CHAPTER

# APPLICATION MANAGEMENT IN THE CLOUD

# 4

*I've looked at clouds from both sides now*
*From up and down, and still somehow*
*It's cloud illusions I recall*
*I really don't know clouds at all*
**Joni Mitchell, "Both Sides Now"**

"The Cloud"—nearly everyone has heard of it, many have used it in some form, and at least among IT professionals, nearly everyone thinks that they know what it is. In truth, most only have limited insight into a single variant of **cloud computing**. Before we can dive into the discussion of **application management** in the cloud, we must first establish some common terminology.

Let's begin with a very basic point. There is no such thing as "The Cloud." Cloud computing is not represented by a single entity as some might argue the Internet is. Rather, cloud computing is a technology that exists in hundreds of thousands (possibly millions) of unique instances.

In 2011, the **National Institute of Standards and Technology (NIST)** published a report titled, "The NIST Definition of Cloud Computing." That report became a watershed definition for cloud computing. It is particularly important because it was developed without the biases of those with commercial interests in the technology. Since that report was published, cloud computing continues to evolve. The range of types of cloud computing services must be viewed as a continuum, rather than a set of discrete types. This is partly due to each solution provider approaching the problem in a slightly different way—sometimes for technical reasons and sometimes in an attempt to set their offering apart from competitors. However, the key aspects of a cloud environment are **virtualization**, scalability, flexible resource management, data center automation (to support on-demand services), usage metering, and **security**.

Cloud computing can be broken down into three broad categories: **private cloud**, **public cloud**, and **hybrid cloud**. That seems pretty simple and straightforward and was, until cloud became the new fad topic in IT. That meant marketing organizations around the world began to weave their own messages around the theme, molding and shaping the meaning to suit their purposes. This allowed them to claim that their products were relevant, perhaps even essential, to cloud computing. The intent was not to fault those focus organizations. What they did is simply attempt to supply a definition (that is most favorable to them) to fill a void when there is no consensus about the meaning of a term. While we can't put that particular genie back in the bottle, we can at least establish a set of definitions that we will use throughout this book.

**Application Performance Management (APM) in the Digital Enterprise.** http://dx.doi.org/10.1016/B978-0-12-804018-8.00004-8

## ESSENTIAL CHARACTERISTICS

*On-demand self-service*. A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.

*Broad network access*. Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops, and workstations).

*Resource pooling*. The provider's computing resources are pooled to serve multiple consumers using a multitenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state, or data center). Examples of resources include storage, processing, memory, and network bandwidth.

*Rapid elasticity*. Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time.

*Measured service*. Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized service.[1]

[1]"Characteristics of Cloud Computing." NIST. NIST SP 800-145, The NIST Definition of Cloud Computing csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf.

Types of Cloud Computing
- Private cloud
- Public cloud
  - Software as a Service (SaaS)
  - Platform as a Service (PaaS)
  - Infrastructure as a Service (IaaS)
- Hybrid cloud

## PUBLIC CLOUD

Public cloud services are offered by companies to provide their customers with access to computing resources over a public network (Fig. 4.1). The cloud service provider is responsible for provisioning and managing the hardware and software associated with the service being offered. The public cloud actually consists of three types of services: **software as a service (SaaS)**, **infrastructure as a service (IaaS)**, and **platform as a service (PaaS)**. The main advantage offered by public cloud environments comes from the lower costs that can be realized. On the other hand, there are concerns about security of data and regulatory issues.

