4 CLOUD COMPUTING REFERENCE ARCHITECTURE[13]

The NIST cloud computing definition is widely accepted and valuable in providing a clear understanding of cloud computing technologies and cloud services. The NIST cloud computing reference architecture presented in this section is a natural extension to the NIST cloud computing definition.

The NIST cloud computing reference architecture is a generic high-level conceptual model that is a powerful tool for discussing the requirements, structures, and operations of cloud computing. The model is not tied to any specific vendor products, services, or reference implementation, nor does it define prescriptive solutions that inhibit innovation. It defines a set of actors, activities, and functions that can be used in the process of developing cloud computing architectures, and relates to a companion cloud computing taxonomy. It contains a set of views and descriptions that are the basis for discussing the characteristics, uses, and standards for cloud computing.

The NIST cloud computing reference architecture focuses on the requirements of what cloud service provides, not on a design that defines a solution and its implementation. It is intended to facilitate the understanding of the operational intricacies in cloud computing. The reference architecture does not represent the system architecture of a specific cloud computing system; instead, it is a tool for describing, discussing, and developing the system-specific architecture using a common framework of reference.

The design of the NIST cloud computing reference architecture serves the objectives to: illustrate and understand various cloud services in the context of an overall cloud computing conceptual model; provide technical references to USG agencies and other consumers to understand, discuss, categorize, and compare cloud services; and communicate and analyze security, interoperability, and portability candidate standards and reference implementations.

4.1 OVERVIEW

The Overview of the Reference Architecture describes five major actors with their roles and responsibilities using the newly developing Cloud Computing Taxonomy. The NIST cloud computing reference architecture defines five major actors: cloud consumer, cloud provider, cloud auditor, cloud broker, and cloud carrier (See Figure 1: Cloud Actors). These core individuals have key roles in the realm of cloud computing. Each actor is an entity (a person or an organization) that participates in a transaction or process and/or performs tasks in cloud computing. For example, a Cloud Consumer is an individual or organization that acquires and uses cloud products and services. The purveyor of products and services is the Cloud Provider. Because of the possible service

[13] NIST Special Publication 500-292, *NIST Cloud Computing Reference Architecture*, September 2011

offerings (Software, Platform or Infrastructure) allowed for by the cloud provider, there will be a shift in the level of responsibilities for some aspects of the scope of control, security and configuration. The Cloud Broker acts as the intermediary between consumer and provider and will help consumers through the complexity of cloud service offerings and may also create value-added cloud services. The Cloud Auditor provides a valuable inherent function for the government by conducting the independent performance and security monitoring of cloud services. The Cloud Carrier is the organization which has the responsibility of transferring the data, somewhat akin to the power distributor for the electric grid.

Figure 1 – **Cloud Actors** briefly lists the five major actors defined in the NIST cloud computing reference architecture.

