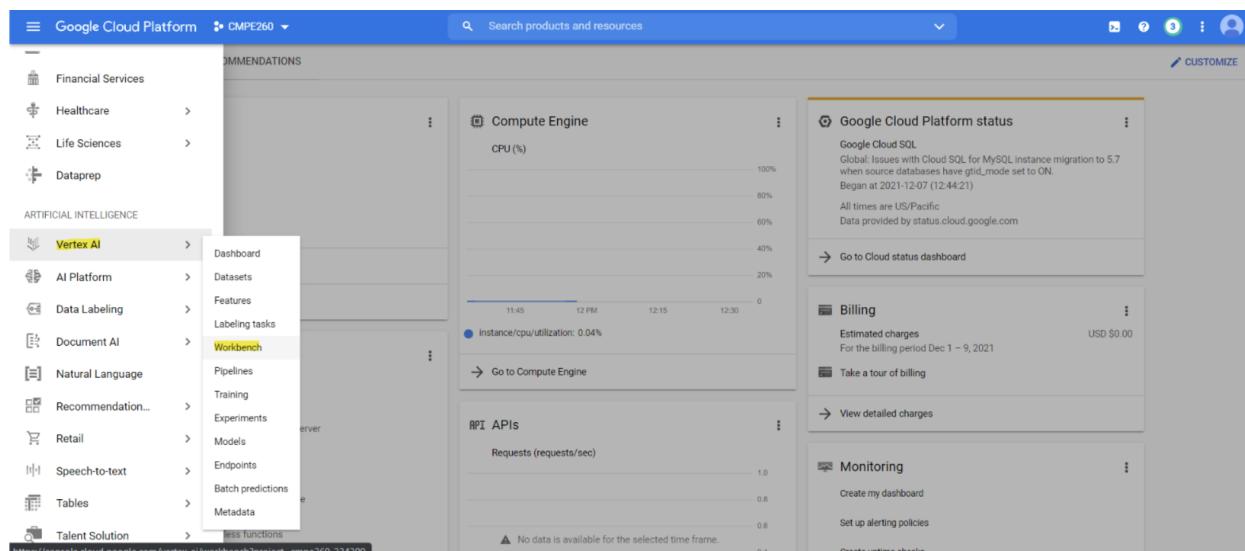
All times are US/Pacific Data provided by status.cloud.google.com

Go to Cloud status dashboard

Billing Estimated charges For the billing period Dec 1 – 9, 2021 USD \$0.00

Take a tour of billing View detailed charges

Monitoring Create my dashboard Set up alerting policies Create uptime checks



**Google Cloud Platform** CMPE260 Search products and resources

**Vertex AI**

- Dashboard
- Datasets
- Features
- Labeling tasks
- Workbench**
- Pipelines
- Training
- Experiments
- Models
- Endpoints
- Batch predictions
- Metadata
- Marketplace

**Notebooks** NEW NOTEBOOK REFRESH START STOP RESET DELETE

MANAGED NOTEBOOKS

Customize...

**Python 3**  
Includes scikit-learn, pandas and more  
Optimized for NVIDIA GPUs

**TensorFlow Enterprise**  
Includes Keras, scikit-learn, pandas, NLTK and more

**PyTorch 1.9**  
Includes scikit-learn, pandas, NLTK and more

**R 4.1**  
Includes basic R packages, scikit-learn, pandas, NLTK and more

**RAPIDS 0.18 [EXPERIMENTAL]**  
Optimized for NVIDIA GPUs

**Kaggle Python [BETA]**  
Python image for Kaggle Notebooks, supporting hundreds of machine learning libraries popular on Kaggle

**Theia IDE [EXPERIMENTAL]**  
IDE with notebook support including scikit-learn, pandas, and more

**Smart Analytics Frameworks**  
BigQuery, Apache Beam, Apache Spark, Apache Hive, and more

SCHEDULES PREVIEW

g an existing environment's JupyterLab 1.x version, disable auto-upgrade (if enabled) and do not manually upgrade the environment by creating specific versions of Notebooks.

