JEPPIAAR INSTITUTE OF TECHNOLOGY

"Self Belief | Self Discipline | Self Respect"

DEPARTMENT OF

COMPUTER SCIENCE AND ENGINEERING

LECTURE NOTES

CS8791 / CLOUD COMPUTING

(2017 Regulation)

Year/Semester: IV / VII

Prepared by

Dr. K. Tamilarasi, Professor / Dept. of CSE.

UNIT III CLOUD ARCHITECTURE, SERVICES AND STORAGE

Layered Cloud Architecture Design –NIST Cloud Computing Reference Architecture –Public, Private and Hybrid Clouds -laaS –PaaS –SaaS –Architectural Design Challenges –Cloud Storage –Storage-as-a-Service –Advantages of Cloud Storage –Cloud Storage Providers –S3.

3.1 Layered Cloud Architecture Design

- The architecture of a cloud is developed at three layers: infrastructure, platform and application as demonstrated in Figure 3.1.
- These three development layers are implemented with virtualization and standardization of hardware and software resources provisioned in the cloud.
- The services to public, private and hybrid clouds are conveyed to users through networking support over the Internet and intranets involved.
- It is clear that the infrastructure layer is deployed first to support IaaS services.
- This infrastructure layer serves as the foundation for building the platform layer of the cloud for supporting PaaS services.
- In turn, the platform layer is a foundation for implementing the application layer for SaaS applications.
- Different types of cloud services demand application of these resources separately.

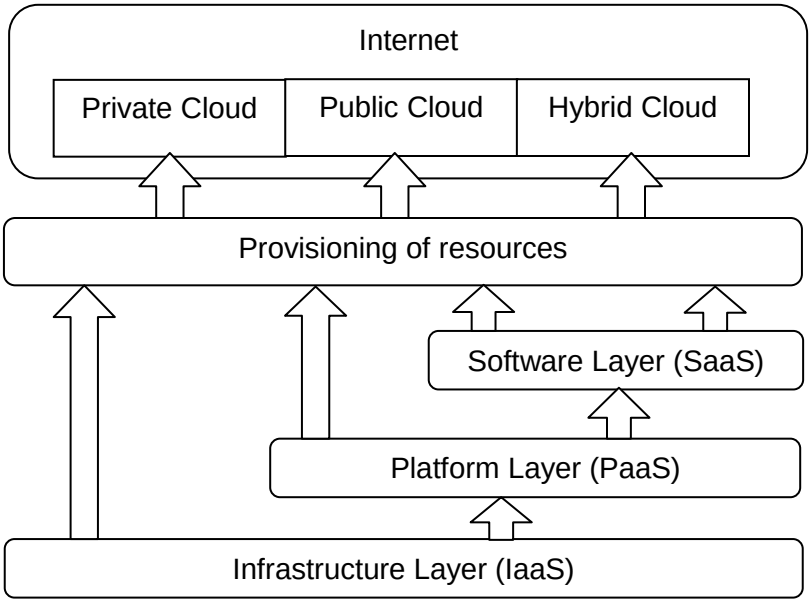


Figure 3.1 Layered architectural development

- The infrastructure layer is built with virtualized compute, storage and network resources.
- The abstraction of these hardware resources is meant to provide the flexibility demanded by users.
- Internally, virtualization realizes automated provisioning of resources and optimizes the infrastructure management process.
- The platform layer is for general purpose and repeated usage of the collection of software resources.
- This layer provides users with an environment to develop their applications, to test operation flows and to monitor execution results and performance.
- The platform should be able to assure users that they have scalability, dependability, and security protection.

- In a way, the virtualized cloud platform serves as a “system middleware” between the infrastructure and application layers of the cloud.

- The application layer is formed with a collection of all needed software modules for SaaS applications.

- Service applications in this layer include daily office management work such as information retrieval, document processing and calendar and authentication services.

- The application layer is also heavily used by enterprises in business marketing and sales, consumer relationship management (CRM), financial transactions and supply chain management.

- From the provider’s perspective, the services at various layers demand different amounts of functionality support and resource management by providers.

- In general, SaaS demands the most work from the provider, PaaS is in the middle, and IaaS demands the least.

- For example, Amazon EC2 provides not only virtualized CPU resources to users but also management of these provisioned resources.

- Services at the application layer demand more work from providers.

- The best example of this is the Salesforce.com CRM service in which the provider supplies not only the hardware at the bottom layer and the software at the top layer but also the platform and software tools for user application development and monitoring.

- In Market Oriented Cloud Architecture, as consumers rely on cloud providers to meet more of their computing needs, they will require a specific level of QoS to be maintained by their providers, in order to meet their objectives and sustain their operations.

- Market-oriented resource management is necessary to regulate the supply and demand of cloud resources to achieve market equilibrium between supply and demand.

- This cloud is basically built with the following entities:

    - Users or brokers acting on user's behalf submit service requests from anywhere in the world to the data center and cloud to be processed.
    - The request examiner ensures that there is no overloading of resources whereby many service requests cannot be fulfilled successfully due to limited resources.
    - The Pricing mechanism decides how service requests are charged. For instance, requests can be charged based on submission time (peak/off-peak), pricing rates (fixed/changing), or availability of resources (supply/demand).
    - The VM Monitor mechanism keeps track of the availability of VMs and their resource entitlements.
    - The Accounting mechanism maintains the actual usage of resources by requests so that the final cost can be computed and charged to users.
    - In addition, the maintained historical usage information can be utilized by the Service Request Examiner and Admission Control mechanism to improve resource allocation decisions.
    - The Dispatcher mechanism starts the execution of accepted service requests on allocated VMs.
    - The Service Request Monitor mechanism keeps track of the execution progress of service requests.

3.2 NIST Cloud Computing Reference Architecture

- NIST stands for National Institute of Standards and Technology

- The goal is to achieve effective and secure cloud computing to reduce cost and improve services

- NIST composed for six major workgroups specific to cloud computing

- Cloud computing target business use cases work group
- Cloud computing Reference architecture and Taxonomy work group
- Cloud computing standards roadmap work group
- Cloud computing SAJACC (Standards Acceleration to Jumpstart Adoption of Cloud Computing) work group
- Cloud Computing security work group

- Objectives of NIST Cloud Computing reference architecture

  - Illustrate and understand the various level of services
  - To provide technical reference
  - Categorize and compare services of cloud computing
  - Analysis of security, interoperatability and portability

- In general, NIST generates report for future reference which includes survey, analysis of existing cloud computing reference model, vendors and federal agencies.

- The conceptual reference architecture shown in figure 3.2 involves five actors. Each actor as entity participates in cloud computing

- Cloud consumer: A person or an organization that maintains a business relationship with and uses a services from cloud providers