In the SaaS model, the customer uses an application that is owned and hosted by the service provider. In the simplest form of the pure public cloud model, there is not any interface between the hosted application and any **applications** running in the customer's organization (or any other SaaS applications hosted by a different service provider). In the public cloud environment, SaaS is the most popular public cloud service offering. Research by Enterprise Management Associates (EMA) found that 67% of enterprises surveyed were delivering at least one production service (i.e., use of an application for

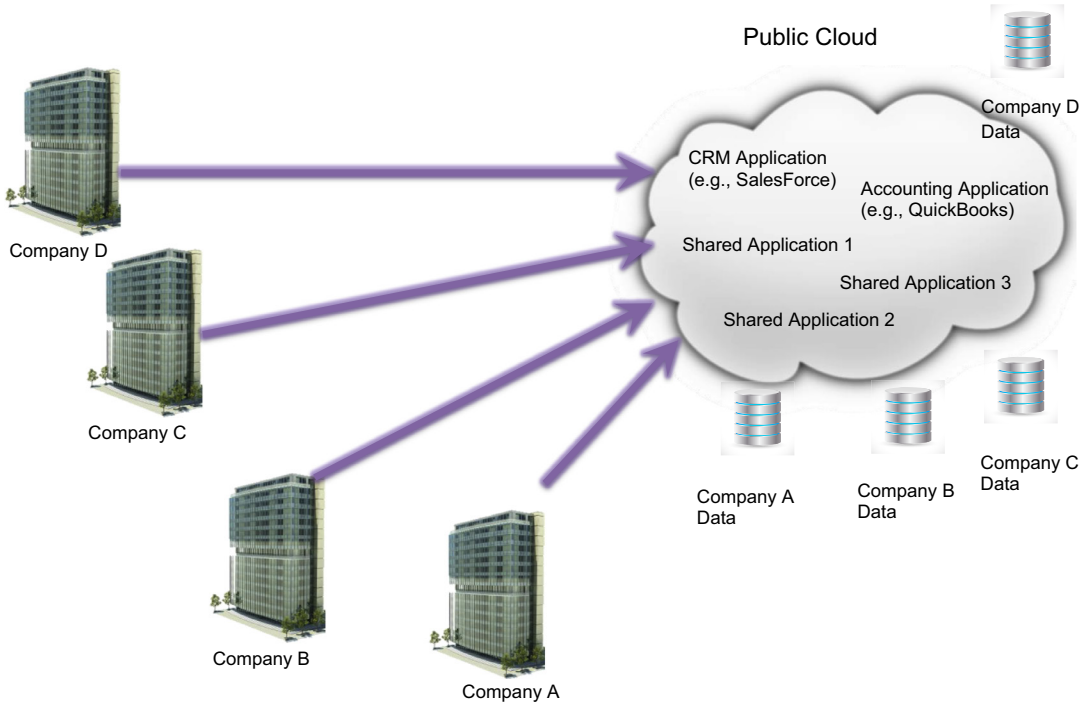


**FIGURE 4.1**

Public cloud.

a routine business activity) via a SaaS-based cloud application. Two examples of widely used SaaS applications are SalesForce.com and Google Docs.

PaaS provides the customers with a complete computing environment that includes **operating systems**, databases, software tools, and support services, often in a multitenant environment. PaaS is quite popular as an application development environment. Developers are freed of the need to address issues of **configuration**, management, and so on, and are able to focus specifically on the development of new software. Not only can the software be developed and tested in a PaaS environment, it can also be put into production in the same PaaS environment.

The popularity of IaaS is similar to that of PaaS. According to that same EMA research report, 55% of organizations surveyed are using IaaS, whereas 56% were using PaaS, although the data indicated that IaaS is being adopted at a faster rate than PaaS. In this service model, the customer is essentially renting computing resources such as computers (servers or mainframe), storage, operating system software, database software, and (in some cases) network capacity. The customer is responsible for configuring and managing those resources. That is simultaneously an advantage of PaaS and a disadvantage. It is particularly suited to rapid prototyping and for applications with workloads that fluctuate widely. It is an alternative to buying or leasing the equipment and associated software. Examples of IaaS are IBM BlueMix, Microsoft Azure CloudBees, and **Amazon Web Services**.

