RED HAT CLOUD FOUNDATIONS: CLOUD 101

TABLE OF CONTENTS

EXECUTIVE SUMMARY

- Clouds are a new way of building IT infrastructures from dynamic pools of virtualized resources that are operated as low-touch IT services and are consumed in a modern, web-savvy way.
- Conceptually, cloud computing can be thought of as building resource abstraction and control on top of the hardware abstraction provided by virtualization.
- In most cases, cloud computing infrastructures will evolve in scope and complexity over time. Functions such as elastic provisioning, metering, and self-service will often be added as the environment matures and goes into full production.
- Cloud computing includes delivering services from multiple levels of the software stack. The most widely-used taxonomy specifies Infrastructure-as-a-Service (e.g. compute and storage), Platform-as-a-Service (e.g. middleware and infrastructure automation), and Software-as-a-Service (applications). These levels may layer on top of each other but can also exist independently.
- Cloud computing can take place either on-premises (private cloud), as a shared, multi-tenant, off-premises resource (public cloud), or some combination of the two (hybrid cloud).
- Cloud computing is not just another name for virtualization. It builds on virtualization, and constructing a virtualized infrastructure will be the first step to a private cloud for many customers.
- A private cloud improves efficiency, helps organizations save money, and improves service levels relative to less flexible and dynamic IT infrastructures.
- Public clouds offer a pay-as-you-go pricing model for computing resources that customers do not need to own or operate themselves.
- Clouds take many forms because different customers or even different business units and applications within a single customer have vastly different requirements. One size doesn't fit all.
- Cloud infrastructure should support interoperability, open standards, and the ability to run existing applications in many different environments and on many different clouds.
- Clouds should provide flexibility for your organization and not lock you into a single vendor's solution.

INTRODUCTION

In a remarkably short time, cloud computing has emerged as a hugely important evolution in the way that businesses and individuals consume and operate computing. It's a fundamental shift to an operational model in which applications don't live out their lives on a specific piece of hardware and in which resources are more flexibly deployed than was the historical norm. It's also a fundamental shift to a development and consumption model that replaces hard-wired, proprietary connections among software components and the consumers of those components with lightweight web services and web-based software access.

In short, cloud computing refers to a convergence of technologies and trends that are making IT infrastructures and applications more dynamic, more modular, and more consumable. That's a big change that has implications that touch on just about every aspect of computing.

For end-user customers, cloud computing provides the means to ramp up new services or reallocate computing resources rapidly, based on business needs. It means having the ability to run an application either on-premises or off-premises (or a combination of the two) based on cost, capacity requirements, and other factors. For software vendors, cloud computing offers new ways to deliver applications and reduce the friction associated with installing upgrades or additional modules.

However, for something that is in many ways at the fore of where information technology is headed, it can still be a challenge to get your arms around what cloud computing is exactly. The biggest stumbling block is that while cloud computing as a high-level business concept and technology approach can be described succinctly, it takes a variety of different forms that aren't always obviously related to each other. Furthermore, the pains that cloud computing relieves aren't necessarily the same for datacenter operators, developers, and end-users; thus, what's most relevant about cloud computing to you depends to some degree on who you are and where you are located in an organization. Plus, of course, cloud computing is young, and it continues to develop along many axes.

This whitepaper aims to make sense of it all for audiences that haven't been deeply involved in the details of cloud computing as it has rapidly evolved. It lays out the characteristics of a cloud computing infrastructure. It discusses some of the things that cloud computing isn't, even if they're often conflated in customers' and prospects' minds. It describes the forms that cloud computing can take and how different types of technology providers and consumers of technology relate to each other.

CLOUDS TODAY AND TOMORROW

The National Institute of Standards and Technology (NIST) defines cloud computing as "a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." This definition, together with associated service and deployment models, has emerged as a tough industry consensus of where cloud computing is headed. (http://csrc.nist.gov/groups/SNS/cloud-computing/)

Characteristics such as the following will become the norm over time in mature, production cloud implementations. However, not all will necessarily be present with early implementations—and, indeed, not all will necessarily be valued by a given customer. For example, while pay-as-you-go utility pricing is often associated with cloud computing, organizations that implement internal clouds may not care about highly granular chargeback to individual departments or users. Ultimately, "cloud computing" is the set of capabilities that solves a customer's problems or provides them with new business value rather than something described by an inflexible, abstract definition.

