Brought to you by:

Informatica

Hybrid Cloud Data Management

for dummies®
A Wiley Brand

Move workloads to AWS® and Azure®

Extend on-premise data warehouses to AWS and Azure

Use iPaas for hybrid data management

Michael Wessler

2nd Informatica Special Edition

About Informatica

Informatica, the Enterprise Cloud Data Management leader, accelerates data-driven digital transformation by enabling companies to become more agile, realize new growth opportunities, and create new inventions. With 100 percent focus on everything data, the Informatica Intelligent Data Platform is the industry's most complete and modular solution, built on a microservices architecture. With more than 7,000 customers worldwide, Informatica allows organizations to unleash the power of data. For more information, call +1 650-385-5000 (1-800-653-3871 in the U.S.), or visit www.informatica.com.

Hybrid Cloud Data Management

2nd Informatica Special Edition

by Michael Wessler

for dummies®
A Wiley Brand

Hybrid Cloud Data Management For Dummies®, 2nd Informatica Special Edition

Published by
John Wiley & Sons, Inc.
111 River St.
Hoboken, NJ 07030-5774
www.wiley.com

ISBN: 978-1-119-45828-9 (pbk); ISBN: 978-1-119-45827-2 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

Publisher's Acknowledgments

Some of the people who helped bring this book to market include the following:

Project Editors: Carrie A. Burchfield, Martin V. Minner

Business Development Representative: Karen Hattan

Senior Acquisitions Editor: Amy Fandrei

Production Editor: Magesh Elangovan

Editorial Manager: Rev Mengle

Special Help from Informatica: Ann Rosen, Ajay Gandhi

Table of Contents

Introduction

Cloud computing is all the rage in the IT world and for very good reason; its impact is massive. Virtually unlimited computing capacity, provisioning new environments in minutes, exploring new technologies, and all at reduced costs! The exciting technical benefits are only exceeded by what the cloud does for business process agility and flexibility.

The immense benefits of the cloud come at a price: disruption of existing enterprise IT. The role of data centers is in flux as workloads shift to the cloud, reducing infrastructure costs. Computing is now a hybrid of traditional on-premise applications and data integrating with modern or next-generation cloud-based services.

Hybrid cloud computing (often called *hybrid cloud*) poses significant challenges for data management and integration between on-premise and cloud environments. Fortunately, this book examines techniques, architectures, and toolsets to turn the challenge of hybrid cloud data management into an asset by using integration Platform as a Service (iPaaS) cloud services. By leveraging iPaaS, cloud architects perform the most common use cases of

- Migrating workloads to cloud (also known as *lift and shift*)
- Performing hybrid app integration
- Supporting a hybrid data warehouse

All three of these use cases use Amazon Web Services and Microsoft Azure.

About This Book

The focus of this book is learning what hybrid cloud data management is, why it's important, and how to design these environments. A great deal of attention is given to explaining data management and integration in simple terms and describing how iPaaS enables your cloud architecture.

Cloud computing is a disruptive paradigm shift with substantial technical impact and business benefit. Many companies aren't

prepared for the resulting hybrid cloud environments where applications and data are distributed both on-premise and within the cloud. Data management and integration are major challenges for companies as they move to hybrid cloud environments. Data management includes data and application integration, as well as data mastering, cleaning, and validation.

Icons Used in This Book

Throughout this book, you occasionally see special icons to bring your attention to a point that needs emphasis:


TIP

Tips indicate information that you may find useful. Often, they relate to an experience I had (or I wish I had at the time), or they add context to a topic.


REMEMBER

If you see this icon, it's probably something that will help you later. You won't find the meaning of life here, but you may find some advice that makes your life easier.


WARNING

Warning means just that; be careful! I use warnings to alert you to common mistakes and serious issues for you to avoid.


TECHNICAL STUFF

I am a technical person at heart and love to understand how and why things work (or don't). Yes, this is a *For Dummies* book, but sometimes I delve deeper into a subject so you understand the "why" and "how" for a key topic.

Beyond the Book

This book can't teach you everything about cloud computing and its benefits, but I do cover the fundamentals. Beyond those basics, I encourage you to research cloud architectures and current cloud service providers on your own. Cloud computing represents the future of the industry; everyone needs to understand it! Find more resources on data management for hybrid cloud (and Informatica's approach to it) at `www.informatica.com/hybridcloud`.

IN THIS CHAPTER

- Understanding the evolution of data centers
- Explaining cloud computing
- Describing cloud service and deployment models

Chapter 1
Identifying Cloud Computing

Computing requirements and the ways companies meet those requirements are continually evolving. One of the most exciting and substantial changes recently is the advent of cloud computing. In many ways, cloud computing changes how IT supports the business and provides increased capabilities at lower costs to a degree that was previously unobtainable. At the same time, cloud computing can be a disruptive paradigm shift for IT. To further complicate matters, many misconceptions exist about what cloud computing is, why it's important, and how cloud computing is operationalized effectively in companies.

To help you fully understand the cloud and its impact, I must dispel common misconceptions and clearly define cloud computing and its major components. In this chapter, you discover how cloud computing has evolved, its core tenets, and fundamental service and deployment models.

Understanding the Evolution of Data Centers

Throughout much of the history of enterprise computing, a company's data center was the focal point of computing power — data, servers, networking, and so on. Ranging from small closets to vast purpose-built secured facilities, data centers were the pride of many IT departments and played a critical role in the success of their companies.

Powerful data centers spanning the globe to support users in every corner of the planet 24/7/365 became the status quo for many companies. Elaborate offsite, disaster recovery data centers sprang up to ensure continuous operations for users.

Connectivity and performance between locations continually improved to the point where in many situations it didn't really matter where a user resides versus where the data or application resides. In many cases, key components of applications are located in different locations and are provided by different sources outside the corporate infrastructure. Distributed computing allowed systems to span multiple data centers and involved partner systems, suppliers, and third parties, and all of this was transparent to the end-user.

As data centers grew in capability, so did their costs, complexity, and connectivity requirements. Technologists reveled in their capability, but CIOs cringed at their costs, while operations folks dreaded their complexity. *Server sprawl* — the unmanaged expansion of servers — increased the number of components to manage further increasing the costs and complexity of data centers. Frequently, the size and complexity of IT environments exceeded the capacity of technical staffs to manage those environments. Overworked IT staffs grew in size (and cost) and were often lucky to "just keep the lights on" to meet existing requirements; they had little ability to meet new business requirements as fast as business opportunities emerged.

Sensing an opportunity, IT vendors attempted to alleviate these challenges through evolutionary enhancements to reduce IT complexity, improve management tools, and reduce server sprawl via virtualization. These efforts helped to slow expanding costs and complexity, but the concept of dedicated data centers for each

company has inherent inefficacies that require a more fundamental (and disruptive) change to resolve. That needed change came in the form of *cloud computing.*


TIP

Many large companies and governmental organizations establish data center consolidation initiatives to reduce the number of data centers they must operate. These efforts can reduce costs and complexity while improving security and standardization. However, these efforts are often themselves complex in nature and can take years to plan and implement.

Introducing Cloud Computing

Cloud computing is a distributed computing architecture that addresses many shortcomings of a traditional data center centric architecture. Cloud computing provides computing resources to IT organizations as externally provided computing *services*. Rather than shouldering the burden with providing every IT resource themselves, IT organizations simply buy whatever resource they need (as a service) from an external cloud service provider. In a nutshell, you supplement your internal IT services with external IT resources from the cloud. However, cloud computing encompasses much more than simply outsourcing services as you explore throughout this book. Unfortunately, like many other hyped technologies, cloud computing is misunderstood by the general public at business and executive levels and even by technical staffs.

The National Institute of Standards and Technology (NIST) provides an unbiased baseline understanding of cloud computing: Cloud computing is a model for enabling universal, convenient, on-demand network access to a shared pool of computing resources (such as networks, servers, storage, application, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

Why the Cloud?

A quick glance at any technology publication or conference would lead one to believe that cloud computing is the only computing architecture available today. That position is somewhat over-exaggerated, but the cloud is a major focus of the IT

industry today. Cloud computing is significant and its importance is rapidly growing.

