**Lesson No. 19**

## TRUST ISSUES IN CLOUD

**Module No – 094: Brief overview (~~more in Lesson 39~~):**

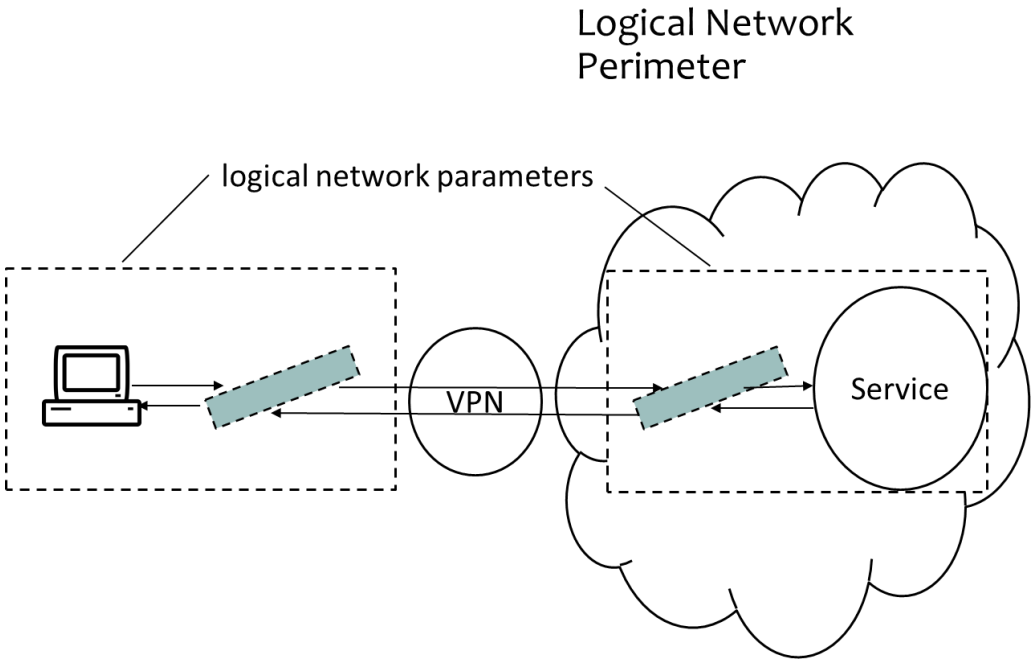
- Link between Privacy, Security and Trust:
  - Privacy: The confidentiality of data related to a person or organization.
  - Security: The preservation of confidentiality, integrity and availability of data.
  - Trust: The state of accepting a vulnerability on the base of positive expectations.
- Privacy issues of Cloud Computing:
  - Lack of user control
  - Lack of training and expertise
  - Possibility of secondary (/unauthorize) use of consumer data
  - Legal compliance
- Security issues of Cloud Computing:
  - Overlapping security boundaries
  - Unauthorized access
  - Lack of interoperability of security policies
  - Uncertainty of data deletion
  - Compromise of management console
  - Backup vulnerabilities
  - Isolation failure in multi-tenant applications
  - Inadequate monitoring and audit
- Trust in Cloud: The consumer's trust in Cloud is affected by the privacy and security vulnerabilities of Cloud as discussed before.
  - Further, due to lack of transparency the blame of responsibility is difficult to be placed if the provider is outsourcing the IT resources from a chain of outsourcing.
  - The pay-as-you-go and *on-demand* provision of cloud resources may be subject to low level of trust.
  - The lack of trust is the key factor for user reluctance to use Cloud services.
  - Consumer feels a *lack of control* in shifting to Cloud.
  - The companies shifting from on-premises setups to public Clouds are more concerned about data security and health than of the servers.
  - Concerns are present regarding foreign governments' access to consumers' data on Cloud.
  - The analysis of tradeoffs of Cloud privacy, security, cost and benefits determines the decision of Cloud usage.
- Conclusion: The consumers' trust can be assured through the safeguarding of personal/confidential/sensitive data. The existence0enhancement of transparency and accountability can increase the trust. Research should be conducted to quantify and model the trust and trust management, so that approaches for strengthening the consumers' trust can be proposed, tested, and/or enhanced.

**Lesson No. 20**

## MECHANISMS RELATED TO CLOUD INFRASTRUCTURE

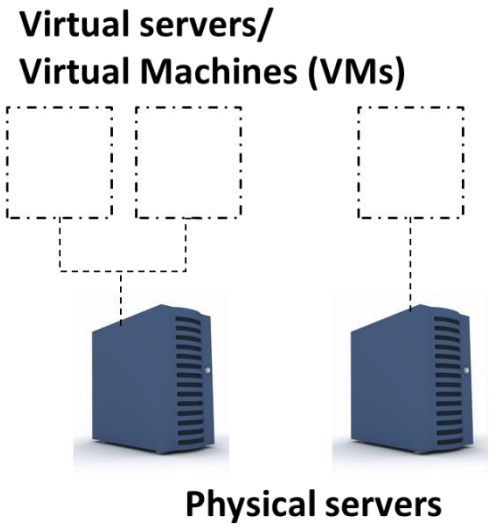
### Module No – 095: Logical Network Perimeter:

- It establishes the boundary of virtual network to hold with in and isolate a set of related cloud-IT resources that may be distributed physically. Implemented as virtual environment, it has the following components:
  - *Virtual Firewall* to filter the traffic of isolated network to and from Internet.
  - *Virtual Network* consisting of virtual nodes and virtual links.