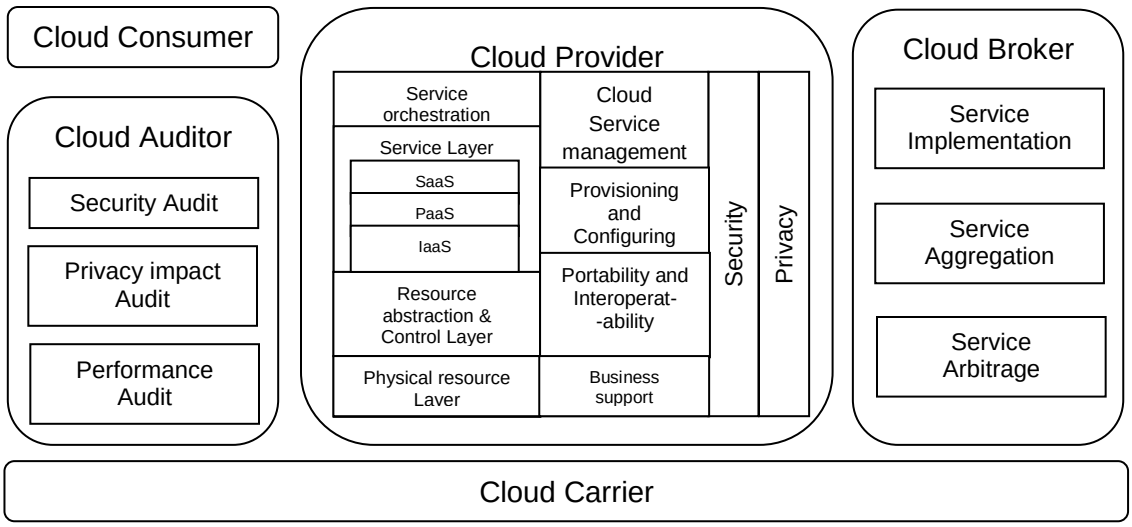


Figure 3.2 Conceptual reference model

- Cloud provider: A person, organization or entity responsible for making a service available to interested parties

- Cloud auditor: A party that conduct independent assessment of cloud services, information system operation, performance and security of cloud implementation

- Cloud broker: An entity that manages the performance and delivery of cloud services and negotiates relationship between cloud provider and consumer.

- Cloud carrier: An intermediary that provides connectivity and transport of cloud services from cloud providers to consumers.

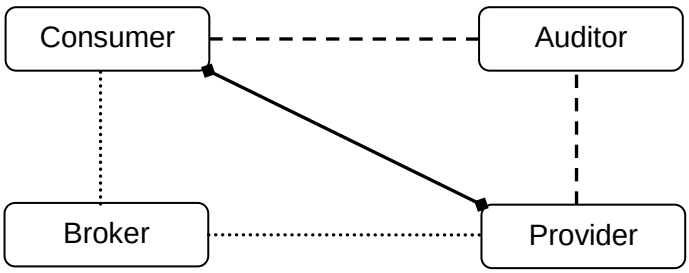


Figure 3.3 Interaction between actors

- Figure 3.3 illustrates the common interaction exist in between cloud consumer and provider where as the broker used to provide service to consumer and auditor collects the audit information.

- The interaction between the actors may lead to different use case scenario.

- Figure 3.4 shows one kind of scenario in which the Cloud consumer may request service from a cloud broker instead of contacting service provider directly. In this case, a cloud broker can create a new service by combining multiple services.

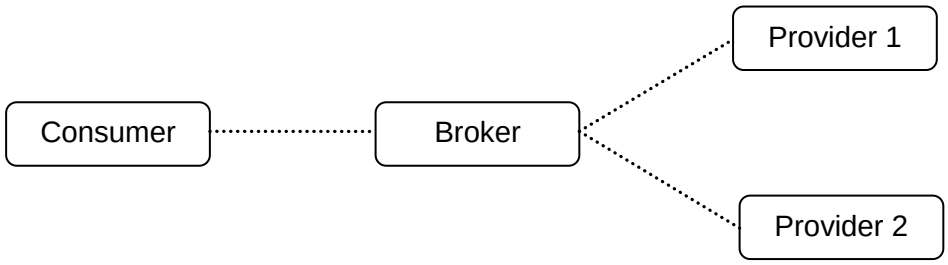


Figure 3.4 Service from Cloud Broker

- Figure 3.5 illustrates the usage of different kind of Service Level Agreement (SLA) between consumer, provider and carrier.

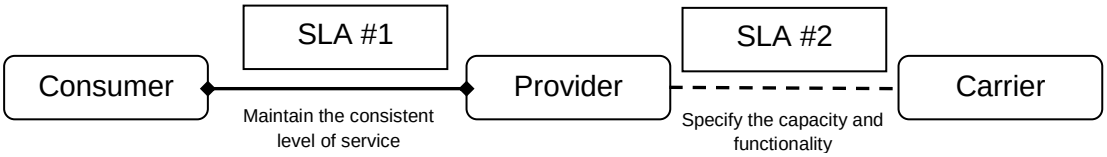


Figure 3.5 Multiple SLA between actors

- Figure 3.6 shows the scenario where the Cloud auditor conducts independent assessment of operation and security of the cloud service implementation.

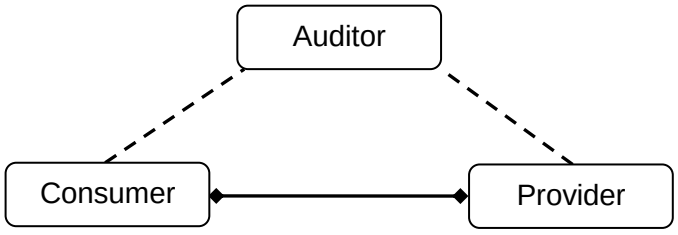


Figure 3.6 Independent assessments by cloud auditor

- Cloud consumer is a principal stake holder for the cloud computing service and requires service level agreements to specify the performance requirements fulfilled by a cloud provider.

- The service level agreement covers Quality of Service and Security aspects.

- Consumers have limited rights to access the software applications.

- There are three kinds of cloud consumers: SaaS consumers, PaaS Consumers and IaaS consumers.

- SaaS consumers are members directly access the software application. For example, document management, content management, social networks, financial billing and so on.

- PaaS consumers are used to deploy, test, develop and manage applications hosted in cloud environment. Database application deployment, development and testing is an example for these kind of consumer.

- IaaS Consumer can access the virtual computer, storage and network infrastructure. For example, usage of Amazon EC2 instance to deploy the web application.

- On the other hand, Cloud Providers have complete rights to access software applications.

- In Software as a Service model, cloud provider is allowed to configure, maintain and update the operations of software application.

- Management process is done by Integrated Development environment and Software Development Kit in Platform as a Service model.

- Infrastructure as a Service model covers Operating System and Networks.

- Normally, the service layer defines the interfaces for cloud consumers to access the computing services.

- Resource abstraction and control layer contains the system components that cloud provider use to provide and mange access to the physical computing resources through software abstraction.

- Resource abstraction covers virtual machine management and virtual storage management.

- Control layer focus on resource allocation, access control and usage monitoring.

- Physical resource layer includes physical computing resources such as CPU, Memory, Router, Switch, Firewalls and Hard Disk Drive.

- Service orchestration describes the automated arrangement, coordination and management of complex computing system.

- In cloud service management, business support entails the set of business related services dealing with consumer and supporting services which includes content management, contract management, inventory management, accounting service, reporting service and rating service.

- Provisioning of equipments, wiring and transmission is mandatory to setup a new service that provides a specific application to cloud consumer. Those details are described in Provisioning and Configuring management.

- Portability enforces the ability to work in more than one computing environment without major task. Similarly, Interoperatability means the ability of the system work with other system.

- Security factor is applicable to enterprise and Government. It may include privacy.

- Privacy is one applies to a cloud consumer's rights to safe guard his information from other consumers are parties.

- The main aim of Security and Privacy in cloud service management is to protect the system from vulnerable customers.

- Cloud auditor performs independent assessments among the services and cloud broker act as intermediate module.

- Service intermediation enhances a given service by improving some specific capability and providing value added services to cloud consumers,

- Service aggregation provides data integration. Cloud broker combines and integrate multiple service into one or more new services.

- Due to Service arbitrage, cloud broker has a flexibility to choose services from multiple providers.

- Cloud carrier is an intermediary that provides connectivity and transport of cloud service between cloud consumer and cloud provider.

- It provides access to cloud consumer with the help of network, telecommunication and other access devices where as distribution is done with transport agent,

- Transport agent is the business organization that provides physical transport of storage media.