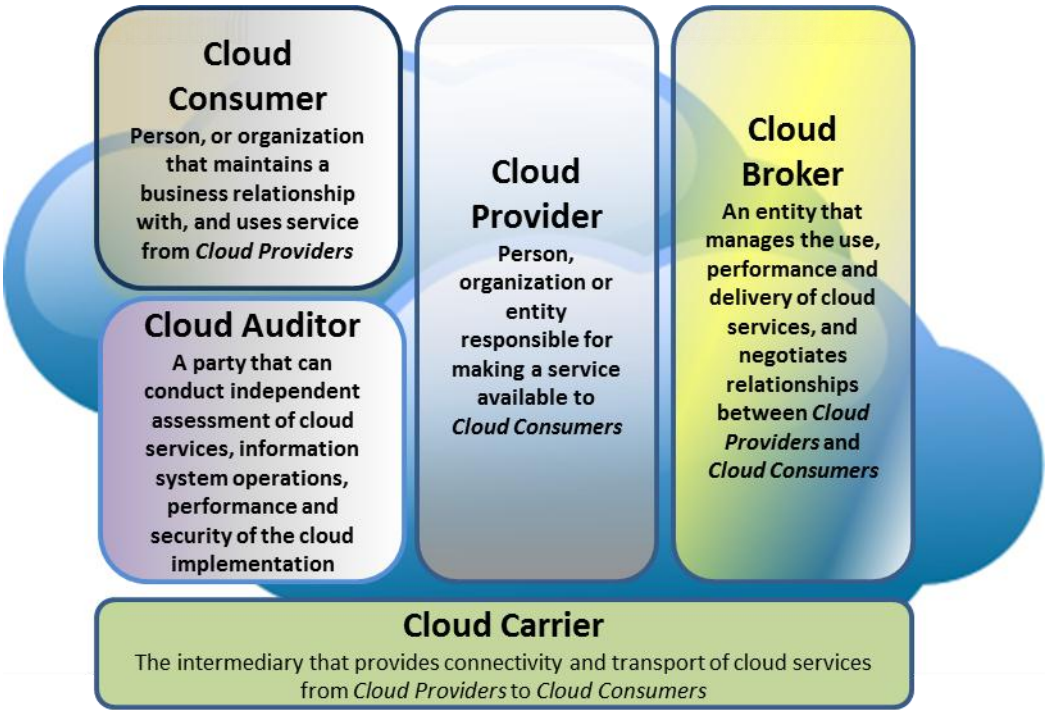

Figure 1 – Cloud Actors

Figure 2 – Interactions between the Actors in Cloud Computing shows the interactions among the actors in the NIST cloud computing reference architecture. A cloud consumer may request cloud services from a cloud provider directly or via a cloud broker. A cloud auditor conducts independent audits and may contact the others to collect necessary information. The details will be discussed in the following sections and be presented as successive diagrams in increasing levels of detail.

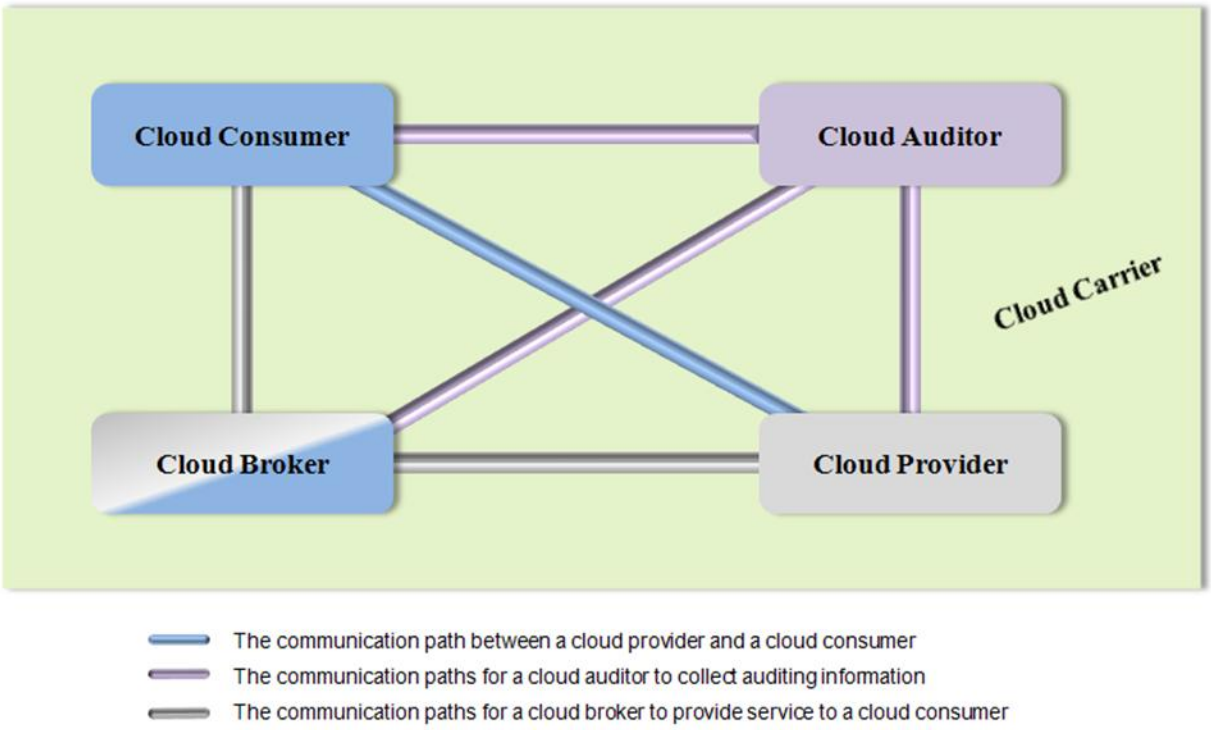

Figure 2 – Interactions between the Actors in Cloud Computing

4.2 CLOUD CONSUMER

The cloud consumer is the ultimate stakeholder that the cloud computing service is created to support. A cloud consumer represents a person or organization that maintains a business relationship with, and uses the service from, a cloud provider. A cloud consumer browses the service catalog from a cloud provider, requests the appropriate service, sets up service contracts with the cloud provider, and uses the service. The cloud consumer may be billed for the service provisioned, and needs to arrange payments accordingly. Depending on the services requested, the activities and usage scenarios can be different among cloud consumers, as shown in Table 1. Some example usage scenarios are listed in Figure 3.

Service Models	Consumer Activities	Provider Activities
SaaS	Uses application/service for business process operations.	Installs, manages, maintains, and supports the software application on a cloud infrastructure.
PaaS	Develops, tests, deploys, and manages applications hosted in a cloud system.	Provisions and manages cloud infrastructure and middleware for the platform consumers; provides development, deployment, and administration tools to platform consumers.
IaaS	Creates/installs, manages, and monitors services for IT infrastructure operations.	Provisions and manages the physical processing, storage, networking, and the hosting environment and cloud infrastructure for IaaS consumers.

