Security and high availability in cloud computing environments

Executive summary
Cloud computing has become an increasingly popular means of delivering valuable, IT-enabled business services. Adopting cloud technology can be an affordable way to get access to a dynamically scalable, virtualized computing environment. Optimal IT hardware, software, expertise and infrastructure management resources that may not otherwise be available from a cost perspective can be rapidly deployed and easily scaled. Processes, applications and services can be available on demand, regardless of the user location or device. The cloud provider is responsible for the environment, so organizations can make use of resources for short periods of time without having to maintain the environment when it is not being used.

While cloud computing models are attractive because of their flexibility and cost effectiveness, certain challenges must be addressed in order to provide a viable option to traditional data services. First and foremost is the issue of security. The externalized aspect of outsourcing can make it harder to maintain data integrity and privacy, support data and service availability, demonstrate compliance, and secure highly available access to applications and information. In short, cloud computing can present an added level of risk.

Organizations must therefore establish trust relationships with their cloud computing providers and understand risk in terms of how these providers implement, deploy, and manage security on their behalf. Whether delivered as part of the service or as specific components added in, your cloud provider should address the fundamentals of security and risk management through a comprehensive approach.
This white paper discusses security and availability-related challenges in cloud computing environments. In particular, the paper introduces the business-driven IBM Security Framework and how it can be leveraged to address cloud security in a more holistic approach. The paper also takes a closer look at the shared security responsibilities that exist between consumer and provider. Finally, the paper investigates high availability concerns and demonstrates how to help improve the resilience of your virtual servers in a cloud computing environment.

Cloud computing—overview and benefits
Each type of cloud computing model—public, private or hybrid—faces different levels of IT risk. In the private cloud delivery model, the cloud owner does not share resources with any other company. Private clouds are owned and operated by a single organization, delivering IT services within the constraints of their own network perimeter.

In the public cloud computing model, IT activities and functions are provided as a service that can be billed on a pay-per-use or subscription basis via the Internet from external suppliers, using resources not owned by the consumer. The sharing of IT resources in a public, multitenant environment can help improve utilization rates and can reduce costs significantly while maintaining access to high quality technology. In a public cloud, an organization rents IT resources instead of having to invest in their own physical IT infrastructure or maintain under-utilized equipment to service peak loads. Instead, they can scale usage up or down, according to need, with costs directly proportional to need.

Many organizations embrace both public and private cloud computing by integrating the two models into hybrid clouds. These hybrids are designed to meet specific business and technology requirements, so that activities and tasks are allocated to traditional IT, external or internal clouds, as appropriate, helping to optimize security and privacy with a minimum investment in fixed IT costs.

In addition to the different cloud computing models, there are distinctions among the most common cloud service models as shown in Figure 1. Available to anyone with Internet access, cloud service models include:

- Software as a Service (SaaS) cloud model—Enables software to be delivered from a host source over a network as opposed to installations or implementations
- Platform as a Service (PaaS) cloud model—Enables operating systems and middleware services to be delivered from a managed source over a network
- Infrastructure as a Service (IaaS) cloud model—Enables the entire infrastructure to be delivered as a service over a network, including storage, routers, virtual systems, hardware and servers.

In this paper we will focus on the IaaS cloud computing models.

---

**Figure 1. Cloud computing models**
Cloud computing and security: the grand challenge

Although cloud computing can shift some control over data and operations from the client organization to its cloud provider—much in the same way organizations entrust part of their IT operations to outsourcing companies—operating a secure cloud-based IT environment is, after all, a shared responsibility. Even basic tasks, such as applying patches and configuring network security, are the responsibility of both the cloud service provider and consumer. Let us examine an example for both public and private cloud.

