Chapter 1

Understanding Cloud Fundamentals and the Cloud Continuum

In This Chapter

▶ Understanding the essentials of cloud computing
▶ Examining foundational delivery services
▶ Exploring the cloud continuum

How quickly things change. Cloud computing has evolved from a risky and confusing concept into a strategy that organisations, large and small, are beginning to adopt as part of their overall computing strategy. Companies are now starting to ask not whether they should think about cloud computing but what types of cloud computing models are best suited to solve their business problems. Many organisations are not only using the cloud for services such as e-mail and customer relationship management but also are utilising a set of important cloud foundational services – Infrastructure as a Service (IaaS) and Platform as a Service (PaaS) – to develop and deploy applications to support their business and open up new, innovative opportunities and new revenue streams.

The kind of cloud deployment you should consider depends on your own performance, security requirements and specific business goals. In this chapter, you look at the fundamentals of cloud computing and the insights into the range of cloud services and models that you need to understand. It’s important to be aware that there’s a continuum of cloud services. These services range from an open and shared public
environment to a private cloud that’s tightly managed with the highest level of security and service. You also discover two key foundational cloud delivery models: IaaS and PaaS. In addition, Software as a Service (SaaS) provides packaged business process offerings that live in the cloud and leverage both IaaS and PaaS services.

**Discovering Cloud Basics**

*Cloud computing* is a method of providing a set of shared computing resources that includes applications, computing, storage, networking, development and deployment platforms as well as business processes. Cloud computing turns traditional siloed computing assets into shared pools of resources that are based on an underlying Internet foundation.

Clouds come in different versions, depending on your needs. There are two primary deployment models of cloud: public and private. Most organisations use a combination of private computing resources (data centres and private clouds) and public services as a hybrid environment. These clouds are covered in more detail throughout this chapter.

The cloud doesn’t exist in isolation to other corporate IT investments. The reality is that most companies use a combination of public and private cloud services in conjunction with their data centre. Companies use different methods, depending on their business requirements, to link and integrate these services. The way you construct your hybrid computing environment is determined by the complexity of the workloads and how you want to optimise the performance of those workloads to support your constituents.

**Foundational Cloud Delivery Services**

Understanding the foundations of cloud computing calls for an understanding of the three main cloud delivery models:
IaaS: The delivery of services such as hardware, software, storage, networking, data centre space and various utility software elements on request. Both public and private versions of IaaS exist.

- In the public IaaS, the user needs a simple sign-up mechanism to acquire resources. When users no longer need the resources, they simply de-provision them.

- In a private IaaS, the IT organisation or an integrator creates an infrastructure designed to provide resources on demand to internal users and sometimes partners. IaaS is the fundamental element used by other cloud models. Some customers bring their own tools and software to create applications.

PaaS: A mechanism for combining IaaS with an abstracted set of middleware services, software development and deployment tools that allows the organisation to have a consistent way to create and deploy applications on a cloud or on-premises environment. A PaaS environment supports coordination between the developer and the operations organisation, typically called DevOps. A PaaS offers a consistent set of programming and middleware services that ensure developers have a well-tested and well-integrated way to create applications in a cloud environment. A PaaS requires an infrastructure service.

SaaS: A business application created and hosted by a provider in a multi-tenant (shared) model. The SaaS application sits on top of both a PaaS and a foundational IaaS. In fact, a SaaS environment can be built directly on an IaaS platform. Typically these underlying services aren’t visible to end users of a SaaS application.

A hybrid cloud combines private cloud services with public cloud services where one or several points of contact are shared between the environments. What does this mean? If a few developers in a company use a public cloud service to prototype a new application that’s completely disconnected from the private cloud or the data centre, the company doesn’t have a hybrid environment. On the other hand, a
cloud is hybrid when a company uses public cloud services for tasks such as prototyping or testing a new application. When the application is completed, it may be moved to the private cloud. In another situation, the Web servers are on a public cloud service that’s integrated with payment systems that are run in a private cloud.

