1.4 Classifying cloud layers: different types for different uses

First, let's learn a little more about how SaaS evolved and established itself, to set the context for discussing the other classes of clouds.

In the earliest days of commercially practicable computing, computer resources were scarce, and the primary model for their use was much like a utility. But this was different from the sense of utility that cloud computing offers today; it was more akin to the community well in a village during a drought. Members of the community had access to and were allocated a fixed amount of water. In the case of cloud computing today, we've returned to the notion of computing being available as a utility, but without the scarcity.

The cloud movement was presaged by the shift in business model toward SaaS that took over the software industry at the turn of the century. Before it was called SaaS, it was an application rented from an Application Service Provider (ASP); here, the traditional enterprise license model was turned on its head, and you purchased in a pay-as-you-go manner, with costs scaling with usage instead of having a large up-front capital investment. You didn't need to provision hardware and software; instead, the services were turned on when needed. After this approach was renamed SaaS, it evolved into several new kinds of offerings that we'll explore next.

We can classify cloud computing several ways. In this book, we present a taxonomy where cloud services are described generically as "X as a Service," where X can take on values such as Hardware, Infrastructure, Platform, Framework, Application, and even Datacenter. Vendors aren't in agreement about what these designations mean, nor are they consistent in describing themselves as belonging to these categories. Despite this, we'll reproduce one interesting hierarchy that illustrates the use of these terms, with representative vendors (some at this point only historical) populating the diagram in figure 1.6.

A more simplified representation of the cloud types shown in figure 1.7 highlights important aspects and key characteristics of different kinds of cloud offerings.

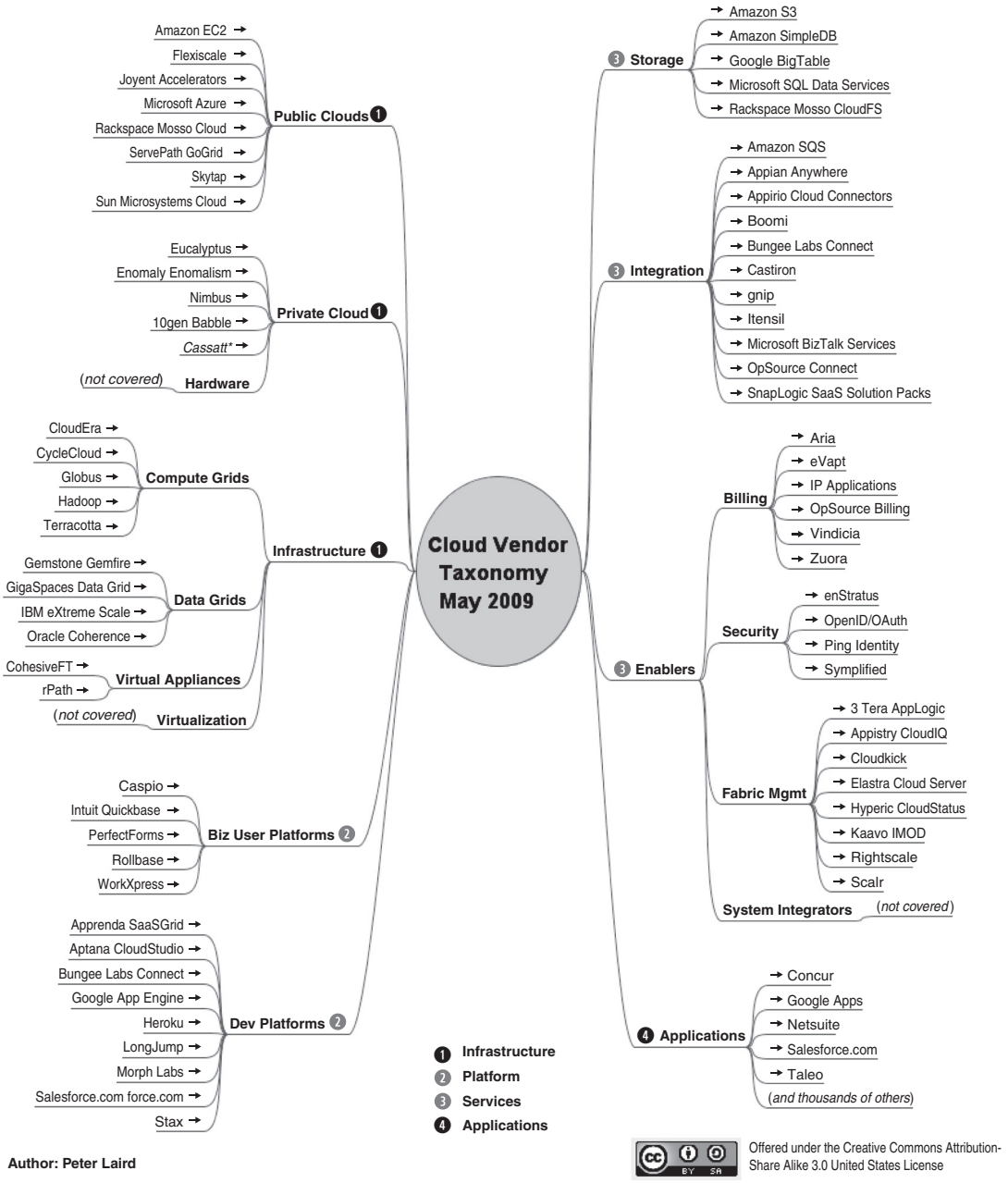

Figure 1.6 Cloud technologies are evolving as various vendors attempt to provide services populating the cloud ecosystem. These services run the gamut from the hardware systems used to build cloud infrastructure to integration services and cloud-based applications. Source: Peter Laird, http://peterlaird.blogspot.com.

