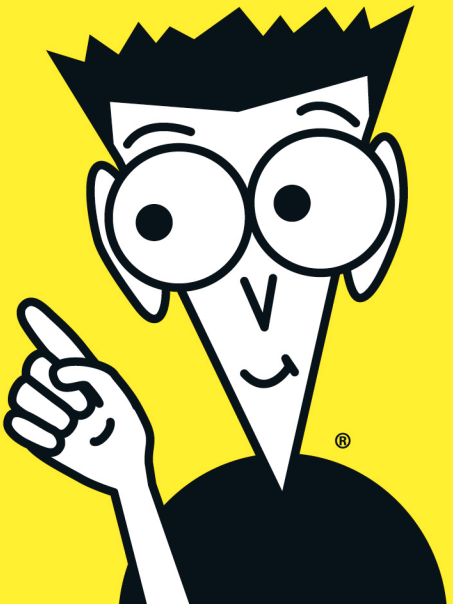
Compliments of IBM

IBM Limited Edition

Cloud Services FOR DUMMIES®

Learn to:

- Support business and IT objectives with cloud services
- Understand the economic benefit of cloud services
- Consider the right level of cloud security and reliability for your organization
- Identify ideal use cases for public, private, and hybrid cloud models



Judith Hurwitz
Marcia Kaufman
Dr. Fern Halper

Cloud Services FOR DUMMIES®

IBM LIMITED EDITION

by Judith Hurwitz, Marcia Kaufman, and Dr. Fern Halper

John Wiley & Sons, Inc.

Cloud Services For Dummies®, IBM Limited Edition
Published by
John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030-5774
www.wiley.com

Published by John Wiley & Sons, Inc., Hoboken, New Jersey

ISBN: 978-1-118-33891-9 (pbk) 978-1-118-34012-7 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

Table of Contents

Chapter 6: Starting Your Cloud Journey 55

Introduction

Welcome to *Cloud Services For Dummies,* IBM Limited Edition. Whether public, private, or hybrid, cloud computing is becoming an increasingly integral part of many companies' business and technology strategy. Cloud services help companies turn IT resources into a flexible, elastic, and self-service set of resources that they can more easily manage and scale to support changing business needs.

While many different delivery models for cloud computing services exist, two foundational services are a requirement for making cloud computing into a strategic part of an overall computing infrastructure. These include Infrastructure as a Service (IaaS) and Platform as a Service (PaaS). IaaS is the services that enable you to gain access to compute and storage resources in an on demand model. PaaS is the services that sit on top of IaaS and enable you to build applications to support the business.

No matter what your plan for your evolving IT infrastructure may be, you can take advantage of these services in combination with on-premises platforms to create flexibility for the business. Cloud computing serves different needs for different constituents within your organization. For business leaders, cloud computing is a cost-effective way to leverage IT resources to prototype and implement strategic change. For your IT organization, the cloud is a platform that allows it to be significantly more proactive and responsive when it comes to supporting strategic business imperatives. While IT is leading the charge in focusing on best practices that support the balanced use of public, private, and data center resources — the emerging world of hybrid computing — don't lose sight of the fact that cloud is just as much about business model transformation as it is about technology transformation. In fact, many companies find that the cloud helps to support increased collaboration between business and IT leaders enabling them to more quickly adjust to changing market dynamics.

About This Book

This book gives you some insights into what it means to leverage both IaaS and PaaS in both public and private cloud environments and how these environments work with data center services. Companies leveraging these foundational cloud services need to keep all the models in context with business requirements for performance, security, and portability.

This book helps put the foundational cloud services — IaaS and PaaS into context. In addition, the topics covered in this book are critical to the success of hybrid environments.

Foolish Assumptions

This book is useful to many people, but we have to admit that we did pick a segment of the world to focus on. Here's who we think you are:

- You're already using various forms of cloud computing and are planning a long-term strategy. Perhaps we're preaching to the choir. You understand that the benefits of using all kinds of flexible cloud computing models represent sources of sustainable competitive advantage.
- You're a business leader who wants IT resources to be a utility that's optimized to leverage what you've already paid for. You want IT to serve your business needs — you want to be able to execute your strategy on your timetable. You want IT to be your partner in innovating for the future.
- You're an IT leader who knows a lot about technology but aren't sure precisely how cloud computing — public, private, or a combination of the two as a hybrid model — works. You need to understand how cloud computing changes IT and what you need to do to support the business with cloud computing as an important enabler.

How This Book Is Organized

This book isn't intended to be an exhaustive technical manual on implementing and managing cloud computing. Instead, we give you a taste of the concepts and approaches you need to consider when embarking on your journey to the hybrid cloud.

We've organized this book into six chapters:

- Chapter 1 gives you an overview of the business case for foundational cloud services — what it means to the business and how these services support the overall IT approach.
- Chapter 2 provides you with an understanding of the technical foundation for IaaS and PaaS. The chapter includes use cases that explain the business benefits to the organization.
- Chapter 3 delves into the economics of cloud services. The chapter explains the type of economic benefit you gain from using foundational cloud services and how they support changing business requirements.
- Chapter 4 provides an overview of managing cloud workloads and services. Many different types of workloads need to be supported in the cloud. This chapter presents the different workloads and how they need to be managed to support IT and business needs.
- Chapter 5 provides insights into the important issue of security, governance, and cloud reliability. What does it mean to have a secure cloud and how do you ensure that your assets are safe? How do you have the right level of support for governance rules that keeps your company safe and in compliance?
- Chapter 6 gives you a roadmap for planning your journey to the hybrid cloud from a best practices perspective.

Icons Used in This Book

The following icons are used to point out important information throughout the book:

Tips help identify information that needs special attention.

Pay attention to these common pitfalls of managing your foundational cloud.

This icon highlights important information that you should remember.

This icon contains tidbits for the more technically inclined.

Chapter 1

Understanding Cloud Fundamentals and the Cloud Continuum

In This Chapter

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

How quickly things change. Cloud computing has evolved from a risky and confusing concept to a strategy that organizations 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. Not only are organizations using the cloud for services such as e-mail or customer relationship management, but also many are utilizing 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 the business and open up new innovative opportunities and new revenue streams.

The kind of cloud deployment you should consider depends on your own particular performance, security requirements, and your 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 understand that there's a

continuum of cloud services. These services range from an open and shared public environment to 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 organizations use a combination of private computing resources (data centers 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 center. 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 optimize the performance of those workloads to support your constituents.

Foundational Cloud Delivery Services

Understanding the foundations of cloud computing calls for understanding three main cloud delivery models:

- **IaaS:** The delivery of services such as hardware, software, storage, networking, data center 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 organization 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 allow the organization 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 organization, 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 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 touch points are 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 center, 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, they may choose to keep data about prospects on a public cloud. However, after those prospects become customers, the companies may begin storing that data on their own premises in their own servers, which is the hybrid cloud model.

Core Cloud Capabilities

Regardless of the model that 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 (you can use just the resources you want when you need them), so the users/customers provision resources, such as computing or storage resources, that they pay for 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 organizations 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 expenditures.

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 appropriate resources needed to build or operate an application. This just-in-time model is a more efficient way to ensure that the IT organization can be responsive to business change.

Billing and metering of service usage

A cloud service has to provide a way to measure and meter a 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 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 that these resources be optimized to work as though 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 organization'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 services. 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 models available to support the business. Your decision of 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 gray.

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 center. Some public clouds may offer a sophisticated level of security offerings while other public clouds have virtually no security at all.

Ultimately, you need to select the type of cloud service that provides 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 organization.

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

Figure 1-1: The cloud services continuum.

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 that 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 any one 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 the open community sites. These services generally have a written service level agreement (SLA) — an agreement outlining the obligation of the provider to the consumer of the service.

Contractual open

Public cloud vendors sometimes productize offerings. Here the user can't simply create login credentials, provide a credit card, and start using the service. Instead, the user actually signs a contract for service. The term can be as short as a month, or more typically a year. Vendors are also offering public IaaS and PaaS public platforms that are based on a formal contractual basis.

