Cloud Computing Foundation White Paper

Contents

White Paper Overview

This white paper provides information required for the Cloud Computing Foundation exam that is not covered by either the NIST Definition of Cloud Computing or the text "The Cloud at Your Service." It is intended to provide supplemental material to address areas in the exam necessary for complete topic coverage. It does not cover all the Foundation content, only that not covered in the other exam materials. For that reason the white paper contains a number of partial topics to address material missing from the exam main resources.

Table of Acronyms

This white paper uses a number of acronyms. To aid reader comprehension, they are listed here:

Acronym	Stands For	Meaning
IaaS	Infrastructure-as-a-Service	Please see NIST Cloud Computing Definition for full description
PaaS	Platform-as-a-Service	Please see NIST Cloud Computing Definition for full description
POC	Proof of Concept	A limited trial of a product to establish its capabilities and functionality
SaaS	Software-as-a-Service	Please see NIST Cloud Computing Definition for full description
CSP	Cloud Service Provider	

1. Security and Compliance Cloud Computing Challenges

Survey after survey identifies security as people's number one concern about cloud computing. IT organizations decide to continue existing on-premise deployment practices (often using a private cloud environment) because they have higher confidence in the security of their own environment.

However, a curious thing emerges when one engages in a discussion on this topic. When IT professionals are asked what specific concerns about cloud computing security they have, responses like these are common:

- What's to prevent a cloud service provider employee from sticking a thumb drive into a server and downloading my data?
- How do I know if my company's data is kept locally or stored in another country?
- What guarantees do I have about my application's uptime when it runs in a cloud computing environment?
- What financial compensation can I receive if my application is unavailable?
- How can I know if the cloud service provider is applying appropriate patches to the hypervisor it uses to provide its cloud computing environment?
- How can my company govern who is able to manage resources in the cloud computing environment?

What's clear from these questions is that the word 'security' is used to represent a range of concerns, only some of which focus on security. Some concerns appear to be associated with security but are actually related different topics.

For example, the topic of financial compensation in the event of downtime is not security-related, but rather falls into the area of risk. Questions regarding change management with respect to infrastructure resources in cloud computing environments relate to governance, while questions regarding hypervisor patching practices rest squarely in traditional security issues.

This conflating of concerns under the term 'security' represents a significant challenge for those seeking to understand how to translate existing practices used for on-premise environments to the new world of cloud computing.

More troubling is what underpins these questions: the assumption that the responsibility for the area under discussion lies entirely with the cloud service provider. The fact is that cloud computing represents a shared responsibility. The demarcation line for the division of responsibility between the user and the cloud provider varies according to the area and also according to what delivery mode is being evaluated.

The demarcation line of responsibility is sometimes referred to as a "trust boundary", which illustrates that for those areas which fall into the cloud provider's area of responsibility users must trust the implementation and execution of the provider. There are techniques of evaluation which will be discussed in detail in this Paper but, at the end of the day, users must trust providers to uphold their responsibility.

As a starting point, it is useful to understand how the three different areas of security, compliance, and risk interact in cloud computing. Even more important is to understand how the responsibility for these is shared between the cloud user and the cloud provider. These two matters are illustrated in Figures 1 and 2.

1.1. The Relationship between Security, Compliance and Risk

To understand the interaction between security, compliance and risk, please refer to Figure 1, which represents how the three areas together form a whole. Significantly, both security and compliance have a boundary within their areas, with both the cloud provider and cloud user retaining responsibility for a portion of that area.

When it comes to risk, however, no such boundary exists. This lack of a boundary represents the fact that all service agreements shift primary responsibility for risk to the user. Should a cloud application fail in availability, cause a financial loss to the cloud user, or even fail to comply with important compliance requirements, the cloud provider limits its responsibility significantly, typically to a refund of fees.

This asymmetric risk arrangement may seem unfair – after all, the cloud provider's decisions and operations may cause a failure in compliance, which results in a financial penalty (i.e. a risk outcome borne by the cloud user). So why should the cloud user have the financial responsibility and not the cloud provider that caused the compliance failure?

Despite this seeming unfair, the careful assignment of risk responsibility to the cloud user is universal throughout the cloud computing world. There may be slight differences in what individual cloud providers will provide in compensation for service failures - for example, one provider may offer a credit for service unavailability on a one-to-one basis (i.e. if the service is down for one hour, the cloud user will receive one hour's credit to the monthly service cost), while another will refund a week's service costs for an outage of one hour – but, despite these differences, every provider limits its financial exposure due to service unavailability.

It's important to understand that this risk limitation is not unique to cloud computing. Outsource providers (e.g., firms that take over operating a company's IT data center) also limit their financial responsibility in the event of an outage. Therefore, it is important not to regard this risk limitation as a complete restriction on using a cloud

provider. If a company regards any risk limitation by a service provider as unacceptable, it should continue to operate its own computing environment and forego use of an external cloud provider.

The important thing to take away from this discussion is that when cloud computing security is raised as an issue, other issues are often being addressed. It is important to distinguish what type of issue is of concern, as that will change the method of evaluating the issue, the demarcation of the trust boundary, and the appropriate actions to be taken by the cloud user.

To help distinguish which issue is being evaluated and how to identify the trust boundary appropriate for the issue, Figure 1 can be used as an aid.

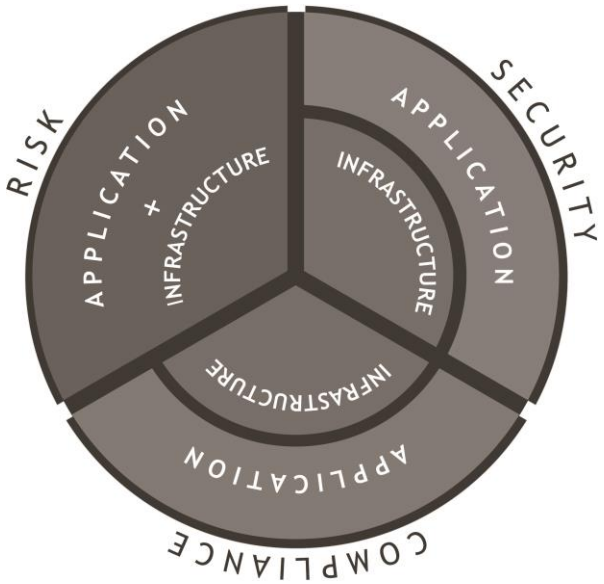

Figure 1

1.2. Understanding the Trust Boundary

As noted, in the areas of compliance and security, the cloud user and cloud provider both hold some of the responsibility. The interface between where one party's responsibility ends and the other begins may be referred to as the trust boundary. While the existence of a trust boundary intuitively makes sense, two questions arise:

1. How can a cloud user know where the trust boundary lies? After all, the service of a SaaS provider is quite different from that of an IaaS provider.

2. Once the boundary is defined, what can the cloud user do to verify that the cloud provider is adhering to its responsibility? In other words, what are the appropriate actions to take to ensure the cloud provider lives up to its commitments?

So, what does the trust boundary represent? In its basic form, the trust boundary represents a demarcation line: on one side of the line, the cloud provider possesses responsibility for security measures; on the other, the cloud user possesses responsibility.

For example, in an IaaS environment it is clear that the cloud provider has responsibility for physical security of the computing facility. It is also clear that the cloud user is responsible for the application code.

The cloud provider is in control of what security practices are followed; the cloud user can only audit what information about those practices the cloud provider offers, and evaluate whether the practices are sufficient for the user's needs. The cloud user can determine what the correct security practices are and can take active steps to implement those practices.

In short, on the cloud provider's side of the trust boundary, the user is a passive assessor of what the cloud provider implements in terms of security practices. On the cloud user's side of the trust boundary, the user is an active implementer of security practices.

The location of the trust boundary varies according to what model of cloud computing is being used: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), or Software as a Service (SaaS). (For definitions and explanations of these terms, please refer to the NIST cloud computing definition document, used as one of this course's resources). In each model the cloud provider takes on different levels of responsibility for the total application, and thereby affects where the trust boundary is located.

To better understand the trust boundary of the various models, and for guidance on how to ensure security for your cloud applications, see Figure 2.

Security Responsibility

		IaaS	Paas	Saas
Application	Responsibility	User	User	Provider
	Action	Best practices And certification	Best practices And certification	Evaluation and certification
Middleware	Responsibility	User	Provider	Provider
	Action	Best practices And certification	Evaluation and certification	Evaluation and certification
Infra-structure	Responsibility	Provider	Provider	Provider
	Action	Evaluation and certification	Evaluation and certification	Evaluation and certification

Figure 2

Figure 2 is a chart of security responsibilities for each of the three cloud delivery models, along three key areas of responsibility: infrastructure, operating system and middleware, and application.

- **Infrastructure:** This area addresses both physical security and software security. Physical security refers both to the physical infrastructure of the cloud computing environment (i.e. the data center itself) and the security practices surrounding the physical infrastructure. For example, this area covers whether the data center has redundant Internet access methods, as well as what practices are in place regarding access to the physical facility (such as requiring both identification documents and biometric scanning as prerequisites for entering the data center facility). Physical infrastructure also refers to the hardware within the data center such as backup generators, etc. Finally, infrastructure also refers to the software infrastructure used to implement the cloud computing environment. Most cloud computing (though not all) uses virtualization as a foundation for the cloud computing environment, so this security area would cover the virtualization hypervisor, including security practices related to controlling access to logging in to administer the virtualization and ensuring that proper security patches are installed.

- **Operating System and Middleware:** This area refers to software components that provide the operating environment within which the application runs. Key security issues for this area of responsibility include whether appropriate security software is installed within the OS, patch installation practices, administrative access to manage these components, and so on.
- **Application:** This area refers to the software used to provide the actual functionality of the application itself. Falling under this area are topics such as software component version verification, patch installation practices, application identity management, etc.

In Figure 2, you will see a thick black zigzag line. This indicates where the trust boundary lies for each cloud delivery model. As an example, in a PaaS environment, the cloud provider is responsible for the security of the infrastructure and the middleware, while the cloud user retains responsibility for the security of the application itself. As you can see from the descriptions, this means that the cloud user would need to audit and evaluate whether the security measures of the provider in its areas of responsibility are sufficient.

One final point to be made about the trust boundary: every cloud provider offers a security framework into which users integrate their application. As every provider has a somewhat different framework, it is crucial that users to understand the framework and ensure that they integrate with it properly. Without understanding the security framework presented by the cloud provider, it's likely that the cloud user will fail to configure its usage properly and thereby leave security vulnerabilities that may be exploited by attackers.

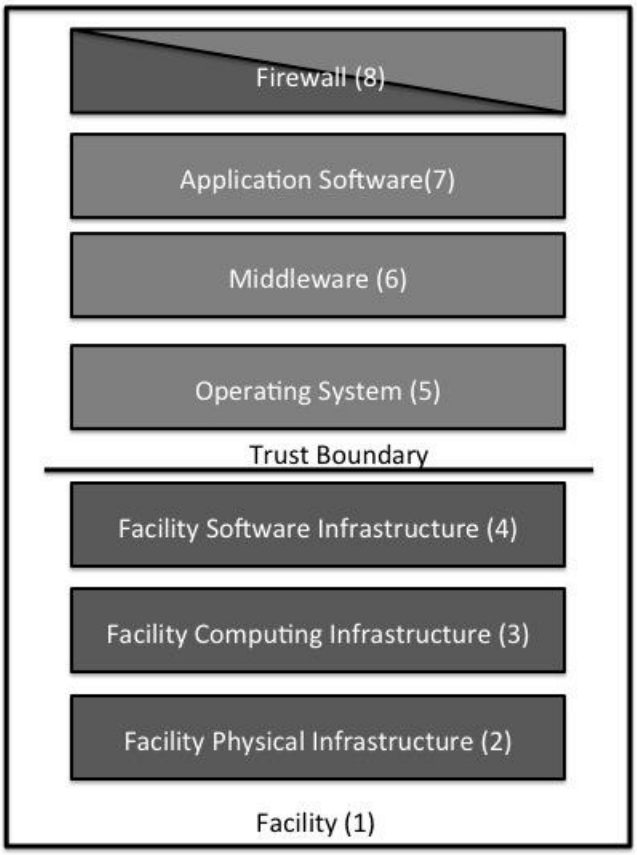

Figure 3

Let's examine the trust boundary and the elements contained within the overall computing environment in more detail. Figure 3 is a diagram containing all major elements that are required to deliver application functionality in an IaaS environment.

The figure includes the computing facility, the internal data center infrastructure (i.e. those components necessary to operate the facility, such as fire suppression and UPS), the hardware infrastructure and network, and then the components of individual servers: computer, hypervisor, operating system, middleware, and application software.

At the top of the figure is the perimeter of the facility, depicted as the firewall (Item (8) in Figure 3) which represents all measures to filter traffic from the Internet and ensure that applications only receive appropriate network traffic.

By looking at this diagram, one can see that there are many security elements present in a cloud computing environment. These elements are comprised of:

- Those delivered by the cloud service provider, which maintains responsibility for the element;
- Those delivered by the user, who is responsible for their implementation and management;
- Those delivered by the cloud service provider but configured by the user. For example, many cloud providers present a firewall to each user account to manage traffic in and out of the applications installed by the user. The user is responsible for ensuring that the firewall passes only appropriate traffic into virtual machines instances within the account's environment. This class of elements can be the most challenging for users to understand, as many users will assume that correct operation of the element (e.g., firewall) is the responsibility of the cloud provider, failing to recognize that it is the user's responsibility to ensure the element operates properly. Continuing the firewall example, many users will assume the cloud provider must make the firewall operate properly, since the provider delivers the firewall, rather than recognizing that configuration (and responsibility for proper operation) must remain with the user – after all, how can the cloud provider know what ports must be open for the user's application?