3.3 Cloud Deployment Model

- 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.

- The differences are based on how exclusive the computing resources are made to a Cloud Consumer.

3.3.1 Public Cloud

- 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.

- Figure 4.7 presents a simple view of a public cloud and its customers.

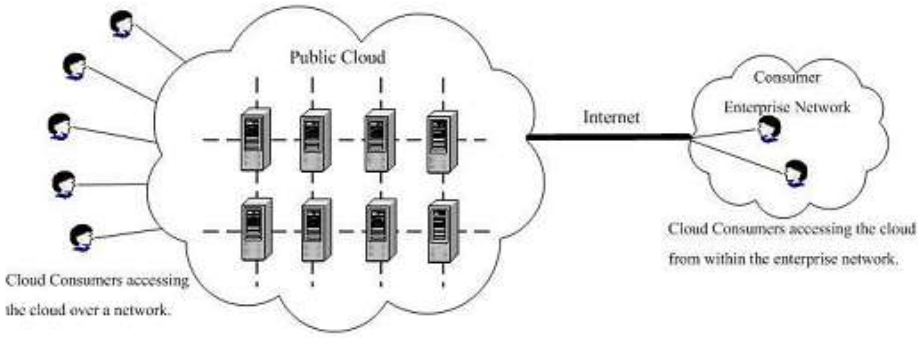


Figure 3.7 Public Cloud

3.3.1.1 Benefits of choosing a Public Cloud

- One of the main benefits that come with using public cloud services is near unlimited scalability.

- The resources are pretty much offered based on demand. So any changes in activity level can be handled very easily.

- This in turn brings with it cost effectiveness.
- Public cloud allows pooling of a large number of resources, users are benefiting from the savings of large scale operations.
- There are many services like Google Drive which are offered for free.
- Finally, the vast network of servers involved in public cloud services means that it can benefit from greater reliability.
- Even if one data center was to fail entirely, the network simply redistributes the load among the remaining enters making it highly unlikely that the public cloud would ever fail.
- In summary, the benefits of the public cloud are:
    - Easy scalability
    - Cost effectiveness
    - Increased reliability

3.3.1.2 Disadvantages of choosing a Public Cloud

- There are of course downsides to using public cloud services.
- At the top of the list is the fact that the security of data held within a public cloud is a cause for concern.
- It is often seen as an advantage that the public cloud has no geographical restrictions making access easy from everywhere, but on the flip side this could mean that the server is in a different country which is governed by an entirely different set of security and/or privacy regulations.

- This could mean that your data is not all that secure making it unwise to use public cloud services for sensitive data.

3.3.2 Private Cloud

- A private cloud gives a single Cloud Consumer's organization the exclusive access to and usage of the infrastructure and computational resources.

- It may be managed either by the Cloud Consumer organization or by a third party, and may be hosted on the organization's premises (i.e. on-site private clouds) or outsourced to a hosting company (i.e. outsourced private clouds).

- Figure 3.8 presents an on-site private cloud and an outsourced private cloud, respectively.

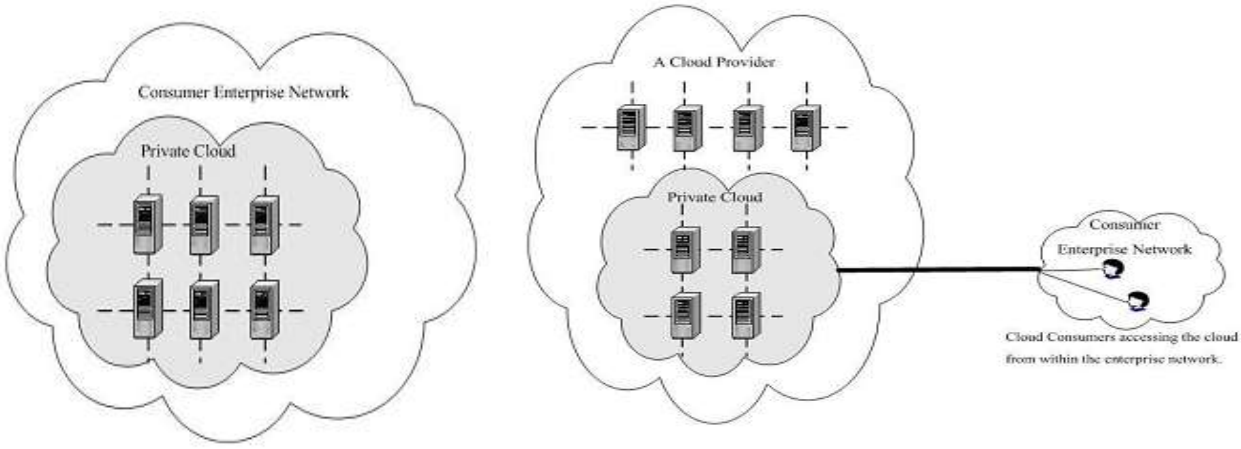


Figure 3.8 (a) On-site Private Cloud (b) Out-sourced Private Cloud

3.3.2.1 Benefits of choosing a Private Cloud

- The main benefit of choosing a private cloud is the greater level of security offered making it ideal for business users who need to store and/or process sensitive data.

- A good example is a company dealing with financial information such as bank or lender who is required by law to use secure internal storage to store consumer information.

- With a private cloud this can be achieved while still allowing the organization to benefit from cloud computing.

- Private cloud services also offer some other benefits for business users including more control over the server allowing it to be tailored to your own preferences and in house styles.

- While this can remove some of the scalability options, private cloud providers often offer what is known as cloud bursting which is when non sensitive data is switched to a public cloud to free up private cloud space in the event of a significant spike in demand until such times as the private cloud can be expanded.

- In summary, the main benefits of the private cloud are:

    - Improved security
    - Greater control over the server
    - Flexibility in the form of Cloud Bursting

3.3.2.2 Disadvantages of choosing a Private Cloud

- The downsides of private cloud services include a higher initial outlay, although in the long term many business owners find that this balances out and actual becomes more cost effective than public cloud use.

- It is also more difficult to access the data held in a private cloud from remote locations due to the increased security measures.

3.3.3 Community Cloud

- A community cloud serves a group of Cloud Consumers which have shared concerns such as mission objectives, security, privacy and compliance policy, rather than serving a single organization as does a private cloud.

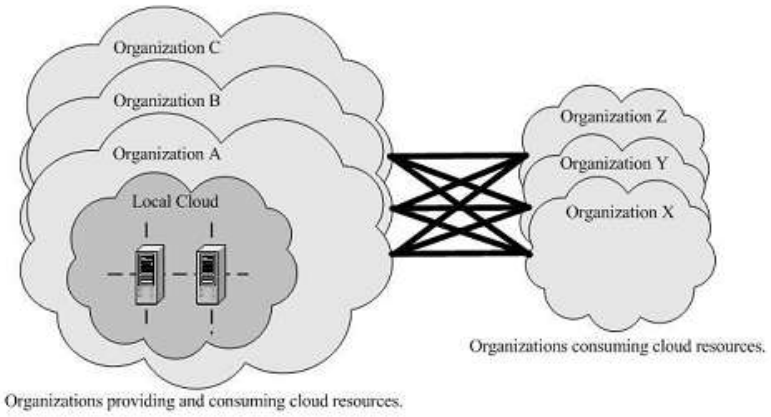
- Similar to private clouds, a community cloud may be managed by the organizations or by a third party and may be implemented on customer premise (i.e. on-site community cloud) or outsourced to a hosting company (i.e. outsourced community cloud).

- Figure 3.9 (a) depicts an on-site community cloud comprised of a number of participant organizations.