Table 1 – Cloud Consumer and Cloud Provider

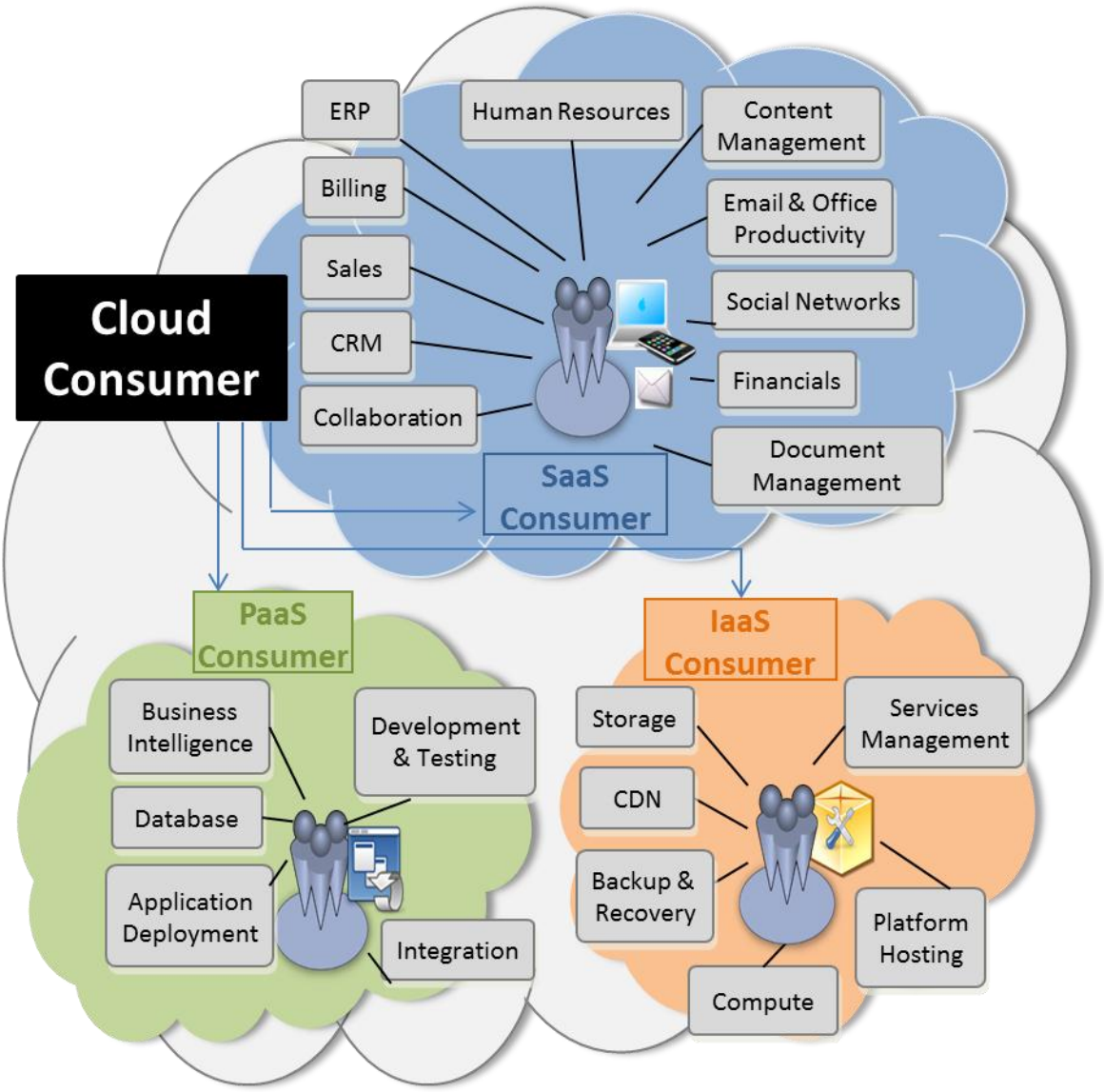

Figure 3 – Example of Services Available to a Cloud Consumer

SaaS applications are usually deployed as hosted services and are accessed via a network connecting SaaS consumers and providers. The SaaS consumers can be organizations that provide their members with access to software applications, end users who directly use software applications, or software application administrators who configure applications for end users. SaaS consumers access and use applications on demand, and can be billed on the number of consumers or the amount of consumed services. The latter can be measured in terms of the time in use, the network bandwidth consumed, or the amount/duration of data stored.