A company with a private cloud may choose to combine some public services for capabilities that are commodities with private services based on the ability to deliver fast innovation to their ecosystem. For example, companies are increasingly discovering that it’s practical to pay a per-user, per-year price for customer relationship management (CRM) and leave the day-to-day management to a trusted vendor. But many companies also want to keep control over some of their most sensitive data. Therefore, a company may choose to keep data about prospects on a public cloud, but after those prospects become customers, the company may begin storing its data on its own premises in its own servers; this is the hybrid cloud model.

Core Cloud Capabilities

Regardless of the model you use, some core capabilities that are essential in the cloud environment include the areas covered in this section.

Elasticity and self-service provisioning

A key feature of a cloud environment is that it provides a platform that’s designed to be elastic (in that you can use only the resources you want, only when you need them), so the users/customers provision resources, such as computing or storage resources, for which they pay on a per-unit basis. When the user no longer needs that resource and stops paying, the resource is released back into the pool of resources. This helps organisations avoid the cost of idle computing resources. Instead of purchasing, managing and maintaining
a server environment, for example, a business can purchase computing on demand, avoiding capital expenditure.

The term self-service is important here too. With self-service, the developer of an application, for example, is able to use a browser or portal interface to acquire the appropriate resources needed to build or operate an application. This just-in-time model is a more efficient way to ensure that the IT organisation can be responsive to business change.

**Billing and metering service usage**

A cloud service has to provide a way to measure and meter the service. Consequently a cloud environment includes a built-in service that tracks how many resources a customer uses. In a public cloud, customers are charged for the units of resources consumed. In a private cloud, IT management may implement a charge-back mechanism for departments leveraging services.

**Workload management**

The cloud is a federated (distributed) environment that pools resources so they can work together. Making this happen requires these resources to be optimised to work as if they were an integrated, well-tuned environment comprised of a variety of workloads. A workload is an independent service or collection of code that can be executed. It’s important in the cloud that workloads be designed to support the right task with the right cloud services. For example, some workloads need to be placed in a private cloud because they require fast transaction management and a high level of security. Other workloads may not be so mission critical and can be placed in a public cloud.

**Management services**

Many management services are mandatory for ensuring that cloud computing is a well-managed platform. Security and
governance are key services to ensure that your applications and data are protected. Data management is also critical because data may be moving between cloud environments. All of these services have to be managed and monitored to ensure that an organisation’s level of service is maintained.

**Understanding the Cloud Continuum**

Meeting the needs of businesses requires that IT provide a variety of different types of cloud service. Understanding the characteristics of a continuum of cloud services helps you understand what’s required to meet certain business goals.

All cloud environments aren’t equal. Therefore, you need to understand the different types of cloud model available to support the business. Your decision about what type of cloud service to select is based first and foremost on your security and service level requirements. It may be straightforward to assume that all public clouds are the same and all private clouds work in the same way. But in reality there are shades of grey.

For example, you may have a public cloud service that’s only available to customers who sign a long-term agreement. You may have a private cloud that’s an evolution of your data centre. Some public clouds may offer a sophisticated level of security, while other public clouds have virtually no security at all.

Ultimately, you need to select the type of cloud service that provides you with the use of the right resources at the right time with the right level of security and governance.

The continuum of cloud services, depicted in Figure 1-1, includes both public and private services that meet different needs within an organisation.
### Figure 1-1: The cloud services continuum.

<table>
<thead>
<tr>
<th>Model</th>
<th>Examples</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open Community</td>
<td>Facebook, Twitter, LinkedIn, MyFitness, Google Groups</td>
<td>No SLA, No Contract, Simple Password Protection, No governance model</td>
</tr>
<tr>
<td>Controlled Open Mode</td>
<td>IBM SmartCloud Enterprise, Amazon Web Services, RackSpace, OpSour</td>
<td>Simple SLA, No Contract, More security, No explicit governance</td>
</tr>
<tr>
<td>Contractual Open</td>
<td>Salesforce.com, Workday, MailChimp, QuickBooks Online</td>
<td>SLA with no indemnification, Contract, High security provided, Governance in place</td>
</tr>
<tr>
<td>Public/Private Hybrid</td>
<td>IBM SmartCloud, HP Cloud Service, Microsoft Azure</td>
<td>SLA guaranteeing uptime, Contract, Highest level of security, Explicit governance</td>
</tr>
<tr>
<td>Private Closed</td>
<td>Internal but can be implemented by a third-party vendor</td>
<td>Explicit SLA, Capital expense with ongoing maintenance, Secure platform, Explicit governance</td>
</tr>
</tbody>
</table>
Open community clouds

