Serverless Machine Learning - Lab 6 : Scaling up ML using Cloud ML Engine v1.3

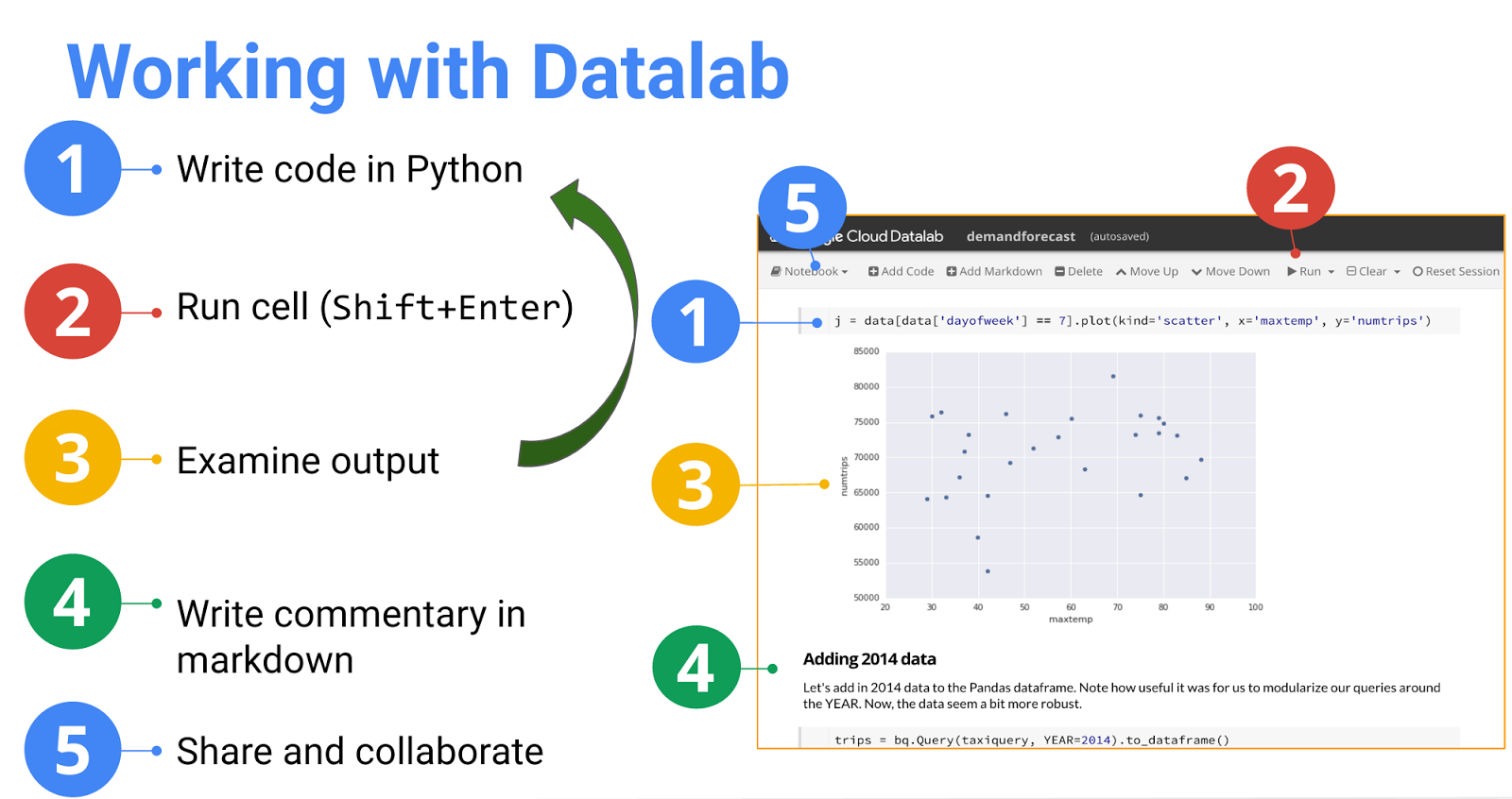
## Overview

In this lab, you will perform the following tasks:

* Package up the code
* Find absolute paths to data
* Run the Python module from the command line
* Run locally using gcloud
* Submit training job using gcloud
* Deploy model
* Prediction
* Train on a 1-million row dataset

## Setup Datalab

You will be using Datalab. The following is a graphical cheat sheet for the main Datalab functionality:



* Create and run a query
* Modify the query to add clauses, subqueries, built-in functions and joins.

## Task 1. Launch Cloud Datalab

To launch Cloud Datalab:

1. In **Cloud Shell**, type:

datalab create dataengvm --zone us-central1-a

1. Datalab setup will prompt you to continue. Enter 'Y'.

**Example**

Connecting to dataengvm.

This will create an SSH tunnel and may prompt you to create an rsa key pair. To manage these keys, see https://cloud.google.com/compute/docs/in

stances/adding-removing-ssh-keys

Waiting for Datalab to be reachable at http://localhost:8081/

This tool needs to create the directory

[/home/yourprojectid\_student/.ssh] before being able to generate SSH

keys.

Do you want to continue (Y/n)? Y

1. Datalab setup will ask you for a passphrase. You can press **Enter** twice.

**Example**

Enter passphrase (empty for no passphrase):

Enter same passphrase again:

1. Datalab will take about five minutes to start. Datalab is ready when you see a message prompting you to do a "Web Preview".

**Example**

The connection to Datalab is now open and will remain until this command is killed.