So, why cloud computing? The answer is it appeals to the needs of many across different positions, ranging from the CIO, to business users, and the IT technical staff. Key drivers include the following:

- Lower total cost of ownership for hardware, software, and IT infrastructure
- Pay based on what you actually use; don't pay for unused capacity
- Faster provisioning of IT environments; new IT capability and capacity delivered in minutes rather than months
- Access to external technologies and capacity that would be impossible to host internally
- Increased focus on meeting business challenges rather than expending resources on commoditized technical problems

Cloud computing offers benefits to many different audiences, but to understand how these benefits are possible, it's necessary to understand more about cloud computing itself.

Defining Cloud Characteristics

To fully understand the cloud and make the most of your cloud investment, you must understand some core concepts. This understanding helps you separate important cloud attributes from fluffy IT vendor marketing. NIST defines the five characteristics of cloud computing as follows:

- **On-demand self-service:** A consumer unilaterally provisions computing resources as needed automatically without human interaction.
- **Resource pooling:** Computer resources are pooled to transparently serve multiple consumers.
- **Rapid elasticity:** Capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out and rapidly released to quickly scale in.

- **Measured service:** Cloud systems automatically control and optimize resource use via a metering capability. Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the service.
- **Broad network access:** Capabilities are accessed over the network and accessed through standard mechanisms that promote heterogeneous thin or thick client platforms.

The power and benefits of the cloud comes from these characteristics. Evaluate any "cloud" solution you see against these characteristics; if that solution is lacking in these areas, you should consider another solution.

REMEMBER

I use NIST definitions frequently to establish baseline terms and concepts because they're unbiased for specific vendors and are stable over time.

Understanding Foundational Cloud Service Models

Cloud computing provides different types of IT *services*; those services can be applications software, prebuilt operating environments, or even hardware and infrastructure. Depending on the type of service provided, cloud offerings are categorized into the following NIST-defined service models:

- **Software as a Service (SaaS):** *Consumers* are provided their software applications from the cloud service *provider.* The consumer gains access to application software while the provider manages the underlying software and infrastructure used to provide that application software. Applications are frequently delivered to the customer via web browsers or program interfaces in SaaS architecture.
- **Platform as a Service (PaaS):** Consumers use programming languages, libraries, and tools from the provider as an application development and deployment platform. The platform may include databases and application middleware in addition to application development tools. Providing virtualized operating environments to consumers is a key component of PaaS architecture.

- **Infrastructure as a Service (IaaS):** Cloud service provider manages the underlying physical cloud infrastructure (servers, networking, storage, operating systems) while the consumer deploys and runs their own application software and provisions additional resources (memory, CPU, storage) as necessary. Self-service virtualization is a key component of IaaS architecture.

As cloud computing has evolved, additional cloud service models have evolved to more precisely support business needs. In particular, building on the foundations of SaaS and PaaS, several more specific models have emerged:

- **Integration PaaS (iPaaS):** Integrates data, applications, and processes in a distributed environment where some services are provided by clouds and other components exist in on-premise data centers. iPaaS is critical to making systems distributed across hybrid environments work seamlessly and efficiently.
- **Database as a Service (DBaaS):** Database storage and management software are provided as a cloud service. The consumer stores their data in the DBaaS cloud and accesses that data through Structured Query Language (SQL) and programmatic interfaces.
- **Data Warehouse as a Service (DWaaS):** Similar to DBaaS, data warehouse storage and management software are provided as a cloud service.
- **Big data and mobility specific cloud offerings:** Access to big data, powerful processing tools, and software to support mobile applications is provided by cloud providers. Many companies want the capability provided by big data and mobile applications, but they're reluctant to host the necessary infrastructure internally.

For hybrid cloud data management, the iPaaS service model is particularly well suited. In Figure 1-1, you see an iPaaS example.

In Figure 1-1, iPaaS, along with on-premise integration software deployed in public cloud, provide integration of SaaS data sources, DBaaS and DWaaS deployed in a public cloud with on-premise databases, data warehouses, and applications. At its core level, iPaaS integrates the data services in a hybrid environment with applications that require distributed data.

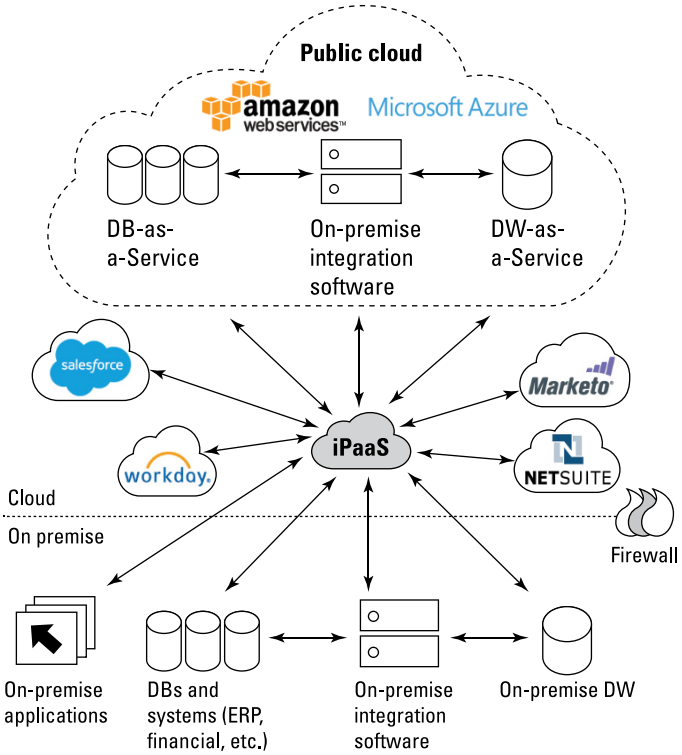

FIGURE 1-1: Deploying iPaaS in a hybrid cloud architecture.

LEVERAGING INTEGRATION COMPETENCY CENTERS

Many large Fortune 2000 companies extend their on-premise data management architecture, often referred to as an *integration competency center* (ICC), to the iPaaS cloud, in order to deliver the following:

- Self-service IT
- Governed self-service for business users
- Global data management
- Geographic/department deployment

(continued)

(continued)

- High velocity
- Best practices and standards
- IT as a service
- Unlimited elasticity for bursting scenarios
- Data stewards empowerment
- Controlled costs
- SaaS end-points integration

Exploring Cloud Deployment Models

If you've been reading this chapter straight through from the beginning, you should have a better understanding of the different services provided via the cloud. Now what? You discover how those services are *deployed.*

Cloud computing is economical because the costs and underlying infrastructure are shared among its consumers. However, it's the details of that sharing that many consumers want to understand for a myriad of security, policy, and technical reasons. Because cloud consumers have differing requirements, several different methods to organize this shared cloud architecture have been developed.

Cloud deployment models determine who has access to the services and where and how the infrastructures for those services are maintained. NIST has defined the following deployment models:

- **Private clouds:** Used exclusively for a single, private organization or company, the cloud infrastructure is frequently hosted in a private data center. Private clouds provide the highest level of control by the consumer and are perceived as providing the highest degree of privacy and security because the consumer owns and manages the infrastructure.
- **Public clouds:** Used by multiple, unrelated organizations or companies on a shared basis. The cloud infrastructure is hosted by an outside entity providing the cloud as a service. Public clouds are the least expensive and incur the least

amount of overhead on the consumer, but the consumer also has the least control of the cloud services and supporting infrastructure.

- **Community clouds:** Used by organizations or companies with a common purpose, mission, or audience. These cloud services aren't available to the general public but are available to members of the defined community. Common examples include governmental or military organizations or specific industries to include partners, suppliers, and authorized customers.
- **Hybrid clouds:** Used when a consumer combines services provided by multiple, distinct cloud providers or by a cloud provider in conjunction with their own internally hosted services. For example, a company has its own private data center for its traditional IT computing, but takes advantage of a new capability from a public SaaS provider rather than attempting to provide that capability itself internally.