RESOURCE ABSTRACTION AND POOLING

Pooled computing resources serve multiple consumers using a multi-tenant model (whether different internal groups within one company or different organizations within a shared, public resource) with physical and virtual resources dynamically assigned and reassigned depending on demand. The consumer of the service generally has limited control over or knowledge of the exact location of the provided resources but may be able to specify location at a higher level of abstraction—such as to create high-availability domains or to meet regulatory requirements around data location. Examples of resources include storage, processing, memory, network bandwidth, and virtual machines.

NETWORK-CENTRIC

Whether implemented within a single organization or at a public cloud provider, cloud computing is network-centric. Services are made available over the network and accessed through standard mechanisms, typically lightweight web protocols.

SIMPLE, FAST PROVISIONING OF RESOURCES

One of the ways that cloud computing makes an IT infrastructure more agile is by enabling new resources to be brought online quickly. This may include over time a degree of self-service—meaning that a user can provision computing capabilities, such as server capacity and storage, as needed without having to interact with a human. (Typically, policies that limit total resource limits or credit without an additional level of authorization would limit the scope of self-service requests.)

RAPIDLY AND ELASTICALLY PROVISIONED RESOURCES

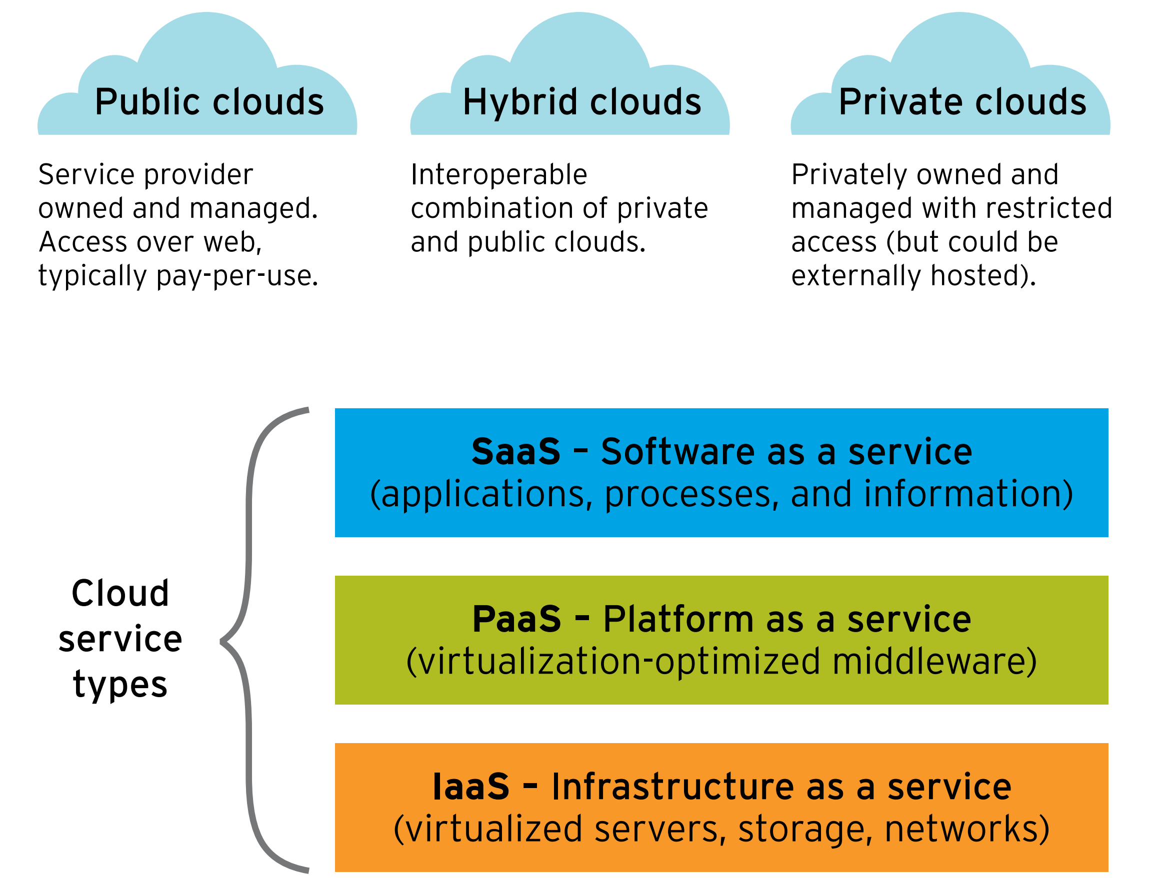
In a cloud environment, resources can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out based on pre-set policies and the demands of an application. Just as importantly, resources can also be rapidly decreased when they are no longer needed, avoiding the familiar situation of unused servers sitting idle after the task they were initially purchased for ends.