For PaaS, cloud consumers employ the tools and execution resources provided by cloud providers for the purpose of developing, testing, deploying, and managing applications hosted in a cloud system. PaaS consumers can be application developers who design and implement application software, application testers who run and test applications in various cloud systems, application deployers who publish applications into a cloud system, and application administrators who configure and monitor application performance on a platform. PaaS consumers can be billed by the number of consumers, the type of resources consumed by the platform, or the duration of platform usage.

For IaaS, consumers are provisioned with the capabilities to access virtual computers, network-accessible storage, network infrastructure components, and other fundamental computing resources, on which consumers can deploy and run arbitrary software. IaaS consumers can be system developers, system administrators, and information technology (IT) managers who are interested in creating, installing, managing and monitoring services for IT infrastructure operations. IaaS consumers are provisioned with the capabilities to access these computing resources, and are billed for the amount of resources consumed.

4.3 CLOUD PROVIDER

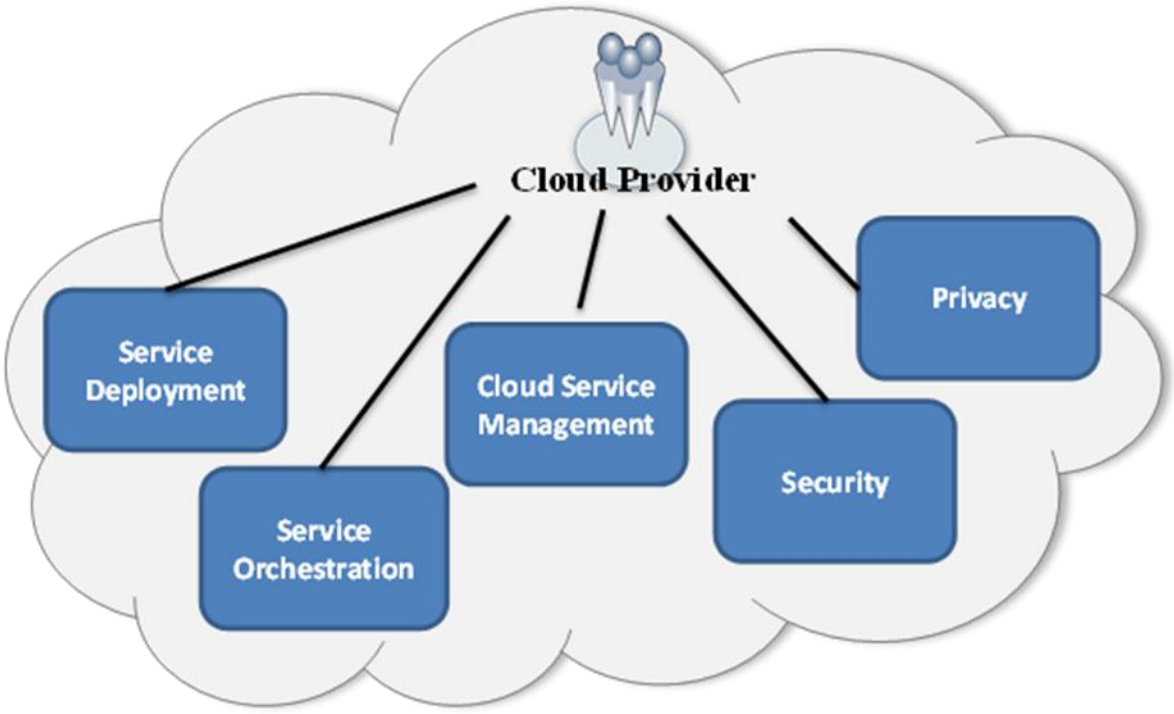

Figure 4 – Cloud Provider: Major Activities

A cloud provider can be a person, an organization, or an entity responsible for making a service available to cloud consumers. A cloud provider builds the requested software/platform/ infrastructure services, manages the technical infrastructure required for providing the services, provisions the services at agreed-upon service levels, and protects the security and privacy of the services. As illustrated in Figure 4 – Cloud Provider: Major Activities, cloud providers undertake different tasks for the provisioning of the various service models.

For SaaS, the cloud provider deploys, configures, maintains, and updates the operation of the software applications on a cloud infrastructure so that the services are provisioned at the expected service levels to cloud consumers. The provider of SaaS assumes most of the responsibilities in managing and controlling the applications and the infrastructure, while the cloud consumers have limited administrative control of the applications.

For PaaS, the cloud provider manages the cloud infrastructure for the platform, and provisions tools and execution resources for the platform consumers to develop, test, deploy, and administer applications. Consumers have control over the applications and possibly the hosting environment settings, but cannot access the infrastructure underlying the platform including network, servers, operating systems, or storage.

For IaaS, the cloud provider provisions the physical processing, storage, networking, and other fundamental computing resources, as well as manages the hosting environment and cloud infrastructure for IaaS consumers. Cloud consumers deploy and run applications, have more control over the hosting environment and operating systems, but do not manage or control the underlying cloud infrastructure (e.g., the physical servers, network, storage, hypervisors, etc.).

