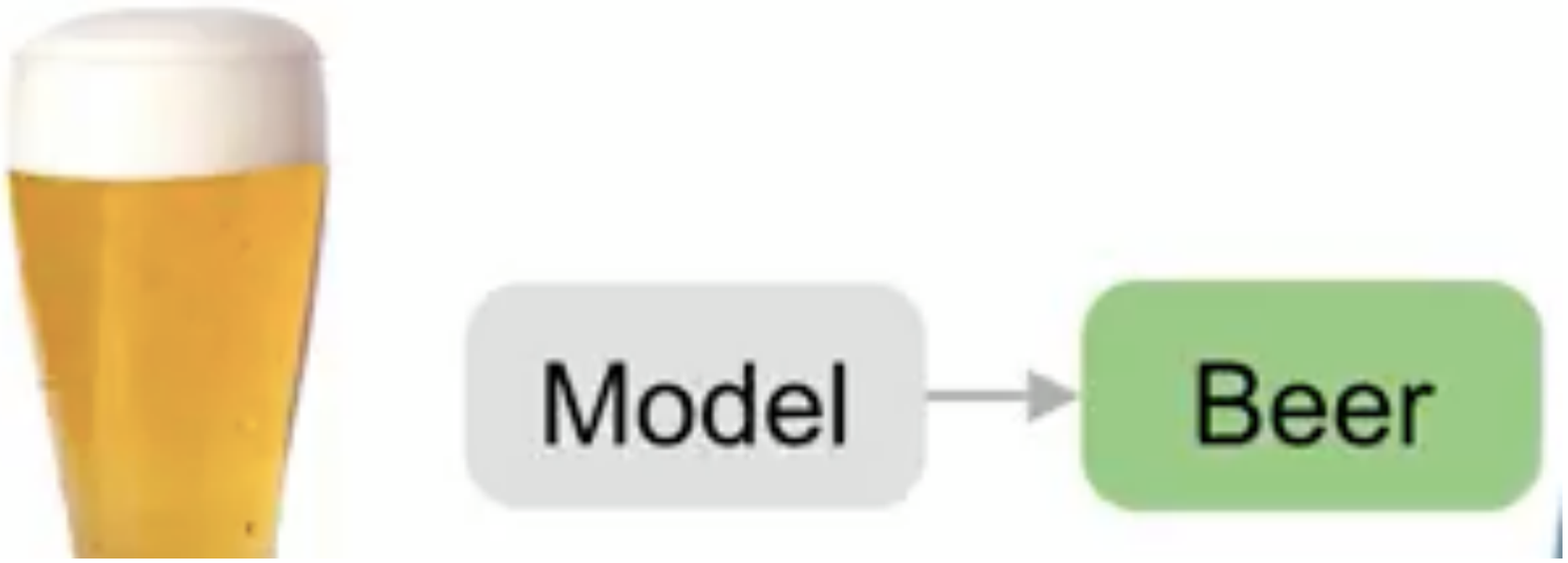
# Google Cloud Machine Learning

Telung Pan Ph.D.
telung@mac.com


Model
Beer

COLOR

13.5% Alc/volume

ALCOHOL

# Could ML TensorFlow Tasks

Google Could ML

TensorFlow Task

# Verify the Google Cloud SDK components

- To verify that the Google Cloud SDK components are installed:

- List the models to verify that the command returns an empty list

  ***gcloud ml-engine models list***

- Verify that the command returns an empty list:

  Listed 0 items.

  After you start creating models, you can see them listed by using this command.

# Google Cloud Shell

- Navigate to the cloudml-samples-master > census > estimator directory. The commands in this walkthrough must be run from the estimator directory.

  ***cd cloudml-samples-master/census/estimator***

# Get your training data

- The relevant data files, adult.data and adult.test, are hosted in a public Google Cloud Storage bucket.You can read them directly from Cloud Storage or copy them to your local environment. For purposes of this sample you will download the samples for local training, and later upload them to your own Cloud Storage bucket for cloud training.

- Download the data to a local file directory and set variables that point to the downloaded data files.