UTILITY PRICING

Cloud computing is often associated with "utility pricing" or pay-per-use, although today, this is something that is largely specific to public cloud providers. Metering at a level of abstraction appropriate to the type of service (such as storage, CPU usage, bandwidth, or active user accounts) will become more widespread over time as organizations learn the types of data that are most useful. Even if this information is not used to directly charge for use, it can be applied to capacity planning and other purposes.

THE CLOUD TAXONOMY

FIGURE 1



The figure above shows what's come to be a widely accepted framework for viewing and discussing the cloud computing space. It shows the different types of cloud services and the different types of cloud deployments.

INFRASTRUCTURE-AS-A-SERVICE (PUBLIC CLOUD EXAMPLES: AMAZON WEB SERVICES, IBM CLOUD)

IaaS lets the user or creator of the service provision processing, storage, networking, and other computing resources on which they can then run operating systems and applications. The deployer of the service does not manage or control the underlying compute infrastructure except select networking configurations or perhaps physical location of the resources at a gross geographical level. On the other hand, the deployer is responsible for configuring and maintaining their own operating systems and software and, to a large degree, scaling their application and provisioning all the services required to run it.

PLATFORM-AS-A-SERVICE (PUBLIC CLOUD EXAMPLES: GOOGLE APP ENGINE, FORCE.COM)

PaaS usually describes an additional level of services layered on top of an IaaS foundation. These services effectively provide an additional level of abstraction, taking care of tasks so that a developer or operator doesn't have to. Examples include widely used middleware, such as application servers and databases that would otherwise have to be added to a base-level infrastructure on a case-by-case basis. Alternatively, grid software middleware serves to automate common tasks such as scaling applications. The lines between what is a platform and what is just infrastructure and what is end-user hosted software can blur. But platforms are generally a higher level of abstraction than infrastructure while still not being an application—that is something directly interacted with by a line-of-business user.

SOFTWARE-AS-A-SERVICE (PUBLIC CLOUD EXAMPLES: SALESFORCE.COM, INTUIT QUICKBOOKS ONLINE)

SaaS is perhaps the most familiar face of cloud computing, referring as it does to the direct consumption of a cloud resource—an application—by end users rather than by developers or operators. There is no inherent relationship between the hosted application and the nature of the infrastructure on which it runs. However, SaaS applications can and often do run on cloud infrastructure and middleware; after all, cloud computing virtues such as flexibility and efficiency are ultimately in service of applications. Thus, although SaaS does not necessarily imply IaaS or PaaS, the underlying infrastructure very much affects how efficiently the application runs and how it is experienced by users.

Cloud computing can also be delivered in different ways.

With **public clouds**, the cloud infrastructure or platform is made available to the general public or a large industry group and is owned and operated by the organization selling cloud services. (To be absolutely true to the described framework, any hosted SaaS application is technically a public cloud, but in practice the term is usually reserved for infrastructure and platform offerings.)

A **private cloud**, by contrast, is operated solely for an organization, either behind an internal firewall or operated by a third party for the exclusive use of that organization. It may be managed by the organization or a third party and may exist on premises or off-premises. (In other words, this definition really refers to the boundaries of control and trust rather than who employs the infrastructure's operators or who holds title to the equipment.)