The activities of cloud providers can be discussed in greater detail from the perspectives of *Service Deployment, Service Orchestration, Cloud Service Management, Security* and *Privacy*.

4.3.1 SERVICE DEPLOYMENT

As identified in the NIST cloud computing definition, a cloud infrastructure may be operated in one of the following deployment models: *public cloud*, *private cloud*, *community cloud*, or *hybrid cloud*. For the details related to the controls and management in the cloud, we refer readers to the NIST Special Publication 800-146, *NIST Cloud Computing Synopsis and Recommendations*.

A public cloud is one in which the cloud infrastructure and computing resources are made available to the general public over a public network. A public cloud is owned by an organization selling cloud services and serves a diverse pool of clients.

For private clouds, the cloud infrastructure is operated exclusively for a single organization. A private cloud gives the organization exclusive access to and usage of the infrastructure and computational resources. It may be managed either by the organization or by a third party, and may

be implemented at the organization's premise (i.e., *on-site private clouds*) or outsourced to a hosting company (i.e., *outsourced private clouds*).

Similar to private clouds, a community cloud may be managed by the organizations or by a third party, and may be implemented at the customer's location (i.e., *on-site community cloud*) or outsourced to a hosting company (i.e., *outsourced community cloud*). However, a community cloud serves a set of organizations that have common security, privacy, and compliance considerations, rather than serving a single organization as does a private cloud.

A hybrid cloud is a composition of two or more cloud deployment models (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability. As discussed in this section, both private clouds and community clouds can be either implemented on-site or outsourced to a third party. Therefore, each constituent cloud of a hybrid cloud can be one of the five variants.

4.3.2 SERVICE ORCHESTRATION

Service orchestration refers to the arrangement, coordination, and management of cloud infrastructure to provide the optimizing capabilities of cloud services, as a cost-effective way of managing IT resources, as dictated by strategic business requirements. Figure 5 shows the general requirements and processes for cloud providers to build each of the three service models.

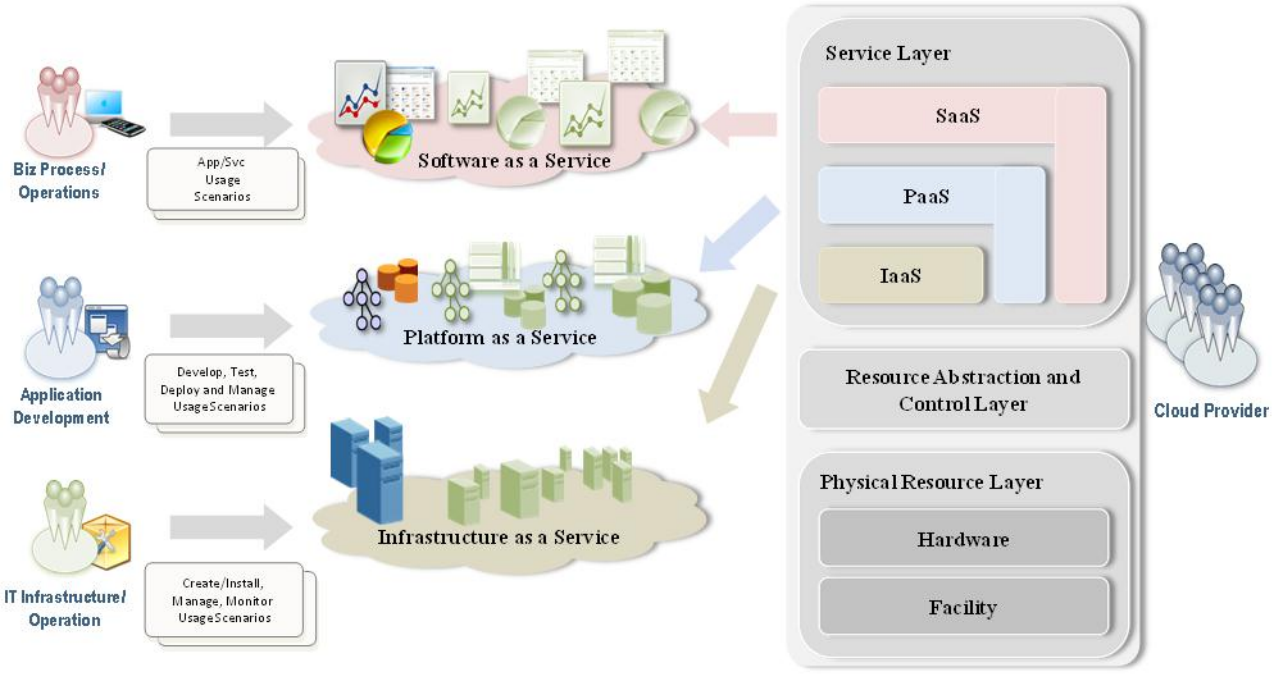

Figure 5 – Cloud Provider: Service Orchestration

A three-layered framework is identified for a generalized cloud system in Figure 5. The top layer is the service layer, where a cloud provider defines and provisions each of the three service models. This is where cloud consumers consume cloud services through the respective cloud interfaces.