In an IaaS model it is the cloud provider's responsibility to offer a number of pre-configured virtual machines that constantly have to be kept updated with the latest security patches. When clients provision the virtual machines, they have to trust the cloud provider to deliver secure systems. The clients do not have access to the hypervisor layer—the underlying operating system that manages one or more virtual machines that run on a physical machine—which typically does not share the virtual network segment with any of the other hosted virtual machines to avoid network-based intrusions. The cloud provider may also separately offer an optional virtual private network (VPN) capability so that the client can help ensure a protected network that is not directly visible for Internet-based attackers. It is the client's responsibility to maintain the patch level for all provisioned virtual machines after the initial deployment as well as a properly configured VPN to shield their valuable data and infrastructure. If a client, however, chooses to refrain from purchasing a VPN option or does not patch any of their web-facing virtual machines, the machines can be vulnerable.

If an organization decides to implement these mechanisms in a private cloud using their internal IT department, they will rely on their business policies to govern aspects like data confidentiality, access control to applications and systems, and so on. Employees will be entrusted to handle the IT infrastructure because they are recertified against the business policies on a regular basis. In this environment organizations have to deal with the risk of unauthorized privileged user access, data loss prevention, malicious intrusions and unintentional user errors. They also have to observe and comply with internal or regulatory guidelines.

If the organization decides to implement these mechanisms in a public cloud using the cloud provider discussed in the example above, they will rely on written business agreements to govern the exact same aspects, as explained in the private cloud example. They also have to deal with the same risks and comply with the same internal or regulatory guidelines.

To implement this or a similar cloud-based approach, organizations must establish trust relationships with their providers and understand the risk in terms of how these providers implement, deploy and manage security on their behalf. This “trust-but-verify” relationship between cloud service providers and consumers is critical because the cloud service consumer is still ultimately responsible for compliance and protection of their critical data, even if that workload moves to the cloud.

Other aspects about cloud computing also require a major reassessment of security and risk. Inside the cloud, it is often difficult to physically locate where data is stored. Security processes that were once visible are now hidden behind layers of abstraction. This lack of visibility can create a number of security and compliance issues and may prohibit certain IT use cases to be implemented using a strictly public cloud environment. Clients need to ensure that they can select a physical location even for a public cloud deployment, and that contracts are in place to guarantee localized data storage.
In addition, the massive sharing of infrastructure with cloud computing creates a significant difference between cloud security and security in a more traditional IT environments. Users spanning different corporations and trust levels often interact with the same set of computing resources. At the same time, workload balancing, changing service level agreements and other aspects of today’s dynamic IT environments create even more opportunities for misconfiguration, data compromise and malicious conduct.

Infrastructure sharing calls for a high degree of standardized and process automation, which can help improve security by eliminating the risk of operator error and oversight. However, the risks inherent with a massively shared infrastructure mean that cloud computing models must still place a strong emphasis on isolation, identity and compliance.

Another critical requirement is to protect a production-grade application environment against failure. This requires that an application be highly available and accessible without interruption.

**Using the business-driven IBM Security Framework for cloud environments**

As an organization secures its business processes, a business-driven approach needs to become the guiding influence for ensuring that all the different security domains work together in a holistic and synergistic manner, and in alignment with the overarching business objectives. Otherwise, the organization’s risk stance becomes vulnerable, due to misalignment of priorities between IT and the business strategy. Using a standards-based approach to map business drivers to IT security domains can be difficult and is often an afterthought.

IBM developed a comprehensive framework (as shown in Figure 2) to better understand enterprise security in terms of the business resources that need to be protected and by examining different resource domains from a business point of view. In the next several sections, we take a closer look at this framework to understand the different aspects of a holistic security architecture and how they can apply to enterprise-class cloud computing security requirements.

---

**Figure 2. The IBM Security Framework**
Security Governance, Risk Management and Compliance
Organizations require visibility into the security posture of their cloud. This includes broad-based visibility into change, image and incident management, as well as incident reporting for tenants and tenant-specific log and audit data.

Visibility can be especially critical for compliance. The Sarbanes-Oxley Act, the Health Insurance Portability and Accountability Act (HIPAA), European privacy laws, and many other regulations require comprehensive auditing capabilities. Since public clouds are by definition a black box to the subscriber, potential cloud subscribers may not be able to demonstrate compliance. (A private or hybrid cloud, on the other hand, can be configured to meet those requirements.)