Let's examine the entire stack given in Figure 3 and assign security responsibility for each layer.

1.2.1. Facility: Cloud Provider Responsibility

Item (1) in Figure 3: The facility is the physical premises of the cloud infrastructure and is always the responsibility of the cloud provider. This entails ensuring that the facility is securely closed, and that no-one enters without being identified. A common practice for ensuring facility security is the provision of guards, who match personal identification with visual examination. Other measures used to control access may include biometric identification (e.g., retina scans), which is ordinarily used in addition to physical identification by a human guard. Perhaps the most extreme version of facility security is that practiced by SuperNAP, which has ex-military armed guards on security duty.

1.2.2. Facility Physical Infrastructure: Cloud Provider Responsibility

Item (2) in Figure 3: Facility physical infrastructure refers to the mechanical devices that are housed within a facility and provide services necessary to operate a cloud infrastructure. The reliability of these devices, along with their management, may have security implications. Common devices include:

- **Power input and distribution**. Obviously, a reliable power supply is critical for running a data center. Usually, a high-quality source of power with a good record of availability will be used to provide this. In addition, within the facility,

electricity will be distributed by power equipment. Increased investment in power distribution will allow more protection against power outages caused by the use of redundant components.

- **Generators and Uninterruptible Power Supply (UPS)**. Despite the use of a reliable electricity supplier, sometimes power will be cut off. Data centers ordinarily have on-site protection against power losses. Most high-quality data centers have on-premise generators that can be used to supply power in the event of an electricity provider outage. In addition, most highly-available data centers will have battery backup to supply power for the period between electricity cut-off and the on-premise generators getting up to speed and delivering power.
- **Air conditioning and distribution**. Servers throw off significant amounts of heat, and most data centers require air conditioning to keep temperatures within operating tolerances.
- **Infrastructure partitioning and security**. Many cloud data centers will have sections separated from one another for security purposes. These security mechanisms often take the form of separate rooms or cages, each of which has its own access control. Another new form of partitioning is "containerization," a form of data center deployment in which computing devices and racks (see below) are placed into a container. This may be either an actual shipping container such as those used to ship goods throughout the world, or a custom-designed container that maintains the same general rectangular dimensions but is built to better support computing form factors by, say, allowing back-to-back racking. The advantage of containers is that they allow easier deployment as there is no need to construct an entire building; all that is needed is a level location to install the container. Some companies have even installed computing containers outside in secure parking lots.
- **Racks and/or Cabinets**. Every data center has some kind of physical structures that contain computing devices. These may be racks or a cabinet (cabinets often have protective doors, while racks expose the computing devices). Surprisingly, significant change and innovation is going on in this area, with different form factors supporting better installation of blade or pizza box servers, and some even supporting vertical installation of standard PC motherboards.

1.2.3. Facility Computing Infrastructure: Cloud Provider Responsibility

Item (3) in Figure 3: Facility computing infrastructure refers to the computing infrastructure necessary for the facility to operate as a data center. This infrastructure includes:

- **External network and Internet connectivity**. For high availability, cloud infrastructure will ordinarily have two Internet connections, each from a different provider. In addition, good design practice dictates that each connection comes into the building from a different side, thereby precluding a single backhoe excavation cutting both connections. Besides Internet connectivity,

the facility may have additional network connections that provide direct, private connectivity between the facility and specific locations or customers. Both types of connectivity are crucial to provide highly available computing, which reduces risk.
- **Internal network infrastructure**. Network traffic from individual computing devices runs across a shared network infrastructure. The performance and security of this shared infrastructure is the responsibility of the facility provider.
- **Storage devices**. All storage within a cloud provider's facility is shared and responsibility for the storage infrastructure lies with the cloud provider.

In addition to the general infrastructure equipment, the facility computing infrastructure also includes the equipment used to perform computing services. According to the NIST definition of cloud computing, these devices serve as a pooled resource, shared among all users, with application workloads migrated throughout the shared pool as necessary. This infrastructure includes:

- **Servers**. These are the devices that perform computing; one may hear them referred to a performing processing. Ordinarily in a cloud computing environment, a server supports a number of virtual machines which are commonly being used by different users. Stated another way, servers are a shared resource used by whatever virtual machines the cloud management software (see below) assigns to them. The cloud provider is responsible for the servers in a cloud environment.
- **Storage.** Each virtual machine uses storage upon which it places its data. While the data is the responsibility of the user, the storage itself is the responsibility of the cloud provider. This storage usually resides in the storage infrastructure, although some cloud providers will use locally attached disk drives to store user data.
- **Network connectivity**. Each server has one or more Network Interface Cards (NIC), which transmit and receive data to and from the network infrastructure.

1.2.4. Computing Software Infrastructure: Cloud Provider Responsibility

Item (4) in Figure 3: Computing software infrastructure refers to the software that is used to operate the cloud computing functionality that supports applications placed in the infrastructure. This infrastructure includes:

- **Hypervisors**. Hypervisors are a software layer that resides between physical servers and the virtual machines running on those servers. Hypervisors form the basis of virtualization, which in turn forms the basis of cloud computing. The cloud provider's hypervisor responsibility includes initial installation, configuration, subsequent upgrades, installing patches, and perhaps code customization.
- **Cloud orchestration software**. Cloud computing represents virtualization married to automation, which allows much faster provisioning times. Cloud computing

automation is accomplished via orchestration software, which translates a high-level command (e.g., create a new server instance) into the individual tasks necessary to accomplish that command (e.g., create new virtual machine, attach two terabytes of storage, assign a network address, etc.) and ensures they are accomplished as a single transaction. It is no exaggeration to say that orchestration represents the single most important operation in a working cloud computing environment.

- **API software**. Cloud providers offer access to their services via service interfaces. While often referred to as API (application programming interface), these may not be a programming interface but rather an online service that is accessed via the Internet using a common protocol like HTTP. For most cloud providers, these online interfaces are the foundation of interacting with their service, and their availability and reliability are paramount in terms of keeping the service up and running.
- **Portal software**. Most cloud providers also offer an online browser interface that people use to interact with the functionality of the provider's cloud infrastructure. This interface will have the ability to start, stop, and suspend virtual machines, assign IP addresses to specific virtual machines, assign load balancers, etc. The best practice for cloud providers is that their portal software will use the service's API, becoming, in effect, an equal partner with all other programs accessing the service's functionality. While many users will choose to use an external tool to manage their cloud resources (or indeed, will have their application directly interact with the CSP environment to obtain and release resources), many others will use the CSP's own portal offering.
- **System services**. Many cloud providers offer software components or services as part of the cloud environment. For example, Amazon Web Services offers a message queue service for applications to use. These services can simplify application development, as the application creator does not need to take responsibility for these system services. Indeed, judicious leveraging of these services allows the application creator to focus on the unique functionality of the application and avoid effort devoted to lower-value "plumbing."
- **Monitoring and management software**. A CSP environment is a large and complex collection of computing resources – many servers, storage hardware, network devices, many software components, etc. – that must be monitored to help ensure optimum uptime and performance.
- **Audit and staff monitoring software**. The previous items in this list pertain to the functionality of the cloud environment and each is necessary for its successful use. This final item pertains to monitoring the people who work for the CSP and interact with the cloud environment. As the Roman poet Juvenal put it: Quis custodiet ipsos custodes? Loosely translated, this equates to "who will guard the guardians?" End users depend upon the cloud provider to protect their applications from access or penetration by other entities. However, end users also have questions about protecting their applications from inappropriate access by the cloud service provider's employees. Consequently, many cloud

providers implement monitoring software that tracks interactions with key system components to ensure that only authorized personnel are accessing them. By reviewing the tracking software records, the cloud provider can be sure that no unauthorized activity is taking place in its cloud computing environment.

In IaaS environments, the hypervisor represents the demarcation of the trust boundary. The cloud provider is responsible for "the hypervisor and everything below it," while the user is responsible for elements above the hypervisor. The cloud computing components for which the user is responsible are examined next.

It's critical to remember that one's application is running in someone else's computing environment, and therefore practices that were acceptable in one's own environment are no longer applicable or appropriate.

On the other hand, from the user's perspective, the management task is far simpler, since so many elements of the complete environment are another person's responsibility – the list of items described above shows how much work is offloaded to the cloud provider.

So, in an IaaS environment provided by an external party, here are the portions of the overall system the user is responsible for.

1.2.5. Operating system: User Responsibility

Item (5) in Figure 3: The operating system forms the foundation for the user's application. From an application execution perspective, there is no difference between an operating system running in one's own data center on a physical server and an operating system running in a cloud service provider's environment.

The operating system is a set of software services that enables user code to operate. Common operating system services include process launch, process scheduling, and file system storage. A number of other system services that enable application code execution also are typically included in the operating system, for example, logging services that store entries reflecting system events. The aggregation of this software is referred to as the operating system.

In a cloud computing environment, security of the operating system is the responsibility of the user. This means that all responsibility for upgrades, patches, configuration, and operations falls upon the user. To a certain degree, this is common sense. The user decides what operating system to use, installs the software, and is responsible for running the application so of course they are responsible for managing the application and responsible for the security of the operating system.

However, many application groups are accustomed to receiving preconfigured virtual machines from a central operations group, with subsequent management being the responsibility of operations.

In a cloud environment, the applications group may implement an application with no interaction with an operations group and no subsequent involvement or support by that group. Or, the application implementer may mistakenly believe that the cloud provider will take responsibility for operating system management. This could result in no-one ensuring that the operating system is upgraded, patched, and managed appropriately.

To confirm, in a public IaaS environment, security of the operating system falls to the user.

1.2.6. Middleware: User Responsibility

Item (6) in Figure 3: Middleware is a vague term, combining the latter part of the word "software" with the location identifier "middle." It was originally coined to represent software installed in an operating system that provides services to applications, but is not part of the application itself. Examples include database software, message queues, application servers or frameworks like Spring, and caching software.

As with the operating system, this software is under the control of the cloud user, and responsibility for its security also lies with the user, not the cloud provider.

1.2.7. Application code: User Responsibility

Item (7) in Figure 3: Application code refers to the software components that provide the actual functionality of the application. This code may take the form of java components, web pages, or standalone executable binaries. Responsibility for the security of application code resides with the user of the cloud service, not the provider.

1.2.8. Other Application Components: User Responsibility

Item (7) in Figure 3: In a complex application topology, components of the application may reside in other virtual machines. Common examples of these components are load balancers, caching software like memcached, and billing software used to charge the ultimate end users of the application for its use. These components provide application-level functionality, in contrast to middleware which provides operating system-level functionality. The cloud user may install these components as software packages, or they may be provided by the component creator in a preconfigured virtual machine template.

Regardless of whether the user installs the component themselves or launches it from a creator-created template, responsibility for its security resides with the user, unless the provider explicitly provides the service and offers support for it.

This implies that the end user is ultimately responsible for the security of these components, despite the fact that the code may be created and released by another company (in fact, the same thing is true for middleware components sourced from a third party). This further implies that an organization that uses third party components in its application must have some mechanism to manage necessary security issues in these components. The mechanism may take the form of having possession of the component source code (as in the case of an open source product) or having a support contract from the component supplier.

No matter how the component is obtained, responsibility for its security lies with the user; no provider will assume responsibility for security of these components. The next section on Templates contains additional discussion about the use of externally sourced components.

1.2.9. Templates: User Responsibility

Item (7) in Figure 3: As the operating system is running in a virtualized environment, there can be significant differences regarding the provenance of the operating systems used by the application. Virtualization supports the concept of a template – in effect, a base image that may be leveraged by a user to form the foundation of a new operating system to be executed in the virtualized environment. This is somewhat analogous to using a standardized form in a word processing program which then has document-specific information filled in for a particular use.

This template concept is extended in cloud computing. Most cloud providers offer the ability to create a template so that a standardized operating system build is used by, say, every application group in an organization. This ensures consistency among all the organization's applications, which is useful for operational simplicity. It also reduces work for people creating new applications: rather than having to create an application environment from scratch, developers can select an appropriate template, clone it, and immediately begin productive work on their project. The ability to create templates to be shared within an organization is often called "private templates," as their use is limited to a single organization.

Many cloud providers also offer the ability for templates to be more widely shared in a format referred to as public templates. In this variant, templates are made available to be used by any user of the cloud environment. The providers of these templates are commonly (but not exclusively) software vendors, who pre-populate the template with their installed and configured product. Using a software vendor's template offers the

same kind of operational efficiency alluded to above; from the vendor's perspective, the practice induces trial, which may lead to increased sales.

The crucial issue relating to public templates is that such a template may be deliberately or inadvertently compromised with viruses, malware, etc. Moreover, if the template is poorly configured, security breaches may occur.

As there is a possibility that a public template might be compromised, one might assume that organizations would avoid their use. Unfortunately, that assumption may not be correct. Someone may have begun using a public template as the basis for an application without realizing the security exposure it presents. Or someone may have begun using a public template with the intention of replacing its use later, but neglected to do so.

An important concern for any organization using an external provider is to ensure that any templates used are appropriate and secure. As noted above, in a public cloud computing environment, security above the hypervisor is the user's responsibility, not the provider's.