- A cloud consumer can access the local cloud resources, and also the resources of other participating organizations through the connections between the associated organizations.

- Figure 3.9 (b) shows an outsourced community cloud, where the server side is outsourced to a hosting company.

- In this case, an outsourced community cloud builds its infrastructure off premise, and serves a set of organizations that request and consume cloud services.



(a) On-site Community Cloud

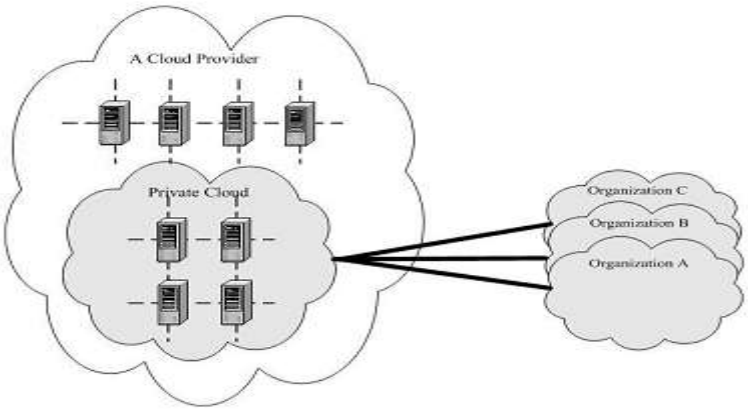


Figure 3.9 (b) Outsourced Community Cloud

3.3.3.1 Benefits of Choosing a Community Cloud

- Ability to easily share and collaborate
- Lower cost

3.3.3.2 Disadvantages of Choosing a Community Cloud

- Not the right choice for every organization
- Slow adoption to date

3.3.4 Hybrid Cloud

- A hybrid cloud is a composition of two or more clouds (on-site private, on-site community, off-site private, off-site community or public) that remain as distinct entities but are bound together by standardized or proprietary technology that enables data and application portability.

- Figure 3.10 illustrates a simple view of a hybrid cloud that could be built with a set of clouds in the five deployment model variants.

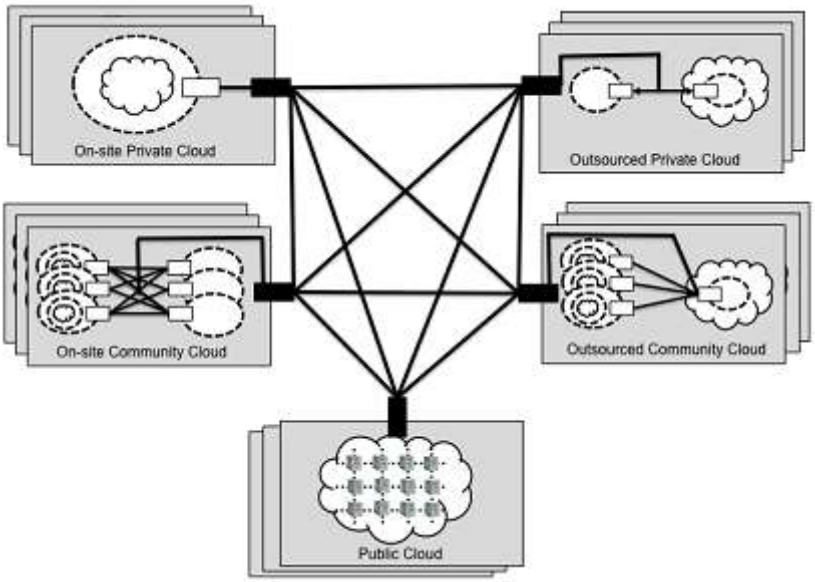


Figure 3.10 Hybrid Cloud

3.4 Cloud Service Model

- The development of cloud computing introduces the concept of everything as a Service (XaaS). This is one of the most important elements of cloud computing

- Cloud services from different providers can be combined to provide a completely integrated solution covering all the computing stack of a system.

- IaaS providers can offer the bare metal in terms of virtual machines where PaaS solutions are deployed.

- When there is no need for a PaaS layer, it is possible to directly customize the virtual infrastructure with the software stack needed to run applications.

- This is the case of virtual Web farms: a distributed system composed of Web servers, database servers and load balancers on top of which prepackaged software is installed to run Web applications.

- Other solutions provide prepackaged system images that already contain the software stack required for the most common uses: Web servers, database servers or LAMP stacks.

- Besides the basic virtual machine management capabilities, additional services can be provided, generally including the following:

    - SLA resource based allocation
    - Workload management
    - Support for infrastructure design through advanced Web interfaces
    - Integrate third party IaaS solutions

- Figure 3.11 provides an overall view of the components forming an Infrastructure as a Service solution.

- It is possible to distinguish three principal layers:

    - Physical infrastructure
    - Software management infrastructure
    - User interface

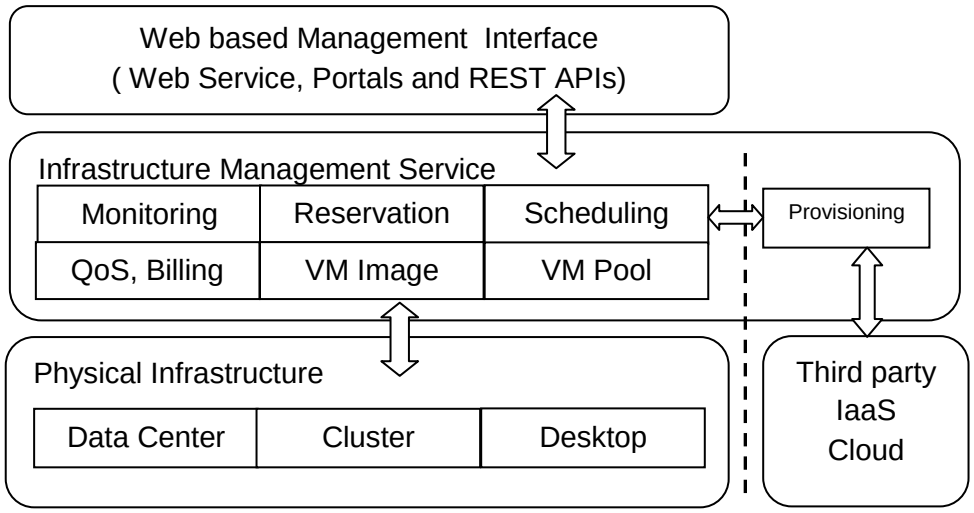


Figure 3.11 IaaS reference implementation

- At the top layer the user interface provides access to the services exposed by the software management infrastructure.

- Such an interface is generally based on Web 2.0 technologies: Web services, RESTful APIs and mash ups.

- Web services and RESTful APIs allow programs to interact with the service without human intervention, thus providing complete integration within a software system.

- The core features of an IaaS solution are implemented in the infrastructure management software layer.

- In particular, management of the virtual machines is the most important function performed by this layer.

- A central role is played by the scheduler, which is in charge of allocating the execution of virtual machine instances.

- The scheduler interacts with the other components such as

    - Pricing and billing component
    - Monitoring component
    - Reservation component
    - QoS/SLA management component
    - VM repository component
    - VM pool manager component
    - Provisioning component

- The bottom layer is composed of the physical infrastructure, on top of which the management layer operates.

- From an architectural point of view, the physical layer also includes the virtual resources that are rented from external IaaS providers.

- In the case of complete IaaS solutions, all three levels are offered as service.

- This is generally the case with public clouds vendors such as Amazon, GoGrid, Joyent, Rightscale, Terremark, Rackspace, ElasticHosts, and Flexiscale, which own large datacenters and give access to their computing infrastructures using an IaaS approach.