The middle layer is the resource abstraction and control layer. This layer contains the system components that a cloud provider uses to provide and manage access to the physical computing resources through software abstraction. The layer typically includes software elements such as hypervisors, virtual machines, virtual data storage, and other resource abstraction and management components needed to ensure efficient, secure, and reliable usage. While virtual machine technology is commonly used at this layer, other means of providing the necessary software abstractions are not precluded. This layer provides "cloud readiness" with the five characteristics defined in the NIST definition of cloud computing.

The lowest layer in the framework is the physical resource layer, which includes all the physical computing resources. This layer includes hardware resources, such as computers (CPU and memory), networks (routers, firewalls, switches, network links, and interfaces), storage components (hard disks), and other physical computing infrastructure elements. It also includes facilities resources, such as heating, ventilation, and air conditioning (HVAC), power, communications, and other aspects of the physical plant.

Note that in this framework, the horizontal positioning of layers implies a stack in which the upper layer has a dependency on the lower layer. The resource abstraction and control layer build virtual cloud resources on top of the underlying physical resource layer and support the service layer where cloud services interfaces are exposed. The three service models can be built either on top of one another (i.e., SaaS built upon PaaS and PaaS built upon IaaS) or directly upon the underlying cloud infrastructure. For example, a SaaS application can be implemented and hosted on virtual machines from IaaS or directly on top of cloud resources without using IaaS.

4.3.3 CLOUD SERVICE MANAGEMENT

Cloud Service Management includes all of the service-related functions that are necessary for the management and operation of those services required by or proposed to cloud consumers. As illustrated in Figure 6, cloud service management can be described from the perspective of *business support, provisioning and configuration,* and from the perspective of *portability and interoperability* requirements.