TensorFlow Enterprise 1.15 (with LTS)

TensorFlow Enterprise 2.1 (with LTS)

**TensorFlow Enterprise 2.3 (with LTS)**

TensorFlow Enterprise 2.6 (with LTS)

TensorFlow Enterprise 2.7

Without GPUs

Version	Machine type	GPUs	Permission	Last modified
TensorFlow Enterprise 2.3 (with LTS)	With 1 NVIDIA Tesla T4	None	Service account	Dec 9, 2021, 11:57:12 AM

SHOW INFO PANEL

**Google Cloud Platform** CMPE260 Search products and resources

**Vertex AI**

- Dashboard
- Datasets
- Features
- Labeling tasks
- Workbench**
- Pipelines
- Training
- Experiments
- Models
- Endpoints
- Batch predictions
- Metadata
- Marketplace

**Notebooks** NEW NOTEBOOK

MANAGED NOTEBOOKS PREVIEW

As of the M80 DLMV release, all environments are now pre-installed and available to a new environment version.

Notebooks have JupyterLab pre-installed and are optimized for many machine learning frameworks. [Learn more](#)

Filter Enter property name or value

Notebook name ↑

tensorflow-2.7-20211209-091022

Region us-central1 (Iowa) Zone us-central1-a

**New notebook**

Notebook name \* tensorflow-2-3-20211209-130458

63-char limit with lowercase letters, digits, or \_ only. Must start with a letter. Cannot end with a \_.

Environment TensorFlow Enterprise 2.3 (with LTS and Intel® MKL-DNN/MKL)

Machine type 4 vCPUs, 15 GB RAM

Boot disk 100 GB Standard persistent disk

Data disk 100 GB Standard persistent disk

Subnetwork default(10.128.0/20)

External IP Ephemeral(Automatic)

Permission Compute Engine default service account

Estimated cost \$102.70 monthly, \$0.141 hourly

ADVANCED OPTIONS CANCEL CREATE

SHOW INFO PANEL

**Google Cloud Platform** CMPE260 Search products and resources

**FEATURES & INFO** SHORTCUT DISABLE EDITOR TABS

**Explorer** + ADD DATA

Type to search

Viewing pinned projects.

- cmpe260-334300
  - dataset\_5716044294586245...
  - export\_evaluated\_data\_items\_...
  - bqquery-public-data

**EDITOR** RUN SAVE SCHEDULE MORE

1

PERSONAL HISTORY PROJECT HISTORY SAVED QUERIES

**Create dataset**

Project ID cmpe260-334300 CHANGE

Dataset ID cc\_default

Letters, numbers, and underscores allowed

Data location us-central1 (Iowa)

**Default table expiration**

Enable table expiration

Default maximum table age Days

**Encryption**

Google-managed encryption key No configuration required

Customer-managed encryption key (CMK) Manage via Google Cloud Key Management Service

CREATE DATASET CANCEL

Google Cloud Platform CMPE260

Search products and resources

FEATURES & INFO SHORTCUT DISABLE EDITOR TABS

Explorer + ADD DATA

Type to search

Viewing pinned projects.

- cmpe260-334300
  - cc\_default
  - dataset\_57160442934586245...
  - export\_evaluated\_data\_items...
- bigrquery-public-data

Query "UNSAVE..."

CANCEL SAVE SCHEDULE MORE

Query running (5.2 sec - Stage: Preprocess)

```
1 CREATE OR REPLACE MODEL
2   `cmpe260-334300.cc.default.logistic_model` OPTIONS(model_type='LOGISTIC_REG',
3   input_label_cols=['default.payment_next_month']) AS
4   SELECT
5     limit_balance,
6     sex,
7     education_level,
8     marital_status,
9     age,
10    pay_0,
11    pay_1,
12    pay_2,
13    pay_3,
14    pay_4,
15    pay_5,
16    pay_6,
```

Query results

Job information Results Execution details

Elapsed time Slot time consumed Stages

3.9 sec — Preprocess

"cc\_default" created. GO TO DATASET

PERSONAL HISTORY PROJECT HISTORY SAVED QUERIES

Google Cloud Platform CMPE260

Search products and resources

FEATURES & INFO SHORTCUT DISABLE EDITOR TABS

Explorer + ADD DATA

Type to search

Viewing pinned projects.

- cmpe260-334300
  - cc\_default
    - Models(1)
      - logistic\_model
  - dataset\_57160442934586245...
  - export\_evaluated\_data\_items...
- bigrquery-public-data

Query "UNSAVE..."

RUN SAVE SCHEDULE MORE

This query will process 471 KiB (ML) when run.

```
1 CREATE OR REPLACE MODEL
2   `cmpe260-334300.cc.default.logistic_model` OPTIONS(model_type='LOGISTIC_REG',
3   input_label_cols=['default.payment_next_month']) AS
4   SELECT
5     limit_balance,
6     sex,
7     education_level,
8     marital_status,
9     age,
10    pay_0,
11    pay_1,
12    pay_2,
13    pay_3,
14    pay_4,
15    pay_5,
16    pay_6,
```

Query results

Query complete (41.5 sec elapsed, 471 KB (ML) processed)

Job information Results Execution details

This statement will create a new model named cmpe260-334300.cc\_default.logistic\_model. Depending on the type of model, this may take several hours to complete.