3.4.1 IaaS

- Infrastructure or Hardware as a Service (IaaS/HaaS) solutions are the most popular and developed market segment of cloud computing.

- They deliver customizable infrastructure on demand.

- The available options within the IaaS offering umbrella range from single servers to entire infrastructures, including network devices, load balancers, database servers and Web servers.

- The main technology used to deliver and implement these solutions is hardware virtualization: one or more virtual machines opportunely configured and interconnected define the distributed system on top of which applications are installed and deployed.

- Virtual machines also constitute the atomic components that are deployed and priced according to the specific features of the virtual hardware: memory, number of processors and disk storage.

- IaaS/HaaS solutions bring all the benefits of hardware virtualization: workload partitioning, application isolation, sandboxing and hardware tuning.

- From the perspective of the service provider, IaaS/HaaS allows better exploiting the IT infrastructure and provides a more secure environment where executing third party applications.

- From the perspective of the customer, it reduces the administration and maintenance cost as well as the capital costs allocated to purchase hardware.

- At the same time, users can take advantage of the full customization offered by virtualization to deploy their infrastructure in the cloud.

3.4.2 PaaS

- Platform as a Service (PaaS) solutions provide a development and deployment platform for running applications in the cloud.

- They constitute the middleware on top of which applications are built.

- A general overview of the features characterizing the PaaS approach is given in Figure 3.12.

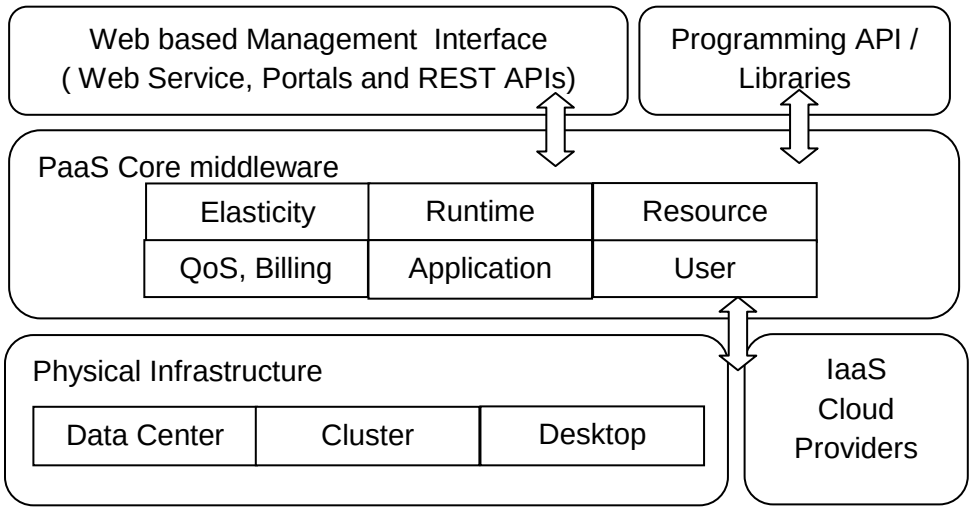


Figure 3.12 PaaS reference implementation

- Application management is the core functionality of the middleware.

- PaaS implementations provide applications with a runtime environment and do not expose any service for managing the underlying infrastructure.

- They automate the process of deploying applications to the infrastructure, configuring application components, provisioning and configuring supporting technologies such as load balancers and databases and managing system change based on policies set by the user.

- The core middleware is in charge of managing the resources and scaling applications on demand or automatically, according to the commitments made with users.

- From a user point of view, the core middleware exposes interfaces that allow programming and deploying applications on the cloud.

- Some implementations provide a completely Web based interface hosted in the cloud and offering a variety of services.

- It is possible to find integrated developed environments based on 4GL and visual programming concepts or rapid prototyping environments where applications are built by assembling mash ups and user defined components and successively customized.

- Other implementations of the PaaS model provide a complete object model for representing an application and provide a programming language-based approach.

- Developers generally have the full power of programming languages such as Java, .NET, Python and Ruby with some restrictions to provide better scalability and security.

- PaaS solutions can offer middleware for developing applications together with the infrastructure or simply provide users with the software that is installed on the user premises.

- In the first case, the PaaS provider also owns large datacenters where applications are executed

- In the second case, referred to in this book as Pure PaaS, the middleware constitutes the core value of the offering.

- PaaS implementation classified into three wide categories:

    - PaaS-I
    - PaaS-II
    - PaaS-III

- The first category identifies PaaS implementations that completely follow the cloud computing style for application development and deployment.

    - They offer an integrated development environment hosted within the Web browser where applications are designed, developed, composed, and deployed.
    - This is the case of Force.com and Longjump. Both deliver as platforms the combination of middleware and infrastructure.

- In the second class focused on providing a scalable infrastructure for Web application, mostly websites.

    - In this case, developers generally use the provider's APIs, which are built on top of industrial runtimes, to develop applications.
    - Google AppEngine is the most popular product in this category.
    - It provides a scalable runtime based on the Java and Python programming languages, which have been modified for providing a secure runtime environment and enriched with additional APIs and components to support scalability.

- AppScale, an open source implementation of Google AppEngine, provides interfacecompatible middleware that has to be installed on a physical infrastructure.

- The third category consists of all those solutions that provide a cloud programming platform for any kind of application, not only Web applications.

  - Among these, the most popular is Microsoft Windows Azure, which provides a comprehensive framework for building service oriented cloud applications on top of the .NET technology, hosted on Microsoft's datacenters.
  - Other solutions in the same category, such as Manjrasoft Aneka, Apprenda SaaSGrid, Appistry Cloud IQ Platform, DataSynapse, and GigaSpaces DataGrid, provide only middleware with different services.

- Some essential characteristics that identify a PaaS solution:

  - Runtime framework: This framework represents the software stack of the PaaS model and the most intuitive aspect that comes to people's minds when they refer to PaaS solutions.
  - Abstraction: PaaS solutions are distinguished by the higher level of abstraction that they provide.
  - Automation: PaaS environments automate the process of deploying applications to the infrastructure, scaling them by provisioning additional resources when needed.
  - Cloud services: PaaS offerings provide developers and architects with services and APIs, helping them to simplify the creation and delivery of elastic and highly available cloud application.

3.4.3 SaaS

- Software as a Service (SaaS) is a software delivery model that provides access to applications through the Internet as a Web based service.

- It provides a means to free users from complex hardware and software management by offloading such tasks to third parties, which build applications accessible to multiple users through a Web browser.

- On the provider side, the specific details and features of each customer's application are maintained in the infrastructure and made available on demand.

- The SaaS model is appealing for applications serving a wide range of users and that can be adapted to specific needs with little further customization.

- This requirement characterizes SaaS as a one-to-many software delivery model, whereby an application is shared across multiple users.

- This is the case of CRM and ERP applications that constitute common needs for almost all enterprises, from small to medium-sized and large business.

- Every enterprise will have the same requirements for the basic features concerning CRM and ERP and different needs can be satisfied with further customization.

- SaaS applications are naturally multitenant.

- Multitenancy, which is a feature of SaaS compared to traditional packaged software, allows providers to centralize and sustain the effort of managing large hardware infrastructures, maintaining as well as upgrading applications transparently to the users and optimizing resources by sharing the costs among the large user base.

- On the customer side, such costs constitute a minimal fraction of the usage fee paid for the software.

- The analysis carried out by Software Information and Industry Association (SIIA) was mainly oriented to cover application service providers (ASPs) and all their variations,

which capture the concept of software applications consumed as a service in a broader sense.

