# [ML on GCP C10] Collaborative Filtering on Google Analytics data

1 hour 30 minutes      1 Credit                  Rate Lab

## Overview

This lab shows you how to do collaborative filtering with Weighted Alternating Least Squares (WALS) matrix refactorization approach.

### Objectives

In this lab, you learn to perform the following tasks:

- Prepare the user-item matrix for use with WALS

- Train a WALSMatrixFactorization within TensorFlow locally and on Cloud ML Engine

- Visualize the embedding vectors with principal components analysis

## Introduction

In this lab, you'll be providing article recommendations for users based on the Kurier.at data. Recall that collaborative filtering doesn't need to know anything about the content. We are only interested in the user-item matrix which defines their relationships.

## Setup

For each lab, you get a new GCP project and set of resources for a fixed time at no cost.

1. Make sure you signed into Qwiklabs using an **incognito window**.

2. Note the lab's access time (for example, `02:00:00` and make sure you can

finish in that time block.

There is no pause feature. You can restart if needed, but you have to start at the beginning.

3. When ready, click **START LAB** .

4. Note your lab credentials. You will use them to sign in to Cloud Platform Console.

**Open Google Console**

**Caution:** When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. **Learn more.**

**Username**
google2876526_student@qwiklabs.n

**Password**
TG959yrKDX

**GCP Project ID**
qwiklabs-gcp-0855e773352d3560

**New to labs? View our introductory video!**

5. Click **Open Google Console**.

6. Click **Use another account** and copy/paste credentials for **this** lab into the prompts.

If you use other credentials, you'll get errors or **incur charges**.

7. Accept the terms and skip the recovery resource page.

Do not click **End Lab** unless you are finished with the lab or want to restart it. This clears your work and removes the project.
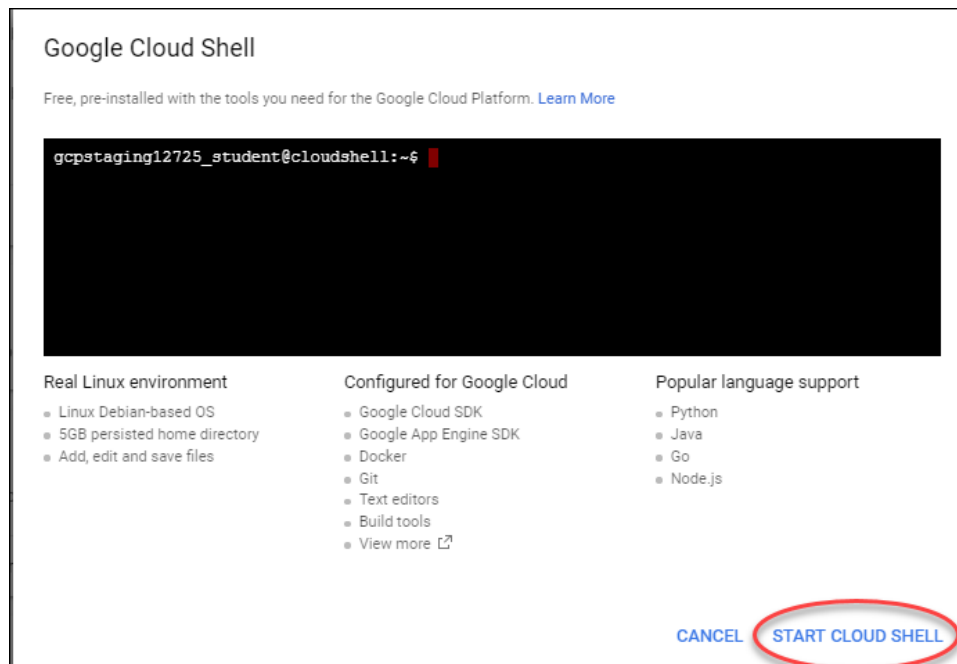
## Activate Google Cloud Shell

Google Cloud Shell provides command-line access to your GCP resources.

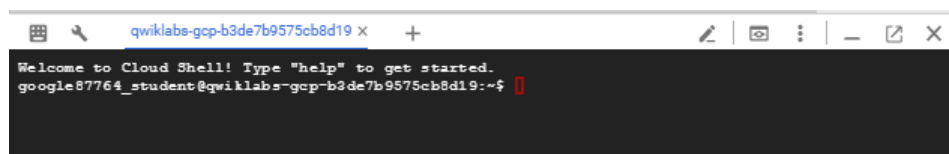From the GCP Console click the **Cloud Shell** icon on the top right toolbar:



Then click **START CLOUD SHELL**:



You can click **START CLOUD SHELL** immediately when the dialog comes up instead of waiting in the dialog until the Cloud Shell provisions.

It takes a few moments to provision and connects to the environment:



The Cloud Shell is a virtual machine loaded with all the development tools you'll need. It offers a persistent 5GB home directory, and runs on the Google Cloud, greatly enhancing network performance and authentication.