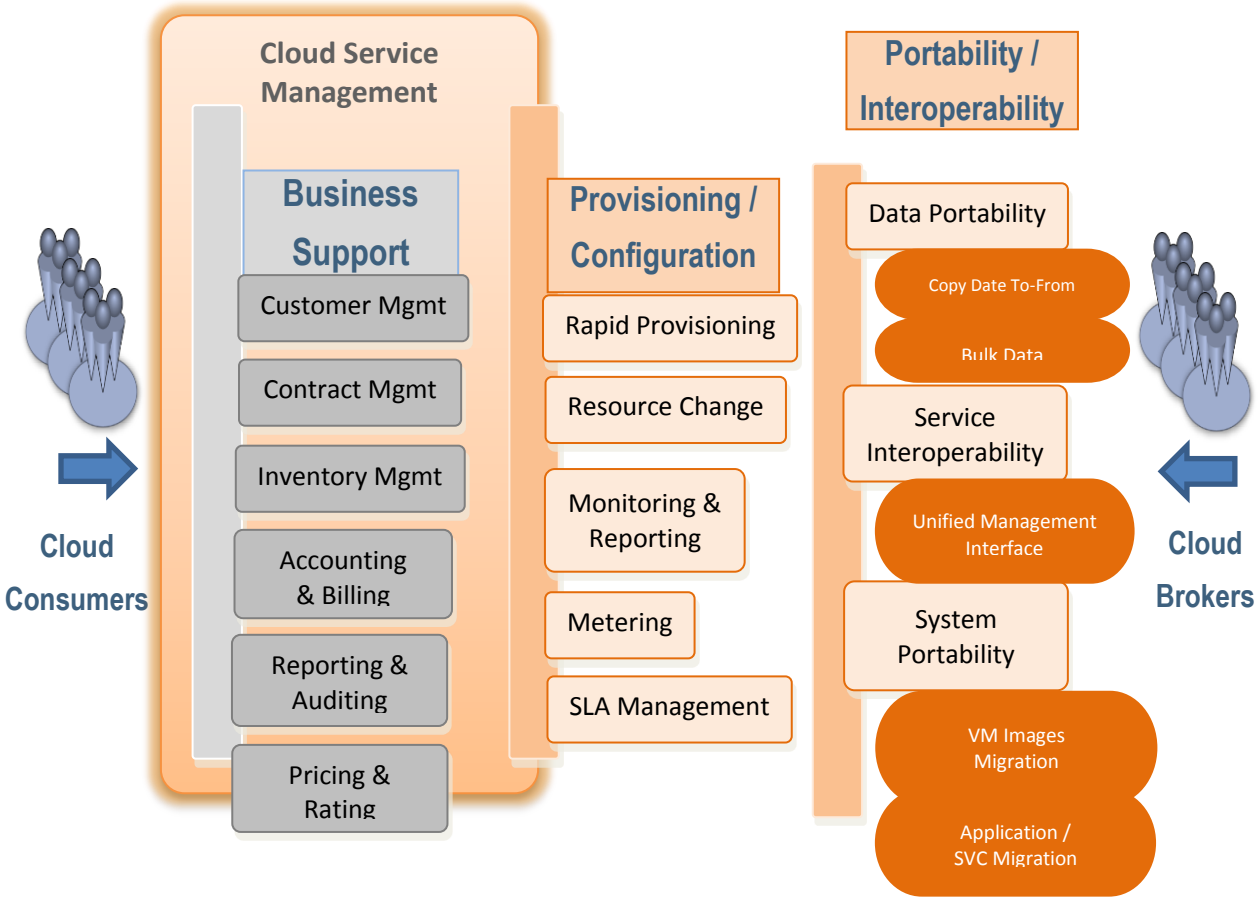

Figure 6 – Cloud Provider: Cloud Service Management

4.3.4 SECURITY

"As the Federal Government moves to the cloud, it must be vigilant to ensure the security and proper management of government information to protect the privacy of citizens and national security" (by Vivek Kundra, Federal Cloud Computing Strategy, February 2011.) In July 2012, the U.S. Department of Defense released a Cloud Computing Strategy, which stated "the Department has specific cloud computing challenges that require careful adoption considerations, especially in areas of cybersecurity, continuity of operations, information assurance (IA), and resilience." Also, in November 2012, NIST published a White Paper – *Challenging Security Requirements for U.S. Government Cloud Computing Adoption*. This document provides an overview of the high-priority security challenges perceived by federal agencies as impediments to the adoption of cloud computing.

Security is a cross-cutting function that spans all layers of the reference architecture (see Figure 12 – The Combined Conceptual Reference Diagram), involving end-to-end security that ranges from physical security to application security, and in general, the responsibility is shared between cloud provider and federal cloud consumer. For example, the protection of the physical resource layer (see Figure 5 – Cloud Provider: Service Orchestration) requires physical security that denies unauthorized access to the building, facility, resource, or stored information. Cloud Providers should ensure that the facility hosting cloud services is secure and that the staff has proper background checks. When data or applications are moved to a cloud, Cloud Consumers ensure that the cloud offering satisfies the security requirements and enforces the compliance rules. Several U.S. government agencies provide computer security guidance, and that the cloud system should support the most up-to-date guidance. It is also important to note that security, compliance, and policy requirements are a function of the legal jurisdiction of the country in which the cloud services are provided and can vary from country to country. An independent audit (see Section 3.4) should be conducted to verify the compliance with regulations or security policies.

4.3.5 PRIVACY

Cloud providers should protect the assured, proper, and consistent collection, processing, communication, use, and disposition of personal information (PI) and personally identifiable information (PII) in the cloud system. PII is the information that can be used to distinguish or trace an individual's identity, such as name, social security number, biometric records, etc., alone, or when combined with other personal or identifying information that is linked or linkable to a specific individual, such as date and place of birth, mother's maiden name, etc. The CIO Council – Privacy Committee[14] has identified privacy and protection of collected PII as one of the federal government key business imperatives. Though cloud computing provides a flexible solution for shared resources, software, and information, it also poses additional privacy challenges to consumers using the clouds.

The Digital Government Strategy[15] issued by the Federal Chief Information Officer (CIO) on May 23, 2012 sets forth a new vision of how government is to connect with and provide services to the American people, harnessing the power of digital technology and enabling citizens and the federal workforce to securely access government digital information, data, and services anywhere, and

[14] https://cio.gov/about/committees/privacy-committee/

[15] *Digital Government: Building a 21st Century Platform to Better Serve the American People* (May 23, 2012), (Strategy) http://www.whitehouse.gov/sites/default/files/omb/egov/digital-government/digital-government.html

anytime (Recommendations).[16] The Federal CIO Council released *Recommendations for Standardized Implementation of Digital Privacy Controls* (Recommendations), which discusses three fundamental privacy controls: PII Inventory, Privacy Impact Assessment (PIA), and Privacy Notice. The Recommendations are that agencies identify and consider all PII that may be collected or otherwise exposed through a particular digital technology, analyze the privacy risks through the data life cycle by conducting and updating a PIA (as needed), and provide notice to individuals of when and how their PII will be collected, used, retained, and disclosed.

Furthermore, federal agencies should be aware of the privacy concerns associated with the cloud computing environment where data are stored on a server that is not owned or controlled by the federal government. Privacy impact assessment (PIA) can be conducted, as needed, to measure how well the cloud system conforms to applicable legal, regulatory, and policy requirements regarding privacy. A PIA can help federal agencies comply with applicable privacy laws and regulations governing an individual's privacy, and to ensure confidentiality, integrity, and availability of an individual's personal information at every stage of development and operation.