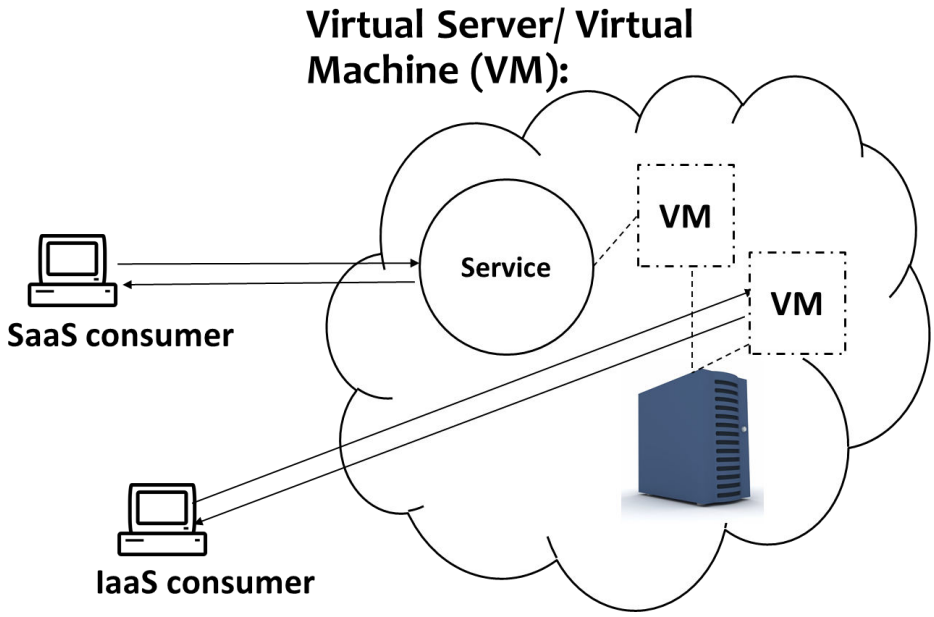



### Module No – 096: Virtual Server or Virtual Machine (VM):

- Virtual Server: Virtual servers or Virtual Machines (VMs) emulate the physical servers.
  - Each virtual server can host numerous IT resources, cloud-based solutions and other cloud computing mechanisms. Depending upon the capacity, a physical server may host multiple virtual servers.




Virtual Servers or Virtual Machines(VMs)




Cloud service consumers, Cloud service and VM relationship

- In order to rapidly provision the VMs with installed and preconfigured software such as OS, programming platforms etc., the virtual servers are *cloned* by templates.
- A *template* is a master copy of virtual server. It contains the configuration, installed software, any configured virtual devices and disk contents.
- A consumer can:
  - Connect to a self-service portal of Cloud provider.
  - Choose a suitable template.
  - Instantiate a virtual server through administrative portal which works with the help of virtual infrastructure manager (VIM) module.
  - Customize the virtual server through usage and administrative portal.