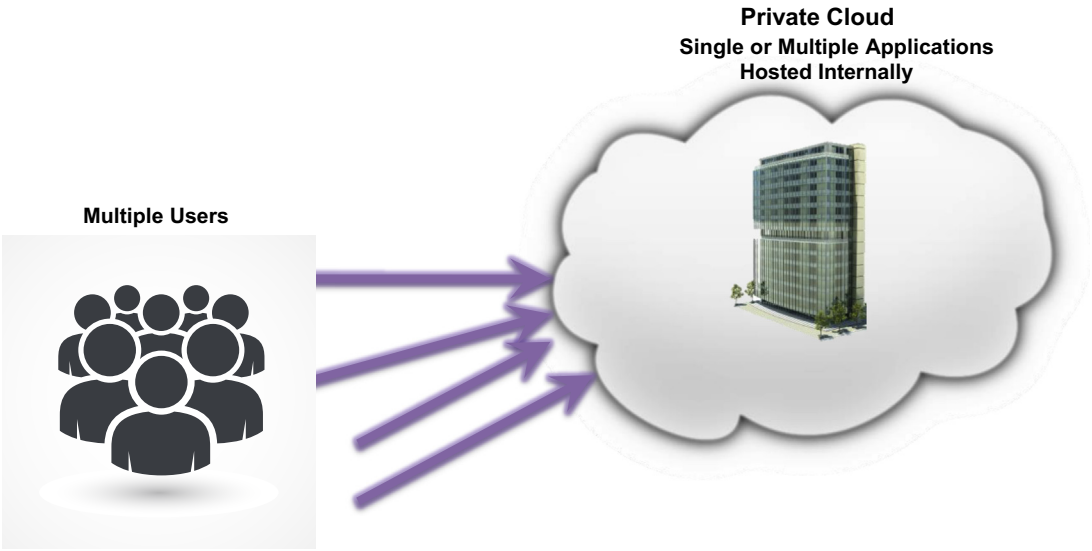


**FIGURE 4.2**

Private cloud.

However, PaaS does require the customer to have IT staff to configure the resources and provide ongoing management. This offsets some of the financial benefits of the PaaS model.

## PRIVATE CLOUD

A private cloud has many of the same characteristics as a public cloud (Fig. 4.2). The difference is that the resources used for this service belong to the enterprise (usually the IT department). Since a private cloud is physically located within the enterprise, concerns about security are greatly reduced. A private cloud offers nearly the same level of security that is available for other corporate systems. Likewise, regulatory compliance becomes a nonissue with a properly configured and secured private cloud environment.

## HYBRID CLOUD

Hybrid clouds are where the greatest ambiguity about the nature of this particular beast lives (Fig. 4.3). Much has been written about it and there are many differing opinions. For this book, we define a hybrid cloud as one that enables the sharing of data between two or more clouds operated by different enterprises. This sharing could be between two public cloud SaaS applications operated by different companies and used by a third company (the customer). For example, the customer might want to share data between SalesForce and Marketo. More common is the **integration** of data between a public cloud application and a private cloud application. However, hybrid cloud is commonly viewed as a mix of public and private clouds.