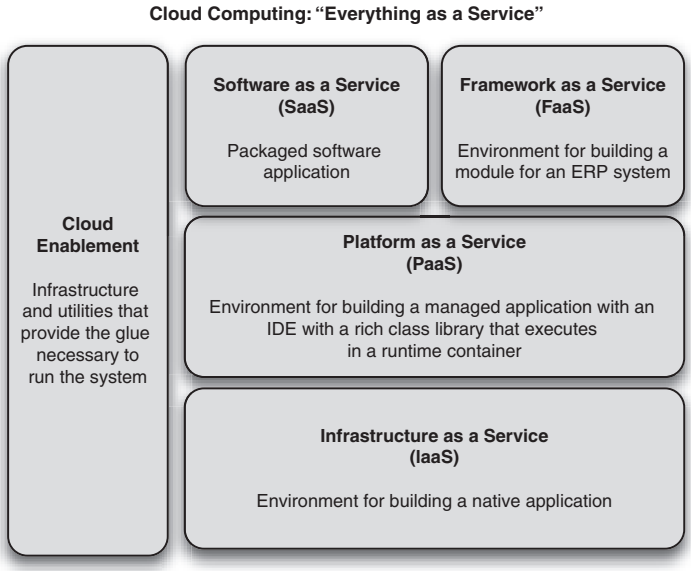

Figure 1.7 In the X-as-a-Service taxonomy, cloud services are classified by the level of prepackaging offered to the consumer of the specific service. An IaaS provides computing capabilities in the rawest form and hence offers the greatest flexibility. At the highest layers, there is less flexibility but also less complexity to be managed.

What does XaaS mean generically? It means on demand, requiring little or no capital expenditure. It means consumable remotely and across any mode of access over the internet, and in a metered billing model. Let's now go through the boxes representing the different classes of clouds in figure 1.7. First up is IaaS.

1.4.1 Infrastructure as a Service (IaaS)

The lowest level of XaaS is known as IaaS, or sometimes as Hardware as a Service (HaaS). A good example of IaaS is the Amazon Elastic Compute Cloud (EC2).

A user of IaaS is operating at the lowest level of granularity available and with the least amount of prepackaged functionality. An IaaS provider supplies virtual machine images of different operating system flavors. These images can be tailored by the developer to run any custom or packaged application. These applications can run natively on the chosen OS and can be saved for a particular purpose. The user can bring online and use instances of these virtual machine images when needed. Use of these images is typically metered and charged in hour-long increments.

Storage and bandwidth are also consumable commodities in an IaaS environment, with storage typically charged per gigabyte per month and bandwidth charged for transit into and out of the system.

IaaS provides great flexibility and control over the cloud resources being consumed, but typically more work is required of the developer to operate effectively in the environment. In chapter 2, we'll delve into IaaS and see how it works in greater detail.

1.4.2 Platform as a Service (PaaS)

PaaS's fundamental billing quantities are somewhat similar to those of IaaS: consumption of CPU, bandwidth, and storage operates under similar models. Examples of PaaS include Google AppEngine and Microsoft Azure. The main difference is that PaaS requires less interaction with the bare metal of the system. You don't need to directly interact with or administer the virtual OSs. Instead, you can let the platform abstract away that interaction and concentrate specifically on writing the application. This simplification generally comes at the cost of less flexibility and the requirement to code in the specific languages supported by the particular PaaS provider.

1.4.3 Software as a Service (SaaS) and Framework as a Service (FaaS)

SaaS, as described earlier in the chapter, refers to services and applications that are available on an on-demand basis. Salesforce.com is an example. FaaS is an environment adjunct to a SaaS offering and allows developers to extend the prebuilt functionality of the SaaS applications. Force.com is an example of a FaaS that extends the Salesforce.com SaaS offering.

FaaS offerings are useful specifically for augmenting and enhancing the capabilities of the base SaaS system. You can use FaaS for creating either custom, specialized applications for a specific organization, or general-purpose applications that can be made available to any customer of the SaaS offering. Like a PaaS environment, a developer in a FaaS environment can only use the specific languages and APIs provided by the FaaS.

1.4.4 Private clouds as precursors of public clouds

In addition to the classifications we discussed earlier, we should introduce some important concepts relative to the different classifications of clouds. *Private clouds* are a variant of generic cloud computing where internal data-center resources of an enterprise or organization aren't made available to the general public—that is, these pooled computing resources are actually not available to *any* subscribing users but are instead controlled by an organization for the benefit of other members of that organization. The public clouds of providers such as Amazon and Google were originally used as private clouds by those companies for other lines of business (book retailing and internet search, respectively).

If an organization has sufficient users and enough overall capacity, a private cloud implementation can behave much like a public cloud, albeit on a reduced scale. There has been a tremendous amount of capital investment in data-center resources over the past decade, and one of the important movements is the reorienting of these assets toward cloud-usage models.

Hybrid clouds combine private and public clouds. You can use them in cases where the capacity of a private cloud is exhausted and excess capacity needs to be provisioned elsewhere.