Go to model

PERSONAL HISTORY PROJECT HISTORY SAVED QUERIES

Google Cloud Platform CMPE260

Search products and resources

FEATURES & INFO SHORTCUT DISABLE EDITOR TABS

Explorer + ADD DATA

LOGIST... X

logistic\_model

DETAILS TRAINING EVALUATION SCHEMA

COMPOSE NEW QUERY

QUERY MODEL DELETE MODEL EXPORT MODEL

Type to search

Viewing pinned projects.

cmpe260-334300 cc\_default Models (1) logistic\_model dataset\_57160442934586245... export\_evaluated\_data\_items... bigquery-public-data

Model type LOGISTIC\_REGRESSION Data location US

**Model Details** EDIT

Model ID	cmpe260-334300.cc_default.logistic_model
Description	
Labels	
Date created	Thursday, December 9, 2021 at 7:17:35 PM GMT-08:00
Model expiration	Never
Date modified	Thursday, December 9, 2021 at 7:17:35 PM GMT-08:00
Data location	US
Model type	LOGISTIC_REGRESSION
Loss type	Mean log loss
Training data	TEMPORARY TRAINING DATA TABLE
Evaluation data	TEMPORARY EVALUATION DATA TABLE

Training Options

PERSONAL HISTORY PROJECT HISTORY SAVED QUERIES

Google Cloud Platform CMPE260

Search products and resources

FEATURES & INFO SHORTCUT DISABLE EDITOR TABS

Explorer + ADD DATA

LOGIST... X

logistic\_model

DETAILS TRAINING EVALUATION SCHEMA

Export model to Google Cloud Storage

Select GCS location \* BROWSE

SUBMIT CANCEL

Type to search

Viewing pinned projects.

cmpe260-334300 cc\_default Models (1) logistic\_model dataset\_57160442934586245... export\_evaluated\_data\_items... bigquery-public-data

Model type LOGISTIC\_REGRESSION Data location US

**Model Details** EDIT

Model ID	cmpe260-334300.cc_default.logistic_model
Description	
Labels	
Date created	Thursday, December 9, 2021 at 7:17:35 PM GMT-08:00
Model expiration	Never
Date modified	Thursday, December 9, 2021 at 7:17:35 PM GMT-08:00
Data location	US
Model type	LOGISTIC_REGRESSION
Loss type	Mean log loss
Training data	TEMPORARY TRAINING DATA TABLE
Evaluation data	TEMPORARY EVALUATION DATA TABLE

Training Options

PERSONAL HISTORY PROJECT HISTORY SAVED QUERIES

The screenshot shows the Google Cloud Platform interface. On the left, the 'Explorer' sidebar lists projects: 'cmpe260-334300' and 'cc\_default'. Under 'cc\_default', there is a 'Models (1)' folder containing a 'logistic\_model' file. The main panel displays the 'logistic\_model' details, including its type as 'LOGISTIC\_REGRESSION' and location as 'US'. Below this, the 'Model Details' section provides specific model parameters like Model ID, Description, Labels, Date created, Model expiration, Data location, Model type, Loss type, Training data, and Evaluation data. At the bottom, 'Training Options' tabs are visible. To the right, a 'Select object' dialog box is open, showing a list of buckets: 'artifacts.cmpe260-334300.appspot.com', 'cloud-ai-platform-65473780-0280-46f1-a6ec-d80226927f2b', and 'cmpe260-334300-bucket'. A red circle highlights the 'Create new bucket' button at the top right of the dialog.

This screenshot shows the continuation of the bucket creation process. The 'Create a bucket' dialog is open, with the 'Name your bucket' field set to 'cc\_logistics'. A tip below the field advises against including sensitive information. The 'LABELS (OPTIONAL)' section is collapsed. The 'CONTINUE' button is visible. The right side of the screen shows the same Google Cloud Platform interface as the first screenshot, with the 'Explorer' sidebar and the 'logistic\_model' details page. The 'logistic\_model' page has identical content to the first screenshot, including its details and training options.

Upgrade your account to avoid a break in service (\$151.76 credit and 5 days left in your trial).

