CHAPTER 1

What Is Cloud Computing?

Cloud computing is the practice where computing services such as storage options, processing units, and networking capabilities are exposed for consumption by users over the Internet (the cloud). These services range from free to pay-as-you-use billing.

The central idea behind cloud computing is to make aggregated computational power available for large-scale consumption. By doing so, the microeconomics principle of economies of scale kicks into effect where cost per unit output is minimized with increasing scale of operations.

In a cloud computing environment, enterprises or individuals can take advantage of the same speed and power of aggregated high-performance computing services and only pay for what they use and relinquish these compute resources when they are no longer needed.

The concept of cloud computing had existed as time-sharing systems from the early years of the modern computer where jobs submitted from different users were scheduled to execute on a mainframe. The idea of time-sharing machines fizzled away at the advent of the PC. Now, with the rise of enterprise data centers managed by big IT companies such as Google, Microsoft, Amazon, IBM, and Oracle, the cloud computing notion has resurfaced with the added twist of multi-tenancy as opposed to time-sharing. This computing model is set to disrupt the way we work and utilize software systems and services.

In addition to storage, networking, and processing services, cloud computing provides offer other product solutions such as databases, artificial intelligence, and data analytics capabilities and serverless infrastructures.

E. Bisong, *Building Machine Learning and Deep Learning Models on Google Cloud Platform*,
https://doi.org/10.1007/978-1-4842-4470-8_1

Categories of Cloud Solutions

The cloud is a terminology that describes large sets of computers that are networked together in groups called data centers. These clustered machines can be interacted with via dashboards, command-line interfaces, REST APIs, and client libraries. Data centers are often distributed across multiple geographical locations. The size of data centers is over 100,000 sq. ft. (and those are the smaller sizes!). Cloud computing solutions can be broadly categorized into three, namely, the public, private, and hybrid cloud. Let's briefly discuss them:

- Public cloud: Public clouds are the conventional cloud computing model, where cloud service providers make available their computing infrastructure and products for general use by other enterprises and individuals (see Figure 1-1). In public clouds, the cloud service provider is responsible for managing the hardware configuration and servicing.

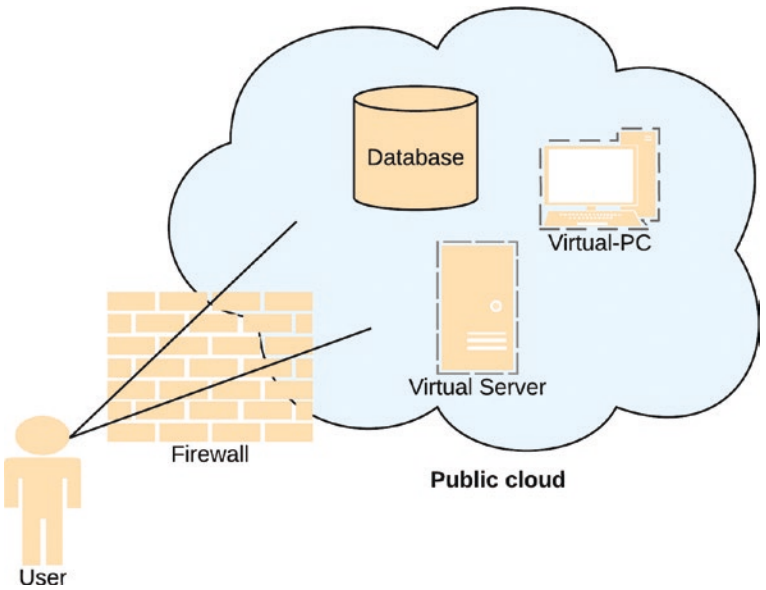

Figure 1-1. *The public cloud*