A **hybrid cloud** blends the public and private cloud models. It assumes some level of interoperability between a private cloud and public cloud implementation. This approach might be of interest to a customer who wants to own and operate a cloud infrastructure sized for typical loads but reserve the option of renting additional capacity on a pay-as-you-go basis to handle load spikes.

Possible variants on these basic approaches include **community clouds**, which are shared by several organizations and support a specific community that has shared concerns such as security, compliance, or other needs that could be specific to a particular industry.

WHAT CLOUD COMPUTING ISN'T

Before getting into the benefits of cloud computing, it's worth briefly discussing what cloud computing isn't but to which it's often equated.

CLOUD COMPUTING IS NOT JUST VIRTUALIZATION OR GRID

Cloud computing also abstracts resources and provides self-service and automated facilities to control the allocation of resources. Does this mean virtualization and grid are irrelevant to cloud computing? Hardly. Hardware abstraction is a key enabler of cloud computing and, in fact, virtualizing/consolidating is a logical first step in a cloud infrastructure project. Virtualization also means that there's an evolutionary path to the cloud.

NOR IS CLOUD COMPUTING JUST ABOUT USING EXTERNAL RESOURCES

Books have popularized the public cloud, presenting it as a revolutionary way to consume computing, akin to the development of the electric grid. Public and hybrid clouds do offer a great way to deal with unpredictable workload demands and a way to get applications up and running and scaling quickly without CAPEX or time-consuming hardware procurement cycles. However, for regulatory, compliance, and a host of other reasons, many organizations will not want to move their applications – or at least all of their applications – to a public service provider. And, for them, private cloud infrastructures can provide many of the same benefits as a public cloud without the loss of control.

CLOUD COMPUTING ISN'T JUST A BUZZ WORD HYPING COMPUTING AS USUAL

Yes, like any hot technology, there is hype around cloud computing. And various aspects of cloud computing do have historical antecedents. Accessing applications over the network recollects application service providers and remote storage storage service providers. Utility pricing for rented compute resources recalls timesharing. However, collectively, the technologies that make up cloud computing represent the infusion of characteristics – like abstraction, elasticity, automated, and dynamic – into traditional computing, which is better described by words like hardwired, fixed, manual, and static.

WHY CLOUD COMPUTING

A cloud computing environment can react quickly and, in some cases, automatically to changes in workload demand, and new applications can be provisioned with much less time lag and effort than with traditional computing infrastructures.

An internally implemented cloud computing infrastructure improves efficiency and helps organizations save money relative to less flexible and dynamic IT infrastructures. Better provisioning processes make it easier to reclaim servers that are no longer being used for a project. These private clouds are often the evolution of a virtualized infrastructure into something that's more dynamic and automated.

Public cloud providers offer a way to buy computing capacity as it's needed, which can save money, especially where the alternative is to over-purchase capital equipment to protect against demand spikes or greater-than-anticipated capacity requirements. The fact that so many public cloud infrastructures are based on Linux and other open source software also helps keep costs down. They also allow organizations to pay for their IT as an operating expense rather than making capital purchases.

However, cloud computing isn't just about cost cutting. It's also about using IT to drive innovation and respond to changes in the business. A more flexible infrastructure, whether hosted internally or externally, encourages more experimentation and iteration, which in turn lets businesses introduce new technologies and services more quickly and more often.

Clouds based on Red Hat technologies leverage infrastructure products that are used by thousands of enterprises worldwide for some of the most demanding applications. And Red Hat Enterprise Virtualization, the on-ramp for Red Hat clouds, is designed from the ground up for the needs of cloud computing with great performance, quality-of-service guarantees, and advanced security features. All this is made possible by an open source development model that brings to bear more development resources than any single company can marshal, while augmenting it with proven enterprise support.

WHO CREATES CLOUDS AND WHO USES THEM

Earlier, we covered the different types of clouds and delivery models. Now, we'll take a look at what that means in terms of who sells what to whom.