- ASPs already had some of the core characteristics of SaaS:
    - The product sold to customer is application access
    - The application is centrally managed
    - The service delivered is one-to-many
    - The service delivered is an integrated solution delivered on the contract, which means provided as promised.
- ASPs provided access to packaged software solutions that addressed the needs of a variety of customers.
- Initially this approach was affordable for service providers, but it later became inconvenient when the cost of customizations and specializations increased.
- The SaaS approach introduces a more flexible way of delivering application services that are fully customizable by the user by integrating new services, injecting their own components and designing the application and information workflows.
- Initially the SaaS model was of interest only for lead users and early adopters.
- The benefits delivered at that stage were the following:
    - Software cost reduction and total cost of ownership (TCO) were paramount
    - Service level improvements
    - Rapid implementation
    - Standalone and configurable applications
    - Rudimentary application and data integration
    - Subscription and pay as you go (PAYG) pricing

- With the advent of cloud computing there has been an increasing acceptance of SaaS as a viable software delivery model.

- This lead to transition into SaaS 2.0, which does not introduce a new technology but transforms the way in which SaaS is used.

- In particular, SaaS 2.0 is focused on providing a more robust infrastructure and application platforms driven by SLAs.

- SaaS 2.0 will focus on the rapid achievement of business objectives.

- Software as a Service based applications can serve different needs. CRM, ERP, and social networking applications are definitely the most popular ones.

- SalesForce.com is probably the most successful and popular example of a CRM service.

- It provides a wide range of services for applications: customer relationship and human resource management, enterprise resource planning, and many other features.

- SalesForce.com builds on top of the Force.com platform, which provides a fully featured environment for building applications.

- In particular, through AppExchange customers can publish, search and integrate new services and features into their existing applications.

- This makes SalesForce.com applications completely extensible and customizable.

- Similar solutions are offered by NetSuite and RightNow.

- NetSuite is an integrated software business suite featuring financials, CRM, inventory, and ecommerce functionalities integrated all together.

- RightNow is customer experience centered SaaS application that integrates together different features, from chat to Web communities, to support the common activity of an enterprise

- Another important class of popular SaaS applications comprises social networking applications such as Facebook and professional networking sites such as LinkedIn.

- Other than providing the basic features of networking, they allow incorporating and extending their capabilities by integrating third-party applications.

- Office automation applications are also an important representative for SaaS applications:

    - Google Documents and Zoho Office are examples of Web based applications that aim to address all user needs for documents, spreadsheets and presentation management.
    - These applications offer a Web based interface for creating, managing, and modifying documents that can be easily shared among users and made accessible from anywhere.

3.5 Architectural Design Challenges

3.5.1 Challenge 1: Service Availability and Data Lock-in Problem

- The management of a cloud service by a single company is often the source of single points of failure.

- To achieve HA, one can consider using multiple cloud providers.

- Even if a company has multiple data centers located in different geographic regions, it may have common software infrastructure and accounting systems.

- Therefore, using multiple cloud providers may provide more protection from failures.

- Another availability obstacle is distributed denial of service (DDoS) attacks.

- Criminals threaten to cut off the incomes of SaaS providers by making their services unavailable.

- Some utility computing services offer SaaS providers the opportunity to defend against DDoS attacks by using quick scale ups.

- Software stacks have improved interoperability among different cloud platforms, but the APIs itself are still proprietary. Thus, customers cannot easily extract their data and programs from one site to run on another.

- The obvious solution is to standardize the APIs so that a SaaS developer can deploy services and data across multiple cloud providers.

- This will rescue the loss of all data due to the failure of a single company.

- In addition to mitigating data lock-in concerns, standardization of APIs enables a new usage model in which the same software infrastructure can be used in both public and private clouds.

- Such an option could enable surge computing, in which the public cloud is used to capture the extra tasks that cannot be easily run in the data center of a private cloud.

3.5.2 Challenge 2: Data Privacy and Security Concerns

- Current cloud offerings are essentially public (rather than private) networks, exposing the system to more attacks.

- Many obstacles can be overcome immediately with well understood technologies such as encrypted storage, virtual LANs, and network middle boxes (e.g., firewalls, packet filters).

- For example, the end user could encrypt data before placing it in a cloud. Many nations have laws requiring SaaS providers to keep customer data and copyrighted material within national boundaries.

- Traditional network attacks include buffer overflows, DoS attacks, spyware, malware, rootkits, Trojan horses, and worms.

- In a cloud environment, newer attacks may result from hypervisor malware, guest hopping and hijacking or VM rootkits.

- Another type of attack is the man-in-the-middle attack for VM migrations.

- In general, passive attacks steal sensitive data or passwords.

- On the other hand, Active attacks may manipulate kernel data structures which will cause major damage to cloud servers.

3.5.3 Challenge 3: Unpredictable Performance and Bottlenecks

- Multiple VMs can share CPUs and main memory in cloud computing, but I/O sharing is problematic.

- For example, to run 75 EC2 instances with the STREAM benchmark requires a mean bandwidth of 1,355 MB/second.

- However, for each of the 75 EC2 instances to write 1 GB files to the local disk requires a mean disk write bandwidth of only 55 MB/second.

- This demonstrates the problem of I/O interference between VMs.
- One solution is to improve I/O architectures and operating systems to efficiently virtualize interrupts and I/O channels.
- Internet applications continue to become more data intensive.
- If we assume applications to be pulled apart across the boundaries of clouds, this may complicate data placement and transport.
- Cloud users and providers have to think about the implications of placement and traffic at every level of the system, if they want to minimize costs.
- This kind of reasoning can be seen in Amazon's development of its new CloudFront service.
- Therefore, data transfer bottlenecks must be removed, bottleneck links must be widened and weak servers should be removed.

3.5.4 Challenge 4: Distributed Storage and Widespread Software Bugs

- The database is always growing in cloud applications.
- The opportunity is to create a storage system that will not only meet this growth but also combine it with the cloud advantage of scaling arbitrarily up and down on demand.
- This demands the design of efficient distributed SANs.
- Data centers must meet programmer's expectations in terms of scalability, data durability and HA.

- Data consistence checking in SAN connected data centers is a major challenge in cloud computing.

- Large scale distributed bugs cannot be reproduced, so the debugging must occur at a scale in the production data centers.

- No data center will provide such a convenience. One solution may be a reliance on using VMs in cloud computing.

- The level of virtualization may make it possible to capture valuable information in ways that are impossible without using VMs.

- Debugging over simulators is another approach to attacking the problem, if the simulator is well designed.

3.5.5 Challenge 5: Cloud Scalability, Interoperability, and Standardization

- The pay as you go model applies to storage and network bandwidth; both are counted in terms of the number of bytes used.

- Computation is different depending on virtualization level.

- GAE automatically scales in response to load increases or decreases and the users are charged by the cycles used.

- AWS charges by the hour for the number of VM instances used, even if the machine is idle.

- The opportunity here is to scale quickly up and down in response to load variation, in order to save money, but without violating SLAs.

- Open Virtualization Format (OVF) describes an open, secure, portable, efficient and extensible format for the packaging and distribution of VMs.

- It also defines a format for distributing software to be deployed in VMs.

- This VM format does not rely on the use of a specific host platform, virtualization platform or guest operating system.

- The approach is to address virtual platform is agnostic packaging with certification and integrity of packaged software.

- The package supports virtual appliances to span more than one VM.

- OVF also defines a transport mechanism for VM templates and the format can apply to different virtualization platforms with different levels of virtualization.

- In terms of cloud standardization, the ability for virtual appliances to run on any virtual platform.

- The user is also need to enable VMs to run on heterogeneous hardware platform hypervisors.

- This requires hypervisor-agnostic VMs.

- And also the user need to realize cross platform live migration between x86 Intel and AMD technologies and support legacy hardware for load balancing.