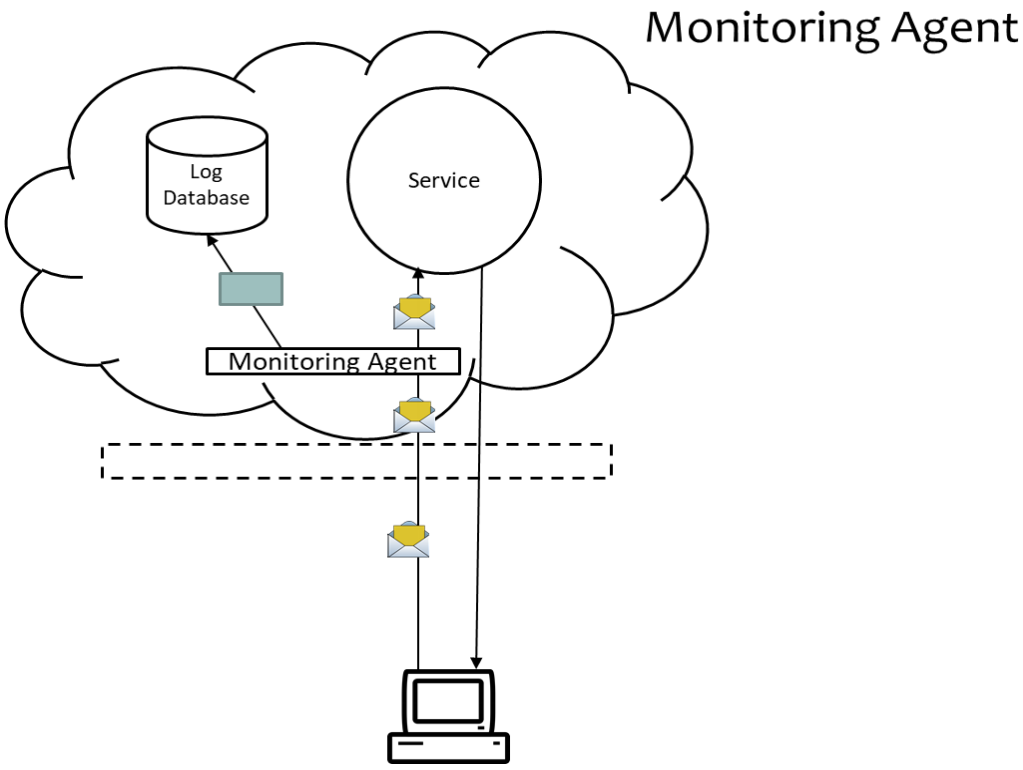
 

## Module No – 097: Cloud Storage Device:

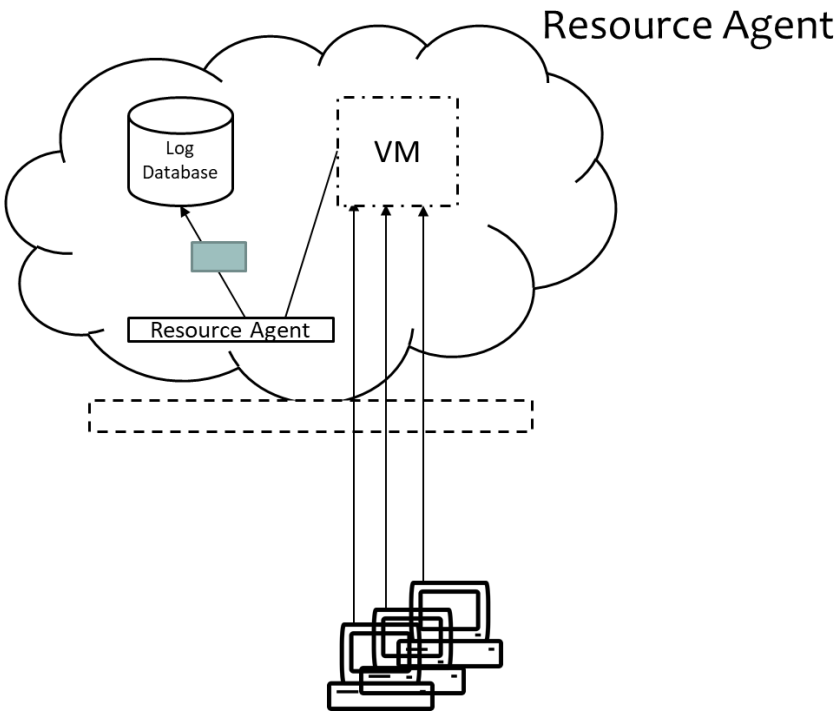
- It represents the storage mechanisms devised specifically for cloud-based provisioning.
  - Instances of these devices can be virtualized.
  - Support dynamic scaling
  - Can be accessed remotely by Cloud storage services.
  - The cloud storage mechanisms support the following (but not limited to) logical units of data storage:
    - Files (data grouped into files that are located in folders)
    - Blocks (the smallest unit of data that is individually accessible)
    - Datasets (such as data arranged in databases)
    - Objects (data and associated meta data)
      - Each of these levels is associated with a certain type of technical interface consisting of a specific type of cloud storage device with a Cloud storage service used to use its API.
    - Network Storage Interface: For file and block storage
    - Object Storage Interface: Based upon technologies that support a range of data and media types. The storage mechanism can be accessed by REST or SOAP based web services.
    - Database Storage Interface: Supports the relational (SQL based) and non-relational databases (NoSQL storage).
    - ~~Database Storage Interface:~~ Data stored in relational database is more structured and normalized than non-relational database. The relational databases have higher processing overhead. While the non-relational have high data-redundancy. Also, transactions and joins are not supported. The relational databases have higher processing overhead than non-relational database. The non-relational databases storage have high data-redundancy and non-structured data. The relational-database functions such as transactions and joins are not supported in non-relational database storage.

## Module No – 098: Cloud Usage Monitor

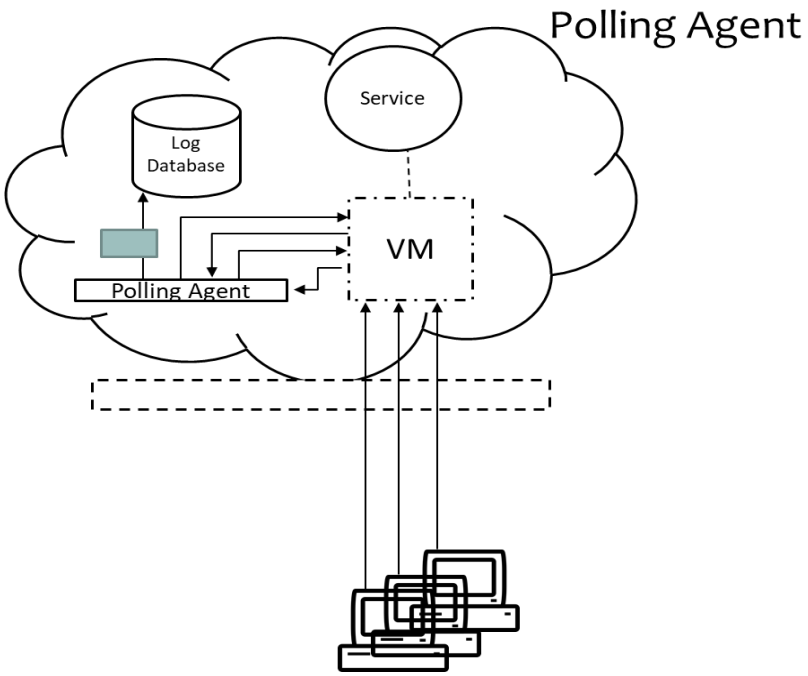
- It is a software used to collect and process the data related to Cloud-based IT resources.
  - The reporting and analysis requirements of the Cloud usage module determines the scope and volume of data collected/extracted.
- There are a few generic types or formats of Cloud usage monitors:
  - Monitoring Agent: It transparently monitors and analyzes the dataflow over communication paths. It measures the network traffic and messages.




- Resource Agent: Collects the resource usage data related to certain events such as initiating, suspending, resuming and vertical scaling. It interacts with the Cloud resource management module.



- Polling Agent: Collects the Cloud service usage data after periodic polling to IT resources. For example the uptime/downtime of a Cloud service. Records the updated status of the resource.




## Module No – 099:Resource Replication

- It is a technique by which multiple copies of the IT resources are created to increase the availability and productivity of the IT resources. Virtualization technology is used for Cloud IT resources' replication.
- For example, due to a physical server failure and in order to over come the resultant downtime of a Cloud service deployed over a VM hosted by that physical server, the entire VM along with the software (Cloud service implementation) is replicated to another server.
- Another example is the horizontal scaling of IT resources such as increasing or decreasing of Cloud service instances by replication of VM hosting the service instance, corresponding to workload.
- The resource replication process yields the IT resources which are monitored under the Cloud usage monitor mechanism.
- Resource replication is also essential for pay-as-you-go type of usage & billing.