1.2.10. Licensing: User Responsibility

Item (7) in Figure 3: One last item tangentially related to security, but extremely important to applications, is software licensing. Software licenses govern the use of software components, and complying with their conditions is an extremely important aspect of IT operations. The issue of software licensing in cloud computing environments can be quite complex. Nevertheless, it is clear that whoever is responsible for the provision of the resource that contains the component is also responsible for complying with the licensing conditions of software components.

So, for example, the cloud provider is responsible for complying with the licensing conditions of the hypervisor software, while the cloud user is responsible for complying with the licensing conditions of user-installed middleware.

This rather neat division of licensing responsibility breaks down, however, when the cloud provider delivers the software component. For example, many cloud providers offer templates of virtual machines that have a Windows operating system pre-installed; the user pays an hourly or monthly fee for use of the Windows operating system within its application topology, while the cloud provider pays Microsoft a licensing fee for the use of the operating system by the user. In a situation such as this, compliance with the software licence is the responsibility of the cloud provider. In other words, while the user pays a fee to use the Windows operating systems, it is the cloud provider's responsibility to comply with all requirements and restrictions imposed by Microsoft relating to deployment of its operating system in the cloud provider's environment.

2. Evaluating a Provider's Security

As discussed above, cloud computing security is different from traditional approaches to computing security. Among the reasons previously described:

- **Virtualization**: Many traditional security solutions rely on examining network traffic. In virtualized environments, network traffic often goes from one virtual machine to another without leaving the physical server, rendering network-attached security devices ineffective.
- **Dynamic environments**: Virtualization environments support dynamic placement and relocation of virtual machines to enable hardware failure resiliency and better application performance. The side effect of this is that security practices that assume a static environment are challenged to operate effectively in a dynamic infrastructure.
- **Multi-tenant environments**: Many security products and practices were designed for environments that are controlled by a single entity. In such environments, examining all network packets or performing port scanning is perfectly acceptable. In a multi-tenant cloud environment, such approaches are often a violation of the provider's terms of service.

In cloud computing environments, the reality is that security is no longer the duty of a single entity. Instead, security is a shared responsibility: part of the responsibility lies with the cloud provider, and part lies with the cloud user. The question is, how does one assess the security of those portions of the cloud computing environment that lie within the responsibility area of the provider?

2.1. Assessing cloud security

One approach, of course, would be to rely on the word of the provider. The provider would assert that everything in the environment is fine, security-wise, and the user would accept the assertion as definitive.

The problem with that approach is that, while the user relies on the provider to implement security correctly, nearly all of the risk for any security issue falls upon the user of the service. This is made quite clear in Figure 1: while providers and users share responsibility for security and compliance, risk associated with failure lies primarily with the user.

In some sense, it may seem unfair that cloud users bear risk responsibility for infrastructure elements out of their control, but that is the reality of how most cloud service contracts are drawn up. However, just to reinforce a point made earlier, this asymmetric risk responsibility is not unique to cloud computing. Nearly every type of

technology outsourcing contract is carefully drawn up by the provider to limit its risk exposure. In fact, one might infer that these contracts are deliberately drawn up to minimize the risk providers face and to shield themselves from any consequences that might arise from their actions.

Given this state of affairs, fair or not, users must recognize that the quality of the provider's security is important to them, and measures beyond blind faith are required. Failing to implement measures designed to minimize security or compliance issues located within the provider's area of responsibility raises the likelihood that the user will increase its overall risk.

2.2. Peeking below the Trust Boundary

The Trust Boundary represents the interface that identifies where a user's responsibility for direct security implementation ends and where the provider's begins. Figure 2 identifies the actions below the Trust Boundary as "Evaluation and Certification."

What this means is that, in the absence of the ability to perform direct security actions (e.g., place a traffic sniffer on the internal data center network), one must evaluate the provider's measures for securing the areas of its responsibility.

On its face, evaluation is a very clear concept: one observes the actions taken by another party and considers whether they are sufficient to implement one's objectives. This practice is present in every situation in which one relies on another party to perform activities. For example, when one calls a plumber to fix a leaking pipe, one evaluates whether he has fixed the pipe properly before paying him.

2.3. The Challenge of Evaluation

Evaluation, however, can be a complex task. Even for a situation which appears straightforward, evaluation can sometimes be very difficult. In the example above, while it might seem trivial to assess whether a plumber fits a pipe properly, this task carries with it complexities such as:

- **Was the task performed correctly?** It might not leak now, but did the plumber use techniques that will prevent leaks in the future? Did he use sealant that will stand up over time, or did he use cheap sealant that will break down quickly?
- **Did he follow all applicable rules and regulations?** Even for something as simple as a pipe (basically, a tube that allows a fluid to be sent from one place to another – what could be simpler?) there can be a number of regulations that are applicable – size, pipe composition, bracing (if necessary), etc.

- **Does he have all necessary licences and certifications?** While many deride the over-regulation that occurs in many domains, licenses and certifications are a fact of life. And the lack of a plumber's necessary licensing might become important in the event of an insurance claim or lawsuit.

Consequently, evaluation can be a very important and difficult proposition, and many organizations (and homeowners, in the case of plumbing!) find it difficult to conduct an appropriate evaluation. Some of the reasons are:

- **Lack of domain expertise**: While it's easy to discern whether a pipe is leaking after the plumber concludes his work, a homeowner may not know all the applicable laws and regulations. For more complex matters like cloud computing, organizations may not have the knowledge base to discern whether a cloud provider is doing a good job or not.
- **Applying expertise to a new domain**: Even detailed knowledge about something fairly similar to a given domain may be insufficient for a new domain. Knowing a lot about water-oriented plumbing may not help in knowing enough about natural gas plumbing. In the case of cloud computing, knowing a lot about running a data center may not be sufficient to understand compliance needs and best practices for operating a multi-tenant data center environment.
- **Creating time to conduct a thorough evaluation**: Everyone is busy, and IT organizations have been systematically squeezed of headcount. Most would find it difficult to divert staff to conduct a thorough cloud security evaluation, just as most of us would find it difficult to find enough time, given our hectic lives, to fully assess the quality of a plumber's work.
- **Getting attention from the vendor**: How can you get a rich enough interaction to ensure you obtain sufficient information to make the necessary evaluation? It's an unfortunate fact of life that cloud providers pay attention to their largest customers and minimize interaction with customers who represent small revenue opportunities – even if a thorough assessment is critical for that small opportunity customer. To be fair to cloud providers, it's challenging for them to deal with an environment in which every customer or potential customer wants to conduct an evaluation. Every evaluation requires time and attention from provider personnel, and most of the evaluations will, ultimately, be nearly identical in 90% of the aspects they assess. Clearly, this repetitive evaluation process is inefficient. There should be a better way. Fortunately, there is.

2.4. The Role of Certification

Many industries confront the evaluation problems outlined above. How can buyers assess the quality of industry vendors, given the complexities and time constraints present in all businesses? How can vendors reduce the time devoted to evaluation in an environment in which many buyers want to perform similar evaluation processes?

The solution for most industries is to develop a set of recommended best practices promulgated by an impartial trade association, usually comprised of vendor and user representatives. The recommended best practices are typically characterized as a standard, indicating that all vendors that wish to be recognized as high quality must meet the requirements laid out in the standard.

Since IT systems often have financial implications due to their use in company operations or as the system of record for financial transactions, the accounting industry often participates in these association efforts.

Alternatively, the accounting industry may itself develop and promulgate the recommended practices. In doing so, it would ordinarily create a body tasked with developing the standard. This body would include representatives from all interested parties, including the accounting industry, vendors, users, and perhaps government agencies and regulatory bodies.

Once a standard is in place, it simplifies the evaluation process. Instead of each participant in an industry developing its own criteria, everyone can use the standard as the basis for evaluation.

Of course, having a common set of criteria doesn't solve the problem of each user conducting its own evaluation. If each potential customer seeks to conduct its own evaluation against the criteria, there is clearly an issue with redundancy of effort, which is inefficient and expensive. Moreover, the existence of a standard doesn't necessarily solve the problem of lack of user expertise – just because criteria have been established doesn't mean that users have the knowledge base to understand the criteria or the judgment to assess a vendor's compliance with the standard.

2.5. Certifications, Audits, and the Role of a Security Auditor

The standards process has taken the further step of establishing an audit process, designed to standardize the evaluation. A vendor can undergo an audit, whereby an external party will assess the vendor's compliance with the relevant standard, and pass judgment as to the level of compliance present in the vendor's environment.

The audit process simplifies things enormously. Instead of repetitive evaluations conducted by multiple customers, the vendor can undergo one evaluation process, at the end of which it receives a certificate of compliance with the standard. Any subsequent customer who wishes to establish the vendor's compliance with the standard can accept the certification as proof that the vendor meets the standard's requirements.

Of course, this raises the question: who conducts such audits and how can that organization be trusted to do a good job?

The answer, naturally enough, is that the body that promulgated the standard certifies an auditing organization to serve as an approved security auditor. The security auditor goes through a process designed to ensure that it has the expertise and thoroughness to fully evaluate a vendor's compliance with a security standard. At the end of the process, the auditor is certified as being capable of performing audits against the standard and, crucially, is able to provide the vendor being evaluated with a certificate of standard compliance.

A security auditor is a separate, impartial entity certified to perform an audit of an organization to assess whether it meets the criteria established by a standards body. The auditor performs the audit and evaluates whether the organization being audited meets the criteria set out by the security standard.

Security auditors are typically accounting, consulting, or system integration companies with significant security expertise. In addition to security auditing, these firms commonly consult with organizations to recommend measures to improve their security; in fact, many auditing firms will identify areas within an organization that need to be improved to enable them to meet the necessary conditions associated with a security audit. Of course, having worked with an organization in a consulting capacity typically precludes a security auditing company from performing a subsequent audit, as that would represent a conflict of interest.

2.6. How Certifications Work

Certification refers to a process whereby a provider has an external auditor examine a given domain and evaluate it against a formal set of criteria. These criteria are typically created by an independent industry body, thereby assuring objectivity and impartiality in the criteria definition.

The external auditor gains access to the provider's infrastructure and assesses its practices according to the criteria defined in the particular certification domain. After the audit is complete, if the provider's practices meet the requirements of the audit, the auditor issues a report detailing the provider's compliance with the certification criteria. The shorthand phrase for this process is that the provider is certified against a particular audit process.

Once a provider has undergone the audit process and received a certification, it can present itself as certified for a particular domain. In turn, customers or potential customers of that provider can accept the certification in place of performing their own audit.

As can be seen, the certification process neatly addresses the issues regarding user evaluation identified earlier in this chapter. Rather than attempting to create and perform its own audit, a customer or potential customer can identify the key audit/certification domains relevant to its needs and request those from service providers.

Many companies considering using an external provider go even further: they insist that providers, in order to be considered as a potential choice, must have undergone specific audit processes and offer proof of successfully achieving certification.

As one can easily imagine, the ability to rely on impartial audit requirements and associated certificates can simplify the evaluation process enormously.

Common certification processes applicable to cloud computing include:

- **COBIT (IT Governance and Control)**. This set of requirements relates to how IT organizations are managed, including how internal processes are defined and enforced.
- **HIPAA (Health Insurance Portability and Accountability Act)**. This Act sets out a number of requirements relating to information privacy. As suggested by its name, HIPAA is focused specifically on health care and affects many different types of organizations involved in the medical field.
- **SP 800-53**. This NIST (National Institute of Standards and Testing) Special Publication focuses on security controls for IT environments. NIST published this as a guide to help government organizations prepare to undergo audits performed under the Federal Information Security Management Act (FISMA). While the titles of the organization and the Act suggest this affects only government bodies, service providers that seek to host federal government applications need to meet FISMA requirements, so this standard is applicable to CSPs as well as government agencies.
- **FedRAMP**. The Federal Risk and Authorization Management Program is an initiative sponsored by the federal government to make the audit process easier for government agencies and CSPs. The FISMA standard outlined above requires every application to undergo a security audit, part of which assesses the infrastructure upon which the application resides. Obviously, if a large number of applications need to be audited in a process that is largely repetitive with respect to the infrastructure, this is inefficient and costly. FedRAMP is an effort to streamline this process and enable a single audit process to be used by multiple FISMA efforts.
- **PCI-DSS**. The Payment Card Industry Data Security Standard is a standard associated with taking electronic payments. Obviously, there are important requirements in the area of security and privacy with respect to financial transactions, and PCI sets out the requirements in this domain.
- **ISO 27001 (full title: ISO/IEC 27001:2005 - Information technology -- Security techniques -- Information security management systems – Requirements)**. This

standard, published by the International Standards Organization, sets out security controls designed to ensure that IT organizations have a defined and consistent approach to information security.

- **BITS**: BITS is part of the Financial Institution Shared Assessments Program and is designed to help assess the security controls of IT service providers. From the name of the sponsoring organization, it is obvious that this is oriented toward the financial services industry. This initiative is analogous to the FISMA and FedRAMP audit standards described above.
- **SAS 70 (full title: Statement on Auditing Standards (SAS) No. 70, Service Organizations)** Please note that SAS 70 was superseded by SSAE 16 (full title: Reporting on Controls at a Service Organization) in Mid-2011 and will be the appropriate certification going forward. SAS 70 and its successor provide an assessment of the processes a provider follows in delivering a service. This assessment ensures that processes are defined explicitly and that the provider has measures in place to ensure that they are consistently executed.
- **GAPP**: The Generally Accepted Privacy Principles are a set of measures formulated by the American Institute of CPAs (AICPA) directed toward IT privacy practices and policies. As can be seen from its title, GAPP is directed toward some of the same objectives as HIPAA.