  ***mkdir* 學號**

  ***gsutil -m cp gs://cloudml-public/census/data/** 學號/**

- Set the TRAIN_DATA AND EVAL_DATA variables to your local file paths. For example, the following commands set the variables to local paths.

- pwd UNIX 作業系統輸出工作目錄的絕對路徑

  ***TRAIN_DATA=$(pwd)/data/adult.data.csv***
  ***EVAL_DATA=$(pwd)/data/adult.test.csv***

- 39, State-gov, 77516, Bachelors, 13, Never-married, Adm-clerical, Not-in-family, White, Male, 2174, 0, 40, United-States, <=50K

- 50, Self-emp-not-inc, 83311, Bachelors, 13, Married-civ-spouse, Exec-managerial, Husband, White, Male, 0, 0, 13, United-States, <=50K

- 38, Private, 215646, HS-grad, 9, Divorced, Handlers-cleaners, Not-in-family, White, Male, 0, 0, 40, United-States, <=50K

- 53, Private, 234721, 11th, 7, Married-civ-spouse, Handlers-cleaners, Husband, Black, Male, 0, 0, 40, United-States, <=50K

# Install dependencies

- The sample provides a requirements.txt file that you can use to install the dependencies required by the project.

  ***sudo pip install -r ../requirements.txt***

# Run a local trainer

- A local trainer loads your Python training program and starts a training process in an environment that's similar to that of a live Cloud ML Engine cloud training job.

- Specify an output directory and set a MODEL_DIR variable. The following command sets MODEL_DIR to a value of output.

  MODEL_學號=output_g9220812

- 2 It's a good practice to delete the contents of the output directory in case data remains from a previous training run. The following command deletes all data in the output directory.

  rm -rf $MODEL_學號/*

- To run your training locally, run the following command:

```
gcloud ml-engine local train \
    --module-name trainer.task \
    --package-path trainer/ \
    -- \
    --train-files $TRAIN_DATA \
    --eval-files $EVAL_DATA \
    --train-steps 1000 \
    --job-dir $MODEL_DIR \
    --eval-steps 100
```

- By default, verbose logging is turned off. You can enable it by setting the --verbosity tag to DEBUG. You'll enable it in a later example command.

# Inspect the summary logs using Tensorboard

- To see the evaluation results, you can use Tensorboard, which is available as part of the TensorFlow installation.

- Launch TensorBoard:

  ***python -m tensorflow.tensorboard —logdir=output_g9220812 --port=8080***

- Select "Preview on port 8080" from the Web Preview menu at the top of the command-line.

# 補充資料

- https://cloud.google.com/ml-engine/docs/getting-started-training-prediction

# Python

- https://www.python.org/downloads/windows/

# TensorFlow

- The central unit of data in TensorFlow is the **tensor.**

- A tensor consists of a set of primitive values shaped into any array of any number of dimensions.

- A tensor's **rank** is its number of dimensions.

```
3 # a rank 0 tensor; a scalar with shape []
[1., 2., 3.] # a rank 1 tensor; a vector with shape [3]
[[1., 2., 3.], [4., 5., 6.]] # a rank 2 tensor; a matrix with shape [2, 3]
[[[1., 2., 3.]], [[7., 8., 9.]]] # a rank 3 tensor with shape [2, 1, 3]
```

# Installing TensorFlow on macOS

- Install pip and virtualenv:
  sudo easy_install pip
  sudo pip install —upgrade virtualenv

- 建立 virtualenv 的工作目錄
  mkdir tensorflow

- Create a virtualenv environment
  virtualenv —system-site-packages tensorflow

- 啟動虛擬環境
  source ~/tensorflow/bin/activate

```
pandelongde-MacBook-Air:~ telung$ source ~/tensorflow/bin/activate
(tensorflow) pandelongde-MacBook-Air:~ telung$
```

- Ensure pip >= 8.1 is installed:
  *easy_install -U pid*

- Install TensorFlow and all the packages that TensorFlow requires into the active Virtualenv environment:
  *pip install —upgrade tensorflow*

- Validate your installation
  python

- Enter the following program inside the python interactive shell:

```
# Python
import tensorflow as tf
hello = tf.constant('Hello, TensorFlow!')
sess = tf.Session()
print(sess.run(hello))
```