REMEMBER

Cloud deployment models are important because many cloud consumers want to know who they're sharing their cloud with and where their applications and data physically reside. Their requirements drive them to a specific deployment model initially, and often you see those consumers evolve into additional deployment models as their requirements change and their comfort level with cloud computing grows.

IN THIS CHAPTER

- Identifying technical and business benefits of the cloud
- Understanding how the cloud is transforming business and IT
- Describing the journey to the cloud experienced by companies

Chapter 2
Understanding Cloud's Impact on Business

Businesses and enterprise IT departments are experiencing the impact of cloud computing. That impact, while ultimately beneficial, is often initially disruptive, especially to IT departments. Lowering infrastructure costs and increasing business agility, coupled with unlimited scalability, are powerful motivators. However, the shift from enterprise data centers to the cloud leaves many IT departments redefining their role. Companies embarking on the cloud journey face many choices, including public versus private hosting and selecting the right SaaS, PaaS, iPaaS, and IaaS offering. In this chapter, I detail the benefits of cloud computing, and you begin the journey to the cloud.

Exploring the Benefits of Cloud

Business and IT departments experience many benefits from the cloud. Some benefits are obvious, but others aren't as intuitive and become apparent over time. Beyond the mantra "doing more with less," some benefits are merely technical while others are nontechnical, business, and cost related.

Foundational cost shift

At first, companies frequently only look at the low initial cost of cloud services, but in reality, cloud cost savings come in multiple forms. In their totality, these changes have a significant impact to change how IT spending is planned; in fact, it shakes the foundation of IT spending. Key cost reductions occur in these areas:

- Cloud services are usually (but not always) *metered,* which translates to "pay-as-you-go," so you only pay for services and capacity you actually use. Rather than buying a dedicated database server that may sit nearly idle during nonpeak processing periods or a software package that may go unused by users, with cloud you're billed based on your system resource utilization and the number of software licenses actually used. This reduces waste and the risk of "overbuying" capacity and software.
- Reduced capital expenses as new investments in enterprise IT infrastructure. Rather than purchasing new data center infrastructure (servers, storage, networking, software) and the corresponding mirrored disaster recovery infrastructure, that capacity is purchased as a metered service from the cloud provider. This reduces upfront capital costs for companies seeking new capability. Companies no longer have to calculate long-term Return On Investment (ROI) for large storage or database server purchases. As a further benefit, because the cloud infrastructure is owned and managed by the cloud provider, it's the provider's responsibility to support expensive technology upgrade and refresh efforts.
- Reduced operational expenses in terms of staffing, licensing, and support agreements for technology infrastructure. Because the cloud service provider maintains the cloud infrastructure, all the associated operating costs are the responsibility of the provider. Because the capital and operating expenses are the responsibility of the cloud provider, those costs are shared across the multiple cloud consumers at a lower rate (rather than each consumer bearing the full cost themselves).
- Lower Research and Development (R&D) and Subject Matter Expert (SME) costs for new, advanced capabilities. Historically, if a company wanted to invest in a new technology, it had to procure new hardware, software, and expertise

either via training existing staff, hiring new staff with specialized skills, or obtaining expensive outside consultants and contractors. With specialized cloud offerings such as big data and analytics, the expert knowledge and experience are provided by the cloud service provider and not required internally by each company.

Cloud computing reduces costs across a broad spectrum beyond simple cost reduction in a single area. This represents a foundational change in spending for IT that CIOs and IT staffs greatly appreciate.

Enhanced agility

Self-service provisioning of cloud services delivers new PaaS environments, increased IaaS capacity, and more software capabilities via SaaS rapidly. Often, new capacity is available within minutes or hours after the request is made; a massive improvement over the days, weeks, or months often required by enterprise IT. Self-service provisioning is a foundational component of cloud computing enabling rapid delivery of increased capability. Increased agility occurs in several areas:

- Speedy provisioning of new cloud services means consumers aren't sitting idle, waiting for delivery of the new capability. Delays in projects due to enterprise IT's inability to deliver in a timely manner are well known. Rapid provisioning helps resolve this long-standing obstacle.
- New business capabilities are rapidly created through a modular mix and match of cloud services to create the integrated system of what's often a diverse ecosystem of components. Consider just some of the cloud service offerings available: DBaaS, DWaaS, SaaS, and iPaaS; the best of breed for each service is integrated with iPaaS, along with on-premise computing resources to deliver new business capabilities faster than ever possible before.

Enhanced agility's greatest benefit is the ability to capitalize on new business opportunities faster than competitors through a modular building block approach of cloud services. Self-service with rapid provisioning is a foundational component, but the mix-and-match modular approach to building new business capability is the strongest factor of enhanced agility.

Elastic scalability

Cloud computing significantly frees IT staffs from the complex task of capacity planning and to a lesser degree performance tuning due to capacity issues. A foundational building block of cloud computing is *rapid elasticity*, which is the ability to scale up or scale down resources to meet current demand. This elastic ability to scale up or down provides multiple benefits:

- System resources (often CPUs, memory, number of application servers, storage, and so on) are allocated to meet the current system requirements; they're neither undersized nor oversized. That means the exact amount of necessary capacity is allocated and that metered usage of capacity is how the consumer is billed. The consumer doesn't pay for unused, excess capacity nor does system performance suffer from an undersized resource configuration.
- As system resource requirements shrink or expand, the corresponding resources (and also billing) elastically shrinks or expands as appropriate. This allows for intelligent allocation of resources across the shared cloud environment resulting in total lower costs for everyone.
- Enterprise IT staffs don't need to allocate highly trained technical people to perform detailed capacity planning. Accurate capacity planning is a complex Skill. By offloading this responsibility to the cloud provider, these talented individuals can focus on other challenges within IT.

Elastic scalability enables the perception of infinite capacity for cloud computing; if you need more resources they're automatically added as needed and are later deallocated when no longer required. Automatic scalability allows cloud service providers to support more customers with less infrastructure while allowing cloud consumers to use the resources they need at a total lower cost.

WARNING

Don't assume that automatic scaling is in place — and certainly not automatic scaling into infinity to support bad code or runaway processes. Check if scaling upward and shrinking downward is automatic; also identify what controls exist for automatic growth as it should not occur without oversight.

Renewed emphasis on innovation

Managing large data centers and enterprise IT environments is complex, time-consuming work performed by teams of highly trained and dedicated individuals. It's a difficult job that requires a large expenditure of resources. Cloud computing has a positive impact on this reality in multiple ways:

- Transferring (redirecting) workload from enterprise data centers into the cloud reduces the amount of work for already overworked IT staffs to perform. This allows the enterprise IT staff to address tasks that have been delayed and focus on "bigger picture" IT projects rather than daily fire-fighting tasks that don't expand the business.
- Freeing up the best and brightest people in enterprise IT from the daily grind of time-consuming, mundane tasks allows them to refocus on the core competencies of their company. They now have the capacity to innovate and provide greater business value to the organization beyond fire-fighting and trying to keep the lights on.

Enterprise IT departments have many smart people who understand the needs of the business. Leveraging cloud computing to reduce or eliminate IT staff's time spent on mundane operational tasks so they can focus on business innovation is a win-win.

REMEMBER

Don't be hasty and believe that because you moved to cloud computing you can downsize your IT staff because you won't need them anymore; that's a shortsighted decision. The work of your IT staff will shift from managing existing infrastructure to governance, compliance, and innovation for your company that enables you to seize new business opportunities while reducing risks.

How Cloud is Changing Enterprise IT

Given the benefits and paradigm shifting capabilities of cloud computing, coupled with the size, complexity, cost, and inability to respond quickly to new requirements, a major disruption in enterprise IT and traditional data centers is underway. Cloud computing is redirecting workload from enterprise IT data centers and into cloud computing environments; that's a major change in

how IT and ultimately businesses operate. How much is this happening? Consider these estimates from industry analysts:

- Cloud computing was the bulk of new IT spending in 2016.
- $204 billion was spent on public cloud computing in 2016.
- Cloud spending will reach $315 billion by 2019.

In Figure 2-1, you see this workload shift into the cloud.