2.7. Dealing with multiple compliance standards

There is an old joke in the IT industry that standards are great and that's why there are so many of them. The joke illustrates what must be obvious from the list in the last section: there are lots of compliance requirements, some generally applicable and others focused on specific industries. Obviously, there must be significant overlap among the different compliance regimes.

Consequently, while in the abstract the existence of audit requirements and certifications would appear to make life simple for IT users, it would be more appropriate to say that it makes life *simpler*.

While audits and certifications may address the detailed specifics of security and privacy, which is certainly a huge benefit, they still present an IT organization with the need to assess which compliance standards are appropriate for its specific environment.

Moreover, if multiple compliance requirements are relevant for the IT organization, the obvious question is how to ensure that every element of each requirement is addressed without having to repetitively run through each audit standard.

Finally, even if an IT organization is able to understand how to manage each of the audit standards appropriate to its specific situation, it is still faced with the need to understand how to apply them to a CSP's cloud offering.

2.7.1. The Cloud Security Alliance

Fortunately, companies seeking to sort out cloud computing security and compliance don't have to fight the battle alone. The Cloud Security Alliance (typically referred to as CSA) is an organization that provides a locus of research, expertise, and recommendations in the area of cloud security. The CSA is located at https://cloudsecurityalliance.org/.

Founded in 2008, the CSA is a "not-for-profit organization with a mission to promote the use of best practices for providing security assurance within Cloud Computing, and to provide education on the uses of Cloud Computing to help security of all other forms of computing."

The CSA is the center of gravity with respect to the topic of cloud computing security. It comprises end users, vendors, and service providers, all of whom interact in the interest of defining what cloud computing security requires and how to achieve it.

CSA has an international scope, and has both national and local chapters. Anyone seeking to understand the nuances of cloud security and to learn current best practices should consider getting involved in the CSA. It sponsors a yearly cloud security conference in November; if you are someone seeking deeper involvement in cloud computing security, this conference is a must-attend event.

2.7.2. Leveraging the CSA

In addition to its overall security focus, the CSA has published research documents focused on easing the audit and compliance complexity faced by cloud users.

As a general overview of how security and compliance relate to cloud computing, the CSA published the *Guidance for Critical Areas of Focus in Cloud Computing*. This document discusses how cloud computing relates to 13 different security and compliance areas, including Governance and Risk Management, Interoperability and Portability, and Data Center Operations, to name but three. It provides an excellent overview and jumping-off point to better understand the landscape of cloud computing security.

With respect to specific aspects of assessing the security of a CSP, two further CSA documents are highly valuable for IT organizations. The documents also provide guidance on how to map the most relevant audit standards to one another.

The first document is the Consensus Assessments Initiative Questionnaire (CAI), published in late 2010. This questionnaire "provides a set of questions a cloud consumer and cloud auditor may wish to ask of a cloud provider." It provides a series of 'yes or no' control assertion questions which can then be tailored to suit each unique cloud customer's evidentiary requirements.

The second document is the CSA Cloud Controls Matrix (CCM), which provides a "controls framework that gives detailed understanding of security concepts and principles that are aligned to the Cloud Security Alliance guidance in 13 domains." The CCM was first published in April 2010, and has subsequently been updated, with the latest version released in August 2011.

All three of these documents may be found at https://cloudsecurityalliance.org/research/.

2.8. Overview of the CAI and CCM

The purpose of both the CAI and CCM is to provide guidance and recommendations in a number of specific areas for IT organizations wishing to audit CSPs and assess their security practices.

The two documents share a common taxonomy of security aspects, which makes it easy to cross-correlate the security requirements (the controls) with assessing how well the CSP implements them (the questionnaire).

The security aspects of the two documents are as follows:

- Compliance
- Data Governance
- Facility Security
- Human Resources
- Information Security
- Legal
- Operations Management
- Risk Management
- Release Management
- Resiliency
- Security Architecture

Each of these areas has one or more individual items associated with it. For example, Release Management has five separate items:

- New Development/Acquisition
- Production Changes
- Quality Testing
- Outsourced Development
- Unauthorized Software Installations

The CCM outlines the control requirements for each item in the Matrix. These requirements are assertions about how a CSP should comply with the individual item.

The CAI contains the same areas and items, but lists a set of questions to be posed to a CSP for each item to assist in determining whether the CSP meets the requirements outlined in the CCM. This list of questions helps IT organizations formulate an organization-specific audit – they provide the foundation for a complete set of questions that will reflect the organization's audit and compliance requirements.

A number of compliance standards relevant to cloud computing were described earlier in this white paper. As noted, some of them overlap. Furthermore, for an individual IT organization, figuring out which parts of each standard are relevant and how they relate to the CSA's security areas and items would seem challenging, to say the least.

Fortunately, the CSA has made this challenge much simpler. Both the CAI and CCM contain columns for each security standard. For every security item outlined in the documents, the relevant portions of each standard are given. For example, for the item Ownership/Stewardship within the Data Governance area, the COBIT column notes as relevant areas COBIT 4.1 DS5.1, and PO 2.3.

The benefit this set of columns provides for IT organizations is immense. Rather than having to work through each relevant compliance standard and assess which part of each standard is appropriate to different audit areas, the organization can rely on the list the CSA has put together. This simplifies the audit task enormously. Of course, it is still necessary to interact with individual CSPs and work through the audit process, but an evaluation framework makes the audit process easier.

The CSA and its CAI and CCM provide an excellent set of resources for the very important security issues associated with using a CSP. The CAI and CCM are, as just noted, a foundation for the audit process itself. Naturally, individual organizations may find it necessary to identify additional audit items or may be subject to additional compliance standard requirements, but the two documents offer an excellent starting point.

2.9. Mapping the CAI and CCM to the Security Stack

For many people, particularly those for whom security is not a primary focus or strength, the organization of the CAI and CCM may be problematic. While the taxonomy areas of the documents may make sense to security professionals, IT personnel in the applications and infrastructure/operations areas may be more comfortable evaluating security in terms of a security "stack," much like the one presented in Figure 3 above.

This poses a problem since the items within the CAI and CCM are organized topically (i.e. Information Governance) and not according to where each item falls within the architecture stack. For IT organizations that assign security responsibility according to the part of the stack in which the security item resides, a mapping of the CSA areas and items to the stack is useful.

Table 1 lists each part of the security stack and identifies which item(s) from the CAI and CCM are associated with it. This structure will aid IT organizations as they work through audit processes and assign individuals to participate in the process.

Table 1 • • •

Security Layer	Consensus Assessment Initiative Section(s)
General Security Policy	Compliance 01-08
	Legal 01-02
	Human Resources 01-03
	Information Security 01-16, 22-27, 32
	Operations Management 01
	Release Management 01
	Resiliency 01-02
	Risk Management 01-05
	Security 01
Facility and Facility Computing Infrastructure	Facility Security 01-08
	Resiliency 03-08
Facility Computing Infrastructure	Data Governance 01-08
	Information Security 04, 09-10, 18-19, 23, 28-32
	Operations Management 02-03, 05
	Risk Management 05
	Release Management 01
	Security 03, 05, 10-13,
Facility Software Infrastructure	Information Security 04, 07-10, 17 20-21, 29, 32-34
	Release Management 02-05
	Security 02, 04, 06-09, 14-15
	Operations Management 04

3. The Legal Role of a Cloud Provider as Third Party

Cloud computing can introduce a number of legal challenges for your organization.

When a company utilizes a cloud provider to store its data, it may very well find that different laws and regulations apply than if the company were storing the data itself within its own data centers. This is because the cloud provider is acting as a third party, i.e. an entity separate from the owner of the data itself.

Furthermore, cloud providers commonly include contractual conditions that limit their liability for data breaches, system downtime, legal compliance, and intellectual property responsibilities.

Legal guidance regarding this issue cannot be provided in this white paper, but companies considering using a third party cloud provider should be aware that there may be different rights and responsibilities regarding the data and should be prepared to take legal advice to determine how best to proceed.

It is critical that cloud computing users recognize and understand that cloud providers operate as a third party and therefore have different responsibilities from the users themselves. Users must be aware of how this may affect their legal and regulatory responsibilities.

3.1. Geography: How Location Affects Legal Responsibility

The challenges outlined in (3) above are compounded when considered in the context of different nations or regions in the world. Legal and regulatory responsibilities vary widely across the world; acceptable or common practice in one nation may be explicitly forbidden -- or even be illegal -- in another.

Here are areas where using cloud computing services in different nations may significantly affect a company's legal and regulatory responsibilities:

- **Data location**. As just described, having a third party hold data may affect the rules and laws concerning how that data must be governed. When a company uses a cloud provider located in another country, its responsibilities are made more complex because (1) it must understand and conform with a second set of laws and regulations, and (2) it may find that there is interaction between the two country's laws, making it more difficult to define a set of processes which adhere to both sets of laws. In fact, the company may find outright contradiction between the two country's laws, making it impossible to follow both.

- **Data transfer**. Just as laws governing storage governance vary across nations, so too do the laws and regulations dictating the conditions under which data may leave national jurisdiction. In short, many countries, particularly those located in Europe, forbid transfer of personal information outside the borders of the country. When using a cloud provider it is important to understand where the data stored with a cloud provider is located, what laws govern data management, and whether the cloud provider may transfer the data to other locations as part of its cloud operations.
- **Data privacy**. Many countries, again, most commonly those in Europe, have stringent controls regarding what personal information a company may store. These countries treat the privacy of personal data very seriously and it is paramount that any company using cloud computing ensures it conforms with the privacy laws and regulations applicable to the data it stores.
- **Data retention**. Less commonly recognized as a requirement applicable to companies is the need to ensure that data is retained. Many industries are require to retain data for a certain period of time to facilitate audits on the length of time the data has been stored, access patterns, and so on. This is important for regulatory reasons as well as to address any potential legal requirements, e.g., needing to produce data in response to a subpoena or legal discovery. Naturally, retention laws vary by country and make it incumbent upon a cloud user to understand and comply with applicable retention requirements.

3.2. Cloud Computing User Duty With Respect to Data Breaches

There have been many incidents of data breaches -- unauthorized parties gaining access to data -- recorded over the past few years. In many countries the holder of data, e.g., a retail firm that stores customer data, is responsible for maintaining that data securely and, in the event of a data breach, must notify those whose data has been exposed, so that they may be aware and take any necessary corrective action.

This responsibility does not disappear if the data is stored at a third party location. Even if the data breach occurs as a result of a cloud provider's mistake or negligence, the holder of the data (in the example, the retail firm) still retains responsibility for notifying those associated with the data.

It is important that cloud users understand their responsibility and be prepared to respond as required by relevant laws and regulations. Failure to do so -- even if a third party was actually responsible for the data breach -- is likely to result in legal and financial penalty.

4. Resource Use and Cost Assignment: Chargeback vs. Showback

The fifth cloud computing characteristic identified by NIST is "Measured Service," defined as "Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported providing transparency for both the provider and consumer of the utilized service." Essentially, with this characteristic, NIST is suggesting that because cloud resources are shared and consumed according to application requirements, they should be paid for on a fine-grained basis, with resource use costs directly assigned to the group consuming them.

This characteristic is by far the most controversial, as it represents a huge change to how most IT organizations charge for their services. Traditionally, much of a company's computing infrastructure is purchased on behalf of specific organizations, so that one bank of servers is "owned" by the HR department, while another is "owned" by the Sales department. This reflects the fact that the servers are purchased on behalf of those organizations and they absorb the full capital cost of the purchase before they begin using the servers. Costs for shared resources like a data center building are assigned across the various company departments by some kind of formula, e.g., by the proportion of total company employee count each organization represents.

By contrast, NIST implies that users should only pay for what they use, and that these costs should reflect the total cost associated with delivering the resource. So, for example, when an organization uses a server for an hour, the cost for that server-hour should include not only the amortized cost of the server, but the cost for power used to run that server for an hour, as well as whatever other costs are associated with providing an hour's service by that server. This form of direct cost assignment is typically referred to as "chargeback," because costs are charged back to the consuming organization based on the amount of resource used.

Obviously, moving from the coarse traditional cost assignment mechanisms typical of most IT groups to a fully-loaded, fine-grained cost assignment is enormously challenging to implement. It requires a full understanding of all the costs associated with running IT, as well as a sophisticated cost assignment algorithm to ensure each resource cost accurately reflects the true cost of providing it.

As this type of cost assignment is so difficult, many IT organizations propose to begin with what is referred to as "showback." Showback provides information on how much resource is being consumed, but does not assign cost based on resource

consumption. In a showback environment, cost assignment is still performed in the traditional manner.

4.1. Implications of Measured Service

It is difficult to overestimate the implications of measured service. Simply put, money is one of the most effective means to guide human behaviour – and changing the way money changes hands is extremely unsettling and disruptive.

Cloud computing offers the opportunity to change application computing use from an "always on" environment to a "on when you need it, off when you don't," so the traditional methods of pricing use – full asset allocation, depreciated over time – are inadequate. It's vital to devise a financial accounting method that aligns with the usage model that cloud computing enables.