In addition, providers sometimes are required to support third-party audits, and their clients can be directed to support e-Discovery and forensic investigations when a breach is suspected. This requirement adds even more importance to maintaining proper visibility into the cloud.

In general, organizations often cite the need for flexible service level agreements (SLAs) that can be adapted to their specific situation, building on their experiences with strategic outsourcing and traditional, managed services.

People and Identity
Organizations need to make sure that authorized users across their enterprise and supply chain have access to the data and tools that they need, when they need them, while blocking those who are not authorized for access. Cloud environments usually support a large and diverse community of users, so these controls are even more critical. In addition, clouds introduce a new tier of privileged users: administrators working for the cloud provider. Privileged-user monitoring, including logging activities, becomes an important requirement. This monitoring should include physical monitoring and background checking.

Identity federation and rapid onboarding capabilities must be available to coordinate authentication and authorization with the enterprise back-end or third-party systems. A standards-based, single sign-on capability is required to simplify user log-ons for both internally hosted applications and the cloud, allowing users to easily and quickly leverage cloud services.

Data and Information
Many organizations cite data protection as their most important security issue. Typical concerns include the way in which data is stored and accessed, compliance and audit requirements, and business issues involving the cost of data breaches, notification requirements and damage to brand value. All sensitive or regulated data needs to be properly segregated on the cloud storage infrastructure, including archived data.

Encrypting and managing encryption keys of data in transit to the cloud or data at rest in the service provider’s data center are critical to protecting data privacy and managing compliance mandates. The encryption of mobile media and the ability to securely share those encryption keys between the cloud service provider and consumer is an important and often overlooked need. Because moving large volumes of data quickly and cheaply over the Internet is still not practical in many situations, many organizations must send mobile media, such as an archive tape, to the cloud provider. It is critical that the data is encrypted and only the cloud provider and consumer have access to the encryption keys.
Significant restrictions regarding data co-location can arise with cloud computing, depending on an organization’s location, the type of data it handles, and the nature of its business. Several member states of the European Union (EU), for example, expressly forbid that its citizens’ nonpublic personal information leave their borders.

Additionally, a cloud deployment can raise export-law violation issues relative to encrypted information, and the deployment can potentially expose intellectual property to serious threats. The organization’s legal counsel must perform a thorough review of all these requirements prior to cloud deployment, making sure the organization can maintain control over the geographic location of data in the provider infrastructure.

In areas involving users and data with different risk classes that are explicitly identified (such as public and financial services), organizations need to maintain cloud-wide data classification. The classification of the data will govern who has access, how that data is encrypted and archived, and how technologies are used to prevent data loss.

Application and Process
Clients typically consider cloud application security requirements in terms of image security. All of the typical application security requirements still apply to the applications in the cloud, but they also carry over to the images that host those applications. The cloud provider needs to follow and support a secure development process. In addition, cloud users demand support for image provenance and for licensing and usage control. Suspension and destruction of images must be performed carefully, ensuring that sensitive data contained in those images is not exposed.

Defining, verifying and maintaining the security posture of images in regards to client-specific security policies is an important requirement, especially in highly regulated industries.

Organizations need to ensure that the web services they publish into the cloud are secure, compliant and meet their business policies. Leveraging secure-development best practices is a key requirement.

Network, Server and Endpoint
In the shared cloud environment, clients want to ensure that all tenant domains are properly isolated and that no possibility exists for data or transactions to leak from one tenant domain into the next. To help achieve this, clients need the ability to configure trusted virtual domains or policy-based security zones.

As data moves further from the clients’ control, they expect capabilities like intrusion detection and prevention systems to be built into the environment. The concern is not only intrusions into a clients’ trusted virtual domains, but also the potential for data leakages and for extrusions—the misuse of a client’s domain to mount attacks on third parties. Moving data to external service providers raises additional concerns about internal and Internet-based denial of service (DoS) or distributed denial of service (DDoS) attacks.

In a shared environment, all parties must agree on their responsibilities to review data and perform these reviews on a regular basis. The organization must take the lead in terms of contract management for any risk assessments or controls deployment that it does not perform directly.