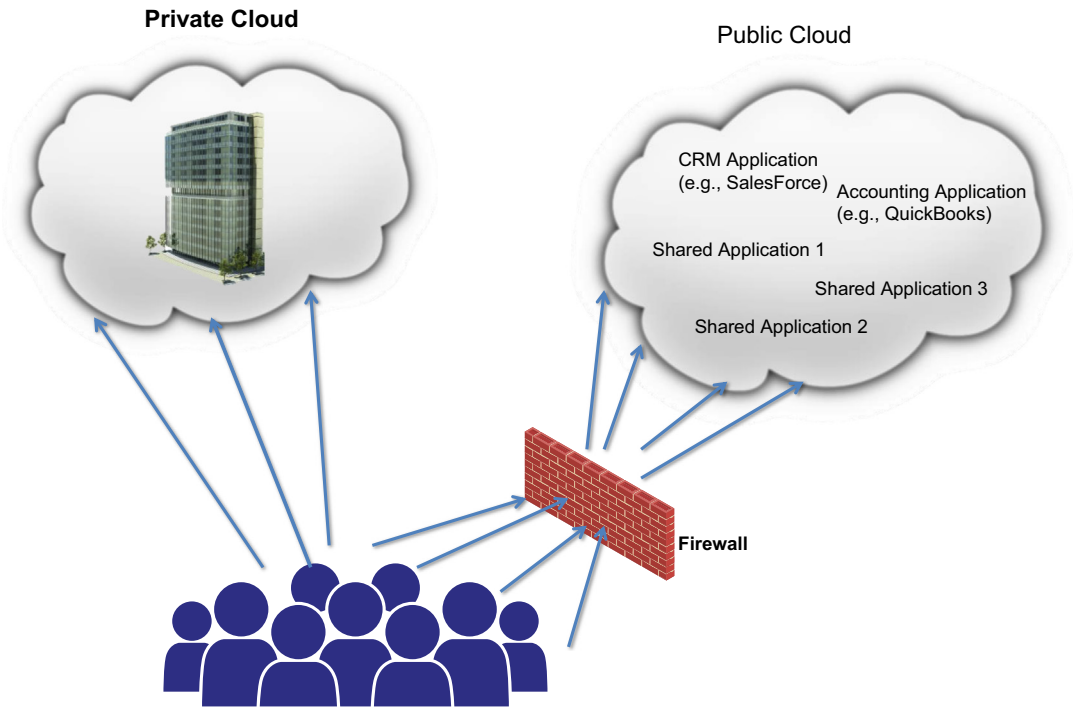


**FIGURE 4.3**

Hybrid cloud.

## MANAGEMENT OF CLOUD APPLICATIONS

Managing applications in cloud environments brings a variety of challenges and those challenges vary, depending on the type of cloud environment. The difficulty in managing cloud-based applications depends largely on the type of cloud environment in which the application is running (Fig. 4.4). There are two sides to the management coin in a cloud environment: service provider and user. As we will see shortly, management can mean very different things for those two groups.

### PRIVATE CLOUD

The simplest environment in which to manage an application is a private cloud. A private cloud is completely within the enterprise and so are any applications running in that cloud. IT (or its equivalent in a business unit) has complete control of all the hardware and software that is enabling the cloud. IT is able to install and use any management tools that it feels will be helpful in monitoring the environment or analyzing the data captured by the management tools. IT can modify the configuration of each of the cloud components to optimize the **performance** of an application. IT's ability to manage an application is limited only by budget and the availability of management tools. In short, IT is able to manage applications in a private cloud as completely and effectively as it can manage any other applications that it

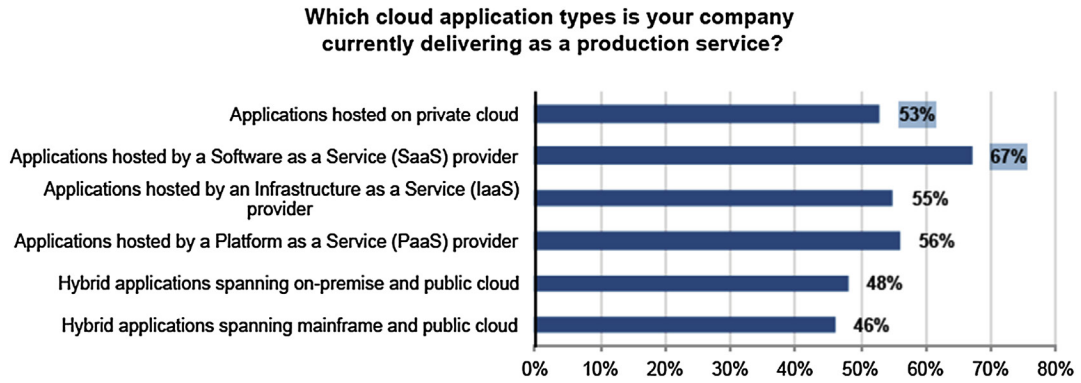


**FIGURE 4.4**

Cloud deployments.

*Figure 1, page 6, "Public Cloud Comes of Age".*

runs in noncloud environments (a list of monitoring techniques available for cloud environments is shown in Fig. 4.5). Yes, if the private cloud is virtualized (as many are), that brings a set of challenges related to operating and managing the virtualized environment. Those are not challenges of application management and therefore not within the scope of this chapter.