The transition from asset billing to measured use pricing is unlikely to be seamless and the following actions will be required:

- All IT costs to be captured into a single financial structure so that the true overall cost of IT can be applied across the resource users. Since it's common that multiple organizations have responsibility for different elements of "IT," integrating the different budget items will require an integration effort among the different organizations. Anyone who has ever drawn up budgets understands how challenging this can be.
- Developing standardized technology offerings associated with the move to measured pricing. Many IT organizations today treat every request for resources as a "one-off," custom designed and manually implemented. Measured services demands standardized offerings so that pricing may be automatically applied, and standardized offerings require a move away from customized service. While a menu-based set of offerings is intuitively attractive, and certainly can reduce provisioning time, it also reduces choice. Developing standardized offerings will present a challenge to both IT and application groups.
- Moving resource consumers to measured service and direct chargeback. While measured service is intuitively attractive, asset cost assignment has the undeniable advantage of simplicity. Migrating application groups who have traditionally made a budget transfer once or perhaps annually to a form of payment that occurs more frequently and also varies will undoubtedly be disruptive. Especially troubling to application groups will be the fact that the costs will vary according to factors that may not be explicitly chosen by the group – in other words, the cost may change according to how heavily used the system is, and that may reflect decisions by end users. Receiving a bill predicated on decisions made by others – who may not even work for the same company – will be enormously disruptive to those paying for the service. IT groups can expect to devote significant resources to explaining the basis

for chargeback costs, and those discussions may be quite heated as chargeback begins its implementation lifecycle.

Long-term, the move to measured service – and the inevitable complementary move to direct chargeback for resource consumption – will profoundly affect the way applications are architected and operated. Just as the increased cost of gasoline and concerns about global warming have resulted in automobiles with engines that shut down when in idling situations, so too will applications begin to reduce resource consumption when dictated by low load levels.

This may prove to be the most profound effect of the move to measured service, as it will require significant changes to application design, with accompanying human resource skill building. In addition, the attempt by application groups to reduce their costs by minimizing use will affect cloud providers, whether internal or public, as they will be forced to devise strategies to raise utilization levels so as to ensure their fixed costs are covered by the measured service charges they present.

5. Rogue/Shadow IT: The challenge of developer/business unit-led cloud adoption

Many discussions about public cloud computing state that it presents a danger in the form of "rogue " or "shadow" IT. However, despite the obvious pejorative nature of the terms, there is often no real definition of them, nor explanation of why rogue or shadow IT occurs and what IT organizations should do to address the issue. In this white paper, the two terms will be referred to as "rogue/shadow IT" to simplify discussion.

As stated above, rogue/shadow IT is a pejorative term; it refers to developers or business units accessing public cloud computing directly and thereby bypassing the IT organization.

5.1. Motivation to Adopt Rogue/Shadow IT

It is critical to understand the motivation for rogue/shadow IT users -- in other words, why do developers and business units choose to bypass IT and directly obtain computing resources via cloud computing?

The primary reason these users pursue a rogue/shadow IT initiative is that they find established IT practices and processes insufficiently responsive to their needs. They are under significant pressure to develop and deploy application extremely quickly, but they often confront IT organizations that require weeks or even months to provision resources. In an effort to accelerate application delivery, they will obtain resources from public cloud computing providers.

An additional reason these groups will use public cloud computing environments is their perception of IT as "the department of 'no'." This phrase refers to a perspective that IT is unhelpful or unresponsive in day-to-day interactions –that requests for help or to take action are not addressed in a timely fashion or are responded to negatively with an accompanying weak rationale.

Compared with this slow or unresponsive behavior, public cloud providers are an attractive alternative. Creating an account can be accomplished in minutes and requires little more than an email address and a credit card. Resources can be obtained in minutes as well; compared with the weeks or months for internal resource provisioning, this is much more satisfactory and much more able to address pressing business needs. Finally, the "pay-as-you-go" metered billing associated with public cloud computing means that it is possible to begin productive work at a much lower cost than the upfront capital investment commonly required for internal IT resources.

Another benefit associated with rogue/shadow IT practices is that the ease of resource access and the freedom of lack of resource commitment make it easy to experiment or pursue potentially innovative business opportunities; the downside of failure is low, since it is easy to return resources to the provider and stop any further cost.

Consequently, the reasons that developers and business units choose public cloud computing for many applications are easily comprehended. The fast resource access and low cost of public cloud providers compared with the extended provisioning times, high cost, and perceived unresponsiveness of internal IT organizations provides a strong contrast and an understandable preference for public cloud computing.

This phenomenon is so widespread that it has given rise to the term "shadow/rogue IT," referring to developers or business unit bypassing IT and directly accessing computing resources via public cloud computing. As earlier stated, the term carries pejorative implications, but, at bottom, means that computing resource users are utilizing new resource providers rather than the long-established internal IT providers.

5.2. The Downside of Rogue/Shadow IT

While the attractiveness of direct use of public cloud computing is easy to understand, it's important to recognize that there can also be drawbacks to rogue/shadow IT. In addition to managing computing infrastructure, IT organizations are commonly required to ensure observation of IT-associated compliance and regulatory requirements. Many times, the provisioning delays that frustrate end users are not the result of installing and configuring computing resources, but instead result from evaluating security and compliance requirements.

When users choose to bypass IT processes in order to more rapidly obtain computing resources, there is a very real possibility that important security and compliance measures can also be bypassed. This poses a danger that the larger organization may face increased risk because the applications deployed by these users fail to address security or compliance requirements.

Many IT organizations resist end user adoption of cloud computing, viewing its benefits as outweighed by these risks. Consequently, cloud computing can cause conflict within companies as users push for its adoption, while IT organizations attempt to prevent its use. The language associated with this direct use reflects this conflict and can escalate it -- "rogue" and "shadow" carry implications of breaking rules or attempting to hide something that is inappropriate.

This conflict can be increased by how IT organizations respond when direct cloud use is discovered. Some in the industry have recommended that IT offer end users

"amnesty" when discussing cloud use that has not included IT. Amnesty is a term that implies forgiveness offered by a superior organization for a transgression of established rules and such an attitude is unlikely to foster collaboration. It may even increase conflict and lead to increased direct cloud adoption by end user organizations, whose frustration with IT may be exacerbated by what is viewed as a negative, unresponsive attitude.

5.3. Responding to Rogue/Shadow IT

Reflexive, antagonistic reaction to rogue/shadow IT does not prevent the phenomenon and may result in an increase. Rather than attempting to stamp out direct user adoption of cloud computing, or responding with a negative and condescending attitude, IT groups are far better served by understanding the motivations underlying direct adoption and addressing them, while implementing practices that still accomplish the very important governance mandates that are IT's responsibility.

Instead of pursuing what it likely to be a futile policy that attempts to prevent direct user cloud adoption, IT organizations should therefore consider implementing the following practices:

1. **Map current governance processes (e.g., resource request evaluation and approval, application security review, resource provisioning) and understand evaluation and approval criteria**. Specifically, define what acceptable "automatic" approvals are and which requests require human intervention for evaluation and approval.
2. **Identify which of these processes are important and must be kept, and which are followed due to inertia or "just to be sure."** Eliminate all unnecessary processes and policies in order to streamline the overall process.
3. **Capture these criteria in a set of formal policies that can be defined and operated in a policy/workflow engine**. This enables resource requests to be mapped against approval policies and enables automated approvals for standard requests.
4. **Create preconfigured cloud resources (e.g., virtual machine templates) that contain approved software components that are configured and patched to ensure appropriate security measures**. Make these resources available for use by application groups, who will find it easier to use preconfigured resources than build these resources manually. Providing an easier-to-use set of resources responds to users' desire for rapid resource availability while ensuring they follow necessary security practices.
5. **Create a service catalog of these preconfigured cloud resources**. The service catalog should include entries made up of aggregated resources (e.g., a preconfigured, multi-tier application topology containing web, application, and

database tiers fully configured and connected), making it easy for developers to launch complex application environments easily. This reduces the effort required for developers to begin productive work, avoids failures due to configuration errors, and ensures proper security configuration requirements are followed.

6. **Create a rapid-response team comprised of all groups involved in governance**. When a cloud application is identified that requires remediation to ensure it complies with company security and compliance requirements, this group can work with the application group to identify necessary remediation steps and help the group implement them quickly. This is a far more productive response than attempting to prohibit use of cloud computing.

One common mechanism to implement these steps is via the use of software products referred to as "cloud management software." These products are used to manage an organization's use of cloud computing environments and are configured to implement compliance and governance policies automatically across any computing resources obtained from a cloud provider.

Use of a cloud management software product ensures the following:

- **Consistent governance**: By capturing all necessary security and compliance policies in a set of rules that are applied whenever new computing resources are provisioned, the cloud management system ensures that necessary policies are applied to every computing resource used in every application. In fact, cloud management software often provides greater consistency than previous processes. This is because traditional IT typically relies on manual interpretation and implementation of governance, making it possible that policies may vary from one application to another. Moreover, relying on manual configuration of resources presents the likelihood of IT personnel making mistakes and failing to implement necessary governance policies accurately.
- **Rapid resource availability**: Cloud management software provides a portal for resource users to request provisioning, which means that resources can be provisioned as soon as requested, rather than waiting for IT personnel to respond to a trouble ticket request. Examples abound of organizations moving from provisioning timeframes of six to eight weeks down to five or 10 minutes.

Use of cloud management software can help reduce developer "shadow/rogue IT" motivation. One of the primary reasons developers and business units embrace direct use of public cloud computing is to gain access to computing resources more rapidly; cloud management software enables IT to deliver the same kind of rapid resource access that direct use of public cloud computing provides.

In addition, cloud management software ensures that all cloud computing resources obtained by users implement all necessary security and governance policies. This

functionality offers central IT assurance that it can fulfil its responsibility regarding compliance. Therefore, many IT organizations implement cloud management software as a mechanism to satisfy developer need for rapid resource availability and IT requirements for security and compliance.

5.4. Rogue/Shadow IT Conclusion

Cloud computing is here to stay and it's critical to understand why cloud adopters are using it and what benefits they achieve by doing so. Attempting to deny cloud computing benefits or prohibit cloud use in the name of "IT policy" is unlikely to be productive and may in fact be counter-productive, causing additional bypassing of IT or even attempts to disguise all cloud computing use.

A far better response is to understand why user organizations are directly adopting cloud computing and bypassing IT, and addressing those reasons directly in a way that makes cloud use more productive for users while ensuring they follow IT security and compliance requirements.

Identifying IT governance requirements and practices and capturing them in a policy engine, thereby ensuring they can be rapidly and automatically applied is the most productive way to respond to rogue/shadow IT. Providing preconfigured resources and automatically applying security and compliance requirements will satisfy the needs of both cloud adopters and central IT.

6. Cloud Computing: Succeeding in a Multi-cloud Environment

One of the most talked-about – and misunderstood – topics in cloud computing is the "hybrid cloud." It's critical to define and understand the topic because, for most enterprises, hybrid cloud computing represents the real future of their computing infrastructure. The reason organizations need to understand all aspects of hybrid cloud computing is that many vendors attempt to skew discussion of the topic in ways that benefit their products or offerings; without having a good knowledge foundation it is easy for end users to pursue a sub-optimal hybrid strategy.

The benefits and challenges of hybrid cloud computing vary widely depending upon the implementation characteristics of a specific hybrid cloud environment. Put another way, one environment that legitimately could be characterized as "hybrid cloud" might be relatively simple to implement and manage but would provide little benefit, while another hybrid cloud environment might provide great functionality richness and benefit, but be quite difficult to implement and manage.

6.1. Hybrid Cloud Computing Definition

By its very name, it is obvious that hybrid cloud computing refers to a mixed environment made up of two or more cloud environments. However, beyond that simple insight, what hybrid cloud computing means rapidly becomes more complex. Essentially, there are two factors that interact to define the specific characteristics of a hybrid cloud computing environment:

- **Location and Operation**: For a given organization, the hybrid environment is affected by where the hybrid environment is located and who is responsible for operating it. At its simplest, a hybrid cloud environment might be two data centers, both of which are owned by the using organization. At its most complex, a hybrid environment might incorporate multiple cloud environments, each of which is provided by external vendors, with the vendors maintaining responsibility for operation of the external environments. And, of course, it's quite likely that for many organizations, their hybrid environment might be a mixture of local self-controlled data centers along with external environments provided by outside vendors.
- **Technology:** Every cloud environment relies on a mix of hardware and software to provide cloud services. The simplest hybrid environment would be multiple execution environments, each of which is based on identical technology. More commonly, most hybrid environments will be made up of multiple technology offerings, each of which provides specific functionality and operational characteristics.

Depending upon the choices an organization makes regarding environment responsibility and technology, its hybrid cloud computing environment is more or less complex, and therefore more or less challenging to operate.

The following table lists hybrid cloud computing options from least complex to most complex:

Complexity	**Environment Responsibility**	**Infrastructure Technology**
Least Complex	Self-controlled	Homogenous
More Complex	Mixed (self and third party control portions of hybrid environment)	Homogeneous
Significantly More Complex	Self-controlled	Heterogeneous
Most Complex	Mixed (self and third party or entirely third party)	Heterogeneous

As one can see, the factor that most significantly affects hybrid cloud computing complexity is the infrastructure technology. This is not to say that hybrid cloud implementations which include environments controlled by both an entity and external providers do not present complexity; for example, when a company implements a hybrid environment containing facilities – even if those facilities use the same infrastructure technology as the company's own – there are additional complexities associated with contractual terms, managing operations and service agreements, and so on.

However, using hybrid environments that mix dissimilar technologies (e.g., an on-premise cloud based on OpenStack and a remote AWS cloud) are far more complex, since even the simplest of actions is complicated by the need to operate in environments that differ in functionality, API, management and monitoring capability, and performance characteristics.