# Exercise

- Gives Python access to all of TensorFlow's classes, methods, …

  *import tensor flow as tf*

- Create tow floating point Tensors node1 and node2 as follows:

# Cloud ML Engine sample file

- Census Income Data Set
- cd cloudml-samples-master/census/estimator
- ls
- mkdir data

# gsutil 工具

- gsutil 是 Python 編寫的應用程式，可以在命令列環境存取 Google Cloud Storage
  - 建立與刪除 bucket
  - 上載、下載與刪除 bucket 中的物件
  - 列出 bucket 物件
  - 移動、複製以及更名物件
  - 編輯物件與 bucket 的 ACL

- Create a bucket:
  *gsutil mb gs://my-0812-bucket/*

- *mkdir data*

- Use the *gsutil cp* command to copy file
  *-m* to perform a parallel (multi-threaded) copy
  Download the data to a local file directory:
  *gsutil -m cp gs://cloudml-public/census/data/* data/*

- gsutil ls 列出 bucket 中的物件
  例如：gsutil ls gs://uspto-pair/applications/0800401*

- cp: Copy files and objects

- 將本機的所有文字檔案複製到雲端 bucket
  gsutil cp *.txt gs://my-bucket

- Set variables point to the downloaded data files, set the TRAIN_DATA and EVAL_DATA variables to your local file paths

  *TRAIN_DATA=$(pwd)/data/adult.data.csv*

  *EVAL_DATA=$(pwd)/data/adult.test.csv*

```
39, State-gov, 77516, Bachelors, 13, Never-married, Adm-clerical, Not-in-family, White, Male, 2174,
0, 40, United-States, <=50K
50, Self-emp-not-inc, 83311, Bachelors, 13, Married-civ-spouse, Exec-managerial, Husband,
White, Male, 0, 0, 13, United-States, <=50K
38, Private, 215646, HS-grad, 9, Divorced, Handlers-cleaners, Not-in-family, White, Male, 0, 0, 40,
United-States, <=50K
53, Private, 234721, 11th, 7, Married-civ-spouse, Handlers-cleaners, Husband, Black, Male, 0, 0,
40, United-States, <=50K
```

# Set up Cloud Storage bucket

- The Cloud ML Engine services need to access Cloud Storage locations to read and write data during model training and batch prediction.

- Create a Google Cloud Storage bucket for reading and writing data during model training and batch prediction:

1. Set a name for your new bucket, 專案名稱附加學號後4碼：
   PROJECT_ID=$(gcloud config list project --format "value(core.project)")
   BUCKET_NAME=${PROJECT_ID}-0812
   或 BUCKET_NAME="你的名稱"

2. 檢查所建立的 bucket 名稱
   echo $BUCKET_NAME

3. 建立新的 bucket
   gsutil mb -l $REGION gs://$BUCKET_NAME

- Upload the data files to your Cloud Storage bucket
  - Use gustily to copy the two files (*adult.data.csv, adult.test.csv*) to your Cloud Storage bucket
    *gsutil cp -r data gs://$BUCKET_NAME/data*
  - 設定讓 TRAIN_DATA 與 EVAL_DATA 分別指向對應的檔案
    TRAIN_DATA=gs://$BUCKET_NAME/data/adult.data.csv
    EVAL_DATA=gs://$BUCKET_NAME/data/adult.test.csv
  - Use gsutil again to copy the JSON test file test.json to your Cloud Storage bucket.
    *gsutil cp ../test.json gs://$BUCKET_NAME/data/test.json*

- Set the TEST_JSON variable to point to that file.

TEST_JSON=gs://$BUCKET_NAME/data/test.json

# Run a singe-instance trainer in the cloud

- JOB_NAME=census_single_1

- 安装 tensorflow-1.2.0

sudo pip install -r ../requirements.txt

# Set up environment

- gcloud beta ml init-project
- Sudo pip install tensorflow==1.2.0 —ignore-installed
- https://cloud.google.com/ml-engine/docs/getting-started-training-prediction#about-data