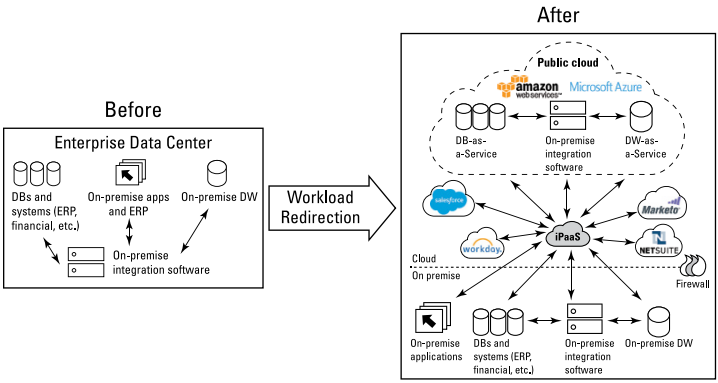

FIGURE 2-1: Redirecting workload into the cloud.

THE ROLE OF ENTERPRISE IT

Just because enterprise IT staffs won't be managing as much traditional infrastructure, it doesn't mean they won't be busy. Rather than focusing on operational technology support, enterprise IT will shift their focus to equally important areas of IT governance, compliance, policy, and licensing. Priorities within IT have evolved where governance and compliance with controls and standards are paramount and grow in importance within hybrid clouds. Well-defined, metadata-driven architectures enable your ability to understand and govern your data, which is a major initiative for many organizations.

Figure 2-1 shows the on-premise workload of the traditional enterprise data center redirecting to cloud service providers. Note that workload can redirect to multiple cloud providers resulting in a multi-cloud environment. Mixing and matching the right cloud service provider in a modular manner is a strength of cloud computing.

Driving to the Public Cloud

Public or private clouds — which to move to? The needs of the company or organization will drive that decision, but unless working with highly secure or governmental organizations, public clouds are the preferred direction of many companies.

Public clouds are still secure; many cloud architects argue that they're more secure because they face greater security compliance and governance than a smaller private cloud can support. In fact, many governmental and defense organizations are allowing vendors to provide community clouds on a large scale.

The list of cloud providers is growing but two powerhouse public cloud providers you can't miss are Amazon Web Services (AWS) and Microsoft Azure. I cover them in greater detail in Chapter 4.

Taking the Journey to the Cloud

Your journey to the cloud is a purposeful process that takes planning, deliberate execution, and time to ensure you reach your destination. Just as with any trip, you need to consider multiple factors. Factors influencing cloud adoption include

- **Newness and maturity of the company and its application:** New companies have fewer applications and less infrastructure to migrate to the cloud.
- **Level of investment on current on-premise applications and infrastructure:** Companies with large on-premise footprints are less likely to migrate to the cloud rapidly.
- **Criticality, complexity, and size of application:** Highly custom, very large, or absolutely mission-critical applications are more challenging to move to the cloud.
- **Perceived risk tolerance of stakeholders:** Applications supporting classified or sensitive data, or those supporting human life, are justifiably subject to greater scrutiny and risk aversion when considering a cloud migration.
- **Existence of mandates to migrate to the cloud:** Policy driven cloud-first mandates or conversely no-cloud directives weigh heavily on cloud planning sessions.

The cloud service model selected is dependent on several factors, including the origin of the request and the expertise and maturity level of the cloud consumer and provider:

- Line Of Business (LOB) organizations frequently seek SaaS to have access to software such as Salesforce and Workday applications to support their users in the field.
- IT initiatives often push for PaaS and IaaS to ease data center pressures and improve their agility to provide capacity.
- Departmental IT groups have more specific requirements such as a database to be deployed to DBaaS or a data warehouse via DWaaS.
- As organizations become more experienced with the cloud, they become more confident and therefore more willing to leverage cloud technology.
- As cloud service providers become more established, they are able to reduce prices and offer increased capacity and more specialized offerings.

The cloud has multiple entry points, each of them reflective of the perspectives for those driving the requests. As companies become more cloud savvy, they expand into different cloud deployment and service models. This evolution is good, but it brings its own unique challenges (I cover those in Chapters 3 and 4).

BORN IN THE CLOUD

Can a company be cloud-only with no internal infrastructure beyond its laptops? Older, more mature companies have a greater challenge migrating existing infrastructures to the cloud. What about younger companies without the burden of infrastructure? Those are more likely candidates for cloud-only footprints because they're "born in the cloud" with a desire to keep initial costs low.

IN THIS CHAPTER

- Understanding the concept of hybrid cloud computing
- Identifying why hybrid cloud computing is present at most companies
- Learning about the challenges of managing data in the hybrid cloud
- Introducing the people playing a key role in the hybrid cloud

Chapter 3
Introducing Hybrid Cloud Data Management

If you have any cloud presence at all, odds are you are in a hybrid cloud environment even if you don't realize it yet. Managing applications, data, and processes within an integrated cloud and on-premise architecture is the basis of hybrid cloud management. Hybrid clouds are the reality of most organizations using cloud technology, and they present specific challenges that must be addressed. In this chapter, I explore hybrid cloud computing, the challenges of hybrid architectures, and introduce the cloud architects and other key stakeholders who must navigate this complex environment.

Defining Hybrid Cloud

Hybrid cloud computing is a relatively new architecture as companies begin adopting cloud services. Although new, the concept is relatively simple; hybrid cloud computing is the mix of new cloud computing technologies with existing on-premise resources in traditional data centers. Processes and data exist in

the on-premise data center and those processes and data integrate with processes and data provided by the cloud to support an application. Workload in the form of data and processes is distributed between the on-premise data center and the cloud to create a hybrid cloud environment.

Pretty simple, right? Hybrid cloud environments are the combination of cloud computing with on-premise computing. Hybrid clouds are based on several foundational assumptions:

- Existing workload in the form of data and applications are present in on-premise data centers. This workload has not yet or will not ever move into the cloud.
- Cloud architectures (SaaS, PaaS, or IaaS) provide some but not all of the applications, data, or infrastructure required by the consumer company. Most companies will have a multi-cloud environment with a mix of SaaS, IaaS, and PaaS.
- Integration of data, applications, and infrastructure between cloud and on-premise assets must occur to complete a task or otherwise support an application or user group.

Basically, neither the cloud nor the on-premise computing environment has all the application components, data, or infrastructure necessary; therefore, the combined cloud and on-premise resources must meld into a hybrid cloud environment.

Obviously in a hybrid cloud environment, integration between on-premise and cloud resources is a key requirement for success. At an even higher level than integration, the proper management of all data, applications, processing resources, and the related governance and security policies is paramount. These are the core topics at the center of hybrid cloud management that I discuss in detail.

Technically, if you have multiple distinct cloud infrastructures such as private, community, or public, you also have a hybrid cloud. However, I focus on the hybrid definition of on-premise computing with any cloud infrastructure because this is a highly prevalent architecture in today's companies.

Following the Journey to Hybrid Cloud

Few organizations start their cloud discussion with the statement, "I want a hybrid cloud," but the reality is most organizations evolve into a hybrid cloud environment.

REMEMBER

If you have some workload in the cloud and some on-premise, by definition you have a hybrid cloud. Operating in a hybrid environment isn't inherently bad; it's simply part of the natural progression from entirely on-premise computing to using the vast capabilities the cloud has to offer.

The upper boundaries of on-premise computing have been pushed for many years; moving to the cloud was inevitable to increase processing capacity. As more data and processing are moved from on-premise to the cloud, you see a shift from data center centric computing with some cloud to ultimately cloud-centric computing with some on-premise capability retained. This is a journey, and it doesn't occur overnight. For the foreseeable future, on-premise computing will continue to integrate with cloud services, which means hybrid clouds will flourish in many organizations.

Given that hybrid cloud computing is a fact of life, what are some of the benefits of this architecture?

- **Scalability:** Cloud environments offer near limitless capacity and processing power beyond what a traditional data center could ever provide. You should fully utilize the existing capacity of on-premise environments, but rather than buying new capacity for your data center, you leverage the cloud to meet new capacity scalability requirements. Elastic scalability of cloud resources ensures the right amount of computing power is present when needed and not wasted when no longer required.
- **Flexibility:** Creating new environments and exploring new technologies and capabilities is much faster and less expensive within the cloud. Existing on-premise data centers serve traditional IT needs, but when new requirements emerge, use the cloud rather than building internally. Cloud computing enables you to try out new technologies faster and cheaper than building internally, which greatly increases your organizational flexibility.