It might seem that the solution is therefore obvious: an organization should use a single infrastructure technology solely within environments it controls. Unfortunately, while this is an attractive vision, many (if not all) companies find it unattractive or unattainable, for some of these reasons:

- **Cost**: Controlling all of the infrastructure environments that make up a hybrid cloud requires the company to own and manage them. This can require significant capital investment, thereby negating one attractive aspect of cloud computing: paying only for resources used
- **Insufficient functionality**: Many companies use external cloud providers because they offer richer functionality than that available from environments based on the current on-premise technology. Confining choices to those offering similar functionality as currently available would be too restrictive in terms of application functionality.
- **Insufficient scalability**: One reason companies seek to use hybrid cloud computing is to access the vast capacity offered by some public cloud providers (e.g., AWS and Google). Insisting that external environments use the same technology as that currently on-premise would restrict the options available to a company and perhaps prove inadequate in terms of application scalability.
- **Organizational complexity**: Even if an organization would prefer to utilize a homogenous hybrid cloud environment, organizational realities may prevent this. Many organizations operate in a decentralized decision-making mode, and parts of the organization may make technology choices which differ from those made by the central IT organization. Another reason that organizations may end up with heterogeneous cloud environments relates to corporate strategy: companies commonly grow by acquisition, and even a company that has implemented a homogenous hybrid cloud environment may acquire a company that uses a different infrastructure technology, and thereby inadvertently become a company with a heterogeneous hybrid cloud environment.

Consequently, it is foreseeable that most hybrid cloud computing environments will, in fact, be comprised of heterogeneous technology distributed across environments managed by multiple parties. Companies need to recognize this and use it as the foundation for their planning.

6.2. Hybrid Cloud Computing Capabilities

Of course, all discussion of hybrid cloud computing would be pointless if there were no benefits to using it; after all, it's obviously more complex to manage a mixed, distributed technology environment than a single technology operating in a single environment.

The answer as to why companies adopt hybrid cloud computing, and why one can expect it to be the dominant cloud adoption model in the future is simple: it better

supports a wider variety of application functionality and operational requirements than the simpler alternative of a single technology operated in a single environment.

These are some of the capabilities that hybrid cloud computing offers:

- **Operating applications in the right environment**. For many companies, their on-premise environment is well-suited to support traditional applications -- those with predictable loads used primarily by the company's own employees. However, the on-premise environment does not support other application use cases very well, e.g., a short-lived application associated with a time-limited marketing campaign. Being able to use infrastructure environments with different capabilities offers better support for the wider variety of application requirements now typical for most enterprises.
- **Partitioning applications across multiple infrastructure environments**. It may be convenient or cost-effective to partition an application topology across multiple operating environments. For example, a company might wish to deploy the data tier of an application within its own data center, while placing the web tier in a public cloud environment better suited for rapid elasticity to support erratic workloads. As more companies "front end" legacy applications with APIs to enable more applications to access core business processes, this approach to application deployment will become more common.
- **Bursting applications into secondary environments to support rapid elasticity**. Some organizations envision that they might deploy applications that occasionally experience heavy traffic loads; for those applications they might want to run application infrastructure supporting the baseline load inside their own data center, while using a public cloud environment to deploy additional computing resources to support the occasional heavy application load. This is referred to as "cloud bursting," and is a common vision for hybrid cloud computing initiatives.
- **Locating workloads in the most cost-effective operating environment**. Different computing environments have varying costs associated with them, and organizations may choose to place an application in an environment that offers low-cost computing. Commonly, this may be a public cloud computing environment, but overall application operating costs can be affected by amount of resource consumed, number of hours per month the application operates, the pricing and discount scheme available from the provider, and even application design. This is a complex issue to analyze, but the widely varying costs associated with infrastructure environments make this an important factor in workload placement decisions.

6.3. Technical Requirements for Hybrid Cloud Computing

Assuming an organization wishes to leverage hybrid cloud computing, what are the technical requirements it must implement to gain the greatest benefit? In other words, once an organization decides it is going to use multiple cloud infrastructure environments, what are the necessary technology elements it must have in place to achieve success?

Organizations pursuing hybrid cloud computing must ensure they address these four technical requirements:

Flexible cloud management: system management that can support multiple, heterogeneous cloud environments. Every cloud infrastructure environment provides different functionality as well as a unique API and management console.

Consequently, if an organization uses multiple cloud providers, it can end up using a number of different management mechanisms. This poses issues, in that supporting multiple management interfaces imposes complexity, makes it more difficult to ensure sufficient employee skill support, and raises operating costs – all clearly undesirable.

A common solution to the challenge posed by multiple cloud management systems is to use a "cloud management" software solution. These solutions encapsulate individual cloud environments and present a single unified interface for users. The cloud management solutions map their abstracted functionality onto each of the managed cloud environments. This unified solution enables user organizations to focus skill building on a single product and reduces operating costs, since it is not necessary to support cloud management solutions for each cloud environment being used.

High bandwidth network connectivity: For applications whose topology spans multiple cloud environments, it is critical that sufficient bandwidth be present to enable application components to communicate to one another with sufficient performance. Also, many applications require low latency between components to support acceptable response times.

Organizations must ensure they have the required network connectivity available to enable communication with and between all cloud infrastructure environments.

Application integration support: Many if not most applications today require the exchanging of data between applications or the calling of one application's functionality from another. When the calling and called application reside in the same infrastructure environment, direct calls using local IP addresses are possible.

However, when the calling and called application components are segregated into separate infrastructure environments, direct calling is not possible. Therefore, hybrid cloud computing relies upon integration mechanism availability that can be called from external environments.

The techniques used for remote integration are commonly referred to as a "service-oriented architecture." This term refers to data or functionality that is made available in a service that can be called by an external entity. Arguably, all integration functionality should be service-based, whether accessed from a remote infrastructure environment or from the same infrastructure environment.

Despite wide agreement that service-oriented integration techniques are preferable from a functionality and security perspective, many companies have not put this kind of integration capability into operation to give the functionality necessary for hybrid cloud computing. Consequently, organizations may find that they are less able to leverage hybrid cloud computing than they would prefer, unless they make investment in creating service-based integration capability for necessary application functionality.

Cloud cost analytics: As noted above, analyzing the true cost of operating an application in one infrastructure or another is quite complex. However, it is important to understand the true cost of application placement in order to make appropriate application deployment decisions. Failure to understand what the costs of operating an application in a given environment are may result in very high costs, or much higher costs than might be incurred if the application were run in a different environment.

This task can be made more complex by several factors. First, many cloud providers do not offer easy ways to understand their pricing models. Second, most cloud providers offer no mechanism to access resource utilization metrics and provide only end-of-month aggregated bills. Third, many user organizations use a single user account to manage all of their applications, making it difficult to associate operational costs to individual applications.

To help organizations analyze and predict the cost of running an application, entrepreneurs have launched a number of "cloud analytics" companies. These companies help users to associate resource use with specific applications. Cloud analytics companies also commonly provide comparison capabilities, so that users can understand what an application would cost to run in different infrastructure environments. Finally, many cloud analytics companies offer forecasting functionality, in which the likely cost of running an application can be evaluated.

Cloud analytics is a nascent area of cloud computing but, as more companies embrace hybrid cloud computing, it is likely to become critical as part of a hybrid cloud computing toolkit. Without the ability to evaluate and predict costs, users will experience "sticker shock" as cloud bills come in and, as noted, given the opaqueness of most cloud providers' billing mechanisms, understanding and modifying use patterns will be quite difficult.

6.4. Hybrid Cloud Use Cases

Given the differences in capabilities among different infrastructure environments, it is important that organizations select the most appropriate environment for a given application. Selecting the correct deployment environment can provide the most appropriate match between an application's requirements and the underlying infrastructure, while making a poor selection can consign an application to poor performance and high costs.

While every application is unique, and therefore requires analysis to determine the right deployment environment, some general guidelines regarding deployment are possible.

Here are some good hybrid use cases:

1. **Heterogeneous application requirements**: Some applications require capabilities that are not available from a single infrastructure environment, e.g., an application that must integrate with an on-premise database but which also has significant use at the web tier, such that the infrastructure demand outstrips resource availability on-premise. For an application such as this, spanning the application across multiple infrastructure environments is a good option.
2. **Highly variable application load**: Applications that require variable resource availability are naturally suited for deployment in a public cloud environment. Some organizations may choose to deploy enough infrastructure resource on-premise to handle the baseline application load and "burst" additional traffic to other infrastructure resources located in public cloud environment.
3. **Differing application security requirements**: Many applications have differing security requirements for different elements of the application. For example, the database tier may have significant restrictions regarding access, while the user interface tier may need to be available to a user population located throughout the world. For an application like this, partitioning it so that the database tier resides on-premise and only accepts traffic from known IP addresses while the user interface tier operates in a public cloud environment and accepts traffic from any IP address, may be a good deployment arrangement.

However, there are also applications for which it is inappropriate to consider deployment into a hybrid cloud environment. When an organization confronts constraints that restrict the usefulness of a hybrid cloud deployment model, it must choose a single environment that best meets the overall requirements of the application.

Here are some scenarios in which hybrid cloud computing would be a poor choice of deployment options:

1. **Poor cloud management capabilities**. As one of the hybrid cloud technical requirements above noted, it is critical that each cloud environment that an organization might desire to use provides sufficiently rich management functionality to support the organization's requirements. If a cloud infrastructure environment does not provide such functionality, it is probably not a good option for inclusion in the organization's hybrid cloud portfolio.
2. **Organizational focus on legacy application support**: Many organizations are primarily focused on supporting legacy applications, with few "cloud native" applications in the overall application portfolio. Organizations such as these are not well-served by attempting to adopt hybrid cloud computing; in fact, they may find their path forward very challenging due to the fact that most of their resources are devoted to operating legacy applications.
3. **Lack of cloud computing application architecture skills**: As should be clear, managing multiple, dissimilar cloud environments can be quite challenging. Attempting to architect applications that span multiple, dissimilar environments and operate effectively is significantly more complex than traditional applications. Many organizations that are fully capable of developing and managing traditional applications may find that hybrid cloud applications are too complex for their employee skillset.

6.5. Hybrid cloud technology options

There are a wide variety of technologies and providers that an organization may utilize as it builds out its hybrid cloud environment. Both open source and proprietary-based environments and products are available; for most organizations selecting the right options will significantly affect how successful their hybrid cloud initiatives will be.

Companies deciding which technology platforms to incorporate into their hybrid cloud environment have a number of different choices. The choices they make carry significant implications in terms of cost and complexity and careful evaluation is important. Broadly speaking, the choices relate to the two environments incorporated into a hybrid cloud environment: on-premise and public cloud provider.

The following sections give an overview of the options available for each of these environments.

6.5.1. On-premise: Open Source or Proprietary?

As in the rest of the software industry, cloud environments can be provided via both proprietary and open source software packages. Proprietary software packages are provided by individual software companies which maintain full responsibility for the package; these packages are distributed with restrictive licenses that, typically, require organizations to pay significant license fees to use the package.

By contrast, open source packages are commonly, although not exclusively, created via the collaboration of a number of separate individuals and the software is made available with an expansive use license that does not require payment of a licensing fee in order to use the software.

The question of whether to use proprietary or open source cloud environment software packages is quite important for the following reasons:

- Use of a proprietary software package commits a user to a specific software provider (commonly referred to as lock-in). Should the user be dissatisfied with the functionality of the package or working with the provider, there is little they can do – there are no other providers of the software, so the user is faced with an unpalatable choice: accept the status quo or invest significant time or money to switch to another software package.
- Proprietary software is typically much more expensive than open source alternatives, making its use in a cloud environment potentially quite expensive.
- Use of a proprietary software package commits the user to the provider's product and ensures that product improvements are subject to the provider's pace and decisions. The question of whether this is a benefit or a liability can be quite controversial. Open source advocates suggest that the opportunity for large developer communities means that these products can evolve much more rapidly than proprietary alternatives and enable much faster innovation as a result. Proprietary advocates suggest that a focused company with domain expertise can focus much better than decentralized open source communities and therefore can deliver functionality and innovation more quickly than these communities.

The primary proprietary cloud software package comes from VMware, which is the company behind ESX, the leading virtualization hypervisor product. Given the preponderance of ESX in most user infrastructures, this would indicate a preference toward VMware's vCloud offering -- although, for those users that have adopted Hyper-V, Microsoft's Azure offering holds potential as an on-premise solution.

In the early days of cloud computing, many proprietary software vendors and technology suppliers (e.g., IBM and HP) also created proprietary cloud orchestration solutions. However, these met with mixed results and most of these vendors and suppliers have deprecated their proprietary solutions in favor of open source-based offerings. Consequently, for proprietary on-premise choices, most organizations will decide between VMware and Microsoft.

However, most IT organizations appear to prefer an open source-based solution for their on-premise cloud environment. Reasons for this preference include:

- **Lock-in avoidance**: Many IT organizations are wary of being committed to and dependent upon a proprietary provider. History has shown that many proprietary providers use customer commitment as leverage in negotiations and IT organizations are reluctant to extend lock-in dependence into another segment of their IT infrastructure.
- **Cost**: Many IT organizations are under significant budget pressure from their parent organizations and view open source as a lower-cost alternative for a given software product.
- **Innovation**: Many IT organizations believe that greater innovation is present in the open source world, both in terms of the creation of software (i.e. open source projects indicate a greater speed of innovation compared to proprietary providers) and also in terms of enabling end user innovation. This latter is often under-appreciated but, as IT becomes a critical part of business offerings, the ability for software users to generate innovation is important; thus, using open source as an innovation enabler is another reason user organizations may prefer an open source-based cloud infrastructure environment.