Once connected to the cloud shell, you'll see that you are already authenticated and the project is set to your *PROJECT_ID*:

```
gcloud auth list
```

Output:

```
Credentialed accounts:
 - <myaccount>@<mydomain>.com
(active)
```

**Note:** `gcloud` is the powerful and unified command-line tool for Google Cloud Platform. Full documentation is available on [Google Cloud gcloud Overview](). It comes pre-installed on Cloud Shell and supports tab-completion.

```
gcloud config list project
```

Output:

```
[core]
project = <PROJECT_ID>
```

# Create Storage Bucket

*Duration is 2 min*

Create a bucket using the GCP console:

**Step 1**

In your GCP Console, click on the **Navigation menu** ( ≡ ), and

select **Storage**.

**Step 2**

Click on **Create bucket**.

**Step 3**

Choose a Regional bucket and set a unique name (use your project ID because it is unique). Then, click **Create**.

# Launch Cloud Datalab

To launch Cloud Datalab:

1. In **Cloud Shell**, type:

```
gcloud compute zones list
```

2. Pick a zone in a geographically closeby region.

3. In **Cloud Shell**, type:
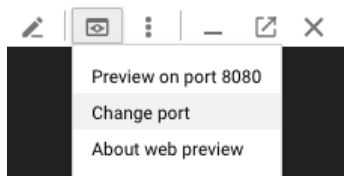
```
datalab create bdmlvm --zone
<ZONE>
```

Datalab will take about 5 minutes to start.

> **Note:** Follow the prompts during this process.

## Checkout notebook into Cloud Datalab

If necessary, wait for Datalab to finish launching. Datalab is ready when you see a message prompting you to do a "Web Preview".

1. Click on the **Web Preview** icon on the top-right corner of the Cloud Shell ribbon. Click on the **Change port**. Switch to port **8081** using the **Change Preview Port** dialog box, and then click on **Change and Preview**.





> **Note:** 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 bdmlvm**' in your new Cloud Shell.

To clone the course repo in your datalab instance:

**Step 1**

In Cloud Datalab home page (browser), navigate into **notebooks** and add a new

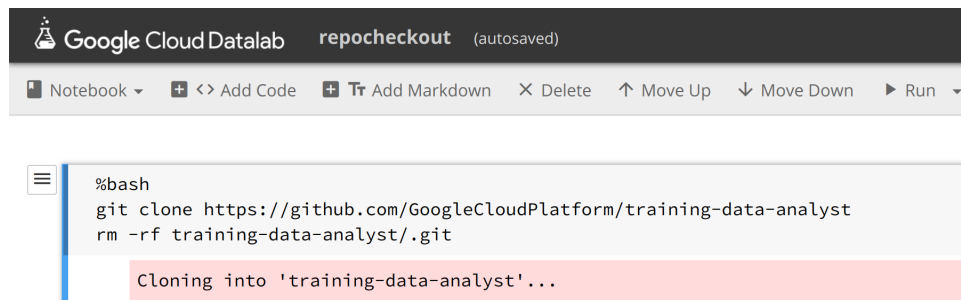notebook using the icon ➕ Notebook on the top left.

**Step 2**

Rename this notebook as **repocheckout**.

**Step 3**

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/GoogleCloudPlat
data-analyst
rm -rf training-data-
analyst/.git
```



**Step 4**

Confirm that you have cloned the repo by going back to Datalab browser, and ensure you see the **training-data-analyst** directory. All the files for all labs throughout this course are available in this directory.
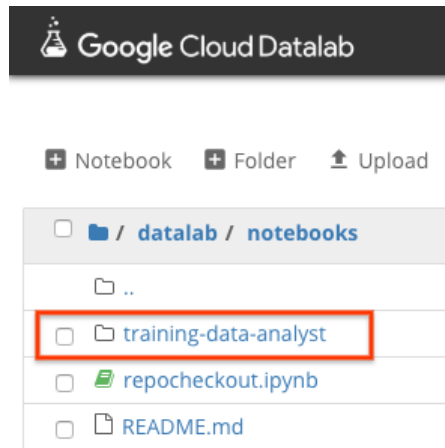
## Open a Datalab notebook

1. In Cloud Datalab, click on the **Home**icon, and then navigate to **datalab > notebook > training-data-analyst > courses > machine_learning > deepdive > 10_recommend > labs > wals.ipynb**.

2. Read the commentary, click **Clear | Clear all Cells**, then run the Python snippets (Use **Shift+Enter** to run each piece of code) in the cell, step by step.

## End your lab

When you have completed your lab, click **End Lab**. Qwiklabs removes the resources you've used and cleans the account for you.

You will be given an opportunity to rate the lab experience. Select the applicable number of stars, type a comment, and then click **Submit**.

The number of stars indicates the following:

- 1 star = Very dissatisfied
- 2 stars = Dissatisfied
- 3 stars = Neutral
- 4 stars = Satisfied
- 5 stars = Very satisfied

You can close the dialog box if you don't want to provide feedback.

For feedback, suggestions, or corrections, please use the **Support**tab.

Last Tested Date: December 17, 2018

Last Updated Date: December 18, 2018