Google Cloud Platform CMPE260 Search products and resources

FEATURES & INFO SHORTCUT DISABLE EDITOR TABS

Explorer + ADD DATA LOGISTI... X logistic\_model

Type to search

Viewing pinned projects.

- cmpe260-334300
- cc\_default
- Models (1)
  - logistic\_model
- dataset\_57160442934586245...
- export\_evaluated\_data\_items...
- bigrquery-public-data

DETAILS TRAINING EVALUATION SCHEMA

Model type LOGISTIC\_REGRESSION Data location US

**Model Details** EDIT

Model ID	cmpe260-334300.cc_default.logistic_model
Description	
Labels	
Date created	Thursday, December 9, 2021 at 7:17:35 PM GMT-08:00
Model expiration	Never
Date modified	Thursday, December 9, 2021 at 7:17:35 PM GMT-08:00
Data location	US
Model type	LOGISTIC_REGRESSION
Loss type	Mean log loss
Training data	TEMPORARY TRAINING DATA TABLE
Evaluation data	TEMPORARY EVALUATION DATA TABLE

**Training Options**

PERSONAL HISTORY PROJECT HISTORY SAVED QUERIES

Location us-central1 (Iowa) CONTINUE

**Choose a default storage class for your data**  
Default storage class: Standard

**Choose how to control access to objects**

Prevent public access  
Restrict data from being publicly accessible via the internet. Will prevent this bucket from being used for web hosting. [Learn more](#)  
 Enforce public access prevention on this bucket

Access control

Uniform  
Ensure uniform access to all objects in the bucket by using only bucket-level permissions (IAM). This option becomes permanent after 90 days. [Learn more](#)

Fine-grained  
Specify access to individual objects by using object-level permissions (ACLs) in addition to your bucket-level permissions (IAM). [Learn more](#)

CONTINUE

**Choose how to protect object data**  
Protection tools: None  
Data encryption: Google-managed key

CREATE CANCEL

Upgrade your account to avoid a break in service (\$151.76 credit and 5 days left in your trial)

Google Cloud Platform CMPE260 Search products and resources

FEATURES & INFO SHORTCUT DISABLE EDITOR TABS

Explorer + ADD DATA LOGISTI... X logistic\_model

Type to search

Viewing pinned projects.

- cmpe260-334300
- cc\_default
- Models (1)
  - logistic\_model
- dataset\_57160442934586245...
- export\_evaluated\_data\_items...
- bigrquery-public-data

DETAILS TRAINING EVALUATION SCHEMA

Model type LOGISTIC\_REGRESSION Data location US

**Model Details** EDIT

Model ID	cmpe260-334300.cc_default.logistic_model
Description	
Labels	
Date created	Thursday, December 9, 2021 at 7:17:35 PM GMT-08:00
Model expiration	Never
Date modified	Thursday, December 9, 2021 at 7:17:35 PM GMT-08:00
Data location	US
Model type	LOGISTIC_REGRESSION
Loss type	Mean log loss
Training data	TEMPORARY TRAINING DATA TABLE
Evaluation data	TEMPORARY EVALUATION DATA TABLE

**Training Options**

Export model to Google Cloud Storage

Exporting a model allows you to use this model in other systems/applications. Please select Cloud Storage path to export this model along with all asset files. [Learn more](#)

Select GCS location \*  cc\_logistics/model-assets BROWSE

SUBMIT CANCEL

Google Cloud Platform CMPE260

Bucket details

cc\_logistics

Location: us-central1 (Iowa) Storage class: Standard Public access: Not public Protection: None

OBJECTS CONFIGURATION PERMISSIONS PROTECTION LIFECYCLE

Buckets > cc\_logistics > model-assets

UPLOAD FILES UPLOAD FOLDER CREATE FOLDER MANAGE HOLDS DOWNLOAD DELETE

Filter by name prefix only ▾ Filter objects and folders

Name	Size	Type	Created	Storage class	Last modified	Public access	Version history	Encryption	Retention expiration date	Holds
assets/	—	Folder	—	—	—	—	—	—	—	—
saved.model.pb	74.2 KB	application/octet-stream	Dec 9, 2021...	Standard	Dec 9, 2021...	Not public	—	Google-managed key	—	None
variables/	—	Folder	—	—	—	—	—	—	—	—

Show deleted data

Google Cloud Platform CMPE260

Search products and resources

Looker

Data Catalog

Data Fusion

Financial Services

Healthcare

Life Sciences

Dataprep

ARTIFICIAL INTELLIGENCE

Vertex AI

AI Platform

Data Labeling

Document AI

Natural Language

Recommendation...