- **Security:** On-premise computing is subject to the security and governance controls of enterprise IT and cloud environments are subject to the security controls of the cloud vendor. Leverage your existing enterprise IT staff to direct their attention to managing the security and governance controls of the complete hybrid environment. Your enterprise IT staff knows your security and governance polices better than anyone else, and as employees, they're vested in the company's success. Expand the scope of enterprise IT security and governance responsibility for the complete hybrid environment, not just their on-premise assets.

Most companies start out with a heavy on-premise architecture and smaller cloud component. As cloud capabilities and confidence in the cloud grows, the focus shifts to a greater utilization of cloud computing for new requirements. Less new work goes into the existing on-premise environments and more emphasis is placed on cloud environments. Eventually, the balance of processing shifts from on-premise data center centric computing to cloud centric computing until only what is necessary remains in the data center and most computing occurs in the cloud.

What is the end-game architecture as processing and data shifts from on-premise to cloud computing? Will we experience a day where *all* computing is migrated from on-premise into the cloud and our enterprise data-centers completely close? The idea of a cloud-only architecture is a big question for many, and there is no clear-cut answer for all situations.

At a minimum, there will seemingly always be a footprint in most companies for some hardware such as printers, workstations, and local connectivity. Legacy and highly customized applications that aren't good cloud candidates will also remain on-premise. Additionally, data and applications that have security or regulatory requirements to not be distributed to the cloud will remain on-premise. However, those applications and data that are good fits for cloud computing will very likely migrate to the cloud eventually. Data centers will greatly shrink, and enterprise IT staff will focus more on governance, integration, and policy, and hybrid cloud architectures will grow in importance.

DATABASE-CENTRIC CUSTOMERS SHOULD CARE

Some database-centric customers take a skeptical view of the cloud and don't think they will be noticeably impacted. Perhaps they have had long careers in IT and seen fads come and go. Or they assume their databases are too big or complex to move and will forever remain on-premise.

In reality, even the most salty database-centric customer needs a degree of cloud expertise. Cloud computing isn't an abstract concept or fad; it's here today for many organizations, and many more will enter it in some form so everyone needs to understand cloud. For those managing large databases and data warehouses, even if they remain on-premise, odds are they will be accessed by cloud services in a hybrid configuration. The message is even if you aren't going to the cloud; *the cloud is coming to you*.

Identifying Challenges of Hybrid Cloud Data Management

Hybrid cloud computing creates distinct data management challenges in orchestrating how data is stored, how it's moved, how it's integrated and transformed, and how it's consumed by applications across the hybrid landscape. Hybrid data management must address governance, risk management, and compliance as major themes in addition to technical aspects. Both hybrid data management and hybrid data integration must occur as parallel services in these environments.

Cloud computing is touted by some as "easy," but in reality it creates a much more complex, global environment when you consider the on-premise versus cloud ramifications of how data is stored, moved, consumed, and processed.

WARNING

The most common challenges of data management in hybrid clouds include the following:

- **Connectivity:** Connecting to all the necessary data
- **Scalability:** Growing upward or shrinking as needed

- **Data visibility:** Seeing the data and lineage of your data
- **Operational control:** Effectively managing your data
- **Data governance:** Enforcing policies and controls
- **Data quality:** Ensuring the quality of your data

Connectivity, scalablity, data visiblity, operational control, data governance, and data quality are major challenges when working within the hybrid cloud. These challenges repeatedly trouble companies, so ensure your data management strategy accounts for them as you move into the hybrid cloud.

REMEMBER

A great focus is on data governance, but don't forget data quality. A metadata-based architecture is the foundation of your data quality, visibility, lineage, governance, and ultimately security.

At the technical level, a hybrid environment introduces complexities regarding network latency, infrastructure, data access, and integration. As a principle, you want to do the processing where the data exists rather than moving data to processing. Because much of the new data exists in the cloud, you need to perform much of your processing in the cloud, which is a paradigm shift for many IT organizations.

With cloud computing, the infrastructure, data, and applications are now deployed around the globe. This distribution allows for constant availability, nearly unlimited capacity, and more nimble deployments than ever possible before, but it comes at the cost of increased complexity for data management. Enterprise IT must ensure their data governance, compliance, and security practices are up to date and ready to meet these new challenges.

For example, consider the legal and compliance implications of how data that was once only located on-premise is now located globally via the cloud:

- Nations have different and frequently changing laws regarding data privacy and ownership.
- "Right to be forgotten" laws in Europe may have implications for U.S. companies.

- Data security requirements and punishment for compromised data vary and have international ramifications.
- Government and law enforcement access to data and encryption laws vary widely and subject data housed in different nations to inconsistent policies regardless of where the company headquarters is based.

These are difficult challenges, but they're manageable when met with a robust approach to governance, security, and compliance by enterprise IT.

TIP

Companies aren't the only ones subject to international laws — so are government agencies moving to cloud architectures. The already complex world of data ownership and privacy becomes even more challenging when dealing with classified government data. As a way to reduce risk and complexity, some government agencies require cloud providers to ensure data only resides in cloud data centers within that nation.

WARNING

Beyond legal and compliance issues, consider the technical challenges of maintaining data lineage. *Data lineage* is the life cycle of data to include its origin and where it moves and how it's transformed along the way; this lineage information has impacts for data quality and data governance, as well as data security. Data lineage is also important for understanding the downstream and upstream impact of any change to data and data flows. Tracking this is important for risk management and development productivity. With cloud computing, maintaining data lineages inherently becomes more challenging:

- Visibility into the data and metadata origin may be reduced without the right tools.
- Cloud infrastructure moves data very fast across multiple boundaries making tracking lineage more complex.

The impact of cloud computing (and therefore hybrid cloud computing) on data lineage translates into greater challenges for data capture, curation, analysis, and usage. This complexity creates challenges at the technical level for data usage and at the non-technical level for data management and overall governance.

DATA MANAGEMENT VERSUS DATA INTEGRATION

Data integration is a subset of the larger world of data management. Data integration is the focused topic of transforming and delivering data from point A to point B in the most efficient manner possible. *Data management* is a broader topic addressing issues of data discovery, governance, cleansing, validation, mastery, and security in addition to integration. This distinction is important when evaluating technologies and the shifting roles of enterprise IT when managing data in hybrid cloud environments.

Focusing on Hybrid Cloud Data Management Principles

Hybrid cloud data management is the solution patterns companies adopt to solve their data management challenges when implementing a hybrid cloud architecture. Several drivers consistent with hybrid cloud data management include

- Minimizing business disruption during migration to the cloud
- Maximizing adoption and integration across multiple applications
- Leveraging a single-pane-of-glass management toolset to manage applications and data regardless of their origin

Single pane of glass refers to having one unified view or toolset of a complex environment. Relying on a myriad of tools without a defined end-to-end management plan creates inefficient redundancies while opening the possibility of visibility gaps. Single-pane-of-glass toolsets simplify management, while reducing redundancies and gaps.

Performing Hybrid Administration and Recovery

Ideally customers shouldn't care where application processing occurs or where data resides; it should be transparent and seamless across the globe. However, experience shows maintaining high availability and disaster recovery is challenging. Performing these same tasks in a hybrid environment introduces additional complexities.

REMEMBER

Quality hybrid cloud data management solutions should make the location of processing and data both transparent and seamless for customers. Your toolsets should embrace this ideal and span the hybrid cloud.

Managing systems in hybrid environments means some servers are on-premise, while others are in the cloud, which you do not directly control. Mismatched Service Level Agreements (SLAs), globally distributed components, and widely disbursed support staffs are obvious factors. However, hybrid cloud environments are more easily managed using these concepts:

AVOIDING THE INTEGRATION HAIRBALL

Point-to-point integration can create a tangled hairball of data integration lines in any environment. Simply integrating components on an as-needed basis with no long-term planning is a recipe for a hairball. The hairball concept predates cloud computing, but with new fast, self-service cloud capabilities you can create your hairballs even faster.