Where image catalogs are provided by the cloud provider, clients want these images to be secure and properly protected from corruption and abuse. Many clients expect these images to be cryptographically certified and protected.
Physical Infrastructure
The cloud’s infrastructure, including servers, routers, storage devices, power supplies, and other components that support operations, should be physically secure. Safeguards include the adequate control and monitoring of physical access using biometric access control measures and closed circuit television (CCTV) monitoring. Providers need to clearly explain how physical access is managed to the servers that host client workloads and that support client data.

IBM SmartCloud Enterprise
The IBM SmartCloud Enterprise offering is an agile, cloud IaaS solution designed to provide rapid access to IT resources on a pay-per-use basis, with features and functions specifically built to meet the needs of enterprise clients. These include reserved capacity pools, account administrator functions, image sharing and VPN connectivity to cloud resources. A broad portfolio of both products and on-demand services help meet diverse client requirements such as cloud management, security and availability, with a proven service management system embedded in the cloud services to provide visibility, control and automation.

Access to the infrastructure self-service portal and application programming interface (API) is restricted to users with an IBM Web Identity. The infrastructure complies with IBM security policies, including regular security scans and controlled administrative actions and operations. Within IBM delivery centers, client data and virtual machines are kept in the data center where provisioned, and the physical security is the same as that for IBM internal data centers. The IBM VPN option lets clients isolate their environments on the IBM SmartCloud Enterprise on a virtual local area network (VLAN) that only the clients can access. See Figure 3 for a conceptual view of the offering.

Cloud computing providers and you: a shared responsibility
As indicated before, every cloud computing business agreement relies on a shared responsibility when it comes to operations and security measures. IBM responsibilities are based on security-rich operational models that are designed to include physical security and management of hardware below the hypervisor. Client responsibilities include managing proper access control for the cloud web portal, operating secure...
and hardened guest operating systems, choosing a secure implementation for a virtual private network, and implementing appropriate data control mechanisms. Figure 4 illustrates the breakdown of these responsibilities.

First, we will take a look at IBM responsibilities.

As mentioned above, IBM holds responsibility for the physical infrastructure that makes up the SmartCloud Enterprise environment and components from the hypervisor level and below.

- **Physical security of IBM cloud components**
  IBM has a long-standing history in the data center and hosting business, with many locations worldwide. Today, the IBM SmartCloud Enterprise offering leverages locations in the United States, Canada, Germany, Japan and Singapore, exploiting the experience, tools and approaches that IBM has in place for physical security. These include but are not limited to:
  - Digital closed circuit television (CCTV) within the data center is either recorded around the clock or event driven (motion activated). CCTV surveillance data is maintained for at least 30 days.
  - Data center access doors are equipped with a local audible alarm.
  - A computer-based controlled access system (CAS) uses badge readers to restrict access to only those with approval to enter controlled areas. All entries and exits to these areas are logged.
  - Biometric and card security is present where appropriate.
  - Anti-pass back (badge-out) function prevents multiple users from using the same badge for data center entry.
  - Facility design and fire protection help prevent cascading failures of other systems.

- **Management of hardware and software below the hypervisor**
  IBM provides management and maintenance of the provisioning environment, including the actual hypervisor, physical network and underlying hardware infrastructure. These systems are managed using Information Technology Infrastructure Library (ITIL)-based processes. IBM internal IT management processes are stringently applied, and regular internal audits are performed. IBM leverages both commercially available products and internal tools to manage the infrastructure. These also include service offerings such as IBM Managed Security Services for intrusion protection and vulnerability scanning.

  By design, the IBM staff and tooling do not access or scan a client's virtual environments. This model allows for a clean separation of duties between IBM and the client.