In furthering the milestone action goal of the Digital Government Strategy for addressing digital privacy, records retention, and security issues, the National Archives & Records Administration (NARA) has issued Electronic Records Management (ERM) guidance for digital content created, collected, or maintained by federal agencies[17]. NARA also serves as managing partner of the E-Government ERM Initiative, coordinating the development and issuance of enterprise-wide ERM tools and electronic information standards, to support the interoperability of federal agency record systems and improve customer service (e.g., digital records access).[18]

[16] *Recommendations for Standardized Implementation of Digital Privacy Controls* (December 2012), https://cio.gov/wp-content/uploads/downloads/2012/12/Standardized_Digital_Privacy_Controls.pdf

[17] http://www.archives.gov/records-mgmt/initiatives/erm-guidance.html.

[18] http://www.archives.gov/records-mgmt/initiatives/erm-overview.html.

4.4 CLOUD AUDITOR

A cloud auditor is a party that can conduct independent assessment of cloud services, information system operations, performance, and the security of a cloud computing implementation. A cloud auditor can evaluate the services provided by a cloud provider in terms of security controls, privacy impact, performance, and adherence to service level agreement parameters.

Auditing is especially important for federal agencies as "agencies should include a contractual section enabling third parties to assess security controls of cloud providers" (*by Vivek Kundra, Federal Cloud Computing Strategy, February 2011*). Security controls are the management, operational, and technical safeguards or countermeasures employed within an organizational information system to protect the confidentiality, integrity, and availability of the system and its information. For security auditing, a cloud auditor can make an assessment of the security controls in the information system to determine the extent to which the controls are implemented correctly, operating as intended, and producing the desired outcome with respect to the security requirements for the system. The security auditing should include the verification of the compliance with regulation and security policy.

4.5 CLOUD BROKER

The NIST Reference Architecture, SP 500-292,[19] defines a Cloud Broker as an entity that manages the use, performance, and delivery of cloud services, and negotiates relationships between Cloud Providers and Cloud Consumers. As cloud computing evolves, the integration of cloud services may become too complex for cloud Consumers to manage. In such cases, a Cloud Consumer may request cloud services from a Cloud Broker instead of directly contacting a Cloud Provider. Cloud Brokers provide a single point of entry for managing multiple cloud services. The key defining feature that distinguishes a Cloud Broker from a Cloud Service Provider is the ability to provide a single consistent interface to multiple differing providers, whether the interface is for business or technical purposes. In general, Cloud Brokers provide services in three categories:

Intermediation: A Cloud Broker enhances a given service by improving some specific capability and providing value-added services to cloud Consumers. The improvement can be managing access to cloud services, identity management, performance reporting, enhanced security, etc.

Aggregation: A Cloud Broker combines and integrates multiple services into one or more new services. The Broker provides data and service integration and ensures the secure data movement between the cloud Consumer and multiple cloud Providers.

[19] http://www.cloudcredential.org/images/pdf_files/nist%20reference%20architecture.pdf

Arbitrage: Service arbitrage is similar to service aggregation except that the services being combined/consolidated are not fixed. Service arbitrage means a Broker has the flexibility to choose services from multiple service Providers.

A Cloud Broker may provide:

1. Business and relationship support services (business intermediation), and

2. Technical support service (aggregation, arbitrage, and technical intermediation), with a key focus on handling interoperability issues among multiple Providers.

4.6 CLOUD CARRIER

A cloud carrier acts as an intermediary that provides connectivity and transport of cloud services between cloud consumers and cloud providers. Cloud carriers provide access to consumers through network, telecommunication, and other access devices. For example, cloud consumers can obtain cloud services through network access devices, such as computers, laptops, mobile phones, mobile Internet devices (MIDs), etc. The distribution of cloud services is normally provided by network and telecommunication carriers or a *transport agent*, where a transport agent refers to a business organization that provides physical transport of storage media such as high-capacity hard drives. Note that a cloud provider will set up service level agreements (SLAs)[20] with a cloud carrier to provide services consistent with the level of SLAs offered to cloud consumers, and may require the cloud carrier to provide dedicated and encrypted connections between cloud consumers and cloud providers.

[20] SLAs are agreements under the umbrella of the overall cloud computing contract between a CSP and a cloud consumer. SLAs define acceptable service levels to be provided by the CSP to its customers in measurable terms. The ability of a CSP to perform at acceptable levels is consistent among SLAs, but the definition, measurement and enforcement of this performance varies widely among CSPs. A cloud consumer should ensure that CSP performance is clearly specified in all SLAs, and that all such agreements are fully incorporated, either by full text or by reference, into the CSP contract. [Source: *Creating Effective Cloud Computing Contracts for the Federal Government – Best Practices for Acquiring IT as a Service* https://cio.gov/wp-content/uploads/downloads/2012/09/cloudbestpractices.pdf