In this kind of environment, the expectation is for a high level 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 center. In these cases, a private cloud provides an environment that sits behind a firewall. Unlike a data center, a private cloud is a pool of common resources optimized for the use of the IT organization. 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 different 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 resides in a vendor's data center and provides 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

organization, and it may also use a SaaS HR public cloud application. 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 optimized 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 that other cloud services will sit upon. IaaS itself is the foundation upon which PaaS can be utilized to build value. It supplies the infrastructure that developers can use to build applications. For example, many organizations are using IaaS and PaaS linked together for the development and operations process — which we will get into later in this book. These organizations may even be using IaaS and PaaS to build actual SaaS services. So, in some ways, IaaS services is the base of a pyramid with the infrastructure at the bottom, the middleware (PaaS) at the center, and the applications on top.

Chapter 2

Digging Deeper into IaaS and PaaS

In This Chapter

- Understanding IaaS
- Diving into PaaS
- Defining the requirements for IaaS and PaaS

In this chapter, you look at Infrastructure as a Service (IaaS) and Platform as a Service (PaaS) — two foundational cloud delivery services. Both of these services provide flexibility that enables companies to respond faster to their constituents' needs.

IaaS can serve two purposes:

- It can be a highly practical solution for companies that want access to resources in an on-demand manner.
- IaaS can also be used effectively to augment data center services, either to increase capacity when needed, to replace aging hardware with cloud-based services, or to provide ongoing access to sophisticated services, such as advanced analytics.

A service model can enable a company to efficiently and effectively gain access to ongoing support as the business changes. In addition, IaaS can allow the business to freely experiment with new innovative software approaches without impacting the budget.

PaaS is a cloud delivery option that sits on top of IaaS to enable companies to take advantage of a unified and abstracted way of developing and deploying applications. PaaS hides the complexity of the underlying computing services from IT practitioners in order to streamline the application development life cycle and deployment process.

Diving into Infrastructure as a Service

IaaS, the most straightforward of the cloud delivery models, is the delivery of computing resources in the form of virtualized operating systems, workload management software, hardware, networking, and storage services. It may also include the delivery of operating systems and virtualization technology to manage the resources.

IaaS provides compute power and storage services on demand. Instead of buying and installing the required resources in their traditional data center, companies rent these required resources as needed. This rental model can be implemented behind a company's firewall or through a third-party service provider. See more information on renting in the section "Listing the characteristics of IaaS."

Virtualization is often used as a foundation of cloud computing. Virtualization separates resources and services from the underlying physical delivery environment. With this approach, you can create many virtual systems within a single physical system. A primary driver for virtualization is consolidating servers, which provides organizations with efficiency and potential cost savings.

Listing the characteristics of IaaS

IaaS has two types of services:

- A *public service* is designed so consumers in any size business can acquire services in a rental model. Some public cloud services are open to anyone with a credit card that pays per use. Other public cloud services are contractual and provide a higher level of service to the buyer.

- In contrast, *private services* are provided inside a company's firewall, enabling IT management to provide a self-service portal for employees and partners to easily access approved services.

Characteristics of both models are covered in this section.

Renting

When you purchase server and storage resources using IaaS services, you gain immediate access to the resources you need. You aren't, however, renting the actual servers or other infrastructure. It's not like a rental truck pulls up to your office to deliver the services. The physical components stay put in the infrastructure service provider's data center.

Within a private IaaS, renting takes on a different focus. Although you may not charge each user to access a resource, in the charge-back model, you can allocate usage fees to an individual department based on usage over a week, month, or year. Because of the flexibility of the IaaS model, the heaviest resource users can pay more than those who use fewer resources.

Self-service provisioning

Self-service provisioning is a key characteristic of IaaS that enables the user to obtain resources — such as servers and networking — through a self-service portal without relying on IT to provision these resources for them. The portal is similar to a banking ATM model that handles repetitive tasks easily through a self-service interface.

Dynamic scaling

Dynamic scaling occurs when resources can be automatically expanded or contracted based on the requirements of the workload or task. If users need more resources than expected, they can get them immediately. A provider or creator of IaaS typically optimizes the environment so the hardware, the operating system, and automation can support a huge number of workloads.

Service levels

Many consumers acquire capacity based on an on-demand model with no contract. In other situations, the consumer signs a contract for a specific amount of storage and/or

compute. A typical IaaS contract has some level of service guarantee. At the low end, a provider may state that the company will do its best to provide good service. Depending on the service and the price, you may contract for 99.999 percent availability. The level of service you require depends on the workloads you're running. We talk more about service levels in Chapter 5.

Licensing

The use of public IaaS has led to innovations in licensing and payment models for software you want to run in your cloud environment (not the license between you and your cloud provider). For example, some IaaS and software providers have created the Bring Your Own License (BYOL) plan so you have a way to use your software in both traditional or cloud environments. Another option is Pay As You Go (PAYG), which generally integrates the software licenses with the on-demand infrastructure services.

Metering

Metering ensures that users are charged for the resources they request and use. This metering to assess the charges for the IaaS services begins when the instance is initiated and ends when the instance is terminated. In addition to the basic per-instance charge, the IaaS provider may include charges for storage, data transfer, and optional services like enhanced security, support, or advanced monitoring.

Considering a private IaaS

A company would choose a private IaaS over a public one for three compelling reasons:

- The company needs to control access because of security concerns.
- The company may require that business critical applications demonstrate predictable performance while minimizing risk.
- The company sees itself as a service provider to its customers and partners.

A company selecting a private approach creates a pool of resources that can be standardized and easily reused by the IT organization to complete projects. Why standardize? In an IaaS service, IT projects are created in predictable ways. For example, a process may be designed to set up a test environment for code or provision storage to support an application. While certain nuances are different, 80 percent of the time the process within IaaS can be standardized. By standardizing these infrastructure services, the organization gains efficiencies, fewer inadvertent errors, and consistency in managing the development lifecycle. This is the same approach used by a public IaaS vendor to control its costs.

Knowing how companies use IaaS

Companies use IaaS for a variety of projects. Here are a few examples:

- A manufacturer needs a development infrastructure for its enterprise resource planning application. It decides to use a public IaaS service to provision development and test environments for the system in an "on demand" fashion. The public IaaS gives it the operational flexibility it needs to provision and de-provision infrastructure instead of constantly having to ask IT for server capacity.
- An insurance company needs a cost-effective compute infrastructure to run quarterly and yearly risk reports. During this peak time, its compute usage may be several times greater than normally would be needed. The insurance company doesn't want to over-invest in capacity that's only sporadically needed. It contracts with a secure public cloud IaaS to manage peak loads.
- A large retailer decides to deploy a private cloud IaaS to provide capacity on demand for its portfolio of retail applications that it offers as a service to a set of small retailers. It has its own in-house staff of application developers who often need more capacity than it has in its development environment for testing purposes. IaaS provides this compute capacity, as required.

The different companies all realized critical benefits from using IaaS:

- Flexibility to dynamically scale the environment to meet their needs
- Reduction in the need to build new IT infrastructure because of increase demands for resources
- Cost savings from eliminating capital expenditures on large systems that may be underutilized much of the year
- Almost limitless storage and compute power

Exploring PaaS

PaaS is another foundational service that provides an abstracted and integrated environment for the development, running, and management of applications. Often the PaaS is tightly integrated with IaaS services because it's utilizing the underlying infrastructure provided by the IaaS.

A primary benefit of a PaaS environment is that developers don't have to be concerned with some of the lower-level details of the environment.

PaaS vendors create a managed environment that brings together a combination of components that would've been managed separately in a traditional development environment. Services integrated in a PaaS environment include middleware (for example, software that allows independent software components to work together), operating systems, and development and deployment services to support software development and delivery. Some enterprises also become, in effect, a PaaS provider to their own internal developers. These organizations follow a similar process of applying best practices to standardize the services developers require to develop and deploy applications.

The goal of the PaaS provider is to create an abstracted and repeatable process for the creation and deployment of high-quality applications. These applications are designed to be implemented in public or private cloud environments.

Variations in PaaS delivery models

PaaS comes in different shapes and sizes. If you're using a *public* cloud-based PaaS, the vendor shoulders the responsibility of managing the middleware software resources and the overall development and deployment environment. If you decide to create your own PaaS environment, your organization is responsible for maintaining the right level of service.

A public PaaS environment looks and acts very differently than your traditional development and deployment platform. For example

- Resources aren't delivered as software in PaaS. Instead the PaaS environment is hosted so the third party is responsible for uptime performance and software updates.
- The development and delivery of services lives in the cloud instead of in a single system.
- Middleware and services have no installation and configuration because they're an integral part of the PaaS platform.
- Because the PaaS is tightly coupled with IaaS services, it offers a consistent way to manage and optimize applications from development to deployment (DevOps).