- **Security of the self-service portal and APIs**
  A user can provision resources in the IBM SmartCloud Enterprise environment in two ways: the self-service portal, and the application APIs. The infrastructure used to implement these entry points is deployed in secured IBM facilities using a zoned, multitier architecture. These resources are subject to stringent IBM internal security requirements and processes, and use recognized offerings and products such as IBM Rational® AppScan to scan, monitor and manage them. Communications to these resources between the customer client (for example, web browser or custom-built application using APIs) are secured using secure sockets layer (SSL) over hypertext transfer protocol (HTTP).
• Public image catalog
IBM provides a public catalog of operating system and middleware images. These images are built to the specification of our vendors and IBM Software Group organizations. Patch management of the base images in the catalog are the responsibility of IBM and adhere to IBM internal standards for deployment of security-related fixes. These images are updated on a regular basis and on an as-needed basis for “hot” fixes. Once the client provisions an instance from the IBM SmartCloud Enterprise catalog, the client is responsible for all patch management of that running instance and any images created from that instance.

As part of the shared responsibility model of the IBM SmartCloud Enterprise offering, the client is responsible for all aspects of security of provisioned resources in the cloud environment. The sections below examine what this encompasses in more detail.

• Identity management and access control
For identity management, IBM SmartCloud Enterprise relies on the standard Web Identity system that IBM has developed and deployed for all users of ibm.com systems. This system allows users to create and manage IDs and includes tools for password maintenance.

Once a client has signed up for the IBM SmartCloud Enterprise service, the ID specified during the sign-up process is assigned as the enterprise account administrator. Through the IBM SmartCloud Enterprise self-service portal, the account administrator has the ability to add, delete and modify additional user IDs that can be used to provision cloud resources (instances, images, storage, and so on). It is the client’s responsibility to manage all account user IDs based upon their own requirements (for example, approval process for adding an ID, revalidation of IDs, and so on).

• Guest operating system
IBM SmartCloud Enterprise provides a self-service environment for provisioning cloud resources. IBM takes a “no-touch” policy for all client-provisioned resources, or in other words, once a client provisions a resource in the cloud, the client is responsible for all security of that resource. When a client provisions an instance (virtual machine) in the IBM SmartCloud Enterprise cloud environment, the client receives full “root” or “administrator” privileges on the guest operating system. With this level of privilege, the client can secure provisioned resources based upon their internal requirements or standards.

The following list contains tasks that every client should include in the management of provisioned resources. (Note: This list is not all encompassing, and clients should manage their cloud operating system resources as they do resources contained in their own labs.)

- Patch management and security fixes—operating system vendors regularly patch their products to meet new threats, and the client fully controls to what extent and when these patches are applied. IBM suggests clients regularly monitor the security bulletins of operating systems vendors and apply updates and fixes to best meet their requirements.
- Secure additional software—when installing, configuring and managing any software on the guest operating system, the client should take care to properly secure the software and any access to the system the software may expose.
- Creating and implementing security policies on the guest operating system, which include but are not limited to:
  o Firewall policies of the guest operating system
  o Protection and distribution of guest operating system secure shell (SSH) keys
  o Encryption of data on the operating system
  o Choice of anti-virus software where appropriate
  o Removal of packages and services that are deemed not needed
**Network access**

By default, each instance (virtual machine) provisioned in IBM SmartCloud Enterprise is assigned one or more publicly routable IP addresses and is accessed via the Internet. An optional offering of IBM SmartCloud Enterprise is a VPN service. Each optional VPN provides an Internet Protocol Security (IPsec)-based VPN tunnel over the Internet between a client's IPsec capable gateway and one IBM SmartCloud Enterprise data center. With the VPN option, the client receives a private Virtual Local Area Network (VLAN). With the VPN option, when the client provisions an instance, the client is able to choose between provisioning the instance on the public VLAN or the private VLAN. The VPN option provides the client with encrypted communication of data over the Internet and an additional level of isolation within the IBM SmartCloud Enterprise virtual network. With the VPN option, a client may provision an instance that spans both the public VLAN and the client's private VLAN. This ability allows for greater flexibility in creating tiered deployment architectures in the cloud. This capability should be protected using software firewalls (either those that come with the operating system or from a third party) to limit both host and port access.

**High availability**

Clients can set up and configure high availability after having provisioned their virtual server environment. High availability is a key area of interest for organizations embracing cloud computing. High availability also has a number of areas that need to be addressed to ensure a high-availability application on the cloud, such as network vulnerability, multisite redundancy and storage failure. But one of the most important areas is IP failover. Utilizing virtual IP addresses for their virtual machines can enable clients to eliminate single points of failures and to design a highly available IT infrastructure.