While a number of open source-based cloud infrastructure projects are available, two of them have emerged as the most widely adopted: CloudStack and OpenStack, with the latter seeming to have greater adoption and a larger community. It is important to note that OpenStack should not be interpreted as a single product: some organizations download the baseline OpenStack software files, while others use one of the more than half-dozen commercial OpenStack distributions available from companies like Red Hat, Mirantis, and Piston Computing. A significant challenge for user organizations regarding OpenStack is the fact that the OpenStack vendor community is likely to shrink in the future as one or two OpenStack distributions emerge as the most widely adopted; companies that have chosen a less successful distribution will be operating an OpenStack product unlikely to have a long lifetime.

In summary, it would appear that the majority of on-premise cloud environments will be based on one of these three choices:

1. VMware vCloud
2. Microsoft Azure

3. OpenStack

6.5.2. Public cloud computing: A multitude of choices

Deciding on a public cloud provider to incorporate into its hybrid cloud strategy is one of the most difficult situations an IT organization confronts today. There are literally hundreds of choices, ranging from tiny providers located in isolated geographies to giant providers available from many regions across the globe

While every IT organization has a company-specific set of objectives and constraints, there is a common set of factors that affect its choice of which public provider (or providers) to incorporate into its hybrid cloud infrastructure:

- **Technology consistency with on-premise environment**. The simplest hybrid cloud infrastructure to manage is one based on a homogenous technology. Being able to use common services delivered with a consistent syntax and operations is very attractive. For this reason, many IT organizations are attracted to public providers based on VMware or Microsoft.
- **Richness of cloud services**. While consistency of technology is attractive, as a general rule, the proprietary-based public providers do not offer the richest set of cloud service functionality. The richest functionality comes from AWS, which is unique to Amazon and very different from both vCloud and Azure. The breadth of services affects the ease of application development and potential for innovation, so, depending upon how important these factors are for a given IT organization, it may direct it toward a non-proprietary cloud provider.
- **Geographic availability**. Many enterprise IT organizations have widely distributed operations and serve customers in many different countries. For these organizations, the ability to deploy applications in a number of locations is important, and therefore how many regions within which a cloud provider offers services is also important.
- **Cost**. Most of the large so-called webscale cloud providers (e.g., AWS and Google) are based on open source software. As they do not have to pay license fees to a proprietary software vendor, it is likely that they will be less expensive than proprietary-based providers. For many IT organizations (especially those under budget pressure), this would direct them toward an open source-based provider. It is important to note that both VMware and Microsoft operate their own public cloud providers, which probably do not incur licensing fees and therefore might be able to match the open source-based providers.
- **Scale**. Some IT organizations may require very large scalability in their applications and would therefore tend to select providers with the most available infrastructure. Today, these are the webscale providers comprising AWS, Google, and Microsoft. They are much larger than any of the other

providers in the industry, with AWS generally considered to be the largest by far, although that scale advantage may diminish over time.

6.6. The Challenge of Choice

As a final note regarding hybrid cloud computing, it is important to reiterate that most enterprise IT organizations fail to settle on a single standard for their environment and inevitably end up with a mix of dissimilar, incompatible products. It is likely that, for most enterprises, the same will be true for both on-premise and off-premise cloud infrastructure; most IT organizations will find that they have a variety of on-premise cloud infrastructure products running in their data centers, along with a number of public providers used for off-premise cloud computing.

Consequently, when planning its hybrid cloud strategy, every IT organization should assume that it will confront the most complex hybrid environment identified earlier. This means that sophisticated cloud management and cost analytics will become core competencies for enterprise IT. Failing to adopt these capabilities will result in IT organizations facing significant issues as cloud computing use outstrips their ability to manage their cloud environments efficiently and securely.

7. Creating a Cloud Computing Action Plan to Ensure Success

Despite being convinced of the benefits of cloud computing, many IT organizations struggle to implement a concrete cloud strategy. Many experience frustration when initial successful cloud initiatives (often under the guise of "shadow IT") fail to develop further and become a mainstream part of the IT application portfolio.

In seeking reasons for this failure to embed cloud computing as a standard method of application deployment, many IT organizations assign blame to a variety of factors:

- "Cowboy" developers who took too much responsibility into their own hands
- Use of the wrong cloud infrastructure environment
- Use of a public cloud provider rather than an internally-hosted environment.

Rarely, though, do IT organizations recognize the primary reason for failure in their cloud computing initiatives: being inadequately prepared when first working with cloud computing and thereby failing to integrate initial cloud efforts into existing technology choices and established processes. This inevitably leads to cloud efforts that fail to build a foundation for ongoing success; worse, it consigns the organization to painful re-work and wasted effort until better alignment between cloud initiatives and existing IT frameworks and processes is in place.

A far better approach is to recognize that initial cloud computing initiatives should be aligned with existing technology frameworks and processes and supported with an action plan designed to treat an initial initiative as the first step in a long-term plan.

7.1. Why an Action Plan is Important

The natural impulse when first exposed to cloud computing is to get started immediately. It seems to hold such promise when contrasted with traditional IT infrastructure capabilities that many developers or business units get started immediately. Moreover, they're often reluctant to involve other groups or consider how cloud computing can be integrated into existing processes, viewing this effort as likely to slow down their path to the cloud.

It's only later, when the developers must engage those parts of the IT organization that were bypassed during the initial cloud adoption effort, that they confront unpalatable reality -- the initial prototype, which seemed to hold such promise, cannot be put into production because it has not passed through a required security assessment. Or the operations group refuses to support the application because it does not align with approved software components and application design. At that point, all the benefits

of bypassing existing arrangements are lost as the application is retrofitted to address corporate IT requirements.

Consequently, planning and implementing a cloud computing action plan can pay significant benefits when the delays imposed by the retrofit process are evaluated.

Organizations that implement a cloud computing action plan achieve the following benefits:

- **Buy-in from the larger organization**. Bypassing established processes and existing organizations might seem like a way to streamline application development and get a project done more quickly. In fact, it may be true that the individual project might be completed sooner. However, failing to engage the larger organization means it is much less likely to view the project positively and therefore is also less likely to appreciate its benefits. Assuming that one objective of the initial group is to generate greater enthusiasm for cloud computing within the larger IT organization, pursuing a "skunk works" approach is unlikely to achieve buy-in.
- **Higher probability of success**. By involving additional groups within IT, the group driving adoption of cloud computing can improve the probability of success for its application. For example, by incorporating a representative of the company enterprise architecture group, the adopting group can leverage the EA group's expertise in terms of application scalability design. By incorporating other groups and their expertise, success is more likely.
- **Alignment with company technology strategy and architecture**. One of the most common problems with initial cloud efforts is that, while the individual project is successful, it fails to act as a catalyst for cloud adoption within the overall IT organization. Frequently, the reason for this is that the initial project is developed independently, with no reference to established IT technology choices or architecture frameworks. Consequently, upon completion, the rest of the IT organization views the cloud application as an interesting but unacceptable experiment. This means that the potential benefits of adopting cloud computing are lost in the argument regarding the failure of the group developing the application to follow approved approaches. Creating an action plan that engages with important IT organizations and incorporates existing technology strategy and architecture will raise the likelihood of success.
- **Documentation of success**. While the benefits of cloud computing are intuitively clear, without documentation many people are skeptical of the potential for improvements to traditional IT practices. A structured action plan allows for comparison of the costs and timeframes of traditional IT versus those achieved with a cloud computing environment.
- **Organizational learning**. One of the limitations of a segregated cloud computing initiative is that only those directly involved with the project learn

important lessons about cloud computing: the capabilities of the environment; new skills that must be learned; vendor interaction patterns; support capabilities, etc. By creating an action plan that incorporates participants from the larger organization, this knowledge can be more widely shared, and thereby makes possible faster cloud uptake by the larger organization when it begins to embrace cloud computing.

When all of these factors are considered, it seems obvious that one is better served by running an initial cloud project with an accompanying action plan to ensure better integration with the larger IT organization, if for no other reason than engaging the larger organization makes it more difficult for the benefits achieved via cloud computing to be denied or rejected.

7.2. Understanding the key components of an action plan

Given the benefits that accrue from creating a cloud computing action plan, the next question is how best to create and implement one, i.e. what people and processes should be part of a cloud computing action plan so as to ensure that the highest benefit is achieved?

Based on analysis of successful cloud computing initiatives across many commercial and governmental organizations, the following sections give key elements of a cloud computing action plan.

7.2.1. Create an evaluation task force

Creating an evaluation task force is critical, and including the right members in it can make the difference between success and failure for a cloud computing initiative. At a minimum, the evaluation task force should include:

- **Application developers**. Obviously, developers are a prerequisite to create an application. Since it is likely that creating a cloud application will require development of new skills, developers with a preference for learning and experimentation are good candidates to participate in a cloud initiative.
- **Infrastructure administrators**. In most IT organizations, responsibility for operating applications in production falls to an operations group, and the employees primarily responsible for day-to-day work are system administrators. Therefore, including system administrators in the evaluation task force will help ensure that the production requirements are addressed.
- **Network administrators**. Because of the dynamic nature of cloud environments and applications, networking groups are often challenged in managing connectivity among infrastructure resources. It is important for network

administrators to be part of the evaluation process so that they may understand the implications of cloud adoption.

- **Storage administrators**. Storage can be quite complex in cloud computing use. It is important that someone from a storage organization, who understands the use and growth of storage in cloud computing use scenarios, participates in a cloud evaluation.
- **IT finance**. One of the most controversial topics in cloud computing is cost evaluation. Many organizations are interested in cloud computing because it holds the potential for cost savings, while others stoutly maintain that cloud computing is more expensive than traditional IT. To take this discussion out of the realm of speculation and anecdote, it is important to have a capable and impartial participant in the task force whose responsibility it is to evaluate the financial outcomes of the evaluation.
- **Executive sponsor**. Adopting a new form of computing not only involves change but also holds the potential for disrupting established organizational structures and power relationships, which can lead to resistance. Without a senior executive sponsoring the initiative, it is unlikely to gain resources or, even if resources are obtained, will fail to generate further adoption.
- **Legal**. Cloud computing operates under very different licensing conditions compared with traditional packaged software products. Moreover, the legal obligations and commitments by cloud providers, as well as the responsibilities assigned to cloud users, differ from traditional vendor/user relationships. For this reason, it is appropriate to have a legal representative assigned to the evaluation task force.
- **Security**. Security is the most commonly raised concern identified as a cloud computing adoption barrier. Most organizations need to understand how security and responsibility for security changes when using a cloud computing environment. As security is such an important topic, a representative of the security group should be part of the task force.
- **Users**. Cloud computing promises to speed application delivery and agility, all in the service of providing greater user satisfaction. Engaging real-world users in the task force makes it possible to evaluate how well cloud computing fulfils this promise.
- **Project manager**. While many cloud initiatives operate as a 'developer-only' project that requires only informal communication, viewing the initiative as the first milestone in an ongoing effort that requires the involvement and coordination of a number of groups makes it clear that project tracking, structured communication, and project documentation are important components of the initiative. Having a project manager who takes responsibility for formal tracking and communication is vital to enable organizational success and ongoing adoption.

7.2.2. Set objectives

Identifying the reasons for engaging in a cloud initiative and establishing criteria by which the initiative's outcomes may be judged is important. Otherwise, the initiative's accomplishments or usefulness for further pursuit may be hazy and relegate it to nothing more than an interesting experiment.

- **Identify IT tasks that require improvement**: Clearly, if all were well with the status quo there would be no need to explore the potential of cloud computing. Consequently, there are undoubtedly reasons why the organization is interested in leveraging this new form of computing. An appropriate first step in the evaluation process is to identify those areas in which the current environment falls short or is less than satisfactory.
- **Documenting current metrics**: Of additional value is to measure current IT metrics that, with improvement, could increase IT productivity or user satisfaction. This provides a baseline against which the cloud evaluation may be measured. With areas for improvement identified, the outcomes of the cloud evaluation will provide data for comparative purposes.
- **Identifying qualitative measures for evaluation**: In addition to specific technical areas that need improvement and metrics that measure current IT performance, it is often useful to identify "soft" areas in which IT falls short of satisfying resource consumers. Soft areas can include items like "denies resource requests citing infrastructure unavailability" and so on. Evaluating whether these qualitative areas improve with cloud computing can be extremely valuable and indicate whether it makes sense to pursue further cloud computing adoption.
- **Identifying other important areas to evaluate**: In addition to the items described above, each company has unique requirements and issues it seeks to address. The objective-setting process should identify any additional areas the company wishes to evaluate during its initial project for inclusion in the overall evaluation process. Common areas that organizations evaluate include the cost of the cloud applications operated during the evaluation period, the ease or difficulty of managing applications in the evaluated cloud environment, and how capable the evaluated cloud environment is in supporting application elasticity.

7.2.3. Identifying the deployment environment to be evaluated:

Clearly, organizations are faced with a plethora of choices in terms of potential cloud computing environments that might be used in an evaluation effort. It is important to identify what characteristics the organization wishes the cloud environment to include and then select one or more environments based on those characteristics. This will ensure that the organization can select evaluation environments that support important elements like support for the organization's technology architecture, availability of packaged software or services necessary for the organization's applications, etc.