- All these issues are wide open for further research.

3.5.6 Challenge 6: Software Licensing and Reputation Sharing

- Many cloud computing providers originally relied on open source software because the licensing model for commercial software is not ideal for utility computing.

- The primary opportunity is either for open source to remain popular or simply for commercial software companies to change their licensing structure to better fit cloud computing.

- One can consider using both pay for use and bulk use licensing schemes to widen the business coverage.

3.6 Cloud Storage

- Cloud storage means storing the data with a cloud service provider rather than on a local system.

- The end user can access the data stored on the cloud using an Internet link.

- Cloud storage has a number of advantages over traditional data storage.

- If the users stored some data on a cloud, they can get at it from any location that has Internet access.

- Workers do not need to use the same computer to access data nor do they have to carry around physical storage devices.

- Also, if any organization has branch offices, they can all access the data from the cloud provider.

- There are hundreds of different cloud storage systems, and some are very specific in what they do.

- Some are niche-oriented and store just email or digital pictures, while others store any type of data. Some providers are small, while others are huge and fill an entire warehouse.

- At the most rudimentary level, a cloud storage system just needs one data server connected to the Internet.

- A subscriber copies files to the server over the Internet, which then records the data. When a client wants to retrieve the data, the client accesses the data server with a web based interface and the server then either sends the files back to the client or allows the client to access and manipulate the data itself.

- More typically, however, cloud storage systems utilize dozens or hundreds of data servers.

- Because servers require maintenance or repair, it is necessary to store the saved data on multiple machines, providing redundancy.

- Without that redundancy, cloud storage systems could not assure clients that they could access their information at any given time.

3.6.1 Storage-as-a-Service

- The term Storage as a Service (another Software as a Service, or SaaS, acronym) means that a third-party provider rents space on their storage to end users who lack the budget or capital budget to pay for it on their own.

- Figure 3.13 illustrates the storage as a service where the data stored in cloud storage.

- It is also ideal when technical personnel are not available or have inadequate knowledge to implement and maintain that storage infrastructure.

- Storage service providers are nothing new, but given the complexity of current backup, replication, and disaster recovery needs, the service has become popular, especially among small and medium sized businesses.

- The biggest advantage to SaaS is cost savings.

- Storage is rented from the provider using a cost-per-gigabyte-stored or cost-per-data-transferred model.

- The end user does not have to pay for infrastructure. They simply pay for how much they transfer and save on the provider's servers.

Figure 3.13 Storage as a Service

- A customer uses client software to specify the backup set and then transfers data across a WAN.

- Examples:

    - Google Docs allows users to upload documents, spreadsheets, and presentations to Google's data servers. Those files can then be edited using a Google application.
    - Web email providers like Gmail, Hotmail, and Yahoo! Mail store email messages on their own servers. Users can access their email from computers and other devices connected to the Internet.
    - Flickr and Picasa host millions of digital photographs. Users can create their own online photo albums.

- YouTube hosts millions of user uploaded video files.
- Hostmonster and GoDaddy store files and data for many client web sites.
- Facebook and MySpace are social networking sites and allow members to post pictures and other content. That content is stored on the company's servers.
- MediaMax and Strongspace offer storage space for any kind of digital data.

- To secure data, most systems use a combination of the listed techniques:

- Encryption: A complex algorithm is used to encode information. To decode the encrypted files, a user needs the encryption key.

- Authentication processes: This requires a user to create a name and password.

- Authorization practices: The client lists the people who are authorized to access information stored on the cloud system. Many corporations have multiple levels of authorization.

- The other concern is reliability.

- If a cloud storage system is unreliable, it becomes a liability. No one wants to save data on an unstable system, nor would they trust a company that is financially unstable.

- Most cloud storage providers try to address the reliability concern through redundancy, but the possibility still exists that the system could crash and leave clients with no way to access their saved data.

3.6.2 Advantages of Cloud Storage

- Cloud storage is becoming an increasingly attractive solution for organizations.

- Cloud storage providers balance server loads and move data among various datacenters, ensuring that information is stored close and thereby available quickly while using the data.

- Storing data on the cloud is advantageous, because it allows the user to protect the data in case there's a disaster.

- Having the data stored off-site can be the difference between closing the door for good or being down for a few days or weeks.

- Which storage vendor to go with can be a complex issue, and how the end user technology interacts with the cloud can be complex.

- For instance, some products are agent based and the application automatically transfers information to the cloud via FTP.

- But others employ a web front end and the user has to select local files on their computer to transmit.

- Amazon S3 is the best known storage solution, but other vendors might be better for large enterprises.

- For instance, those who offer service level agreements and direct access to customer support are critical for a business moving storage to a service provider

3.6.3 Cloud Storage Providers

- There are hundreds of cloud store providers every day.

- This is simply a listing of what some of the big players in the game have to offer and anyone can use it as a starting guide to determine if their services match user's needs.

- Amazon and Nirvanix are the current industry top dogs, but many others are in the field, including some well known names.

- Google offers cloud storage solution called GDrive.

- EMC is readying a storage solution and IBM already has a number of cloud storage options called Blue Cloud.

3.6.4 S3

- The well known cloud storage service is Amazon's Simple Storage Service (S3), which is launched in 2006.

- Amazon S3 is designed to make web scale computing easier for developers.

- Amazon S3 provides a simple web services interface that can be used to store and retrieve any amount of data, at any time, from anywhere on the Web.

- It gives any developer access to the same highly scalable data storage infrastructure that Amazon uses to run its own global network of web sites.

- The service aims to maximize benefits of scale and to pass those benefits on to developers.

- Amazon S3 is intentionally built with a minimal feature set that includes the following functionality:

    - Write, read, and delete objects containing from 1 byte to 5 gigabytes of data each. The number of objects that can be stored is unlimited.
    - Each object is stored and retrieved via a unique developer assigned key.
    - Objects can be made private or public and rights can be assigned to specific users.
    - Uses standards based REST and SOAP interfaces designed to work with any Internet development toolkit.

- Design Requirements Amazon built S3 to fulfill the following design requirements:

    - Scalable: Amazon S3 can scale in terms of storage, request rate and users to support an unlimited number of web-scale applications.
    - Reliable: Store data durably with 99.99 percent availability. Amazon says it does not allow any downtime.
    - Fast: Amazon S3 was designed to be fast enough to support high-performance applications. Server-side latency must be insignificant relative to Internet latency.
    - Inexpensive: Amazon S3 is built from inexpensive commodity hardware components.
    - Simple: Building highly scalable, reliable, fast and inexpensive storage is difficult.

- Design Principles Amazon used the following principles of distributed system design to meet Amazon S3 requirements:

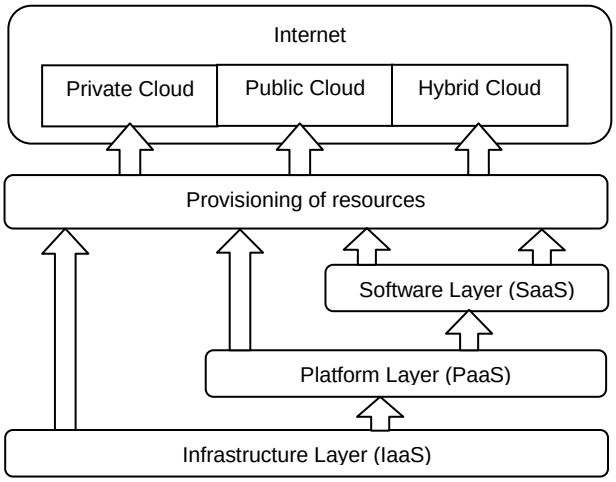
    - Decentralization: It uses fully decentralized techniques to remove scaling bottlenecks and single points of failure.
    - Autonomy: The system is designed such that individual components can make decisions based on local information.
    - Local responsibility: Each individual component is responsible for achieving its consistency. This is never the burden of its peers.
    - Controlled concurrency: Operations are designed such that no or limited concurrency control is required.
    - Failure toleration: The system considers the failure of components to be a normal mode of operation and continues operation with no or minimal interruption.
    - Controlled parallelism: Abstractions used in the system are of such granularity that parallelism can be used to improve performance and robustness of recovery or the introduction of new nodes.
    - Symmetry: Nodes in the system are identical in terms of functionality, and require no or minimal node specific configuration to function.
    - Simplicity: The system should be made as simple as possible, but no simpler.