To avoid creating new hairballs at cloud speeds, you need a disciplined approach to data integration. Self-service capabilities are good, but you must be consistent in how you format and define your data. Good data management practices promote good data integration outcomes. Furthermore, modern iPaaS technology makes integration easier and less likely to create hairballs when coupled with a solid data management approach.

- Have a single pane of glass to monitor and manage the entire infrastructure regardless of where it resides (on-premise or in the cloud). One centralized view of the complete architecture will reduce complexity and improve operations management.
- Create and use "availability zones" where if one zone fails, it will fail over to another zone that is working. This concept is related primarily to disaster recovery operations, but is applicable for maintenance and rolling upgrades.
- Cross cloud computing allows primary work to occur in one cloud, but in the event of a failure the workload fails-over to another cloud. This is somewhat complex, but leverages cloud capabilities rather than using redundant data centers.
- Copy-back from cloud into on-premise data center. This practice copies back data from the cloud into an on-premise data center nightly for disaster recovery. However, as cloud computing capabilities increase and confidence in the cloud grows, this practice is becoming less common.

Hybrid cloud environments introduce inherent challenges to system administration, operations management, and disaster recovery. However, with the proper tools and planning in place, these challenges are overcome by enterprise IT staff.

Exploring Roles of Different IT Personas with Hybrid Clouds

Hybrid cloud computing is disruptive to companies and particularly to enterprise IT departments. However, disruption isn't necessarily bad, and the resulting upheaval often results in greater efficiencies. This impact is felt by many personas and new personas emerge as a result.

One very important emerging persona is the *cloud architect.* This person is responsible for designing and implementing an architecture and overall cloud strategy that will meet the company's business goals and IT objectives. The cloud architect must ensure cloud deployments are beneficial to the organization without introducing new inefficiencies. Implementation plans, timelines, and ultimately project success are responsibilities of the cloud architect.

Additional key personas exist with the distinct interests and roles in cloud adoption and hybrid cloud data management:

- **Chief Information Officer (CIO):** Focused on compliance, governance, and performance while wanting to lower operating costs and minimize disruption to business units
- **Vice President (VP) of Applications:** Maintains existing mission critical, on-premise applications while working with Line of Business (LOB) groups to deploy new cloud applications; focused on lowering costs and minimizing business disruption
- **Integration and data architects:** Technical staff with focus on integration and data locality, access, quality, and connectivity
- **Database and data warehouse staff:** Technical staff with initial focus on existing on-premise data stores, but will become more cloud centric as existing data moves from on-premise to the cloud and new data is delivered from the cloud
- **Business unit citizen integrators:** These people include departmental IT staff and SaaS admins who provide data management capabilities that require governed, self-service toolsets. These benefit directly from SaaS, PaaS, and IaaS cloud deployments.

IN THIS CHAPTER

- Examining common use cases for hybrid cloud data management
- Identifying ways to simplify hybrid cloud implementations
- Exploring hybrid clouds with Amazon Web Services and Microsoft Azure
- Highlighting iPaaS platforms to enable your hybrid cloud

Chapter 4
Using Hybrid Clouds in the Real World

Cloud computing propels companies into hybrid environments at an explosive rate. However, complexities emerge as the practical considerations of hybrid clouds become apparent. As companies dive into hybrid clouds, questions regarding the best use cases and toolsets come to the forefront. The use of integration Platform as a Service (iPaas) with SaaS and PaaS in hybrid clouds become a winning solution. In this chapter, you explore the many choices faced by companies and highlight successful approaches for hybrid cloud data management.

Exploring Use Cases for Hybrid Cloud Data Management

Based on real-world experience, companies find three common use cases where the benefits of hybrid cloud computing are optimized. The most successful hybrid cloud implementations are cloud migration (sometimes called *lift* and *shift*), hybrid data warehousing, and hybrid app integration.

Cloud migration with lift and shift

The most obvious and one of the most beneficial use cases for hybrid cloud is simply redirecting on-premise workload into the cloud. Infrastructure costs are reduced as databases, data warehouses, and applications are lifted out of on-premise and shifted to public cloud and/or to SaaS applications. In addition to lower infrastructure costs, once the components are in the cloud, customers benefit from all the cloud advantages including elastic scalability and paying for only the resources and capacity used.

Figure 4-1 outlines the lift and shift use case.

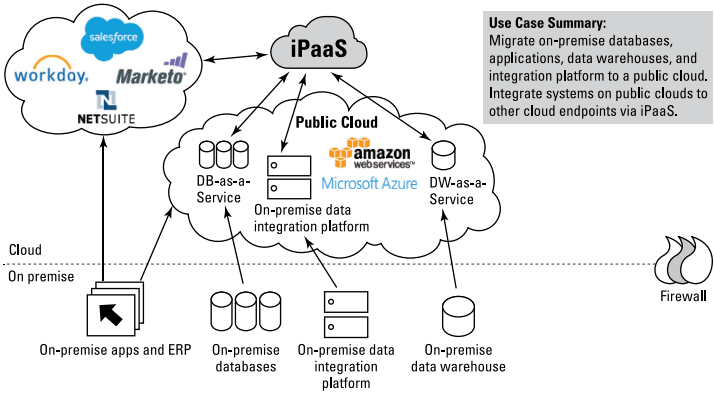

FIGURE 4-1: Lifting and shifting workload into the cloud.

The workload is shifted into the cloud. In this specific example, you move some of your existing databases, data warehouses, and integration tools to a public cloud such as Amazon Web Services or Microsoft Azure. Your on-premise applications move to SaaS cloud offerings or are migrated to public cloud infrastructure such as Amazon Web Services (AWS) or Azure. As there will still be some components on-premise, this architecture represents a hybrid cloud. This is just one example of lift and shift. Other customers may also optimize their data architecture as they're moving their data assets to public cloud. For example, a customer may choose, as part of lift and shift, to restructure an existing on-premise data warehouse as some combination of DW as a Service and Hadoop as a Service and data lake in public cloud, utilizing big data management tools in public cloud.

Lift and shift use cases are what many people envision when they adopt PaaS and IaaS. The lift and shift use case seizes the benefits of cloud computing, while reducing hosting and infrastructure costs within the data center.

Hybrid data warehousing

A second common use case is supporting hybrid data warehousing. Both moving on-premise data warehouses to the cloud or creating new cloud-based data warehouse, for new projects, and populating them with data is the heart of hybrid data warehousing. Data sources for these cloud data warehouses may reside in public cloud, SaaS applications, or on-premise data sources and applications. One of the challenges of a hybrid data warehouse project is the integration of all these data sources in hybrid cloud.

In Figure 4-2, you see hybrid data warehousing.

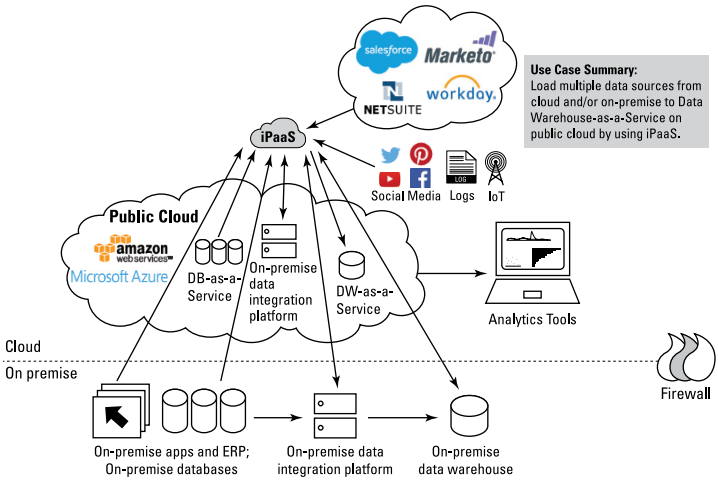

FIGURE 4-2: Leveraging hybrid data warehousing.