Retail

Dashboard

Datasets

Features

Labeling tasks

Workbench

Pipelines

Training

Experiments

Models

Endpoints

Batch predictions

Metadata

Objects and folders

FOLDER MANAGE HOLDS DOWNLOAD DELETE

Show deleted data

https://console.cloud.google.com/vertex-ai/models?project=cmpe260-334300

Google Cloud Platform CMPE260 Search products and resources REFRESH

Models + CREATE IMPORT

Models are built from your datasets or unmanaged data sources. There are many different types of machine learning models available on Vertex AI, depending on your use case and level of experience with machine learning. [Learn more](#)

Region us-central1 (Iowa)

Filter Enter a property name

Name	ID	Status	Data	Endpoints	Region	Type	Created	Notifications	Labels
iowa_daily_2021121002247	5256479219373834240	Ready	iowa_daily	0	us-central1	Tabular Forecasting AutoML	Dec 9, 2021, 4:28:37 PM		
beans-model-pipeline	344177925819465728	Ready	—	1 SHOW ENDPOINTS	us-central1	Imported Custom training	Dec 9, 2021, 1:42:22 PM		
beans-model-pipeline	6992616875725160448	Ready	—	1 SHOW ENDPOINTS	us-central1	Imported Custom training	Dec 9, 2021, 1:40:14 PM		

Upgrade your account to avoid a break in service (\$152.59 credit and 5 days left in your trial).

Google Cloud Platform CMPE260 Search products and resources

Models + CREATE IMPORT

Models are built from your datasets or unmanaged data sources. There are many different types of machine learning models available on Vertex AI, depending on your use case and level of experience with machine learning. [Learn more](#)

Region us-central1 (Iowa)

Filter Enter a property name

Name	ID	Status	Data	Endpoints
iowa_daily_2021121002247	5256479219373834240	Ready	iowa_daily	0
beans-model-pipeline	344177925819465728	Ready	—	1 SHOW ENDPOINTS
beans-model-pipeline	6992616875725160448	Ready	—	1 SHOW ENDPOINTS

**Import model**

1 Name and region  
2 Model settings  
3 Explainability (optional)

IMPORT CANCEL

You can import model artifacts that have been trained outside of Google Cloud. Once your model has been imported, you can serve it for online or batch predictions and compare it against your other Cloud AI models. [More info](#)

Model name \* predict\_default

Region us-central1 (Iowa)

ADVANCED OPTIONS CONTINUE

Upgrade your account to avoid a break in service (\$152.59 credit and 5 days left in your trial).

Google Cloud Platform CMPE260 Search products and resources

**Models** + CREATE IMPORT

Models are built from your datasets or unmanaged data sources. There are many different types of machine learning models available on Vertex AI, depending on your use case and level of experience with machine learning. [Learn more](#)

Region: us-central1 (Iowa) Filter: Enter a property name

Name	ID	Status	Data	Endpoints
iowa_daily_2021121002247	5256479219373834240	Ready	iowa_daily	0
beans-model-pipeline	344177925819465728	Ready	—	1 SHOW ENDPOINTS
beans-model-pipeline	6992616875725160448	Ready	—	1 SHOW ENDPOINTS

**Import model**

1 Name and region (checked)

2 Model settings (selected)

3 Explainability (optional)

**IMPORT** CANCEL

**Import model artifacts into a new pre-built container**  
View the list of [supported runtimes](#) including TensorFlow, scikit-learn and XGBoost versions

**Import an existing custom container**  
Build a custom Docker container. Must be stored in [Container Registry](#) or [Artifact Registry](#)

**Pre-built container settings**

In order to run in a pre-built container, your code needs to be in Python 3.7

Model framework \*: TensorFlow (selected)

Model framework version \*: 2.3

Accelerator type \*: None

Model artifact location (Cloud storage path) \*: gs://cc-logistics/model-assets/ (selected) BROWSE

Path to the Cloud Storage directory where the exported model file is stored (not the path to the model file itself). The model name must be one of: saved\_model.pb, model.pkl, model.joblib, or model.bst, depending on which library you used.

**Predict schema**

Optional. [Learn more about the predict schema](#)

gs:// Instances (selected) BROWSE

Cloud Storage location to a YAML file that defines the format of a single instance used in prediction and explanation requests.

gs:// Parameters (selected) BROWSE

Cloud Storage location to a YAML file that defines the prediction and explanation parameters.