The most open type of cloud environment is an open community cloud – a cloud environment that doesn’t require any criteria for joining other than signing up and creating a password. These environments may be privately or publicly owned and include social networking environments, such as Facebook, LinkedIn and Twitter. There are also open community sites that enable individuals with a common interest to participate in online discussions. For example, there may be a community of professionals in a certain industry who want to share ideas.

These open community sites generally involve a relatively simple sign-up process, although some of the more sophisticated sites request additional information from you.

These sites also generally have a low level of security. Therefore, it’s relatively simple for someone inside or outside the open community to penetrate a user’s secure area. In addition, these sites generally don’t offer service level guarantees to the user. Sites that are advertising-driven typically spend more effort on security and service level management.

Controlled open mode

Some public clouds offer a higher level of service because they’re true commercial environments. Commercial public clouds are those environments that are open for use by anyone at any time, but these clouds are based on a pay-per-use model. For example, a SaaS vendor that charges per-user-per-month (or per-year) is one example of this kind of environment. In addition, vendors can offer analytics as a service to customers on a per-use or per-task basis.

Because companies offering commercial public clouds are providing a commercial service, they provide a higher level of security and protection than do the open community sites. These services generally have a written service level agreement (SLA) outlining the provider’s obligation to the consumer of the service.
Contractual open

Vendors are offering public IaaS and PaaS public platforms that are based on a formal contractual basis. The user can’t simply create login credentials, provide credit card details and start using the service. Instead, she actually signs a contract for service. The term can be as short as a month or, more typically, a year.

In this kind of environment, the expectation is for high levels of security, privacy and governance. These vendors provide a written SLA. Because of the service and security guarantees of this model, some customers may be willing to store critical data in the cloud.

Public/private hybrid clouds

Companies often want the flexibility of the cloud but with the security and predictability of the data centre. In these cases, a private cloud provides an environment that sits behind a firewall. Unlike a data centre, a private cloud is a pool of common resources optimised for the use of the IT organisation. Unlike a public cloud, a private cloud adheres to the company’s security, governance and compliance requirements. Whatever service level is required for the company applies to the private cloud.

There are two types of private clouds:

✓ A private cloud owned and managed by a company for the benefit of its employees and partners
✓ A commercial private cloud residing in a vendor’s data centre and providing a secure connection to the customer’s other IT resources. This approach securely augments a customer’s IT environment.

In some instances, companies use a combination of public and private cloud services. A retail company may have a private cloud to support its highly distributed development organisation, and it may also use a SaaS public cloud application for human resources. In addition, to support its online commerce
system, the company may leverage public commercial cloud services to ensure that customer service remains satisfactory during times of peak use, such as holidays. The same company might also create a private cloud application that it makes available to partners linking to its online sites.

This type of hybrid environment will become the standard way companies run IT in the future. A company will typically use public cloud services such as SaaS to support customer relationship management, IaaS to add capacity on demand and PaaS to support an experiment development process. This development makes sense because, increasingly, companies are looking for a cost-effective, flexible and optimised environment to support internal operations, customers, partners and suppliers.

When a company selects this route, it takes the responsibility for the integration, security, manageability and governance of the composite environment – including the public services that are included.

In other words, IaaS and PaaS are foundational services on which other cloud services will sit. IaaS itself is the foundation upon which PaaS can be utilised to build value. It supplies the infrastructure that developers can use to build applications. For example, many organisations are using IaaS and PaaS linked together for the development and operations process, which we will get into later in this book. These organisations may even be using IaaS and PaaS to build actual SaaS services. So, in some ways, IaaS serves as the base of a pyramid, with the infrastructure at the bottom, the middleware (PaaS) at the centre and the applications on top.