Figure 4-2 is a combination of on-premise data warehouses and Data Warehouse as a Service deployed. There's an iPaaS cloud that in this case loads the data warehouses from a myriad of data sources, such as SaaS applications, social media, logs, and Internet of Things (IoT), as well as on-premise data sources and applications. Some customers may utilize on-premise integration software hosted in public cloud or use it in conjunction with iPaaS. Others may continue to maintain some on-premise data

warehouse for loading their traditional on-premise data sources and implement in public cloud new data warehousing projects, which rely heavily on cloud data, such as IoT or social data.

The primary focus of the hybrid data warehouse use case is loading multiple data sources from cloud and on-premise into a cloud-based data warehouse, which can be rapidly set up and cost-effectively scaled to support big data. As part of this data loading process, iPaaS continues to play an important role, in some cases, in conjunction with on-premise data integration software, which may be hosted in public cloud or be installed on premise.

WARNING

Don't assume that iPaaS is the only way to solve hybrid cloud data management challenges. It's possible to use on-premise data management software (as part of Integration Competency Center) either installed on premise or hosted in public clouds such as Amazon Web Services and Microsoft Azure.

Hybrid app integration

One very exciting use case is hybrid application (app) integration. Unlike lift and shift, which is focused on reducing data center workload and costs, hybrid app integration's purpose is to integrate mission critical on-premise applications with cloud-based SaaS and public-cloud-hosted apps to promote greater agility and end-to-end business process integration within hybrid cloud. This use case is primarily implemented by using iPaaS solutions.

TIP

When I refer to apps in this context, I mean fully featured enterprise-scale applications rather than apps on mobile devices.

In Figure 4-3, hybrid app integration is displayed.

Within Figure 4-3, you see how iPaaS is used to synchronize and integrate data between on-premise applications (such as SAP), SaaS applications such as NetSuite (ERP), Salesforce (sales, service, and marketing), Workday (HR), and Marketo (marketing automation), as well as some apps that may have moved to AWS.

iPaaS is critical with hybrid app integration environments. The iPaaS provides critical integration services so on-premise applications are able to access and integrate with cloud services provided by SaaS and PaaS. The importance of iPaaS in this model can't be overstated.

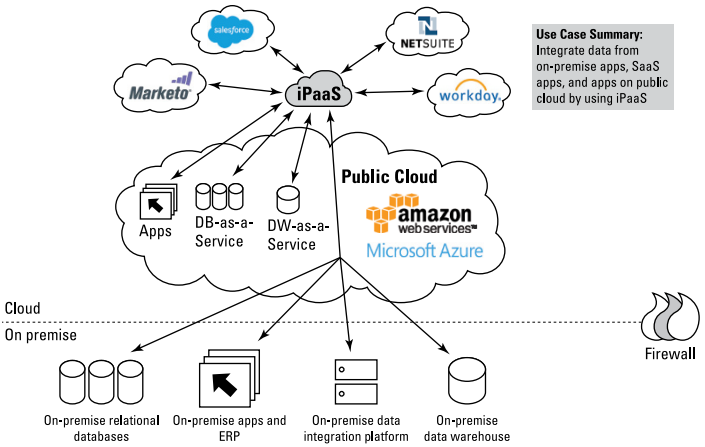

FIGURE 4-3: Deploying hybrid app integration.

Simplifying the Hybrid Cloud

With lift and shift, hybrid app integration, and hybrid data warehouses, we have identified the three most common use cases. Each use case has a different purpose, but for companies new to hybrid clouds, it's beneficial to recognize commonalities. Within these different use cases, commonalities exist as shown in Figure 4-4.

In Figure 4-4, you see the standard on-premise applications, databases, and data warehouses as well as on-premise integration software, common in so many data centers. Within the AWS and Azure clouds, IaaS and PaaS offerings are focused on data services such as data warehouse as a Service. In many implementations, the iPaaS solution is present and provides integration between SaaS services, public cloud services, and on-premise components.

When designing your hybrid cloud, consider cloud integration and data management recommendations:

- **Simplicity:** People like software solutions that provide value and are easy to use. If solutions aren't simple, they will not be widely adopted. Make sure your cloud integration and data management solutions embrace simplicity and ease of use with role-based tools designed for different personas such as developers and citizen integrators.

- **Connectivity:** Use a tool that separates and abstracts integration logic from the underlying data sources. This allows you to easily redirect workloads between data systems, add off-the-shelf connectors, and build connectors yourself for custom systems.
- **Automation:** Ensure you have the ability to automate and monitor the deployment of data integration and end-to-end business process integration once you've developed them with your toolsets.
- **Efficiency:** Deploy solutions that use templates, prebuilt connectors, visual design, reusable components, and Application Program Interfaces (APIs) as much as possible. Leverage prebuilt components to make common tasks both simple and efficient because they're already optimized and reduce your integration work.
- **Flexibility:** Cloud integration solutions must connect to all your on-premise and cloud-based applications and data systems; you need the flexibility to connect to whatever you need when you need it.
- **Security:** Security must be at the forefront of any solution you deploy; it can't be an afterthought. Not deploying secure solutions is negligence in this age of data breaches and cyber-crime. Make sure your integration and data management solutions place security as a top priority.

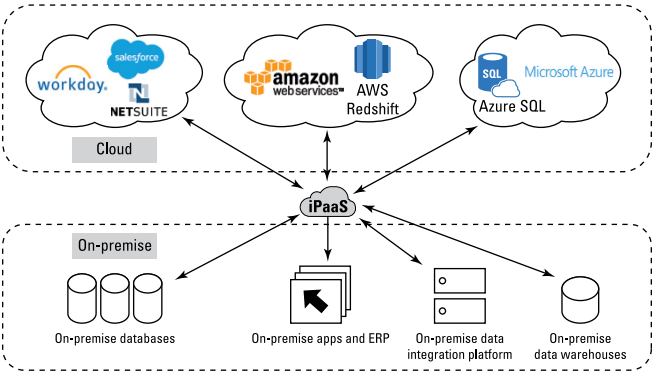

FIGURE 4-4: Recognizing commonalities within hybrid clouds.

Recognizing the commonalties in hybrid cloud solutions, and having guiding principles for cloud integration and data management, your journey to the hybrid cloud is made easier.

REMEMBER

Data management and integration solutions should seamlessly work across all environments, on-premise and cloud. Without this capability, you risk creating disparate silos of data on-premise and in the cloud. Ensure your toolset provides complete data management coverage across your entire hybrid cloud architecture.

Exploring Amazon Web Services

AWS provides both IaaS and PaaS offerings with PaaS as our primary focus for hybrid cloud data management. In particular, these AWS services are part of the hybrid cloud data management landscape:

- **Amazon Elastic Compute Cloud (EC2):** Virtual server hosting to provide capacity on demand
- **Amazon Simple Storage Solution (S3):** Scalable cloud storage service
- **Amazon Redshift:** Cloud-based petabyte-scale data warehouse service
- **Amazon Relational Database Service (RDS):** Cloud-based interface and management tool into 3rd-party relational databases such as Oracle, MySQL, PostgreSQL, Microsoft SQL Server, and MariaDB, in addition to Amazon's Aurora database
- **Amazon DynamoDB:** Cloud-based NoSQL database service
- **Amazon Aurora:** Relational database service, which is MySQL compatible
- **Amazon EMR (Elastic Map Reduce):** Web service using cloud-based Hadoop clusters designed to enable processing and analyzing large amounts of data

These AWS offerings, along with data management software, play an important role in the lift and shift, hybrid data warehousing, and hybrid app integration use cases. These AWS offerings benefit customers in hybrid environments to lift and shift capacity into

the cloud, reduce data center infrastructure costs, and enhance organizational agility and flexibility.