Google Cloud Platform CMPE260 Search products and resources

**Models** + CREATE IMPORT

REFRESH

Models are built from your datasets or unmanaged data sources. There are many different types of machine learning models available on Vertex AI, depending on your use case and level of experience with machine learning. [Learn more](#)

Region: us-central1 (Iowa) Filter: Enter a property name

Name	ID	Status	Data	Endpoints	Region	Type	Created	Notifications	Labels
<b>predict_default</b>	3559748059761999872	Ready	—	0	us-central1	Imported Custom training	Dec 9, 2021, 8:57:56 PM	...	⋮
iowa_daily_2021121002247	5256479219373834240	Ready	iowa_daily	0	us-central1	Tabular Forecasting AutoML	Dec 9, 2021, 4:28:37 PM	...	⋮
beans-model-pipeline	344177925819465728	Ready	—	1 SHOW ENDPOINTS	us-central1	Imported Custom training	Dec 9, 2021, 1:42:22 PM	...	⋮
beans-model-pipeline	6992616875725160448	Ready	—	1 SHOW ENDPOINTS	us-central1	Imported Custom training	Dec 9, 2021, 1:40:14 PM	...	⋮

Google Cloud Platform CMPE260 Search products and resources

**predict\_default** EXPORT

DEPLOY & TEST BATCH PREDICTIONS MODEL PROPERTIES

**Deploy your model**

Endpoints are machine learning models made available for online prediction requests. Endpoints are useful for timely predictions from many users (for example, in response to an application request). You can also request batch predictions if you don't need immediate results.

**DEPLOY TO ENDPOINT**

Name	ID	Status	Models	Region	Monitoring	Most recent monitoring job	Most recent alerts	Last updated	API	Notification	Labels	Encryption
No active endpoints containing this model												

**Test your model** PREVIEW

In order to test your model, you will need to deploy it first. [Pricing guide](#)

☰



Upgrade your account to avoid a break in service (\$152.59 credit and 5 days left in your trial).

Google Cloud Platform CMPE260 Search products and resources

**predict\_default** EXPORT

DEPLOY & TEST BATCH PREDICTIONS MODEL PROPERTIES

**Deploy your model**

Endpoints are machine learning models made available for online prediction requests. Endpoints are useful for timely predictions from many users (for example, in response to an application request). You can also request batch predictions if you don't need immediate results.

**DEPLOY TO ENDPOINT**

Name	ID	Status	Models	Region	Monitoring	Most recent monitoring job
No active endpoints containing this model						

**Deploy to endpoint**

1 Define your endpoint  
2 Model settings  
3 Model monitoring

DEPLOY CANCEL

**Create new endpoint**  Add to existing endpoint

Endpoint name: default\_predict

**Location**

Region: us-central1 (Iowa)

**Access**

Determines how your endpoint can be accessed. By default, endpoints are available for prediction serving through a REST API. Endpoint access can't be changed after the endpoint is created.

Standard Makes the endpoint available for prediction serving through a REST API. AutoML and custom-trained models can be added to standard endpoints.

Private Create a private connection to this endpoint using a VPC network and [private services access](#). Only custom-trained and tabular models can be added to private endpoints. [Learn more](#)

▼ ADVANCED OPTIONS

CONTINUE



Upgrade your account to avoid a break in service (\$152.59 credit and 5 days left in your trial).

Google Cloud Platform CMPE260 Search products and resources

**predict\_default** DEPLOY & TEST BATCH PREDICTIONS MODEL PROPERTIES

**Deploy to endpoint**

- Define your endpoint
- Model settings
- Model monitoring

**DEPLOY CANCEL**

**Deploy your model**

Endpoints are machine learning models made available for online prediction requests. Endpoints are useful for timely predictions from many users (for example, in response to an application request). You can also request batch predictions if you don't need immediate results.

**DEPLOY TO ENDPOINT**

Name	ID	Status	Models	Region	Monitoring	Most recent monitoring job	Most recent alerts	Last updated	API	Notification	Labels	Encryption
No active endpoints containing this model												

**Test your model** PREVIEW

In order to test your model, you will need to deploy it first. [Pricing guide](#)

**Compute resources**

Traffic split \* 100

Choose how compute resources will serve prediction traffic to your model

- Autoscaling: If you set a minimum and maximum, compute nodes will scale to meet traffic demand within those boundaries
- No scaling: If you only set a minimum, then that number of compute nodes will always run regardless of traffic demand (the maximum will be set to minimum)

Once scaling settings are set, they can't be changed unless you redeploy the model. [Pricing guide](#)

Minimum number of compute nodes \* 1

Default is 1. If set to 1 or more, then compute resources will continuously run even without traffic demand. This can increase cost but avoid dropped requests due to node initialization.