## PUBLIC CLOUD: INFRASTRUCTURE AS A SERVICE

The first step beyond the enterprise and a private cloud is IaaS. In the IaaS environment, application is relatively straightforward. The customer (i.e., usually the IT department) can use whatever management tools and strategies they wish (Fig. 4.6). They can install agents to monitor the applications and provide data to management tools. They can install application management tools on the IaaS server. Those tools can be accessed remotely and used to monitor and control the application(s) on that remote server. It is also possible to use remote tools to manage the application(s) running on the IaaS server. Of course, that is constrained by the security measures on the IaaS server and the terms of the license agreements for those management tools. In short, applications on an IaaS server can be managed as if that environment is just another corporate data center, at least to the extent that the service provider will allow.

The fact that an application in an IaaS environment happens to be remote is only a minor consideration. It does introduce the network as a factor that must be taken into account when assessing the application's performance. However, this is true with any application, whether that application is behind the corporate **firewall** or outside of it.

## PUBLIC CLOUD: PLATFORM AS A SERVICE

When an application is moved into a PaaS environment, management of that application becomes more difficult. The customer's IT department no longer has complete control of the environment. It is the service provider's staff that will make and implement the decisions about the configuration of the environment in which the application will run. They will determine what security measures will be put in

| |
|---|
| Agent-based monitoring |
| Synthetic transactions |
| Endpoint (desktop, mobile device, etc.) monitoring |
| Real user (network-focused) monitoring (RUM) |
| Transaction tracing and "stitching" to deliver end-to-end execution visibility |
| Infrastructure instrumentation |
| Browser injection |
| Web server instrumentation |
| Java instrumentation |
| .NET instrumentation |
| Transaction tagging |
| Protocols such as WMI, SNMP, HTTP/S, etc. |
| Network flow tools |
| Third-party agents |
| Log files |
| Unstructured data |
| APIs (for public cloud platforms) |
| Automated discovery data |
| CMDB/CMS data/metadata |

**FIGURE 4.5**

Cloud data collection techniques and data sources.

place (those measures may impact access to the application, or associated agents, by management tools). Some service providers offer some limited management information, but this is not a universal practice.

Unless the customer is able to negotiate an exception to the service provider's policy, the customer will not be able to install management software in the cloud environment. Furthermore, in most situations, the customer will not be able to use remote management tools to see through the firewall and monitor the application. If the service provider allows customers to install and run their own applications in the environment, then some management capability may be able to be included in the application (this approach is closer to IaaS than the usual PaaS offering).