## Module No – 100:Ready-Made Environment

- This mechanism represents the provisioning of preconfigure PaaS instances with ready to use and customizable programming environments. Provide the dependable PaaS instances.
- Time efficient provisioning

- Typically include:
  - Software development tools
  - Databases
  - Middleware
  - Governance tools
- The middleware is provided to support multi-tenant platforms to develop and deploy the complementary web services for SaaS scenarios.
- Overall, the ready-made environment mechanism supports the development and production level deployment of Cloud services.

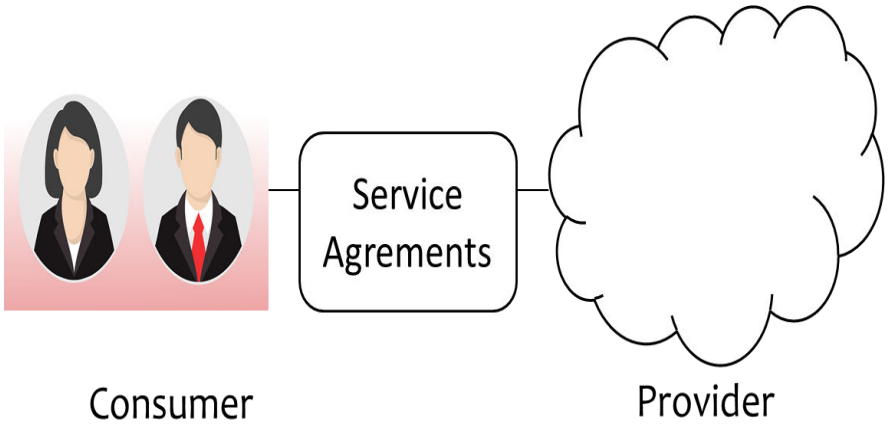
**Lesson No. 21**

## SERVICE AGREMENTS(SAs)

**Module No – 101:**

- NIST identifies that the consumer and provider are under a legal agreement or terms of service.
- The agreement has two parts:
  - Service Agreement
  - Service Level Agreement (SLA)
- Service agreement contains the legal terms of contract.
- The SLA contains the technical performance promises by the provider and the remedies for performance failures.
- Over all called *Service Agreements* by NIST
- The following promises are made to consumer by the provides:
  - Availability:
    - Usually 99.5% to 100% availability is assured.
    - The assurance is for a time intervals of a billing cycle e.g., 15 minute, 1 hour, 1 Year etc. for which the service status will be "up" for sure.
    - But this has to be clarified that for example time period of assurance is 15 minutes and even if the service is "down" for 14 minutes, then it legally means that the service was not "down" for the whole interval.
    - Typically, several failures in subsystems are required to completely "down" a service for the whole period of billing.
    - The provider may adjust the availability promises on case to case basis.
  - Remedies for Failure to Perform:
    - In case of violation of the promise of *availability* (during a time period) by the provider, the customer will be compensated in terms of service credit for future use of Cloud service.
    - A refund is usually not given.
    - Consumer is responsible to monitor the availability of service and claim for compensation.
    - The following situations result in termination of Cloud IT resources usage for a consumer:

- Voluntarily by consumer
- Terminated by the provider for violating the provider's rule of service and/or for non-payment.
    - The providers usually take no responsibility for preserving the data in later case. While in former case, the preservation is done for few days.




    - Legal Care of Consumer Information:
        - The provider assures for not disclosing/viewing/using/sharing the consumer's data except in case of legal requirement.
        - On the other hand the provider retains the right of monitoring the consumer data as well as may demand a copy of consumer's software for monitoring assistance.
- The following limitations are included in the policies by the provider:
    - Scheduled Outages:
        - Will not be considered as service failure.
        - Will be informed in advance.
        - Will be of a limited time period.
    - Force majeure events:
        - Providers do take the responsibility for the events out of their realistic boundary. Such as:
        - Power failure, natural disaster and unreliable connectivity between consumer and cloud service.
    - Service Agreement Changes:
        - The provider usually retain the right to change the terms of contract, billing amount etc. on limited notice.
        - Consumers should keep a regular check for updated service charges
        - Sometimes the provider inform a specific consumer by email or postage.
        - The changes may take effect immediately or after few weeks.
    - Security:

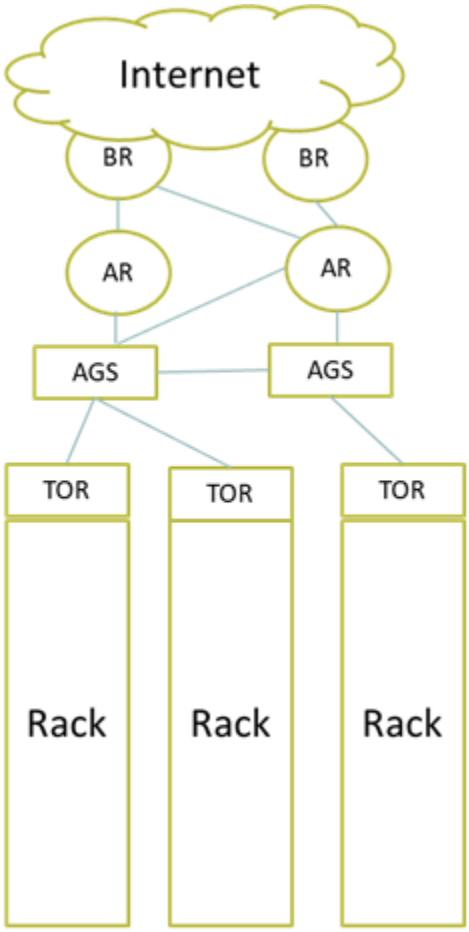
        - The providers do not take liability of data loss, data corruption or unauthorized data usage if they happen due to security breach or due to service interruption caused by a malicious activity.
        - At most, the service credit is compensated in case of data loss.
        - Although the providers promises for best effort security but the responsibility of data security is placed on the consumer.
        - It is difficult for the customer to determine the cause of data loss (malicious activity or some other reason).
    - Service API Changes:
        - The providers generally retain the right to delete or update the service API.
        - Can happen any time and without prior notice.
- Generally the consumer has to agree upon the following **obligations**:
    - Acceptable Use Polices: The consumers are generally required to refrain from:
        - Storing illegal data
        - Conducting security attacks on Cloud infrastructure and/or on any other user.
    - Licensed Software: The provider require the consumer to install and use only the licensed third party software over the Cloud.
    - Timely Payments: The consumer should timely pay the bill from the provider. Otherwise the consumer may get terminated after some time.
- **Recommendations** by NIST:
    - The consumers should carefully study and negotiate the service agreements. Specially take care of the SLA assurances and responsibilities by the provider.
    - Choose the most suitable Cloud provider periodically after review.
    -