Maximum number of compute nodes (optional) Enter a number equal to or greater than the minimum nodes. Can reduce costs but may cause reliability issues for high traffic.

**ADVANCED SCALING OPTIONS**

Machine type \* n1-highcpu-2, 2 vCPUs, 1.8 GiB memory

Service account

A service account determines what Google Cloud resources your service code can

Google Cloud Platform CMPE260 Search products and resources

**predict\_default** DEPLOY & TEST BATCH PREDICTIONS MODEL PROPERTIES

**Deploy your model**

Endpoints are machine learning models made available for online prediction requests. Endpoints are useful for timely predictions from many users (for example, in response to an application request). You can also request batch predictions if you don't need immediate results.

**DEPLOY TO ENDPOINT**

Name	ID	Status	Models	Region	Monitoring	Most recent monitoring job	Most recent alerts	Last updated	API	Notification	Labels	Encryption
default_pred_v1	2407253165666729984	Deploying model	0	us-central1	—	—	—	Dec 9, 2021, 9:02:48 PM	Sample request			Google-managed key

**Test your model** PREVIEW

Your model must be successfully deployed to an endpoint before you can test it.

Your JSON request must contain an `instances` field and an optional `parameters` field if you're using a custom container. No other fields can be present in the JSON request. [Learn how to format your JSON request.](#)

JSON request	Response
{	

Google Cloud Platform CMPE260

predict\_default EXPORT

DEPLOY & TEST BATCH PREDICTIONS MODEL PROPERTIES

**Deploy your model**

Endpoints are machine learning models made available for online prediction requests. Endpoints are useful for timely predictions from many users (for example, in response to an application request). You can also request batch predictions if you don't need immediate results.

**DEPLOY TO ENDPOINT**

Name	ID	Status	Models	Region	Monitoring	Most recent monitoring job	Most recent alerts	Last updated	API	Notification	Labels	Encryption
default_pred_v1	2407253165666729984	Active	1	us-central1	Disabled	—	—	Dec 9, 2021, 9:09:16 PM	Sample request			Google-managed key

**Test your model** PREVIEW

Your JSON request must contain an instances field and an optional parameters field if you're using a custom container. No other fields can be present in the JSON request. Learn how to format your JSON request.

**JSON request**

```
{
  "instances": [
    {
      "sample_key": "sample_value"
    }
  ]
}
```

**Response**

Upgrade your account to avoid a break in service (\$152.59 credit and 5 days left in your trial).

Google Cloud Platform CMPE260

predict\_default EXPORT

DEPLOY & TEST BATCH PREDICTIONS MODEL PROPERTIES

**Deploy your model**

Endpoints are machine learning models made available for online prediction requests. Endpoints are useful for timely predictions from many users (for example, in response to an application request). You can also request batch predictions if you don't need immediate results.

**DEPLOY TO ENDPOINT**

Name	ID	Status	Models	Region	Monitoring	Most recent monitoring job	Most recent alerts	Last updated
default_pred_v1	2407253165666729984	Active	1	us-central1	Disabled	—	—	Dec 9, 2021, 9:09:16 PM

**Test your model** PREVIEW

Your JSON request must contain an instances field and an optional parameters field if you're using a custom container. No other fields can be present in the JSON request. Learn how to format your JSON request.

**JSON request**

```
{
  "instances": [
    {
      "sample_key": "sample_value"
    }
  ]
}
```

**Response**

**Sample Request**

**REST** **PYTHON**

You can now execute queries using the command line interface (CLI).

1. Make sure you have the [Google Cloud SDK](#) installed.
2. Run the following command to authenticate with your Google account.

```
$ gcloud auth application-default login
```

3. Create a JSON object to hold your data.

```
{
  "instances": [
    {
      "instance_key_1": "value", ...
    }
  ],
  "parameters": { "parameter_key_1": "value", ... }, ...
}
```

4. Create environment variables to hold your endpoint and project IDs, as well as your JSON object.

```
$ ENDPOINT_ID="2407253165666729984"
$ PROJECT_ID="cmpe260-33430"
$ INPUT_DATA_FILE="INPUT-JSON"
```

5. Execute the request.

```
$ curl \
-X POST \
-H "Authorization: Bearer $(gcloud auth print-access-token)" \
-H "Content-Type: application/json" \
https://us-central1-aiplatform.googleapis.com/v1/projects/$PROJECT_ID/models/default_pred_v1:predict \
-d @"$INPUT_DATA_FILE"
```

**DONE**

The screenshot shows a Jupyter Notebook environment with a file browser on the left and a terminal window on the right.