- Amazon keeps its lips pretty tight about how S3 works, but according to Amazon, S3's design aims to provide scalability, high availability, and low latency at commodity costs.

- S3 stores arbitrary objects at up to 5GB in size, and each is accompanied by up to 2KB of metadata.
- Objects are organized by buckets.
- Each bucket is owned by an AWS account and the buckets are identified by a unique user assigned key.
- Buckets and objects are created, listed and retrieved using either a REST or SOAP interface.
- Objects can also be retrieved using the HTTP GET interface or via BitTorrent.
- An access control list restricts who can access the data in each bucket.
- Bucket names and keys are formulated so that they can be accessed using HTTP.
- Requests are authorized using an access control list associated with each bucket and object, for instance: http://s3.amazonaws.com/samplebucket/samplekey
- The Amazon AWS Authentication tools allow the bucket owner to create an authenticated URL with a set amount of time that the URL will be valid.
- Bucket items can also be accessed via a BitTorrent feed, enabling S3 to act as a seed for the client.
- Buckets can also be set up to save HTTP log information to another bucket.
- This information can be used for later data mining.

TWO MARK QUESTIONS

1. Illustrate architecture of a cloud is developed using three layers.



2. What is Market-Oriented Cloud Architecture?

- As consumers rely on cloud providers to meet more of their computing needs, they will require a specific level of QoS to be maintained by their providers, in order to meet their objectives and sustain their operations.
- Market-oriented resource management is necessary to regulate the supply and demand of cloud resources to achieve market equilibrium between supply and demand.

3. List the entities involved in the cloud platform.

- Users or brokers and Request examiner
- Pricing mechanism and VM Monitor mechanism
- Accounting mechanism
- Service Request Examiner and Admission Control mechanism
- Dispatcher mechanism
- Service Request Monitor mechanism

4. List the objectives of NIST Cloud Computing reference architecture

- Illustrate and understand the various level of services
- To provide technical reference
- Categorize and compare services of cloud computing
- Analysis of security, interoperatability and portability

5. Mention the major actors involved in NIST reference model.
   - Cloud consumer
   - Cloud provider
   - Cloud auditor
   - Cloud broker
   - Cloud carrier

6. Define service orchestration.

   - Service orchestration describes the automated arrangement, coordination and management of complex computing system.

7. Differentiate between Public cloud and Private Cloud.

   - 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.
   - A private cloud gives a single Cloud Consumer's organization the exclusive access to and usage of the infrastructure and computational resources.
   - It may be managed either by the Cloud Consumer organization or by a third party, and may be hosted on the organization's premises (i.e. on-site private clouds) or outsourced to a hosting company (i.e. outsourced private clouds).

8. Tabulate the merits and demerits of Choosing a Community Cloud.

Merits	Demerits

• Ability to easily share and collaborate • Lower cost	• Not the right choice for every organization • Slow adoption to date

9. What is IaaS or HaaS?

- Infrastructure or Hardware-as-a-Service (IaaS/HaaS) solutions are the most popular and developed market segment of cloud computing.
- They deliver customizable infrastructure on demand.

10. What is PaaS?

- Platform-as-a-Service (PaaS) solutions provide a development and deployment platform for running applications in the cloud.
- They constitute the middleware on top of which applications are built.

11. Classify Pass Implementation

- PaaS implementation classified into three wide categories:
- PaaS-I, PaaS-II, and PaaS-III.

12. What is SaaS?

- Software-as-a-Service (SaaS) is a software delivery model that provides access to applications through the Internet as a Web-based service.
- It provides a means to free users from complex hardware and software management by offloading such tasks to third parties, which build applications accessible to multiple users through a Web browser.

13. What is SaaS 2.0?

- SaaS 2.0 is not a new technology but transforms the way in which SaaS is used.

- In particular, SaaS 2.0 is focused on providing a more robust infrastructure and application platforms driven by SLAs.
- SaaS 2.0 will focus on the rapid achievement of business objectives.

14. List the six architectural design challenges in cloud.

- Service Availability and Data Lock-in Problem
- Data Privacy and Security Concerns
- Unpredictable Performance and Bottlenecks
- Distributed Storage and Widespread Software Bugs
- Cloud Scalability, Interoperability, and Standardization
- Software Licensing and Reputation Sharing

15. What is cloud storage?

- Cloud storage means storing the data with a cloud service provider rather than on a local system. The end user can access the data stored on the cloud using an Internet link.
- Cloud storage has a number of advantages over traditional data storage.
- If the users stored some data on a cloud, they can get at it from any location that has Internet access.

16. What is Storage-as-a-Service?

- The term Storage as a Service means that a third-party provider rents space on their storage to end users who lack the budget or capital budget to pay for it on their own.
- It is also ideal when technical personnel are not available or have inadequate knowledge to implement and maintain that storage infrastructure.

17. List the real time examples for cloud storage.

- Google Docs allows users to upload documents, spreadsheets, and presentations to Google’s data servers.

- Web email providers like Gmail, Hotmail, and Yahoo! Mail store email messages on their own servers.
- Flickr and Picasa host millions of digital photographs. YouTube hosts millions of user-uploaded video files.
- Hostmonster and GoDaddy store files and data for many client web sites.
- Facebook and MySpace are social networking sites and allow members to post pictures and other content.
- MediaMax and Strongspace offer storage space for any kind of digital data.

18. How to secure data in cloud storage?

- Most systems use a combination of following techniques:
    - Encryption
    - Authentication processes
    - Authorization practices

19. List the advantages of cloud storage.

- Storing data on the cloud is advantageous, because it allows you to protect your data in case there's a disaster.
- Having your data stored off-site can be the difference between closing your door for good or being down for a few days or weeks.
- Which storage vendor to go with can be a complex issue, and how the end user technology interacts with the cloud can be complex.

20. What is S3?

- The best-known cloud storage service is Amazon's Simple Storage Service (S3), which launched in 2006.
- Amazon S3 is designed to make web-scale computing easier for developers.
- Amazon S3 provides a simple web services interface that can be used to store and retrieve any amount of data, at any time, from anywhere on the Web.
- It gives any developer access to the same highly scalable data storage infrastructure that Amazon uses to run its own global network of web sites.

21. What are the design requirements considers by Amazon to build S3?

- Scalable
- Reliable
- Fast
- Inexpensive
- Simple

22. What are the design principles considers by Amazon to meet S3 requirements?

- Decentralization
- Autonomy
- Local responsibility
- Controlled concurrency
- Failure toleration
- Controlled parallelism
- Symmetry
- Simplicity

23. How the data stored in S3?

- S3 stores arbitrary objects at up to 5GB in size, and each is accompanied by up to 2KB of metadata.
- Objects are organized by buckets.
- Each bucket is owned by an AWS account and the buckets are identified by a unique, user-assigned key.
- Buckets and objects are created, listed, and retrieved using either a REST-style or SOAP interface.