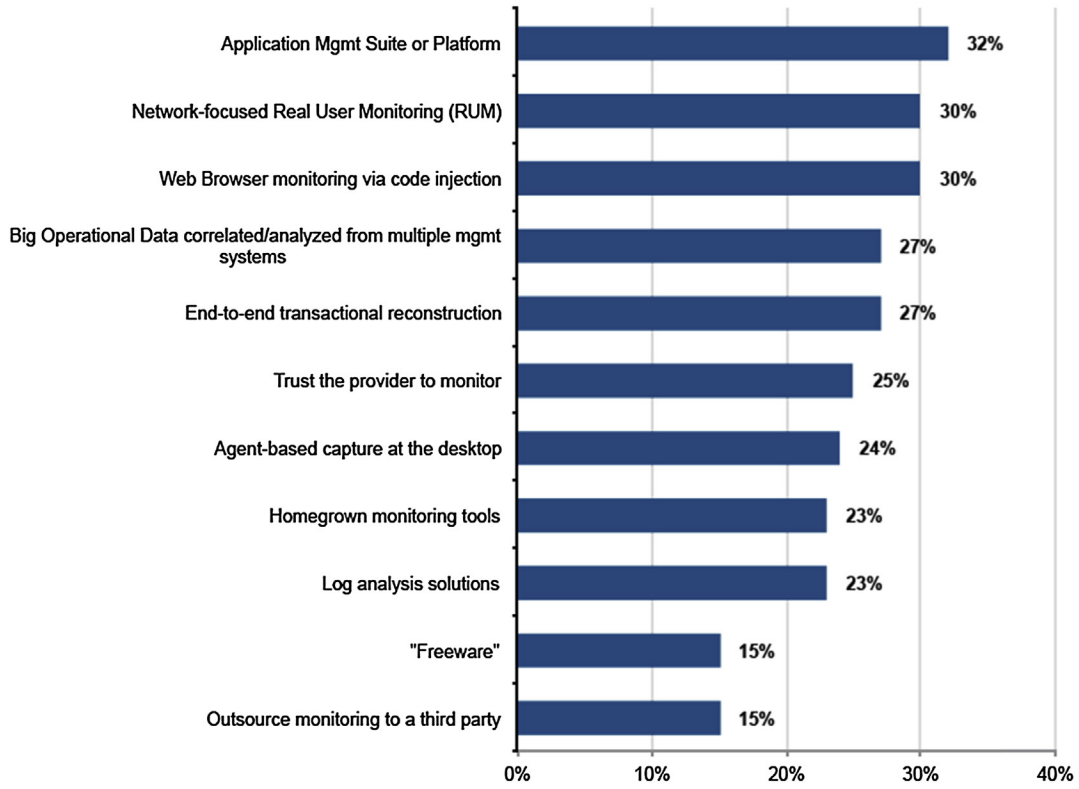


**FIGURE 4.6**

Approaches to managing application performance and availability in public clouds.

*Public Cloud Comes of Age by EMA 2016.*

The primary options are to use **synthetic transactions** and/or **real user monitoring (RUM)** to give an indication of performance and availability of the application. Either approach can be considered as approximating the **end-user experience (EUE)**. It provides a reasonable approximation of the end-to-end response time that is seen by a user.

Some products are specifically described as EUE. This approach inserts instrumentation (a network device or agent) on premise to track every interaction between the user and the service provider. The EUE solution "sees" when a transaction calls out to the cloud and when it receives data back. These EUE tools provide an overall metric on how long it took for the transaction to traverse the **wide area network (WAN)**, the Internet, the processing on the provider side, and the return trip. However, everything outside the organizational borders becomes a black box. So while the tool measures how long execution took outside the organizational borders, it is unable to tell you how long each segment took—or the specific performance on the PaaS side.

A synthetic transaction can give a reasonable estimate of end-to-end response time and a rough approximation of the availability of the application. RUM is a **passive monitoring** technology that

records all user interactions with an application or a website. It provides information similar to what synthetic transactions provide, that is, availability and response time. Either approach will require the deployment of software and in the case of RUM, possibly hardware as well. It is possible to custom build these tools internally, however, most organizations opt to turn to commercial solutions to address these requirements.

However, when performance is degraded, neither synthetic transactions nor RUM can identify the source of the problem. They cannot determine if the problem is originating in the network, if there is an operating system problem, or if it originates in one or more of a myriad of other possible causes. At that point, the customer is primarily limited to contacting the service provider's support desk for help resolving the problem.

Some network providers are willing to give service-level guarantees (particularly to their largest clients). Those guarantees are typically structured around metrics for the performance and availability of the network. Some of the carriers will provide real-time access to data about the network. Others will only give monthly, or even quarterly, reports. The latter have almost no value in the management of PaaS-based applications.

The net result is that in a PaaS environment the customer has little information about the performance or availability of an application and no ability to make changes even if they did have that information. However, the information that the customer has is useful for holding a PaaS service provider accountable and being able to demonstrate that a problem exists.

## PUBLIC CLOUD: SOFTWARE AS A SERVICE

There are thousands of SaaS applications. Therefore, we have to be careful in making sweeping generalizations because there are almost certainly exceptions to any statement that can be made about them. However, while accepting that there are exceptions, we need to take a look at management as it applies to SaaS applications. SaaS applications are useful, and usually cost-effective, solutions tailored to address specific needs. Their functionality tends to be fixed and relatively inflexible.