Exploring Microsoft Azure

Microsoft Azure provides both IaaS and PaaS offerings with PaaS as the primary focus. Azure provides many cloud offerings, but these are especially useful for the support of hybrid cloud computing:

- **Azure Virtual Machines (VM):** Virtual machine server hosting to provide capacity on demand
- **Azure SQL Data Warehouse:** Cloud-based data warehouse service
- **Azure Binary Large Object (BLOB) Storage:** Storage services for BLOBs, such as unstructured data, documents, binary objects, and audio/visual files
- **Azure SQL Database:** Cloud-based database service
- **Azure HDInsight:** Cloud-based Apache Hadoop managed service
- **Azure documentDB:** Managed cloud-based, NoSQL document database service

These Azure offerings, along with data management software, enable hybrid cloud computing use cases of lift and shift, hybrid data warehousing, and hybrid app integration. Azure supports extending your data center capacity into the cloud and enhances your agility and flexibly by supporting new initiatives for your development teams while reducing your data center infrastructure costs.


TIP

You can learn more about Amazon Web Services and Microsoft Azure at `aws.amazon.com` and `azure.microsoft.com`, respectively.

IN THIS CHAPTER

- Ensuring connectivity and data visibility
- Supporting growth and performance
- Maintaining constant security

Chapter 5
Ten Ways to Manage Hybrid Cloud Data

Managing data effectively in your hybrid cloud environment is the key to successful cloud projects. Your cloud approach and the architectural decisions you make are the primary factors leading to success. In this chapter, I identify tips and techniques to enhance your skills managing data and apps in hybrid cloud environments.

Ensure You Have Connectivity

A computer system is only as good as the data that drives it; good data leads to good results. The challenge is getting quick access to that quality data. Any data integration strategy and the accompanying technology must access the data required. In hybrid cloud environments, this challenge is magnified as data is spread across on-premise and cloud data sources.


TIP

Ensure that your data integration and management tools connect out of the box to a wide variety of data sources regardless of their location in cloud or on-premise. Make sure that your connectors are high performance, standardized, easy to use, and reusable

across systems. Also ensure that your data management tool has a connectivity architecture that separates integration logic from underlying data systems.

Understand Your Data Visibility

Data visibility and lineage are critical to high-quality data management. Data lineage is the life story of your data — where it originated, where it has been, and how it presently exists. For example, with great data lineage, you instantly know if you have sent customers' Personally Identifiable Information (PII) to a cloud database or data warehouse in another country, out of compliance with your corporate policy or regulation. Without clear visibility into your data, its lineage, and its metadata, your data governance suffers, impacting compliance with regulations and the ability to support audits.

In cloud environments, data moves fast and spans the globe; establishing a clear view of the origin and journey of your data becomes an especially daunting challenge. Complex data flows obscure your comprehension of the data. Counter these complexities with tools that focus on metadata to uncover the mystery of data on-premise and in the cloud. The payoff is greater governance for your organization as well as increased development productivity and reduced risk associated with code change management. Metadata is also the foundation for data quality mastering and security initiatives.

Prepare for a Leap in Scalability

Sure, clouds are supposed to scale, but have you really considered what that means for your business? Imagine being able to process (and manage) extreme volumes of data. Consider the benefits of scaling without performance problems or delays to add more hardware when a workload threshold is exceeded. Oh, how many problems we could've avoided if we had this capability in years past!

The catch is how do you integrate all this new data (especially with social media, big data, and IoT) now that the sky is the limit? It's not enough to have data systems that scale infinitely, you also

need the tools that load them and manage the data to match that scale. Use iPaaS technologies that are engineered for performance with scale so you can access all the data you need now that you don't have to worry about traditional capacity issues.


TIP

Cloud bursting occurs when companies extend their on-premise workloads onto public clouds to cost-effectively handle short-term, infrequent events such as Black Friday sales events or end-of-year processing.

Redefining Operational Control

With hybrid cloud computing, the good news is we are now free of the limited confines of our data center; our applications and data span the globe. The downside is that with hybrid cloud computing, our data and applications have escaped our data center and are potentially traveling all over the globe.

You need a special set of solutions designed for hybrid to manage data and applications distributed on-premise and within infinitely scalable clouds. Obviously you want to monitor and manage your production applications. Leverage tools that provide a single point of control for your production systems regardless of their location; you want a single pane of glass from where you manage your operations. Ensure your tools efficiently manage your production data flows and alert you to issues before they grow into major problems.

Understand How Security and Compliance Is Managed

Cloud computing, and in particular hybrid cloud computing, changes the landscape for how security, compliance, and governance is managed. While IT has a long track record of data center security and compliance, the world of cloud changes long-established boundaries and controls.

Enterprise IT no longer should view its role as limited to the confines of the data center. The applications and data have expanded into the cloud, so too must the role of enterprise IT security, compliance, and governance. Enterprise IT will no

longer control physical infrastructure when it is cloud based, but it will still own that data and be responsible for ensuring that necessary controls are established and followed. With the advent of the cloud the mission of enterprise IT governance has not ended, but it must *evolve*.

Rethink Special Cloud Cases

Let's be honest; don't believe all the marketing hype that any application can easily move to the cloud because they can't. Cloud computing is optimized for standardized applications and data; highly custom applications that don't fit into cloud service models don't thrive in the cloud.

The cloud is evolving to host more service models (mobile, big data, analytics, IoT, and so on) and applications are becoming more standard. However don't try to force an application not suited for the cloud just to be "cloud enabled." Your business will be better off leaving highly custom applications on-premise to create hybrid cloud environments integrating with applications that are well-suited for the cloud.

Identify Mandates for Cloud

C-level executives are justifiably excited for the cloud; they read books like this too. As a result, "Cloud First" dictates are common both in private companies and even within the U.S. Federal government. In many cases, new applications are mandated to deploy to the cloud unless a Business Case Analysis (BCA) is performed to justify an on-premise deployment. Alternatively, some environments permit cloud, but only with specific security restrictions (think military and intelligence). The takeaway is that you need to understand the guidance of your organization as you plan your cloud strategy.

Target IT Infrastructure to Move

Leading industry analysts estimate $204 billion (yes, that is *billion*) was spent in 2016 on the public cloud; up from $175 billion in 2015. Of that, the largest sector of growth is IaaS showing a jump

of 38.4 percent ahead of strong growth in cloud management and security services, SaaS, and PaaS. This estimate is a clear indicator companies are redirecting large amounts of their infrastructure budgets to cloud-based infrastructure rather than back into existing data centers. Evaluate your strategy in this infrastructure paradigm shift.

Deploy Highly Sensitive Environments to the Cloud

"Cloud isn't secure!" has been repeated many times, but does it have merit? Probably not. If that was true, would the U.S. Department of Defense (DoD) have authorized and promoted the use of vetted commercial cloud service providers for their data? Would the Central Intelligence Agency (CIA) have contracted with Amazon for a $600 million cloud?

Cloud service providers know they must disprove the myth of cloud insecurity, so they work hard to ensure needed controls are in place. Additional stringent controls and governance are applied for customers with specialized security needs.

Ensure Security Is at the Forefront

Regardless of on-premise computing or new cloud architectures, security must be built in from the start rather than bolting in the last phase of system design. Select tools and solutions that support secure cloud services. Leverage mature iPaaS offerings that aid in security and governance as part of the larger topic of data management. For example, make sure your iPaaS solution ensures secure movement of data between on-premise and cloud.

WARNING

Verify the details of what security protocols are applied and how standards are enforced. Don't trust high-level marketing material; fact check the details. You must take these extra steps not just for security itself but also for audit compliance, governance, and due diligence requirements.

Unleash the power of intelligent data

Businesses are relying on cloud computing to deliver services. Many are running cloud services alongside traditional on-premises applications. With businesses pressured to efficiently extract value from their data, you can't afford to get hybrid cloud data management wrong. This book shows you how to move workloads to public cloud with Amazon Web Services and Azure. Learn to better manage and integrate your data by creating a connected, scalable, flexible, and secure hybrid cloud computing architecture.

Inside...

- Understand cloud's impact on business
- Define hybrid cloud data management
- Learn ways to integrate and manage hybrid cloud data with AWS and Azure
- Design for real-world use cases
- Avoid integration hairballs
- Enhance your data management

Informatica

Michael Wessler has managed multiple web applications for the Department of Defense and has consulted at various government and private agencies.

Go to Dummies.com®
for videos, step-by-step photos, how-to articles, or to shop!

ISBN: 978-1-119-45828-9

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley’s ebook EULA.