**File Browser:**

- Path: / cassava / trainer /
- Items:
  - default-pred.json (modified 2 minutes ago)
  - task.py (modified 4 hours ago)
  - Untitled.ipynb (modified seconds ago)

**Terminal:**

```
[1]: %writemode default-pred.json
{
  "instances": [
    {"age": 39,
     "bill_amt_1": 47174,
     "bill_amt_2": 47974,
     "bill_amt_3": 48630,
     "bill_amt_4": 54803,
     "bill_amt_5": 30789,
     "bill_amt_6": 15874,
     "education_level": "1",
     "limit_balance": 50000,
     "marital_status": "2",
     "pay_0": 0,
     "pay_2": 0,
     "pay_3": 0,
     "pay_4": 0,
     "pay_5": "0",
     "pay_6": "0",
     "pay_ant_1": 1800,
     "pay_ant_2": 2000,
     "pay_ant_3": 3000,
     "pay_ant_4": 2000,
     "pay_ant_5": 2000,
     "pay_ant_6": 2000,
     "sex": "1"
   ]
}
Writing default-pred.json

[2]: ENDPOINT_ID="2407253165666729984"
PROJECT_ID="cmpe269-334300"
INPUT_DATA_FILE="INPUT-JSON"
```

Mode: Command Ln 1, Col 1 Untitled.ipynb

The screenshot shows a Jupyter Notebook environment. On the left, there's a file browser with a folder named 'cassava / trainer /' containing files: 'default-pred.json' (4 minutes ago), 'task.py' (4 hours ago), and 'Untitled.ipynb' (in a few seconds). The main area has two tabs: 'Terminal 2' and 'Untitled.ipynb'. The 'Untitled.ipynb' tab contains the following code:

```
[3]: REGION="us-central1"
[4]: !curl \
-X POST \
-H "Authorization: Bearer $(gcloud auth print-access-token)" \
-H "Content-Type: application/json" \
https://us-central1-prediction-aiplatform.googleapis.com/v1beta1/projects/$PROJECT_ID/locations/$REGION/endpoints/$ENDPOINT_ID:predict \
-d "@default-pred.json"

{
  "predictions": [
    {
      "predicted_default_payment_next_month": [
        "0"
      ],
      "default_payment_next_month_values": [
        "1",
        "0"
      ],
      "default_payment_next_month_probs": [
        0.1808152952977791,
        0.81918470470022209
      ]
    }
  ],
  "deployedModelId": "1507694324471562240",
  "model": "projects/196373151126/locations/us-central1/models/3559748059761999872",
  "modelDisplayName": "predict_default"
}
[ ]:
```

The screenshot shows the Google Cloud Platform Notebooks page for the project 'CMPE260'. The top navigation bar includes 'Google Cloud Platform' and 'CMPE260'. Below the navigation, there are tabs for 'Notebooks' (selected), 'NEW NOTEBOOK', 'REFRESH', 'START', 'STOP' (highlighted in yellow), 'RESET', and 'DELETE'. There are also tabs for 'MANAGED NOTEBOOKS' (PREVIEW), 'USER-MANAGED NOTEBOOKS', 'EXECUTIONS' (PREVIEW), and 'SCHEDULES' (PREVIEW). A message at the top states: 'As of the M80 DLVM release, all environments will include JupyterLab 3.x by default. To continue using an existing environment's JupyterLab 1.x version, disable auto-upgrade (if enabled) and do not manually upgrade the environment to a new environment version. To create new Notebooks with JupyterLab 1.x installed, see creating specific versions of Notebooks.' Below this, it says 'Notebooks have JupyterLab pre-installed and are configured with GPU-enabled machine learning frameworks. Learn more'. A filter bar allows entering a property name or value. The main table lists two notebooks:

	Notebook name	Zone	Auto upgrade	Environment	Machine type	GPUs	Permission	Last modified	
<input type="checkbox"/>	tensorflow-2-3-20211209-130458	OPEN JUPYTERLAB	us-central1-a	—	TensorFlow:2.3	4 vCPUs, 15 GB RAM	None	Service account	Dec 9, 2021, 3:35:42 PM
<input checked="" type="checkbox"/>	tensorflow-2-7-20211208-091022	OPEN JUPYTERLAB	us-central1-a	—	TensorFlow:2.7	4 vCPUs, 15 GB RAM	None	Service account	Dec 9, 2021, 3:39:09 PM