**Lesson No. 22**

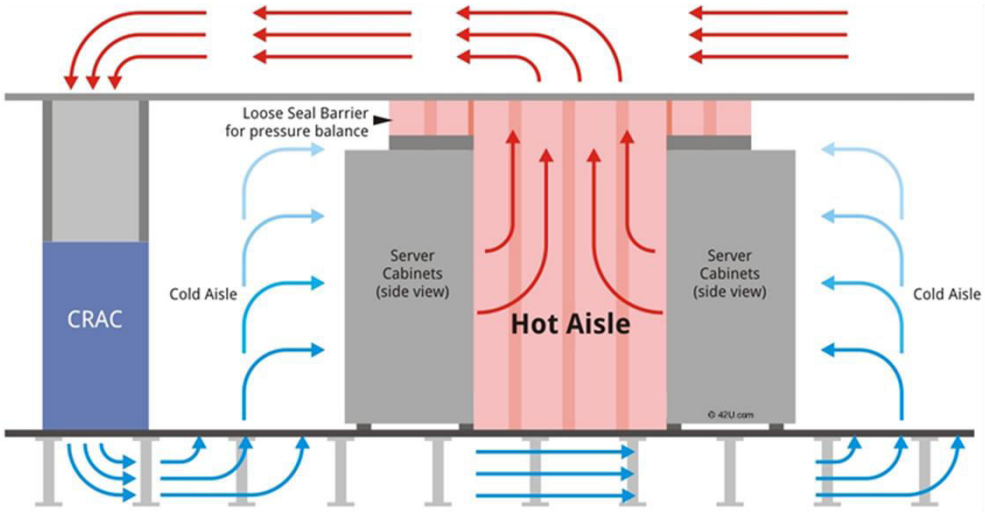
## CLOUD HOSTING DATA CENTER DESIGN

**Module No – 102:**

- Key terms:
    - CRAC: Computer Room Air Conditioning
    - Hot aisle
    - Cold aisle
    - Server cabinets (Racks)
    - Hollow floor
    - Perforated tiles
- Cloud hosting data center has a layered architecture for the Internet access.
- The servers are physically connected to layer 2 switches. There is a top of rack (TOR) in each rack. One server is connected to only one TOR switch.
- The TORs are connected to aggregate switches (AGS).




Cloud hosting data center design






Cloud hosting data center design

- Data centers consume huge amounts of electricity. As much as a small town in USA.
- A large data center can host hundreds of thousands physical servers.
- It is more costly to setup and run a small data center in terms of unit costs (per server, per MB of storage, per GHz, Network bandwidth) and operational costs as compared to larger data centers.
- Google has 900,000 physical servers around the world in its data centers. Together these servers consume 260 million watts of power which accounts to 0.01% of global energy usage.
- Facebook data center servers process 2.4 billion pieces of content and 750TB of data every day.

**Module No – 103:Data center Interconnection Networks**

- The network connecting the data center servers is called *data center interconnection network.*
- It is a core design of data center.
- The network design must support the following features:
  - Low latency
  - High bandwidth
  - Low cost
  - Message-passing interface (MPI) communication support
  - Fault tolerance
  - Must satisfy both point-to-point and collective communication patterns among all server nodes.

- Application Traffic Support: The data center interconnection network must support the MPI communication and high bandwidth.
    - Example: Distributed file access, Map and Reduce functions etc.
    - Some servers can be configured to be master and others be slaves.
- Network Expandability: The interconnection network must be expandable.
    - Should support load balancing and data movement.
        - No bottlenecks
        - Can be expanded in the unit of *data center container* which contains hundreds of servers and is a building block of large data centers.
        -
- Fault Tolerance and Graceful Degradation: Can be implemented through:
    - Replication in software and hardware resources
    - Redundant links among any two servers
    - No single point of failure or critical links
    - Two layered design should be used (a network layer close to servers and the upper layer or backbone) to support modular (container) based expandable design.

## Module No – 104: Modular Data center and Interconnection

- Modular Data Center in Shipping Containers: The modern data centers are a the collection of container based clusters that can be shipped from one location to another through trucks.
- It is an alternative to warehouse based data center.
- Modular Data Center in Shipping Containers:
- For example: The SGI ICE Cube container can house 46,080 processing cores or 30 PB of storage per container.
- Modular Data Center in Shipping Containers:
    - Such a design:
        - Is more energy efficient in terms of cooling cost as compared to warehouse based design.
        - Is more mobile and easily transportable.
        - Is ready to be deployed since it is assembled with servers, networking, power supplies and cooling mechanisms. It is then tested and shipped.
        - Helps in dynamic scalability of data center.
        - Makes the relocation of data center as relatively easier than warehouse based design.
        - Inter-Module Connection Networking requires an extra layer over modular containers to allow dynamic scaling and interconnection.