IT infrastructure vendors like Red Hat sell the products needed to build clouds to end-user organizations and to public cloud providers like Amazon. They may also sell services to assist those organizations with their implementations and certify specific cloud providers as destinations for their products. For example, Red Hat has worked with its Red Hat Certified Cloud Providers to develop an industry-leading certification and testing program, making it simple, safe, and cost-effective to move and manage applications between on-premise and public clouds.

However, IT infrastructure vendors do not generally operate their own public clouds. (IBM is an exception; it both sells hardware and software to build clouds, and it operates a cloud for consumption by end-user customers.)

Public cloud vendors may internally develop value-add cloud services that they offer to their customers and that they use to monitor, manage, and run their implementation. However, especially as you move beyond the very largest public cloud providers, which have large in-house development staffs, service providers are primarily consumers of IT infrastructure software and hardware.

Application vendors/developers may themselves offer their software in the form of SaaS — in which case they will directly use infrastructure software and hardware developed by others — or they may use a third-party cloud or other service provider, in which case they consume PaaS, SaaS, or some form of hosting.

End users — whether consumers or business users — typically will only consume SaaS, whether it's an internally-hosted business application or one that's on the public web (which is why cloud means "software delivered over the Internet" to so many people).

Operators of enterprise IT infrastructure will purchase the infrastructure needed to build private clouds along with infrastructure services from external cloud providers.

ONE SIZE DOES NOT FIT ALL

We see many types of clouds and delivery models for the simple reason that not all customers are alike. They have different security and compliance requirements, which may even be mandated by the government. They have different levels of IT skills. A given application may be more or less central to their core business. They may be large or small. They may have big, sophisticated IT departments, or IT could be a part-time job for one person.

Consider these scenarios.

A medium-sized manufacturing company is implementing a new customer relationship management (CRM) system. This is obviously an important application for the organization, but it's not an application that particularly differentiates it from its competitors. It might therefore be a candidate for using a SaaS application from a third party. On the other hand, if it has extensive integration needs with existing enterprise resource planning applications, it may choose to implement CRM internally within a virtualized environment or as part of a broader private cloud implementation.

A large Wall Street firm has far more constraints when it comes to core trading applications. All sorts of regulatory and other legal requirements would make it difficult, if not impossible, to delegate such computing to a third party. Besides, this is at the heart of its business, and any downtime, or even sub-par performance, will be hugely expensive. Such an application has to be kept in-house under tight control. Given the potential for rapid growth and the ongoing need to bring new applications online quickly, this is fertile ground for a private cloud.

A software development startup, on the other hand, has little capital and wants to be able to iterate quickly and incur costs only as it gains users for an online application that it plans to offer. It really has no idea how many users it will gain and when. This is a tailor-made case for a public cloud platform with its utility pricing model and virtually unlimited scalability. However, as it grows, the company may find it's cheaper to start running some of its computing infrastructure in-house, making a common application platform between public and private environments an important consideration. This makes Linux and open source, the foundation of the vast majority of public clouds around the world, the logical foundation for private clouds as well.

ADOPTION CONSIDERATIONS

The above scenarios suggest factors that play into how cloud computing is deployed. Although some might be thought of as barriers to cloud adoption, in reality, they're better thought of as requirements that make some approaches preferable to another.

Security and compliance is one such set of requirements. Some industries, such as healthcare and financial services, have to comply with regulations regarding things like privacy, security, and compliance. As a result, such organizations will tend to implement private clouds rather than making use of public clouds for their applications covered by such regulations. The issue isn't that public clouds are insecure in some general sense, but rather that very specific requirements for data separation, audit, and so forth may be difficult or impossible to observe in a public cloud where they can't lay hands on the servers where their applications are running or the disks where their data is stored.

Of course, a secure foundation is important for cloud computing as it is for other enterprise applications; indeed, it's even more important because a cloud infrastructure is a shared infrastructure. That's why Red Hat's KVM-based virtualization uses sVirt, which leverages SELinux to enable Mandatory Access Control security to be applied to guest virtual machines and thereby eliminates the possibility of guest operating systems attacking not only each other but the host system. SELinux was developed cooperatively with the U.S. government to handle the most demanding security requirements, such as shared cloud computing environments.

