CHAPTER 5

Google Compute Engine (GCE)

Google Compute Engine (GCE) makes available to users virtual machines (VMs) that are running on Google's data centers around the world. These machines take advantage of Google's state-of-the-art fiber optic powered network capabilities to offer fast and high-performance machines that can scale based on usage and automatically deal with issues of load balancing.

GCE provides a variety of pre-defined machine types for use out of the box; also it has the option to create custom machines that are tailored to the specific needs of the user. Another major feature of GCE is the ability to use computing resources that are currently idle on Google infrastructure for a short period of time to enhance or speed up the processing capabilities of batch jobs or fault-tolerant workloads. These machines are called preemptible VMs and come at a huge cost-benefit to the user as they are about 80% cheaper than regular machines.

Again one of the major benefits of GCEs is that the user only pays for the time the machines are actually in operation. Also, when the machines are used for a long uninterrupted period of time, discounts are accrued to the prices.

In this chapter, we will go through a simple example of provisioning and tearing down a Linux machine on the cloud. The examples will cover using the Google Cloud web interface and the command-line interface for creating VMs on GCP.

Provisioning a VM Instance

To deploy a VM instance, click the triple dash in the top-left corner of the web page to pull out the GCP resources drawer. In the group named 'COMPUTE', click the arrow beside 'Compute Engine' and select 'VM instances' as shown in Figure 5-1.

E. Bisong, *Building Machine Learning and Deep Learning Models on Google Cloud Platform*,
https://doi.org/10.1007/978-1-4842-4470-8_5