No real-time application management information is provided to the users (i.e., the customer) of SaaS applications (remember that there can be exceptions). Like users of a PaaS system, the users may be able to use synthetic transactions, EUE, or RUM technology to gauge the performance and availability of an application. Similarly, data from the network service provider may help to isolate performance issues. Ultimately, the data is useful only for convincing the service provider's service desk personnel that there is a problem and for evaluating whether or not service-level guarantees have been met.

The management of SaaS applications is the responsibility of the service provider's staff. They will perform all the management functions required for those applications. For those people, managing a SaaS application is no different than managing a private cloud application is for an IT department in an enterprise.

## HYBRID CLOUD

In Dante's *Divine Comedy*, one of the characters comes upon a gate that is the entrance to Hell. Above the gate is written, "Abandon all hope, ye who enter here." Such a warning should be given to anyone contemplating the challenge of managing applications in a hybrid cloud environment. In truth, the

situation is not entirely hopeless, but it is a daunting challenge. The management of applications in each of the parts of a hybrid cloud equation, separate from the other parts, is no different and no more difficult than managing those parts outside of a hybrid cloud environment. However, in the case of a hybrid cloud, those parts do not exist in isolation from each other and those parts do not have common ownership or singular, comprehensive, overarching control.

If a SaaS or PaaS application is one part of the equation for application management, it is effectively a black box. The customer's IT department has no control over the application and little or no insight into the workings of the application specifically, or the environment in general. The only semblance of control over a SaaS application is any action that the customer might coax from the service provider's staff.

The lack of management control in a simple SaaS or PaaS environment is enough of a challenge, but when it is joined with a private cloud environment or an IaaS environment to provide greater functionality through a hybrid cloud structure, the complexity soars. There used to be a saying among Central Office technicians in telephony companies: "The trouble is leaving here fine." In a hybrid cloud that combines SaaS or PaaS with IaaS or private cloud, the customer's IT staff responsible for managing those hybrid cloud applications is in the same position as those Central Office technicians. They know what is happening to the applications in the private cloud or IaaS. They have all the tools they need to evaluate and control those pieces. However, they cannot see what is happening in the PaaS or SaaS applications. Furthermore, it is very difficult for them to assess how the interactions between these applications may alter the behavior of each application.

In more sophisticated organizations, a hybrid cloud may not be made up of a single, static pair of applications. Instead, it may be a collection of applications running in a dozen or more clouds. The interactions between them can be dynamic and unpredictable, appearing and disappearing depending on the needs of the various organizations. Management tools, techniques, and data sources that are relevant for one part of a hybrid cloud may be totally irrelevant for another portion of that same hybrid cloud. A significant challenge is in recognizing when relationships appear and disappear. It is only with that awareness that management tools can attempt to apply the relevant management techniques and data collection to all parts of the hybrid cloud.

## SUMMARY

The use of public and private cloud environments by organizations around the world continues to grow at a rapid pace and shows no signs of abating in the near future. The financial benefits of cloud computing will continue to drive its adoption. However, there is a tradeoff for those savings. In the IaaS and PaaS environments, the customers have the ability to monitor and control the environment and the applications running in it. Cloud computing brings unique challenges for managing the applications that run in those environments.

Application management in cloud environments is multifaceted. There are at least four types of cloud environments: private cloud, public cloud (SaaS), public cloud (PaaS), and public cloud (IaaS). In each of the public cloud environments there is a user (i.e., the customer's IT department) dimension to the question of application management. There is also a service provider aspect to application management. The capabilities and responsibilities are different in each instance. Unless the service provider offers some reporting, SaaS or PaaS environments are essentially black boxes. In IaaS and private cloud

environments, the customer's IT department is responsible for the active management of the applications and is able to install the appropriate tools that allow them to do that.

| | Manage Application Performance | Manage Application Availability | Monitor End User Experience |
|---|---|---|---|
| Private | Customer | Customer | Customer |
| SaaS | Service provider | Service provider | Customer |
| PaaS | Service provider | Service provider | Customer |
| IaaS | Customer | Customer | Customer |