A *private* PaaS environment also looks and acts differently than your traditional development platform. Large enterprises may implement well-designed patterns and best practices to achieve efficiency and productivity gains and reduce software development and deployment costs.

Increasingly software developers are becoming strategic partners for the business. To support this strategic role, development organizations are adopting new business practices including the following:

- Application developers no longer operate as disconnected units making individualized selections for hardware and software development tools to fit each new project.
- Enterprise IT standardize on a framework for all developers to use to write their code.

Understanding the benefits of PaaS

Organizations can gain a few different benefits through a PaaS environment. For example, it's possible to architect a private cloud environment so development and deployment services are integrated into the platform. This provides a similar benefit gained from a public PaaS but in a private environment. A private PaaS implementation can be designed to work in concert with public PaaS services.

The benefits to using PaaS include the following:

- **Improving the development life cycle:** Effectively managing the application development life cycle can be challenging. For example, teams may be in different locations, with different objectives, and working on different platforms. When it comes time to integrate, test, and build the application, problems can arise because developers are working on different platforms with a different configuration than the operations team is working on. In another situation, some developers don't have the latest version of the code. These same developers may also be using a different set of tools. A key benefit of an abstracted platform is that it supports the life cycle of the application.
- **Eliminating the installation and operational burden from an organization:** Traditionally, when a new application server or other middleware is introduced into an organization, IT must make sure that the middleware can access other services that are required to run that application. This requirement can cause friction between Development and Operations. With PaaS, these conflicts are minimized. Because the PaaS environment is designed in a modular, service-oriented

manner, components can be easily and automatically updated. When PaaS is provided by a third-party organization, those changes occur automatically without the user having to deal with the details. When PaaS is implemented in a private cloud, the IT organization can automate the process of updating a self-service interface to provision the most current services to the IT organization.

- **Implementing standardization:** PaaS enables development professionals and IT operations professionals to use the same services on the same platform. This approach takes away much of the misunderstanding that happens when the two teams with different responsibilities aren't in sync.
- **Having ease of service provisioning:** A PaaS provides easy provisioning of development services including build, test, and repository services to help eliminate bottlenecks associated with non-standard environments. This in turn improves efficiency, reduces errors, and ensures consistency in the management of the development life cycle. Additionally, PaaS provides ease of provisioning in runtime services that include application runtime containers for staging, and running and scaling applications.

PaaS has two fundamental parts: the platform and the service. The service is what can set the PaaS vendor apart from its competitors. The PaaS vendor continuously services and improves the software. As new updates and new configurations become available, the PaaS vendor can immediately push them to its customers.

Having the Correct Requirements for IaaS and PaaS

In many instances, a blurring of the lines of IaaS and PaaS occurs. In fact, many vendors are already offering IaaS as part of a PaaS solution. That's something to consider when you put together your infrastructure and development plans.

The key requirements for IaaS and PaaS include the following:

- A consistent platform that's optimized to support a variety of workloads needed by customers
- An integrated stack of middleware optimized for automated deployment and management of heterogeneous workloads that dynamically adjusts
- Reliable, highly secure and scalable platform

  The continuum of the cloud rests on the reliability and security of the platforms used and the track record of the cloud provider to support customers' demands.
- A choice of deployment models that support the right service level, quality of service, and security required to support constituents

Chapter 3

Diving into Cloud Economics

In This Chapter

- Forming an economic strategy
- Exploring the costs of the cloud

Many companies today are expanding into cloud computing as a way to reduce the cost and complexity of delivering traditional IT services. But determining the best mix of public and private cloud services and data center services is complicated. You can't simply add up specific costs and make a fast determination of what's the best approach for your organization. Instead, you have to look at your business requirements for performance, availability, and security and the workloads that you need to support. A *workload* is an independent service or collection of code that can be executed. You have to look at services such as Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) in context with your business goals and the impact on the customer experience.

In this chapter, you examine both the indirect value factors that affect the economics behind IaaS and PaaS models as well as the direct cost factors.

Developing an Economic Strategy

An organization typically has many different types of workloads to manage in its data center that may be better suited for a cloud environment. So, to optimize your economic benefit from the cloud, you must first have a good understanding of your workload requirements.

Developing an economic strategy for the cloud can be a balancing act. Some workloads may be more suited to your data center. There are compelling reasons why others belong in the cloud. And, while most organizations can't predict the actual costs of running any given service in a data center, looking at direct and indirect costs of moving to the cloud is important.

Comparing traditional models with the cloud

As companies move to consider a cloud model, they no doubt are looking at their traditional data center. So, here's an important point: The data center doesn't go away. After all, almost all medium-size and large companies run their own data center to support and operate their systems of record, including accounting systems, payroll, and human resource as well as industry specific applications.

When you think about a traditional data center, you generally picture a centrally managed data center that contains all of a company's applications and data. While initially the data center was well planned and architected, it has typically evolved over decades to be less streamlined. Today, the typical data center supports different hardware, architectures, operating systems, applications, and tools.

Although IT has made the data center more efficient, organizations are taking a hard look at what workloads the centralized data center is well suited for. The reality is that the traditional data center is often best suited for a complex line of business applications. These applications are often

transaction-intensive and need to confirm and track the movement of financial transactions among customers, suppliers, and partners. Additionally, large, often highly customized systems of record are and will continue to be data center-based. These applications are typically tightly managed for corporate governance and compliance.

The traditional data center is designed to manage applications, but the cloud is designed to manage homogeneous workload resources. Managing workloads is foundational for the cloud, and we discuss workloads more in Chapter 4.

IaaS and PaaS models are intended to manage a pool of resources, which is a set of shared, configured services that are independent of a physical location. PaaS models can be optimized to manage a particular process, such as application management. In many situations, cloud service providers create a multi-tenant environment to support the deployment of these resources. Multi-tenancy enables the sharing of a service while keeping the data and configurations of individual customers separate, meaning that in an IaaS model, for example, you may be sharing server infrastructure with many other users (either internally in the private cloud or externally in the public cloud). This factor enables economies of scale for the cloud provider.

Finding the value

Operational performance, security, economics, and flexibility all have a great impact on an organization's cloud strategy. Striking the right balance among public cloud services, private cloud, and the data center can come down to a mix of these factors or can be dominated by just one. Finding the right mixture of environments is critical for your organization to achieve the best value when creating a cloud strategy and determining which model may provide economic value.

For example, a move to IaaS is likely to deliver an economic benefit if you have the need for increased capacity. This may take several forms:

- Say your organization is ramping up for a new but short-term initiative and you temporarily need some extra CPU capacity and storage. This may be a good match for an

IaaS service because building out a full infrastructure for unpredictable requirements isn't economical.

- Organizations may have a continuing need for additional compute or storage resources over time and can use a private IaaS to make those resources available on demand for a variety of projects. For example, additional public or private resources may be required as hardware reaches end-of-service life.
- Some organizations are using IaaS for *cloudbursting* when there are unexpected or planned high-load periods. The flexibility of using IaaS means that the company doesn't have to overinvest in hardware. These companies must be able to adapt to higher loads to protect themselves.

In the case of PaaS, companies find that PaaS can offer value during development and deployment, instead of having to purchase many different tools. By providing the underlying software infrastructure, PaaS can reduce organizational costs. PaaS reduces many of the costs involved with the traditional application development and deployment model including server and storage overhead, operating systems, workload and performance management software, network bandwidth, software maintenance, and support personnel. However, there are also some indirect value drivers for PaaS. These include

- **Reducing careless mistakes:** Mistakes, such as someone forgetting to load a configuration file can take a week before the problem is identified. Such mistakes add up to time and money wasted and cause deployment delays. With PaaS, such mistakes are reduced because the platform has been tested and is known to work. These successful and well-understood patterns are an important benefit of a PaaS.
- **Lowering skill requirements:** Perhaps only one or two people in an organization have the skills necessary to work with a certain kind of middleware. By providing the development tools and middleware, PaaS lowers the skill level required to deploy applications and removes the bottleneck that can form waiting for one specific person's assistance.

- **Maintaining speed, flexibility, and agility:** By providing a predictable, heterogeneous application infrastructure, organizations don't get bogged down with a different approach to operations than to developing applications. Therefore, employees, customers, and suppliers can gain better access to the services they need.