**Addressing high availability for cloud environments**

The requirement to protect a production-grade IT system or application against a failure of any node is not new by any standard, and has been addressed in many software products. Yet these software products are, by and large, not compatible with many cloud offerings, and most public cloud providers do not always provide the required functionality. As a result, users need to complement the cloud deployments with high availability constructs that exist outside the cloud.

To address these needs and concerns surrounding high availability, and to do so in a security-rich and expedient manner, the IBM SmartCloud Enterprise offering has added support for virtual IP addresses (vIPs) on the cloud virtual instances. In addition to a regular static IP address, which each virtual machine gets when it is set up, and that never changes, a machine can dynamically assume one or several additional virtual IP addresses. Because it is the application code that controls the association between the vIPs and instances, the application topology can adjust to a node failure very quickly, typically in less than a second. Setup and configuration of high availability for the IBM SmartCloud Enterprise can be performed solely within your provisioned virtual machine environment.
As an example, take a pair of virtual machines, Virtual Machine A and Virtual Machine B, as shown in Figure 5. Each machine has a static IP address. Both virtual machines are also configured to allow the dynamic assignment of a virtual IP address in addition to the existing static one.

If the first machine encounters a problem, slow-down or failure, the virtual IP address is switched to the second machine. Now the second machine serves all the traffic while the first machine, once repaired, becomes the warm standby, as shown in Figure 7.

The static IP addresses are used to administer and maintain the machines. The virtual IP address is the one that is given to the clients as the published application or server IP address.

To begin with, the virtual IP address is attached to the first machine. This first machine handles all the service traffic. The second machine is up and running and acts as a warm standby, as shown in Figure 6.

The published address being the static address means that the failure and switch are virtually seamless to the end user—anywhere from a few seconds to a few minutes. Because the virtual IP address can be transferred between different machines very quickly, with careful programming, you could significantly reduce unplanned outages by having the IP address move at the first sign of trouble.

By supporting virtual IPs for virtual machines, IBM SmartCloud Enterprise can help reduce system downtime in the event of a virtual machine failure. By switching to a backup virtual machine, the consumer can help ensure that disruptions are almost seamless to an end user.

**Conclusion**

Today, security is often listed as the number one concern for clients considering cloud adoption. Cloud security issues persistently rank above cloud reliability, network issues and concerns about the economic payback of cloud. This concern is immediately followed by high availability, a concept that needs to be addressed for almost every production-grade IT environment.1
To fully benefit from cloud computing, clients must ensure that data, applications and systems are properly secured so that the cloud infrastructure won’t expose their organizations to risk. Cloud computing comes with all the usual requirements of traditional IT security and availability, with an added level of risk and complexity because of the external aspects. These risks call for a comprehensive framework to security that helps ensure all the different security domains work together in a holistic and synergistic manner, in alignment with overarching business objectives.

The IBM SmartCloud Enterprise offering builds on the skills, best practices and assets developed through years of experience managing and operating security-rich, reliable enterprise data centers around the world. The infrastructure complies with IBM security policies, including regular security scans and controlled administrative actions and operations. Within our delivery centers, client data and virtual machines are kept in the data center where provisioned, and the physical security is the same as that for IBM internal data centers.

The IBM SmartCloud Enterprise model does not provide a one-size-fits-all solution that can solve all client concerns, but it can provide appropriate levels of security and availability for specific cloud needs. Drawing on a broad portfolio of consulting services, software, hardware, and managed security services, IBM SmartCloud Enterprise can help you successfully implement a security-rich and highly available cloud computing environment.

For more information
To learn more about IBM SmartCloud Enterprise, please contact your IBM marketing representative or Business Partner, or visit the following website: ibm.com/cloud/solutions/enterprise


To learn more about implementing high availability within your IBM SmartCloud Enterprise environment, refer to the IBM developerWorks® article “High availability apps in the IBM Cloud” at ibm.com/developerworks/cloud/library/cl-highavailabilitycloud.