## Module No – 105: Data center Management Issues

- Modern day data centers handle ever larger volumes of data and conduct the processing massive amounts of user requests around the globe.
- In order to maintain user satisfaction and performance, the managing of a data center has become a set of complex tasks. These include (but not limited to):

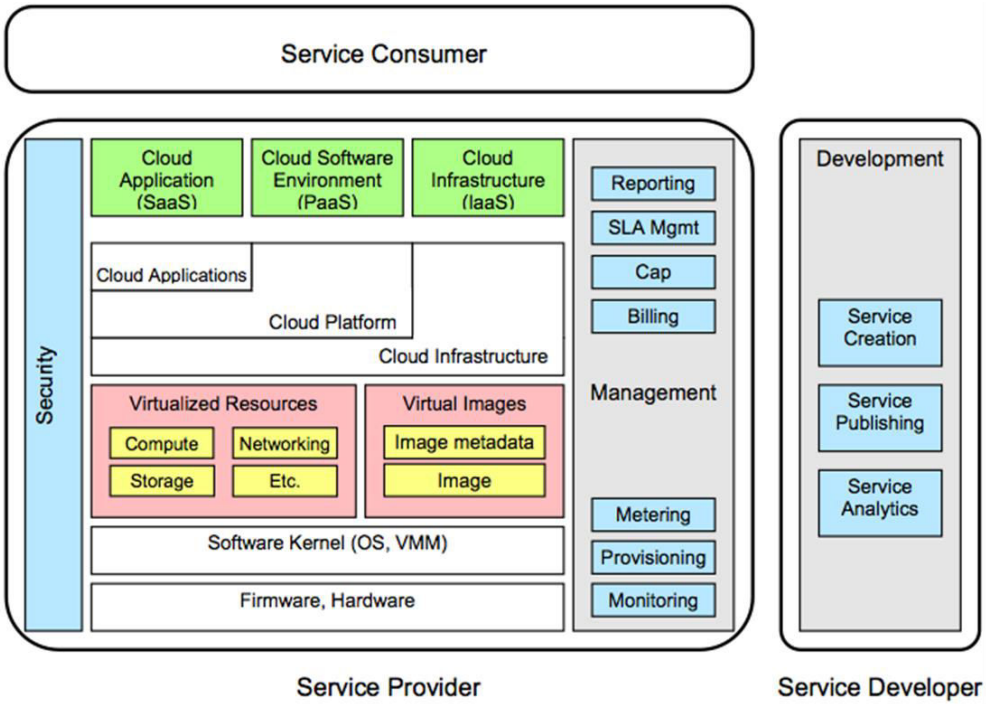
- Making common users happy by providing quality services.
- Ensuring uninterrupted and high availability of (Cloud) services.
- Managing multiple modules concurrently. Such as processing, networking, security and maintenance etc.
- Managing and planning for the scalability of data center.
- Ensuring the reliability of virtual infrastructure through fault tolerant and recovery mechanism to minimize the downtime and data loss.
- Managing and lowering the operational costs and transferring the cost benefit to Cloud providers and consumers.
- Security enforcement and data protection
- Implementation of Green information technology usage to lower the amount of energy consumption.

**Lesson No. 23**

## CLOUD ARCHITECTURE

### Module No – 106: Generic Cloud Architecture Considerations:

- A generic architecture of a (public) Cloud can be envisioned ~~on the basis of technologies we have studied so far.~~
- Major goals of a Cloud platform can be:
  - Scalability
  - Virtualization
  - Efficiency
  - Reliability
- A Cloud management software receives the consumers' requests for IT resources and provisions these resource by using various internal services.
- A Cloud architecture has to deal with certain challenges. A few of them are:
  - Establishment of large scale computing (hardware + software) infrastructure.
  - User friendly and efficient management of Cloud infrastructure.
- Ensuring scalability of IT resources.
- Reliable and fault tolerant implementation for processing and data.
- Implementation of disaster recovery mechanisms.
- Cloud architecture should be expandable by adding more hardware.
- Software, hardware and network technologies have emerged as Cloud enabling technologies.
- Enhancement in the following technologies have contributed towards wide spread establishment of Cloud computing:
- Software: Virtualization, multi-tenancy, web applications, SOA, load balancing, monitoring, billing, data storage
- Hardware: CPU, memory, storage, network
- Connectivity: Web2.0

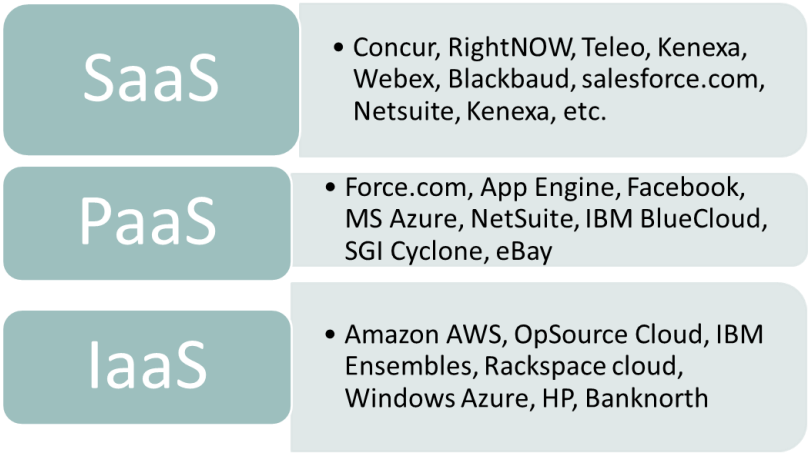





Generic Cloud Architecture

**Module No – 107: Layered Cloud Architecture:**

- Cloud architecture can be considered as consisting of layers and sub-layers of Services with each layer supporting the upper layer.
- In order of dependency, these layers are grouped at high level as:
  - SaaS
  - PaaS
  - IaaS
- Software Service development and deployment requires a platform service.
- A platform service is deployed over a VM provisioned through IaaS.
- Some services may draw resources from multiple layers/sub-layers.
- The scope of support from vendor side is highest for SaaS and lowest for IaaS.
- Services developed on PaaS require the later to provide support for scalability, security, and must be dependable.