Exploring the Costs

When you're looking at the right balance of public cloud, private cloud, and data centers services, you have to take a step back and look at the overall costs of every environment. Start by understanding what it costs you to operate your data center. To do this, look at both direct and indirect costs related to the application or type of workload you want to move to the cloud. Some of these indirect costs are hard to evaluate, making it difficult to accurately predict the actual costs of running any given application in your company.

Here is a fairly comprehensive list of possible costs:

- **Server costs:** With this and all other hardware components, you're specifically interested in the total annual cost of ownership, which normally consists of the cost of hardware support plus some amortization cost for the purchase of the hardware. Additionally, a particular server may be used to support several different workloads. The more disparate workloads a server manages, the higher the support costs.
- **Storage costs:** What are the management and support costs for the storage hardware required for the data associated with this application? Storage costs may be very high for certain types of applications, such as e-mail or complex analytics.
- **Network costs:** When a web application you host internally, such as e-mail or collaboration, is moved to the cloud, this may reduce strain on your network. However, it can substantially increase bandwidth requirements.

- **Backup and archive costs:** The actual savings on backup costs depends on what the backup strategy is when the application moves into the cloud. The same is true of archiving. First, you have to understand who's doing the backup and archiving. Is backup the responsibility of the IT organization or is it handled by the service provider? Will all backup be done in the cloud? If so, do you have a contingency plan if that cloud service is unavailable when you need that backup? Will your organization still be required to back up a percentage of critical data locally?
- **Disaster recovery costs:** In theory, the cloud service has its own disaster recovery capabilities, so there may be a consequential savings on disaster recovery. However, you need to clearly understand what your cloud provider's disaster recovery capability is. For example, does the cloud service provider have mirrored sites in case of a power outage at one data center location? IT management must determine the level of support the cloud provider will offer. This can be an added cost from the provider, or you may seek out a secondary vendor to handle disaster recovery and procedures.
- **Data center infrastructure costs:** A whole series of costs — including electricity, floor space, cooling, and building maintenance — are an integral part of managing any data center. Because of the large investment in data centers, moving workloads to a public cloud may not be financially viable if you're only utilizing as little as 40 percent of the data center's compute power. (Of course, you can deploy a private cloud to take advantage of the underutilized space and the advantages of the cloud.)
- **Software maintenance costs:** What's the annual maintenance cost for the software you may move to a cloud-based service? The answer can be complicated if the software license is part of a bundle or if the application is integrated with other applications. In addition, there's the cost of purchasing the software. Is the organization taking advantage of a "pay-as-you go" licensing model that allows the user to pay only for what's used?
- **Platform costs:** Some applications run only on specific operating environments — Windows, Linux, HP-UX, IBM z/OS, AIX, and so on. The annual maintenance costs

for the application operating environment need to be known and calculated as a part of the overall costs.

- **Support personnel costs:** What are your costs for staff support for day-to-day operations and management of this application? Will some of these costs be transferred to the cloud provider? Your own personnel will still be required to manage and monitor your cloud services in concert with your data center services.
- **Infrastructure software costs:** A whole set of infrastructure management software is in use in any installation in the data center and in a hybrid environment. Needless to say, associated costs are involved. For example, management software is typically used across a variety of data center applications and services. It is typically difficult to separate costs that may be applied to a hybrid cloud environment.

Some of these costs aren't likely to be affected by migrating a single application to the cloud. However, if you move multiple applications to the cloud, you may realize a significant decrease in many of these indirect costs.

The reserved capacity package

Reserved capacity is a part of many cloud initiatives. A *reserved capacity package* is a pool of virtual machine resources, which helps ensure resources are available when you need them. You commit to a certain period of time, such as 6 or 12 months, and can provision and de-provision virtual machines within your pool during that time. A pool has one or more units of capacity. For example, one unit may include

- 64 virtual central processing units (CPUs)
- 96 gigabytes (GB) of memory
- 9,600 GB of storage

The great thing about reserved capacity is that you pay a monthly fee for the reservation, and then pay a discounted rate only for the virtual machines you provision within the pool. You aren't locked into a set fee for the entire pool of virtual machines over the course of the reservation.

If you don't need reserve capacity, you can simply pay as you go for virtual machines you provision. But without a reservation, you aren't guaranteed capacity.

What you save or gain with cloud services

Some hard costs are involved with cloud services, but understanding your business needs and growth strategies helps you put the costs into perspective. Having the ability to pay on an as-needed basis for servers, storage, and other services, for example, can give your company a needed boost for implementing innovative solutions with far less economic risk than buying the required infrastructure upfront.

While business and IT come to the cloud for different reasons and with different goals, both see the cloud's overall potential:

- **Lasting customer relationships:** Implementing an online collaboration for your customers and partners can lead to business innovation and transformation that far outweighs the costs associated with the implementation.
- **IT without traditional boundaries and restrictions:** Whereas traditional IT can trap technology and information in silos, cloud services simplify access and help connect people across the enterprise (and beyond).
- **Improved speed and agility:** Cloud services help you increase the delivery of IT computing resources. Whether your goal is to be first to market with a new product or simply providing the best, most-responsive customer service, speedy yet flexible IT is a necessity.
- **Transformation of the economics of IT:** Self-service and deployment help speed delivery of new offerings and services to your customers.

Cost calculating

One approach to estimating costs is to first examine your expected workloads in detail and then use an estimator tool to calculate real-world costs of running those workloads in the cloud.

Assessing workloads

When considering a migration to a cloud environment, you need to know which applications transition easily and give you the best return on investment (ROI). Assessing potential workloads is key to understanding what you should migrate and what should remain in a traditional IT environment.

Many kinds of workloads are far from static and predictable. Workloads generally have a stable base but experience minor fluctuations and occasional peaks. The peaks may be seasonal in nature or triggered by a business event (such as a product launch), a sudden change in market conditions, or a product recall.

To manage all these capacity requirements without having to invest in excess capacity, you need to plan ahead and decide which workloads can be moved to a dynamic infrastructure. Some vendors, such as IBM, provide workload assessment tools and services, to help prioritize and classify potential workloads for cloud delivery. After selecting inputs in a spreadsheet-like form, the tool gives you a pain versus gain score that reflects a combination of effort (to migrate), investment, and benefit of migration.

Using a cost estimator tool

A number of cloud providers and vendors offer calculators for helping you estimate charges for their services or to help you estimate the cost savings from cloud computing services. Some vendors provide monthly calculators for their web services.

For example, a calculator may ask you a series of questions about the number of compute instances, storage needs, data transfer, load balancing, and IPs needed and then provide you with a monthly estimate. Of course, your monthly usage and, therefore, the charges may vary from the estimates that the calculator provides.

Other vendors may offer total cost of ownership (TCO) calculators. These calculators may ask you a series of

questions about the type of deployment, the number of servers, storage requirements, and load volatility. Then the calculators estimate how much you may be able to save versus a data center deployment over five years. These calculators look at factors such as server utilization, facility, power, and hardware costs, as well as the cost of downtime, reduction in deployment, and provisioning time.

Don't base your decision to move to the cloud simply on these calculators. Some of these calculators are best used to help you better understand your computing requirements including support, training, and migration costs. These types of cost estimation tools help you determine when a public service or a private service provides the optimal solution to your future computing requirements.

Chapter 4

Managing Cloud Workloads and Services

In This Chapter

- Digging into workloads and workload management
- Exploring workload use cases
- Understanding key principles of workload management
- Moving your workloads

A *workload* is an independent service, application, or collection of code that can be executed. We've mentioned workloads in the first three chapters of this book, but managing workloads is so fundamental to the success or failure of your cloud activities (whether it be public, private, and especially hybrid) that it deserves its own chapter.

In cloud computing, workloads are abstracted from their physical implementation, meaning that they're isolated from the hardware they are running on. Therefore managing cloud workloads involves a different approach than companies may be accustomed to in a traditional environment.

Workloads need to be structured and packaged so they can execute most efficiently. But all workloads aren't the same: they come in many forms and flavors, as you see in this chapter, and each also comes with its own management and monitoring needs.