The perception, abetted by books, magazine articles, and, indeed, the strategy of some vendors, is that cloud computing requires a radical rip-and-replace approach. This scares off potential customers. It's true that a fully-realized cloud computing implementation does represent a dramatic step forward from a traditional computing infrastructure where applications are closely tied to specific, physical servers.

However, cloud computing isn't all or nothing. Public clouds enable—encourage even—experimentation with, say, a new application under development. And private clouds can be built out incrementally; virtualization offers an on-ramp to cloud computing and a logical first step. Depending upon the organization, the applications, and even the point in an application's lifecycle (for example, prototype vs. production), it may be preferable to run the app in a private cloud, a public cloud, or even a combination of the two. Thus, an important consideration for customers is to ensure that their cloud technology provider provides the means to interoperate among different types of clouds.

In general, the lack of control and confidence that service level agreements will be met—if indeed they are even explicitly guaranteed—makes some organizations wary of public clouds. This highlights the need for the same sort of due diligence and risk mitigation required whenever a business process or task is outsourced to a third party.

But cloud computing is not much different in this regard from using any other hosted service, or for that matter, contracting out any sort of function. Consider things like the stability and maturity of the provider, support policies and guarantees, whether their platform allows you to move the workload to another provider or to an internal cloud, backup and recovery guarantees, and their contract terms and service levels. And for core applications with stringent SLAs, the right answer may ultimately be to run them in a private cloud. This isn't so much because an internal IT infrastructure is inherently better than that operated by a third party, but it is more visible—and that may be the deciding factor if the application is that important to the business.

MOVING TO THE CLOUD

Some of the characteristics and examples of cloud computing discussed here may not seem particularly new. SaaS brings to mind application service providers (ASP) and other forms of outsourced applications that go back to third-party payroll providers. Public cloud storage likewise has an analog in the storage service providers. What is outsourcing but paying someone else to operate your computing resources? Hosting and ultimately even time sharing represent other ways of renting time on third-party equipment. Virtualization, even if it's not a synonym for cloud computing, plays a big part in building computing infrastructures that have the sort of flexibility and resiliency that cloud computing demands.

Cloud computing does indeed have many antecedents and builds on many existing technologies and approaches, which is why we say that cloud computing is evolutionary rather than revolutionary.

That said, cloud computing represents an important step forward in the way that computers are operated and computing is consumed because of the way it brings together pieces that were previously disconnected or immature. Networks are faster and applications more modular and web-savvy. Server virtualization is both an important component of cloud computing and a major catalyst in changing the way people think about abstraction of which cloud computing is just another type. An increasingly mobile workforce and the proliferation of client form factors make traditional types of enterprise applications and the way they are operated increasingly outmoded. And the sheer scale of today's computing infrastructures demands something new. That something is cloud computing.

Many organizations are ready to start heading there. Assessing the IT and business problems that they're trying to solve will help suggest the best path to follow. For some, the fast adoption and utility pricing of public cloud services will address the most immediate pain points. For other, a private cloud that builds on a virtualized foundation will be the strategic choice that helps IT bring on new business services more quickly. Whichever direction (or hybrid approach), it need not be disruptive but can incrementally build on existing investments and processes while starting to reap the benefits of a new approach to computing.

ABOUT RED HAT

Red Hat was founded in 1993 and is headquartered in Raleigh, NC. Today, with more than 60 offices around the world, Red Hat is the largest publicly traded technology company fully committed to open source. That commitment has paid off over time, for us and our customers, proving the value of open source software and establishing a viable business model built around the open source way. Red Hat provides high-quality, affordable technology to the enterprise. Our solutions are delivered via subscription and range from operating systems and platforms like Red Hat Enterprise Linux and JBoss Enterprise Middleware, to application and management tools, as well as consulting, training, and support.

RED HAT SALES AND INQUIRIES

NORTH AMERICA
1-888-REDHAT1
www.redhat.com

EUROPE, MIDDLE EAST AND AFRICA
00800 7334 2835
www.europe.redhat.com
europe@redhat.com

ASIA PACIFIC
+65 6490 4200
www.apac.redhat.com
apac@redhat.com

LATIN AMERICA
+54 11 4341 6200
www.latam.redhat.com
info-latam@redhat.com