Click on the \*Web Preview\* (square button at top-right), select \*Change port > Port 8081\*, and start using Datalab.

The connection to your Datalab instance remains open for as long as the datalab command is active. If the cloud shell used for running the datalab command is closed or interrupted, the connection to your Cloud Datalab VM will terminate. If that happens, you may be able to reconnect using the command **datalab connect dataengvm** in your new Cloud Shell.

## Task 2. Checkout notebook into Cloud Datalab

1. Click on the **Web Preview** icon on the top-right corner of the Cloud Shell ribbon.



1. Click on **Change Port**.
2. In the **Change Preview Port** dialog, in the **Port Number** box, enter **8081**.
3. Click **Change and Preview**.

### **Continue in the Cloud Datalab tab**

1. In Cloud Datalab **Home** page (browser), open a new notebook using the icon 5fdee4bbcdee4b9a.png on the top left.
2. In the new notebook, enter the following commands in the cell, and click on **Run** (on the top navigation bar) to run the commands:

%bash

git clone https://github.com/GoogleCloudPlatform/training-data-analyst

cd training-data-analyst

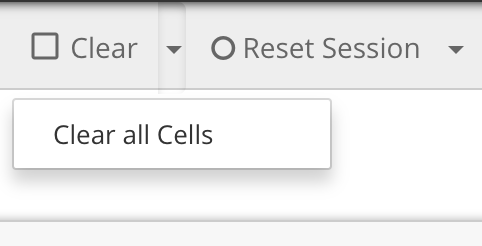
1. Confirm that you have cloned the repo by going back to Datalab browser, and ensure you see the **training-data-analyst** directory.

For the remainder of the labs, you will be loading Datalab python notebooks and performing the lab using the instructions in each notebook.

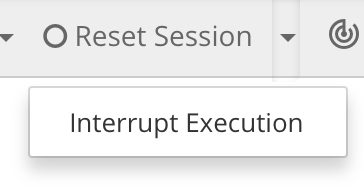
**Troubleshooting**

What do if a lab hangs or fails:

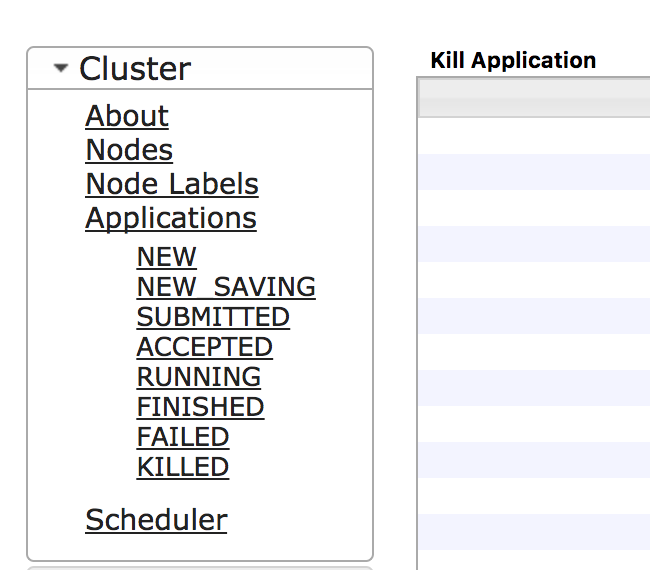
If the lab fails, you can click **Clear** In Datalab to clear the cell output and try again.



In some cases you may need to reset the Python 2 kernel from within the Datalab. Resetting the kernel causes the job in progress to change state to FINISHED and to have its FinalStatus marked as SUCCEEDED.



If a job is stuck in execution, you can browse to the Hadoop Applications interface, and click on a job that is running. In the upper left corner there is a link that says **Kill Application**.



## Task 3: Verify that you have a Cloud Storage bucket

The next lab notebook requires you to provide a **Project ID**, **Bucket Name** and **Bucket Region**.

You can recall the Project ID in Cloud Shell by entering the following:

echo $DEVSHELL\_PROJECT\_ID

You should have a bucket from the previous lab. If you don't you can follow these instructions to create a bucket.

The bucket must be regional. A Multi-Regional bucket will not work with this lab. If your existing buckets are Multi-Regional, you will need to create a new one for this lab.

1. In the Console, on the **Navigation menu** ( 7a91d354499ac9f1.png) click **Home**.
2. **Select and copy** the Project ID. For simplicity you will use the Qwiklabs Project ID, which is already globally unique, as the bucket name. If that is already taken, create a globally unique bucket name.
3. In the Console, on the **Navigation menu** ( 7a91d354499ac9f1.png) click **Storage** > **Browser**.
4. Click **Create Bucket**.
5. Specify the following, and leave the remaining settings as their defaults:

|  |  |
| --- | --- |
| **Property** | **Value (type value or select option as specified)** |
| **Name** | <your unique bucket name (Project ID)> |
| **Default storage class** | [x] Regional |
| **Location** | <Your location> **Valid Regions for this lab are** asia-east1, europe-west1, us-central1, us-east1, us-west1 |

1. Click **Create**.
2. Record the name of your bucket and the location. You will need it in subsequent tasks.

## Task 4: Continue the lab in the notebook

1. The home icon in Datalab looks like this:



1. In Cloud Datalab, click on the **Home** icon, and then navigate to **datalab/training-data-analyst/courses/machine\_learning/cloudmle**.
2. Open **cloudmle.ipynb**.
3. In Cloud Datalab, click on **Clear | Clear all Cells** (click on **Clear** drop-down menu, select **Clear all Cells**).
4. Read the narrative and execute each cell in turn.