Understanding Workloads

Because computing requirements are varied, so too are the workloads. Whether you're using an IaaS for infrastructure or you're developing SaaS applications using a PaaS, here are some of the kinds of workloads you're likely to find in a cloud environment:

- **Batch workload:** These workloads operate in the background and are rarely time sensitive. Batch workloads typically involve processing large volumes of data on a predictable schedule (for example, daily, monthly, and quarterly).
- **Database workload:** These are the most common type of workload, and they affect almost every environment in the data center and the cloud. A database workload must be tuned and managed to support the service that is using the data. A database workload tends to use a lot of Input/Output (I/O) cycles.
- **Analytic workload:** Organizations may want to use analytic services in a cloud environment to make sense of the vast amounts of data across a complex hybrid environment. In an analytics workload, the emphasis is on the ability to holistically analyze the data embedded in these workloads across public websites, private clouds, and the data warehouse. A social media analytics workload is a good example of this. These kinds of workloads tend to require real-time capabilities.
- **Transactional workload:** These are the automation of business processes such as billing and order processing. Traditionally, transactional workloads were restricted to a single system. However, with the increasing use of electronic commerce that reaches across partners and suppliers, transactional workloads must be managed across various partners' computing environments. These workloads are both compute and storage intensive. Depending on the cost-benefit analysis, it's likely that complex transactional workloads are best suited to a private cloud.

- **Test/development workloads:** Many organizations leverage the cloud as a platform for testing and development workloads. Using cloud services can make the process of creating and then testing applications much more cost effective and efficient. In this way, developers have access to a set of common confirmations and development tools. Testing can be accomplished in a more efficient way within a cloud environment.

Of course, some workloads are simply not suited for a cloud implementation. One example may be a workload that needs high performance network storage. Because these workloads may need to be accessed very quickly, they may not be suited for the cloud (say in an IaaS model) where you're dependent on the Internet for network speed. It makes sense to do a cost-benefit analysis that looks at your particular workload and what it costs you to migrate it to the cloud versus the expected benefit of that move (check out Chapter 3 for more information).

Looking at Workload Use Cases

This section gives you two use cases that illustrate the kinds of workloads described in this chapter.

Analytics workload

A major maker of life-science tools and integrated systems for large-scale analysis of genetic variation and functions needed a cost-effective computer infrastructure. It wanted to expand to meet growing demand for processing related to genome research without scaling its IT investment and staffing.

A public cloud solution let this company scale operations in parallel, so it could leverage multiple virtual infrastructures to different work groups at the same time. This permitted it to offer genome processing as a service at a competitive cost per processing run. Its genome analytics application involves truly huge amounts of data, and provides a wide range of analysis in the form of outcome-focused reports and analytics.

For this science services provider, the benefits of a public cloud solution included

- Creation of a custom cloud-hosted software platform designed specifically to meet genome processing needs
- Ability to scale to meet information processing and data handling needs for genome research
- Ongoing access to current infrastructure (systems, software, and communications) without requiring a substantial investment in hardware or software

Batch workloads

A large insurance provider in North America required a cost-effective computer infrastructure to support quarterly and year-end batch processing for capital reserves and risk reporting (required by law at state and federal levels in the U.S.). During such peak periods, computing capacity needs routinely quadruple. Purchasing capacity to meet such needs could cost millions. Naturally, this insurer wanted to meet its peak demands without over-spending on computing capacity that would go underutilized during off-peak times.

A public cloud with IaaS capability provides a stable and reliable platform for provisioning compute capability and related infrastructure when needed. When the peak subsides, this environment enables quick and easy de-provisioning of added capacity as well. For this insurer, using the public cloud let it purchase extra capacity only when added work justified its use and saved it 75 percent as compared to the cost of acquiring such capacity in-house.

The benefits to this insurance provider included the following:

- Scaling computing capacity and infrastructure up and down to match actual demand over time
- Maintaining a current infrastructure, without incurring capital costs for hardware or operating costs for software and services
- Not paying for underutilized computing capacity just to make sure it's available only when peak times occur

Looking at the Principles of Workload Management

Management in this context refers to how resources are assigned in order to process workloads. Assignments may be based on resource availability, business priorities, or event scheduling.

The idea of managing workloads has been around for decades. In the unified mainframe-computing era, workload management was pretty straightforward. When a task had to be executed, a job was scheduled to run on that system. The instructions for running that task or job were typically written in a complex job-control instruction language. This set of commands helped the IT organization carefully plan the execution of workloads. If a mission-critical workload required a huge amount of time to run, a set of instructions could be established to stop that workload and allow another workload to run. When the second workload finished executing its task, the long-running workload could resume.

The challenge in managing any workload is making sure that it can be executed and delivered at the right performance level. It involves understanding processing requirements, modeling resources, and determining capacity. The principle is not that difficult if you're dealing with applications running on one server or even in a homogeneous cloud environment. And, if you're using just a public cloud provider then that provider manages your workloads. However, as IT infrastructures become more complex and heterogeneous, such as in a hybrid environment, this becomes harder to do.

Seeing Workload Management in a Hybrid Cloud

Things get a lot more complicated in a hybrid cloud environment. With the advent of a hybrid cloud, many more applications and services exist across different countries that have to run. Some workloads may be permanent and need to

run constantly, such as an online commerce site or a control system that manages a critical environmental process. Business services and various application models are added into the mix as well.

In a hybrid cloud environment, your workloads may be running on different clouds, and running different kinds of infrastructure using different operating systems. You're bringing together workloads from different environments that often have to behave as though they're a unified system.

Now you may think that all you have to do in a hybrid cloud environment is to get some automation software to automatically schedule resources and to perform some other functions associated with allocating resources and you're done. However, do consider some issues when thinking about how to create a hybrid cloud environment that both performs at a quality level and meets security and governance requirements.

For example, say a workload is being used within a geography that has different rules for where data must be placed. If the data has to be stored within a country, then that workload has to be managed differently than the same workload running in a country without this kind of governance requirement. With fewer restrictions, IT operations is free to move workloads to locations that have the bandwidth or capacity to meet the quality of service the business needs. In fact, the ability to change and move workloads based on business requirements is at the heart of operational issues in the cloud.

Connecting Workloads in the Cloud

There are different kinds of workloads that you work with in the cloud, and while balancing workloads may not be a problem when you have a homogenous workload running in a particular cloud, like a public cloud for e-mail, things can get more complicated as you try to bridge different environments. That's where portability and standards come into play.

Say you're developing and testing an application in the cloud using a public PaaS. However, all your databases reside in your on-premises data center. That can get complicated. Or, say you're building an application using a private PaaS, which needs to access your CRM system that is in the cloud. There needs to be a way for the two environments to share information. The question becomes, how do you manage workloads across potentially incompatible environments?

The importance of APIs

Application programming interfaces (APIs) allow communication to occur between products and/or services. For example, if you've developed a gaming application you can write an API that allows other developers to write to your application. The API specifies how one application can work together with another one. It provides the rules and the interfaces. The developer doesn't need to know the nitty-gritty of your application because the API abstracts the way these programs can work together.

APIs are important for managing workloads in a cloud environment. For example, APIs allow a developer to build an application that runs on top of a public IaaS service. In fact, every company that offers foundational cloud services, such as IaaS, PaaS, and SaaS, provides APIs for its customers.

The sticky point with APIs and the cloud is that different APIs aren't always compatible. For example, you may build an application on top of an IaaS offered by one vendor, but if you want to move it to another cloud provider, it may require extensive re-programming.

A standard workload layer

No standard API exists that enables the developer to work in different cloud models provided by different cloud vendors. What is actually needed is a standard layer that creates compatibility among cloud workloads. Of course, you can always find ways to work around complicated problems.

In hybrid workload model, management companies, such as IBM, create customizable templates that allow developers to make allowances for differences in APIs and are able to deploy and migrate workloads.

Portability of workloads

Discussing APIs and standards is essential because workload management is fundamental to the operation of a hybrid cloud. In a hybrid cloud environment, being able to move workloads around and optimize them based on the business problem being addressed is critical. Despite the fact that workloads are abstracted, they are built with middleware and operating systems. In addition, workloads must be tuned to perform well in a specific hardware environment.

Another challenge is that a majority of workloads in the cloud are virtualized with hypervisors such as KVM, VMware, and PowerVM. Each of these virtualization implementations are different and will impact the portability of workloads. In today's hybrid cloud world, a lot of manual intervention is needed to achieve workload portability. However, in the future standards and well-defined approaches will hopefully make hybrid cloud workload management a reality.

Managing and Monitoring Workloads

To support the needs of the business with a hybrid cloud environment requires that considerable attention be paid to how workloads are managed and monitored. Increasingly, organizations depend on these environments to support their internal teams as well as partners and suppliers. Without careful management and monitoring of the required workloads, the organization can't achieve the right level of support and service the business demands.