Layered Cloud Architecture

- Unless there is interoperability among the Clouds, a Service deployed on a certain platform instance may not be portable to another platform.

**Module No – 108: Virtualization Support and Disaster Recovery:**

- The IT resources and data are prone to disasters (natural and/or human made) which damage them partially or fully and thus may crash the whole computing system of an organization.
- Key terms:
  - Failover: It is process through which a system transfers control (usually automatedly)to an alternate deployment upon failure of primary deployment.
  - Failback: The process of restoring of the system from alternative to primary deployment and restoration of original state.
  - The use of virtualization can implement the failover and brings reduction in failback time.
  - As compared to (for example) a data disaster for data stored on magnetic tapes, days are require for restoration/recovery.
  - The redundant deployment of software solutions, data and IT resources is quite easy by using virtualization.
  - One deployment is considered as primary, while other deployment/s are kept as backup.
  - The backup deployment is either updated periodically or the image/snapshot of the primary deployment (e.g., VMs) can be saved.
  - Upon failure, the backup deployment takes over.
  - The primary deployment is then restored from the most recent snapshot.
  - Virtualization has become the core part of disaster recovery plans of major organizations since last decade.
  - Virtualization even allows the testing of disaster recovery plan through emulation and without disturbing the production/primary deployment.
  - Although the failed physical servers have to be re-purchased/repaired, but the virtualization lowers the additional costs and time related to failback.

- The organizations should mark the critical applications and data and use replication of data in virtualized environments to support effective disaster recovery.

## Module No – 109: Cloud Architectural Design Challenges:

- **Challenge 1: Service availability and Data Lock-in Problem**:
  - Depending upon a single provider for service deployment results in a single point of failure or lock-in.
- ~~**Challenge 1: Service availability and Data Lock-in Problem:**~~
  - High availability of a service can be assured by distributed deployment over multiple Clouds.
  - Requires the interoperability/standardization of API calls on different PaaS platforms.
- **Challenge 2: Data Privacy and Security Concerns:**
  - Due to public access of Clouds, multitenancy and sophisticated attacks/malware, the implementation and assurance of privacy and security of consumers' data is a big challenge.
- **Challenge 3: Unpredictable Performance and Bottlenecks:**
  - The unpredictability of processing and data load over Cloud services introduce I/O bottlenecks such as concurrent read/write access requirements to shared storage for large data volumes by multiple VMs.
  - The providers have to carefully analyze the deployment decisions according to surge in computing/data loads and should tune the bottlenecks.

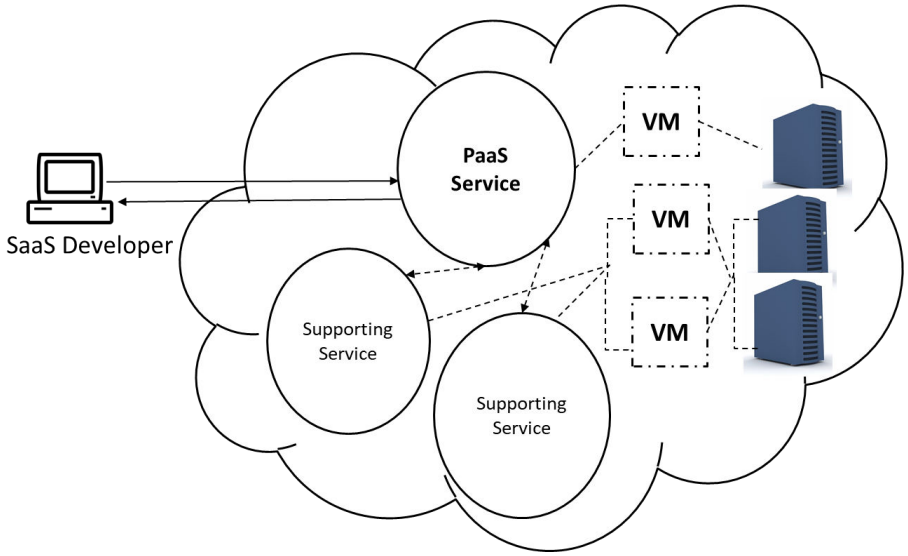
## Module No – 110: Cloud Architectural Design Challenges (continued):

- **Challenge 4: Distributed Storage and Widespread Software Bugs:**
  - Ensuring data consistency, durability and high availability is a challenge when the data is distributed.
  - Debugging of data to remove inconsistencies and errors is important but challenging.
- **Challenge 5: Cloud Scalability, Interoperability and Standardization:**
  - Scalability is one of the basic features of Cloud computing and thus requires (for example) dynamic availability of IT resources (hardware) for scaling up.
  - The heterogeneity in hardware and/or hypervisor makes it challenging to dynamically include more hardware/virtualized IT resources.
  - The open virtualization format (OVF) describes and open, secure, efficient, portable and extensible format for packaging and distribution of VMs and the software to be deployed over VMs.
  - OVF allows hypervisor, guest OS and hardware platform independent packaging of VMs and software.
  - Interoperability should be provided for cross hypervisor and cross platform (intel & AMD) live migration of VMs.
- **Challenge 6: Software Licensing and Reputation Sharing:**

- The fact that the license model of commercial software is not suitable for utility computing, the providers have to rely upon open source software and/or bulk usage license.
- If the reputation of a provider is affected (due to consumers' malicious behavior), then there is no service to safe-guard the provider's reputation.