At a minimum, companies selecting a cloud deployment environment should follow this process:

- **Identify organizational requirements**: There may be certain capabilities that the organization needs from its provider to meet its requirements. For example, the organization may make heavy use of Microsoft technologies, making it imperative that the cloud provider it uses support common Microsoft technologies like .NET and SQLServer. Since most cloud providers offer little or no Microsoft-oriented functionality, this will limit the options the organization has regarding the cloud providers it will use. There may be other requirements as well that constrain the organization's choices, e.g., it may need to operate its cloud environment in certain geographies, or have the provider offer HIPAA compliance. Consequently, the organization should identify and documents its requirements to ensure the pilot project fully meets its long-term needs. Selecting a cloud provider that offers quick access but an inadequate functionality portfolio will, in the long run, mean that the organization will have to go through a second provider search to identify one that fully meets its actual requirements.
- **Identify deployment candidates**: Once the organization has its functionality requirements in hand, it can use that list to evaluate cloud providers that might be used for its pilot project. It is an unfortunate fact that most providers will fail to fully support all of an organization's requirements, and may only partially fulfil those requirements that it does offer. For this reason, a good practice is to assign a score between 1 and 10 for each organizational requirement for every potential cloud provider. This scoring enables the organization to avoid black-or-white judgments and offers the ability to create a more nuanced assessment of each cloud provider, as well as enabling comparison across a larger number of potential cloud providers.
- **Select one or more cloud providers for the POC/Pilot application(s)**. With the requirement scoring available, the organization can calculate total scores for each deployment candidate and identify the highest-ranked ones. From this list the organization can identify one or more providers that offer environments well-suited for the POC/Pilot.

7.2.4. Implement a POC/Pilot Application

This is the most critical part of an evaluation effort, and forms the foundation of the evaluation outcome. The application that is implemented should:

- **Reflect the organization's current or desired future architecture**. This ensures that any cloud environment choice made as a result of the evaluation will support the majority of the organization's applications.
- **Fully exercise the cloud characteristics that the organization seeks to leverage in future applications**. It is important that this evaluation tests the capabilities of

the evaluated cloud so that future application designs can rely on these capabilities being available and robust. Failing to evaluate important characteristics may cause problems in the future when the organization is unable to achieve the necessary application capabilities it desires.

- **Identify the current skills and skill shortages the organization has with respect to using cloud computing**. Any missing skills that the evaluation application identifies will require a training program as cloud computing is diffused throughout the organization.
- **Map the desired cloud characteristics against the capabilities of the cloud environment(s) used in the evaluation process**. Every person in the IT industry has confronted products or services that fail to support promised functionality, and cloud computing is no different. The evaluation process should analyze the desired capabilities and identify those that are less capable than promised by the cloud vendor.
- **Evaluate the other capabilities identified in the objective-setting part of the evaluation process**. These may be organization-specific or cloud-specific, but require evaluation along with the mainstream capabilities typically used during an evaluation project.
- **Costs**. It is important to track and capture total costs of the project, both those of the cloud environment itself as well as any unexpected costs (e.g., new software licenses required as a result of using a software package in a cloud environment). These costs will be relevant to downstream cloud initiatives and will prove useful for project planning, so capturing them during the evaluation process is important.
- **Operational experience**. Many organizations are surprised at the differences in monitoring and managing applications in cloud environments. Any changes in standard operational processes encountered during the initial cloud application deployment should be identified and documented, so that they may serve as guidance during subsequent cloud application initiatives.

7.2.5. Report POC/Pilot results

Once the POC/pilot/ effort is complete, it is appropriate to take stock and document its results. In turn, these results will serve as the basis for the organization to determine whether the POC/pilot outcomes justify further use of cloud computing. These are the items that should be addressed at the end of the POC/pilot:

- **Capturing POC/pilot outcomes**: Part of the preparation for the POC/pilot was identifying current unsatisfactory IT processes and outcomes in the hope that cloud computing could improve upon them. Consequently, it is important to compare achieved outcomes against the list of unsatisfactory items to determine whether cloud computing can improve upon existing infrastructure practices. This needs to be nothing more complex than a simple comparative

chart listing each item with a "before" and "after" column indicating the metrics associated with each item in both the traditional infrastructure environment and the cloud computing environment. Easily comprehensible documentation makes subsequent parts of the reporting process easier and will enable a straightforward go/no-go decision regarding subsequent cloud adoption.

- **Briefing executive sponsor**: Perhaps the critical stakeholder in the POC/pilot process is the executive sponsor. This individual is commonly positively disposed toward cloud computing and made a personal commitment to sponsor the project. Providing him or her with information regarding the POC/pilot outcomes is critical and, if possible, the sponsor should be the first individual outside the core implementation group to learn of the outcomes. Providing early access to the POC/pilot outcomes will enable the sponsor to further prepare the organization for subsequent cloud adoption efforts or, if the POC/pilot is unsuccessful in improving upon current practices, provide organizational support for the POC/pilot team as its members transition back to their everyday jobs.
- **Briefing IT groups**: Every group that contributed personnel or expertise to the POC/pilot should be briefed on the project's outcomes. This briefing should present full findings of the project's outcomes, not merely the outcomes relevant to the group's area of responsibility. This ensures that the group has the full context in which to evaluate the impact cloud computing presents to its area of responsibility. Further, it allows consideration of the benefit cloud computing provides to the overall organization compared with any extra effort or changed processes necessary for it to be implemented. For example, many security groups find that they need to modify the products used for application security when applications operate in a shared environment managed by a third party; however, in light of significantly reduced operational costs or greater business agility, a security organization may conclude that modifying its security practices is justified in light of the overall achieved benefits.
- **Obtaining approval for further cloud use**: Very few organizations undertake a cloud evaluation project as an academic exercise. It is generally the intent that the evaluation project serve as evidence to support further cloud commitment as well as an opportunity to develop skills necessary for further organizational adoption. However, even though the evidence may be clear-cut, it is important to obtain explicit approval for further cloud adoption. Subsequent cloud adoption efforts will require co-operation from additional personnel and groups not involved in the original evaluation. Without explicit approval from a senior executive, it would be easy for these groups to resist or drag their feet with respect to participating in cloud-based projects. This approval will make it clear that they are expected to participate; refusal to participate in an environment with clear senior executive approval invites organizational disapproval. However, even in an environment in which all groups are positive toward engaging in cloud computing efforts, explicit approval can be useful because

these efforts will require budget; senior executive approval can often shift budget priorities and free up moneys for cloud projects.

7.3. Preparing for wider organizational adoption

Once the organization has completed its initial investigation of cloud computing and determined that it holds benefits sufficient to warrant further use, the next step is to create the foundation for broader adoption throughout the organization. Companies most successful in achieving broad adoption of cloud computing typically follow a program with these elements:

- **Creating an internal cloud support group**: Each new group beginning to use cloud computing faces a shortage of skills and knowledge. Rather than making each group go through a low-productivity period of learning and experimenting, it is far more efficient to create a centralized group that can provide guidance and expertise. This enables cloud adoption efforts by new groups to move much more quickly and achieve the benefits of cloud computing in a much shorter timeframe. An internal cloud support group is often referred to as a "Center of Excellence" or "Center of Expertise." A second benefit of using a centralized group as a resource is that it helps to ensure that different parts of the organization use consistent practices, techniques, and technologies. This avoids inefficiency and mistakes, and helps realize incremental benefits associated with common practices such as reduced costs, greater ease of personnel reassignment, etc.
- **Identifying "launch groups"**: One challenge associated with broader cloud adoption is that enthusiasm for new technologies varies among potential user groups. This can cause a significant challenge for cloud success, because unsuccessful adoption efforts after the initial evaluation can cause organizations to rethink their commitment to cloud computing, and even decide to drop further adoption efforts. The reasons for lack of enthusiasm vary but can include lack of resources to explore a new technology, previous failure in attempting to adopt a new technology, or feeling threatened by the new technology. Whatever the reason for tepid enthusiasm, it is extremely problematic for long-term cloud success to start wider adoption efforts with a user group which is unlikely to be successful. A far better approach is to seek groups that are highly motivated to adopt the new technology and able and willing to devote the resources and effort to address the inevitable challenges that accompany taking up a new way of doing things. Identifying "launch groups" that are enthusiastic about taking up cloud computing increases the probability of successful adoption and can create the foundation for ongoing success; moreover, additional insights that can be used throughout the organization can be identified when cloud computing is used outside the original evaluation group. Naturally, having the internal support group involved in these initial launch efforts further increases the probability of success; the combination of

a willing launch group with an internal support group is the best recipe for beginning the broad cloud computing adoption process.

- **Creating training materials**: As mentioned above, spreading consistent practices and tools across the organization carries far-reaching benefits. Many organizations have begun cloud adoption with early success and then encountered long-term problems as they suffer with multiple cloud provider environments, differing levels of provider support, inconsistent application architectures, and uneven levels of knowledge among different IT groups. The best way to avoid these problems is to provide a common foundation among all cloud users via a standardized training program. The internal cloud support group is an obvious candidate to develop and deliver this training. It should create a standardized, easily-accessible training program which, ideally, should be capable of being delivered both on-line and in-person as appropriate. If the internal support group does not have sufficient skills to develop training, the company's HR department or an external training organization may be able to assist in this initiative.
- **Evangelizing cloud computing success**: Generating adoption for a new technology is a challenge in any company; people don't like change and will often resist it, even if it represents an improvement over current approaches. One very effective method to increase cloud adoption is to develop supporting evidence for its benefits and successful outcomes. Developing metrics and success examples (aka anecdotes) provides material that can be presented when attempting to convince a resistant individual or group to adopt cloud computing; moreover, should an individual or group prove recalcitrant to moving forward, the metrics and success examples can be presented to higher level management in order to generate adoption by explicit direction. It is a good idea to capture the data and evangelize as widely as possible; with sufficient awareness it is possible to create an atmosphere of assumed adoption and make resistance to adoption very difficult.
- **Develop a community of interest**. Many of the most successful cloud adopters institutionalize their cloud efforts in a "community of interest." This is a cross-organizational group devoted to sharing knowledge and experience. The benefits of creating a community of interest include faster adoption and greater momentum of cloud computing. In addition, as individual groups learn about cloud computing, they can share their experience, enabling quicker spread of knowledge throughout the organization. This also allows peer-to-peer sharing, relieving the internal cloud support group of some responsibility and avoiding the support group becoming a bottleneck.

7.4. Developing Ongoing Relationships with Public Cloud Providers

The rapid growth of public cloud providers makes it probable that every enterprise will use one or more of them. Moreover, despite the fact that many observers still maintain that enterprises deploy only unimportant applications to these environments, more

and more enterprises are operating important production applications in public cloud computing environments.

This fact means that public cloud providers will be a strategic part of every organization's infrastructure, and it is therefore important that each enterprise implement a working partnership with important providers to ensure necessary support and information flow.

These are the important elements of an ongoing relationship with a cloud provider. An organization should develop a strong relationship with every public cloud provider with whom it has significant application deployments; note further that this list should be regularly reviewed, as providers often become more important over time.

Here are the elements that an enterprise should ensure are in place for each of its important cloud providers:

- **Contract review**. Every cloud provider's contract varies, with significant differences in support commitments, service level agreement, cost, and liability nearly universal. It is important for an enterprise to understand what it can expect from a cloud provider, as well as understand its commitments in terms of using a given cloud provider's services. A good way to achieve this is for the company to put the provider's contract through its standard review process. However, it is very common that enterprises have unrealistic expectations of how flexible a cloud provider will be in terms of an agreement. While there may be some room for negotiation, a company should not expect that it can impose unlimited liability or receive unusual concessions unavailable to other users. Cloud providers are very careful about their contracts and commitments and typically reluctant to modify them, except in the face of extremely large financial outcomes, and sometimes not even then.
- **Support**: Support for use of the cloud provider's environment is critical. It is a universal experience that every organization and every application will confront problems in every cloud provider environment, so having explicit support mechanisms in place is vital for success. Part of the contract review discussed above should be a review of the support commitments made by a provider and the organization should be clear that it understands what it can expect from a provider when a trouble ticket is opened. It is very common that a user organization communicates an important problem to a provider and only then discovers that the provider's support commitments are far less expansive than expected. In addition to reviewing a provider's support commitments, the organization should walk through several scenarios during contract discussions to ensure it has a full understanding of what may be expected when a trouble ticket is submitted. Part of this process should include a review of the provider's escalation process so that critical problems can be prioritised. Finally, any financial penalties for service outages should be reviewed so that there are no

surprises in the event of an outage. Quite often cloud users discover there is little financial penalty to the provider for a service outage even if the user forfeited a significant amount of revenue during an outage.

- **Roadmap exposure**: Cloud computing is an ever-evolving area of technology, and providers commonly improve their offering frequently. As a user, it is important to understand what functionality is available today and what will be coming in the future. Armed with that information, a user can make appropriate choices about where it should deploy an application and what kind of functionality it can deliver in the application. Cloud users should seek a strong technical relationship with each of its providers and should use a provider's willingness to share roadmap information as one criterion in the provider selection process.
- **Finance contact**. Just as provider contracts universally vary in terms of conditions and commitments, so too do their usage metrics and costs. Nearly every large cloud user organization finds that provider bills are complex and difficult to comprehend. This becomes a significant problem as cloud usage grows and bills increase to tens or hundreds of thousands of dollars per month. Enterprises should therefore identify internal resources to become experts in how its providers charge for service and should also expect their providers to offer a contact person or group to help decipher the provider's charges. This often seems unimportant when first using a cloud provider because the bills seem quite small; however, many organizations are surprised at how rapidly their cloud use (and therefore provider bills) grow, leading to "sticker shock" when looking at a provider's charges. Because of this, an enterprise should require every cloud provider to offer a point of contact with which monthly bills can be reviewed to ensure only appropriate charges are placed on its account.