Tracking workloads

Managing workloads in a hybrid cloud environment requires a set of distinctive steps including the following:

- Keep track of dependencies among specific services, such as IaaS, PaaS, and SaaS.
- Workloads need to be monitored and optimized based on the company's service level requirements.
- Governance of workload management is essential for success. The IT organization needs to guarantee that corporate and governmental regulations are adhered to.
- Workload transparency is important, no matter where they are physically located. This includes on-premises systems as part of the overall workload management environment.

In Chapter 5, we provide more information about workload management in terms of security, governance, and reliability.

Asking the right questions

In a traditional data center, workloads tend to be constructed as complete applications instead of independent workloads. The criteria for success in optimizing workloads in the data center are performance, reliability, and security. In a cloud model, and particularly a hybrid cloud model, where the workloads are often not tied to a particular server, the challenge is to provide performance, reliability, and security in a constantly changing world. Balancing an acceptable level of risk and an acceptable level of service is delicate. So before you make decisions related to workload management in a hybrid cloud, ask yourself the following questions:

- What's the purpose of your workloads and how do they support the business?
- What are the legal risks that are unacceptable?

- What's the reputation of the public cloud provider?
- How well does your internal organization understand the various internal and external workloads that need to be supported?

You need to be practical in how you address these questions. If the workloads require real-time performance, you need to keep those as close to the source of the transactions as possible. However, some workloads can run with less stringent performance and can be placed in a less expensive cloud model. Regardless, you never want to risk the reputation of your company. Part of the risk is selecting a cloud provider that is here today and gone tomorrow. Workload management must be viewed as an overall part of your cloud management strategy. Therefore, planning for managing workloads goes hand in hand with a strong cloud strategy plan.

Chapter 5

Improving Security, Governance, and Cloud Reliability

In This Chapter

- Understanding the nuances of cloud security
- Ensuring accountability through proper governance
- Establishing the right service delivery levels for your company
- Asking the right questions

Many companies contemplating the addition of the cloud into their IT strategy are concerned about three key issues: security, accountability, and reliability. This chapter presents best practices that can help you improve security, build a reliable and resilient environment, and put proper controls in place to meet governance requirements.

Finding out Why Cloud Security Matters

Security is top on the list of any IT manager who's thinking about the cloud. Whether you're looking at creating a private cloud, leveraging a public cloud, or implementing a hybrid environment, you must have a security strategy.

Many of the same security risks that companies face when dealing with their own computer systems are found in the cloud, but there are some important twists. With the cloud, you no longer have well-defined boundaries regarding what's internal and what's external to your systems. You must assess whether holes or vulnerabilities exist across servers, networks, infrastructure components, and endpoints and then continuously monitor them.

According to the Cloud Security Alliance (CSA), an organization dedicated to ensuring security best practices in the cloud, significant areas of operational risk in the cloud include the following:

- **Physical security:** Covers security of IT equipment, network assets, and telecommunications infrastructure
- **Human resource security:** Deals with the people side of the equation — ensuring background checks, confidentiality, and segregation of duties (for example, those who develop applications don't operate them)
- **Business continuity:** Ensures that the provider meets its service level agreement for operation with you
- **Disaster recovery:** Ensures that your assets (your data and applications) are protected

  If, for example, you're using a public Infrastructure as a Service (IaaS) to run an application, find out what happens if there's some sort of disaster (natural or otherwise).
- **Incident handling changes in a cloud:** Working with your service provider to control at least part of the infrastructure

  The multi-tenant nature of the cloud often makes investigating an incident more complicated. For example, because information may be commingled, log analysis can be difficult because your service provider is trying to maintain privacy.
- **Application security changes in the cloud:** Uncovering exposed security threats (in a public cloud)

  The CSA divides application security into different areas including securing the software development lifecycle, authentication, authorization, identity management, application authorization management, application

monitoring, application penetration testing, and risk management. So, if you're using a Platform as a Service (PaaS) to develop applications, be concerned about application security. Likewise, if you're running your application in the cloud or using a SaaS provider, application security will be an issue.

- **Identity and access management:** Controls and maintains access to computer resources, applications, data, and services

  In a traditional data center, you may use a directory service for authentication and then deploy the application in a firewall safe zone. The cloud often requires multiple forms of identity to ensure that access to resources is secure.

- **Encryption and key management:** Ensures that only intended recipients receive data and can decrypt it

  Data encryption refers to a set of algorithms that can transform text into a form called cyphertext (an encrypted form of plain text that unauthorized parties can't read). The recipient of an encrypted message uses a key that triggers the algorithm to decrypt the data and provide it in its original state to the authorized user.

Develop a thoughtful approach to cloud security to succeed in mitigating security risks. Part of this involves asking your cloud provider some tough questions. It may also include a visit to the provider's facility. Here are a few suggestions for questions to ask your potential provider:

- What security policies does it have in place? Are they consistent with a recognized framework and control standard?
- Does the provider have any industry certifications?
- How does the provider meet audit standards?
- Does the service provider have documented policies and procedures, including escalation procedures in the event of an incident?
- How does the provider handle identity and access management?
- How does the provider protect data?

Speak to your cloud provider regarding what data controls it provides. In addition, develop and publish a consistent set of rules and policies regarding the creation, capture, management, transmission, storage, and deletion of confidential and business-critical data. Make sure that your provider adheres to data location requirements dictated by certain governments. Data is the lifeblood of your organization, so you don't want it to be compromised in any way. Not only do you want to save your data but also restore that data to ensure that no data is lost.

Establishing a Cloud Governance Strategy

Hand in hand with a security strategy needs to be a governance strategy — a way to ensure accountability by all parties involved in the cloud deployment. Basically, governance is about applying policies related to using services. Governance incorporates the organizing principles and rules that determine how an organization should behave when leveraging cloud services. These policies determine who is accountable for what actions. Cloud governance is a shared responsibility between the users of cloud services and the cloud provider. Understanding the boundaries of responsibilities and defining an appropriate governance strategy for your company requires careful balance.

Governance issues in the cloud

Of course, cloud governance can be tricky. IT governance is hard enough. Cloud governance requires governing your own infrastructure as well as infrastructure you don't totally control. A successful governance strategy in a cloud world requires a negotiated agreement between you and your cloud providers(s). Generally, several goals are involved in cloud governance, including risk and monitoring performance. Your governance strategy needs to be supported in two ways:

- **Understanding the compliance and risk measures the business must follow:** What does your business require to meet IT, corporate, industry, and government

requirements? For example, can your business share data across international borders? These requirements must be supported through technical controls, automation, and strict governance of processes, data, and workflows. Make sure that you have the ability to control how much you're spending for independent cloud services based on financial and business requirements.

- **Understanding the performance goals of the business:** If you have specific performance demands, work with your service provider to establish benchmarks based on business goals. You may have an uptime requirement or a need to monitor who has access to services. Your cloud provider needs to be able to support these goals and help you optimize business performance.

Risks worth noting

Governance has a lot to do with assessing and managing risk. If you're going to hold a cloud provider (public, private, or hybrid) partly accountable for your IT cloud services, you need to consider risks.

As you move into a cloud model, your governance team needs to consider the following risks:

- **Audit and compliance:** Include issues around data jurisdiction, data access control, and maintaining an auditing trail
- **Security:** Includes data integrity, confidentiality, and privacy
- **Other information:** Include protection of intellectual property
- **Performance and availability:** Include the level of availability and performance your business needs to be successful
- **Interoperability:** Associated with developing a service that may be composed of multiple services

  Are you assured that the infrastructure will continue to support your service? What if one of the services you're using changes?

- **Contract:** Associated with not reading in between the lines of your contract
- **Billing:** Associated with ensuring that you're billed correctly and only for the resources you consume

Of course, your governance body has the responsibility of monitoring these and other risks on an ongoing basis.

Making cloud governance work

Effective management of the cloud is going to be part people and processes and part technology. It's really a three-part solution:

- You may want to set up some sort of governing body to deal with all the different environments you now need to manage. Hopefully you already have an existing governance group in place. This group can be part of that group. The governance body has oversight responsibilities and collaborates with the business. This group deals with cloud providers to discuss the issues and negotiate terms and conditions with cloud providers.
- Your organization needs to have governance bodies in the cloud that deal with standardization of services and other shared infrastructure issues. You need some sort of interface to these groups. Of course, your level of involvement depends on your level of involvement in the cloud.
- Your organization also needs to have technology in the mix that helps your organization automatically monitor the performance of the cloud provider.

Managing Service Levels

The third leg of the stool in the security and governance triad deals with managing service levels. Cloud providers need to monitor cloud services to ensure that they meet agreed upon service levels. This means that the performance of the servers, networks, and virtualized images in the cloud providers' environments need to be both measured and monitored — individually and collectively — to ensure that the environment is tuned to satisfy all business requirements.

A standardized and automated system needs to be in place to track, trace, and audit all aspects of performance. For example, bandwidth, connectivity, and scalability are all performance characteristics that should be monitored. In addition, automated systems should be designed to quickly identify the root cause of the hardware or software failures so performance can be restored.

These systems should answer questions such as

- Is the cloud infrastructure performing as expected?
- Are identified performance problems occurring randomly at regular intervals?
- Which performance problems are most severe and need to be given top priority to find the root cause and resolve the issue?
- How can performance be improved?

Monitoring must include the ability to quickly respond to dramatic or unexpected performance fluctuations, especially as a result of malicious activities. Authorized users need visibility into your public, private, and/or hybrid cloud environments.

Defining SLAs

A *service level agreement* (SLA) is a document that captures the understanding between a service user and service provider that defines uptime, availability, and performance. The SLA is also a contractual agreement between the participants in a service delivery contract. In the world of computing, an SLA is typically written based on the expectation that a system could be operational 99.99 percent of the month. It may also specify that the service provider's help desk will respond to an outage in a set amount of time. Also, the service provider is expected not to share a company's information with anyone and that data will be preserved for a set period of time and backed up regularly. In a complex hybrid cloud environment, managing SLAs of all the relationships of all cloud services a company may be dependent on can get complicated.

Think about the Service Level Agreement (SLA) for how and what services are provisioned. This can be complicated because your service provider may be giving you access to a virtualized image that you have little control over. The next time you work with a service, you may be provisioned a different image. This makes the issues of monitoring the service level complicated and difficult to track. There are issues that are more straightforward that you can monitor. Make sure you can access the following:

- A dashboard that provides you with insights across the applications and services that are running in your server rooms and those that are running in the cloud. This dashboard should include a way to monitor when applications are running and whether there are incidents or problems.

  Ask your service provider what kind of visibility you have from its systems. Then, determine the level of risk you're willing to take in terms of what you can't see.

- An SLA across your own services and those provided by cloud providers to get a true picture of the services you're providing to your internal and external customers

Service-level decisions are always about tuning the environment to the business purpose. It's not sufficient to measure and monitor the performance of servers, networks, or virtualized images as individual components of your environment. You need to understand how they all work together to meet business objectives. In addition, you need to establish a process for monitoring and managing service levels that includes an awareness of the relative business priority of each of the business services supported by the environment.

Developing a Secure, Accountable, and Reliable Cloud Environment

One of the first steps in creating a secure and reliable IT computing environment or cloud-computing environment is to assess your risks. At the end of the day, you're responsible for security of the services you deliver to your customers, whether they're delivered from your on-premises servers or utilizing a third-party cloud provider.

Assessing your current state

In a cloud environment, and especially a hybrid one, security, accountability, and reliability start with assessing your current state. Determine if holes or vulnerabilities exist across servers, networks, infrastructure components, and endpoints to properly assess and monitor your business's security posture. You need to be able to trust your own infrastructure as well as that from a potential cloud provider. Begin by answering a set of questions that helps you form both your approach to governance and your security strategy. To get you started, here are a few top questions:

- How do you control access rights to applications and networks — both those within your company and those that are outside your firewall? Who has the right to access IT resources? How do you ensure that only the right people gain access to your applications and information?
- Can you identify web application vulnerabilities and risks and then correct any weaknesses?
- Do you have a way of tracking your security risk over time so you can easily share updated information with everyone with a need to know?
- Are your server environments protected at all times from external security threats?
- Are you able to monitor and quantify security risks in real time?
- How do you adequately monitor, measure, and manage your IT assets across multiple environments?
- Do you have a way to manage assets across your public and private cloud environments?
- Can you implement security policies consistently across all types of on-premises and cloud architectures?
- How do you protect all your data no matter where it's stored?
- Can you satisfy auditing and reporting requirements for data in the cloud?
- Do you have a process for change and configuration that ensures that members of the organization have reliable access to the cloud service configuration information they need to perform their responsibilities?

Implementing security best practices

Knowing your current state helps you on your way to building a comprehensive strategy. Then, you ensure that best practices are followed.

To help reach your goals, check out these tips:

- In a highly distributed environment, manage the identity of who's allowed to access what resources under what circumstances. Clearly defined rules combined with automation provide a path forward.
- Try to create general awareness of security risks by educating and warning staff members about specific dangers. Complacency is easy, especially if you're using a cloud service provider. However, security threats come from employees as well as outside organizations.
- Regularly have external IT security consultants check your company's IT security policy, IT network, and the policies and practices of all your cloud service providers.
- Determine specific IT security policies for change management and patch management, and make sure that policies are well understood by your service management staff and by your cloud service provider.
- Review backup and disaster-recovery systems in light of IT security. Apart from anything else, IT security breaches can require complete application recovery.

Security risks, threats, and breaches can come in many forms and from many places, so companies need to establish a comprehensive approach to security management across IT and the business. As your IT environment moves beyond the internal data center to include public and private cloud services, security, governance, and reliability has to be handled in a cohesive manner based on a coordinated plan.

Chapter 6

Starting Your Cloud Journey

In This Chapter

- Making cloud computing your business strategy
- Getting started with PaaS and IaaS
- Accelerating your momentum
- Looking into business considerations

As you begin your cloud journey, you need to consider what types of cloud models are the best fit for your business and understand the current reality of the computing resources that you own and how effective they are in supporting your business strategy. Do you have the flexibility to scale your resources quickly enough to leverage new business opportunities? Do you have the ability to collect and analyze critical information in real time so you can reduce risk and improve asset management? Have you been able to improve your customer experience fast enough to build stronger and more personal customer relationships? Do you have the ability to support the needs of the business when IT budgets have decreased? If your answers are yes, your business strategy is probably already integrated with your cloud strategy. However, in most situations, you'll identify opportunities for improvements in business flexibility by looking into how your current computing environment supports your business goals.

In this chapter, you discover how you should start planning your cloud journey. You explore both the issues that you need to take into account as well as the implementation considerations.

Integrating Your Business, IT, and Cloud Strategies

Business leaders want to develop more innovative products and services while continuing to reduce operational expenditures. However, given the fast pace of today's complex economic environment, this task is often not easy. In order to remain competitive, business leaders need to approach each new opportunity with a well-conceived strategy that provides speed, flexibility, and scalability.

The typical business is faced with a vast number of challenges:

- Changing customer requirements
- Increasing industry regulations and standards
- The need for a mobile platform

Businesses also need the freedom to use their existing resources in new ways to address these challenges. At the same time they must limit the risk of investing in new capital while speeding up the delivery of IT resources.

When a new business strategy is developed, it's critical to map the business requirements to the IT resources and capabilities. Increasingly, IT components incorporate cloud capabilities. When business and IT leaders collaborate, the organization is in a better position to achieve the greatest flexibility and agility. The results to the business can be significant because this approach allows both business and IT to focus on the customer experience and business outcomes. Based on the specific requirements for performance, security, scalability, and service levels, companies start their cloud journey in different ways.

It's no surprise that companies are bringing business opportunities and cloud computing models together to forge a unified and cohesive business strategy. Business leaders see the potential in utilizing a set of cloud foundational services — Infrastructure as a Service (IaaS) and Platform as a Service (PaaS) — so the right cloud computing model can be selected

at the right time to bring new innovative services to customers and partners with greater speed and at less cost.

Getting Started with IaaS and PaaS

How do companies decide on the best fitting cloud model for their business strategies? To begin their cloud journeys, companies may select a few starting points.

Private IaaS for development and test

An insurance company wants to add a new service for a dedicated team of independent agents. A lot of personal customer information is stored in the application so the company needs to maintain a high level of security and control. They also expect to ramp up this new service very quickly and want to make sure they have the capacity both for ongoing testing and production. The private IaaS model provides the company with predictability and scalability while at the same time minimizing security risk.

Public IaaS for development and test

A leading retail bank needs to evaluate new technologies and solution platforms without investing in costly IT architecture. The testing environment for the new initiative was migrated to the secure public cloud based on a combination of the private IaaS service and the PaaS environment. The team needs to test its application for scalability with an expected one million concurrent users. The extra capacity from the secure public IaaS shortened total build times by enabling testing operations to leverage a highly-scalable, secure, and well-managed infrastructure. The business gets to market faster while keeping budgetary constraints under control.

Public PaaS for architecting new business models

An independent software vendor needs a flexible and secure PaaS environment to support the rapid development of a new innovative mobile shopping platform. To eliminate the need to manage the development and deployment environment internally, the company selected a third-party service provider. The company quickly provisioned a standard set of development services that met the needs of its developers on a pay-per-use basis.

Private PaaS for delivering new services

A company was prototyping a new application that would be a revenue generating service for its business partners. It piloted the application in a public cloud. After the business decided to launch the service, it moved the newly designed application to its private cloud. The private PaaS environment gives the company the right combination of consistency, scalability, and security to support ongoing development and deployment of its new application.

Accelerating the Company's Momentum

By selecting the best starting point, organizations can quickly gain experience and understand the benefits of cloud for their business strategy. Not only does this approach help jump-start a cloud implementation, but also it demonstrates what's possible. After the business understands the potential benefit, it needs to look at the operational issues in making the cloud an integral part of the IT environment.

Gaining IT acceptance

Many IT professionals are concerned about the growing popularity of cloud computing. What's the impact on their jobs? Will a self-service and highly-automated cloud computing environment mean that IT professionals will have little to contribute or manage? It isn't surprising that IT would be concerned about the transition from traditional computing to private and hybrid clouds. However, the reality is that the hybrid world is one vital component in an overall computing strategy.

Overall, whether we're talking about a public, private, or data center service, they're all components of the core IT environment. So organizations can't think about each service in isolation; instead they need to think of it as a set of unified services that have to be managed in concert.

Managing cloud services

Application developers and operations teams will benefit from the best practices approach that can be leveraged from the various cloud deployment models. However, to ensure a successful cloud journey, IT has to manage several key responsibilities:

- **Building shared services based on a service oriented approach:** To enable a hybrid environment to predictably support evolving business requirements, IT needs to be able to codify and then create well-designed business services with well-defined interfaces. These documented services allow the business to create new applications and services quickly and deploy them in many different situations in the new hybrid world.
- **Consistently managing the synchronization of data center systems of record with data stored in cloud environments:** Data is foundational to an effective hybrid cloud strategy. Data resides in all the key applications and systems that are supported in these environments. Therefore, IT needs to be able to provide effective management of common definitions and rules on an ongoing basis.

- **Managing the overall service level of the combination of all computing services inside and outside the firewall:** While IT management has long focused on meeting service levels within its own data center, the requirements are now expanding with the hybrid cloud. As this evolves, IT has to incorporate all these public, private, and data center services into a virtual environment that's managed as though it were a single system through an integrated service delivery approach.
- **Managing configurations, licenses, and the usage requirements for services:** Every environment that becomes part of the hybrid cloud environment includes system requirements that have to be managed in a different way. IT needs to have deeper control over how configurations relate to each other and how systems coordinate their services.
- **Ongoing support of security and governance:** Security and governance issues become more complicated in a hybrid world. IT has to create a security fabric and governance framework that supports both IT and business integrity.
- **Providing cloud integration services between clouds and traditional on-premises applications:** Each business unit tends to focus on the applications that support its business. IT has a unique opportunity to take a holistic perspective on integration across the extended enterprise.

Planning the Successful Journey

After you understand the business strategy and the technology requirements, you need a broad set of cloud options to support changing business needs. You also need an ecosystem of partners that can help you integrate your resources together. Make sure to have a roadmap and plan that takes you from an inflexible, slow-paced environment to a dynamic elastic environment that cloud services provide.

This sounds like common sense, but too often business and IT don't collaborate on planning for the future. Successful companies are able to view IT and business as a strategic partnership. Planning for the cloud contains two parts:

- First, a set of business considerations
- Second, a set of technical considerations

The most effective approach is actually to involve both IT and business teams in both assessments. This facilitates an understanding of issues and considerations. So, what are the key considerations that should be part of the planning and decision making processes? Five business considerations and five technical implementation considerations are covered in this section.

Business considerations

Business considerations are the strategic goals and plans that determine how the business changes over the next five years. Planning your cloud journey is more successful if it's planned in context with the issues driving the company's strategy.

How's the business changing?

A cloud strategy has to be targeted to how well your organization is structured to support changing business and customer requirements. Therefore, you need to be able to understand the anticipated opportunities and the threats from the competitive environment. Are new competitors entering your market? Is your industry changing dramatically and will that cause you to totally restructure how you serve customers and partners? If the answers indicate that major change is coming, that impacts the structure and process of creating and managing your hybrid cloud environment.

How does the company want to provide services in the future?

Delivery of services and customer value can transform a company dramatically. Your strategy may be to continue with traditional methods of servicing your customers. However, many industries are finding new channels and new models to support customers. These business models typically depend on sophisticated and emerging technologies. Understanding these requirements helps determine what technologies need to be incorporated into the cloud plan. For example, analytics may play a more far-reaching role than in the past.

What are the financial constraints for the company?

While it's important to understand new business opportunities, it's also important to understand the constraints that the business is experiencing. Understanding how the business needs to control expenses while increasing productivity and efficiency are essential in the planning for the cloud.

Is the company too siloed for the strategy?

Many business units have acted almost as separate companies — each with its distinctive set of processes, systems, data, and way of working with customers. However, this approach may be holding the company back from leveraging all assets across the company.

If a company is too siloed, the cloud computing strategy can be structured to help create more cohesion across processes, systems, and data.

Is there an easy mechanism to encourage experimentation and innovation?

It isn't always easy to change. However, businesses that are successful make sure that leaders are encouraged to think in creative ways about potential opportunities. Can technology provide a mechanism to support innovation? If this is a business priority, the cloud computing strategy can provide enabling technology to support experimentation.

Implementation considerations

After the business and IT leadership teams have a common understanding of the business drivers, creating a cloud computing strategy will be much more straightforward. Implementation considerations are based on planning for an environment that's long-term thinking and an environment that's not tied to a single project.

Evaluating reference architectures

A *reference architecture* is a best practice approach to creating a private cloud based on a composite of successful implementations. Therefore, reference architecture is a blueprint. While there isn't single reference architecture, most models have many of the same components. These documents can serve as an excellent planning tool.

Focusing on efficiency and flexibility

You don't want to repeat the mistakes of the past. Clouds — whether they're public, private, or hybrid — have to be designed to maximize the ability to standardize and automate. In this way, you have better control over costs and productivity. In addition, this consideration ensures that a standard approach based on best practices is followed consistently.

Planning for a fabric of services

Your environment will incorporate a lot of elements or services that are available across business units. This prepares you to be able to respond across business units as well as across partners. For example, you want consistent fabrics for managing security, data, integration, and business services. These services should be independent of any specific implementation. Creating this type of best practice establishes a foundation for everything that you create, buy, or connect to.

Assuming that you'll plan for a lightweight approach

Don't over engineer your approach to cloud computing. Make sure that any service or application that's a part of your hybrid environment includes well-defined interfaces (APIs) that are as standardized as possible. You want a streamlined approach that allows you to achieve your business goals and support innovation.

Monitored and managing everything you do

Every service that's created or any service that's used needs to be considered part of your overall cloud environment — even components such as public cloud services (IaaS and PaaS). You can't just worry about the services you own. Any service that touches an employee, a customer, or a partner needs to be monitored and managed with a service level. This approach helps the company better serve its customers, partners, and suppliers in a consistently predictable manner.

Transforming IT with Cloud

If your organization expects to be successful in an increasingly interconnected and highly instrumented world you need to fundamentally transform the economics and flexibility of your IT environment. A good place to begin this transformation is

by understanding the diverse requirements for your unique mix of workloads. It isn't a one-size-fits-all IT world anymore. Ensure that your IT environments have the resiliency needed to adapt to the high velocity of business change. In addition, make sure to leverage all opportunities to incorporate the choice and flexibility of a hybrid environment. An IT strategy that leverages all your IT resources, including the dynamic scalability of IaaS and PaaS, gives you the flexibility to take on the challenges ahead.