## Module No – 111: Public Cloud Platforms Architecture Examples:

- We shall look at a few examples of PaaS platforms on public clouds.




PaaS Platform Architecture

- **Google App Engine (GAE):** It is a popular platform for developing Cloud applications.
  - Based upon technologies:
    - Google File System (GFS): Stores large volumes of data
    - MapReduce: Used in parallel job execution on massive data
    - Chubby (Distributed applications' locking)
    - BigTable (Storage service to access structured data)
  - Consumers are allowed to develop applications in popular languages such as Java, PHP, Go and Python. The following are components of GAE:
    - Datastore
    - Application runtime environment (for web applications)
    - Software Development Kit (SDK) (for local application development)
    - Administration console (management of user application development cycles)
    - Web service infrastructure (interfaces for flexible use of storage and networks resources)
  - Well known applications of GAE are Google Search Engine, Google Docs, Google Earth, and Gmail.

- Consumers can create Cloud applications by using GAE which run on Google data centers.

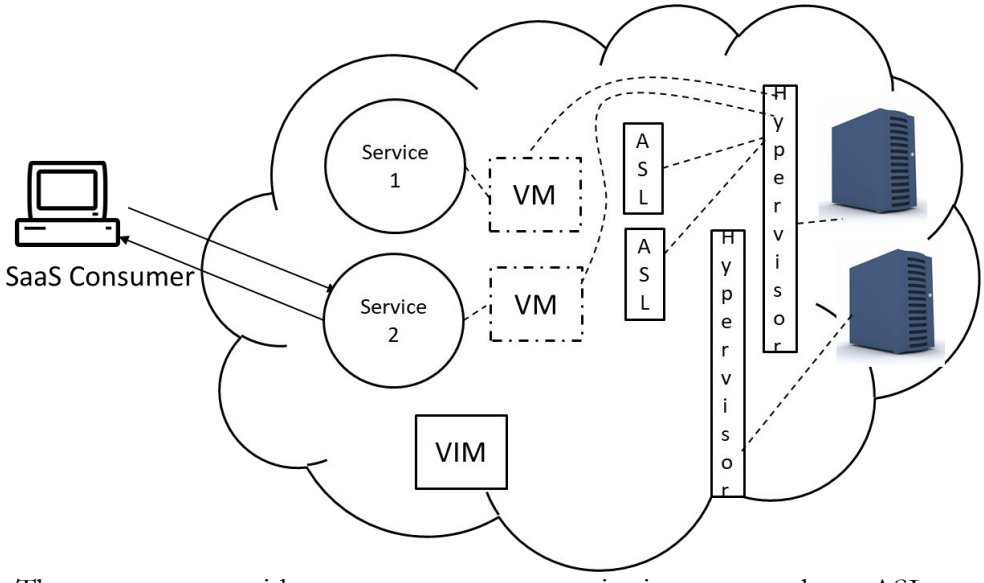
- **Amazon Web Services (AWS):**
  - Amazon provides the SOAP web services and IaaS to the consumers/developers to create and host Cloud services.
  - Amazon Elastic Computing Cloud (EC2) is a web service to provide the VMs for hosting Cloud applications.
  - Simple Storage Service (S3) provides the object-oriented storage service.
  - Elastic Block Service (EBS) provides the block storage interface.
  - Simple Queue Service (SQS) provides inter process message passing.
  - Amazon DevPay service can be used for online billing and account management for the service providers to sell the applications developed and/or hosted on AWS.

**Lesson No. 24**

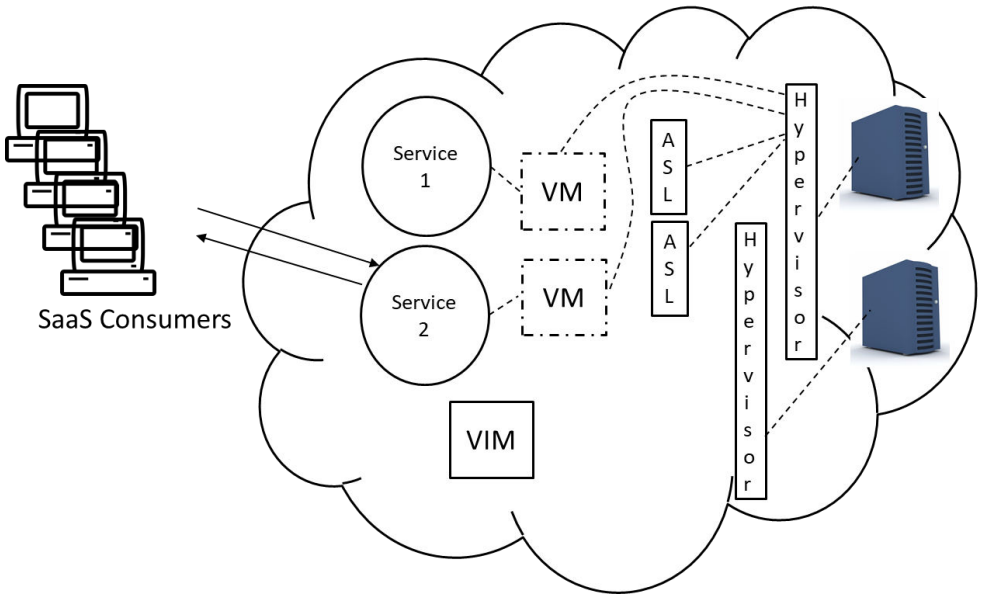
## SPECIALIZED CLOUD MECHANISMS

### Module No – 112: Automated Scaling Listener (ASL)

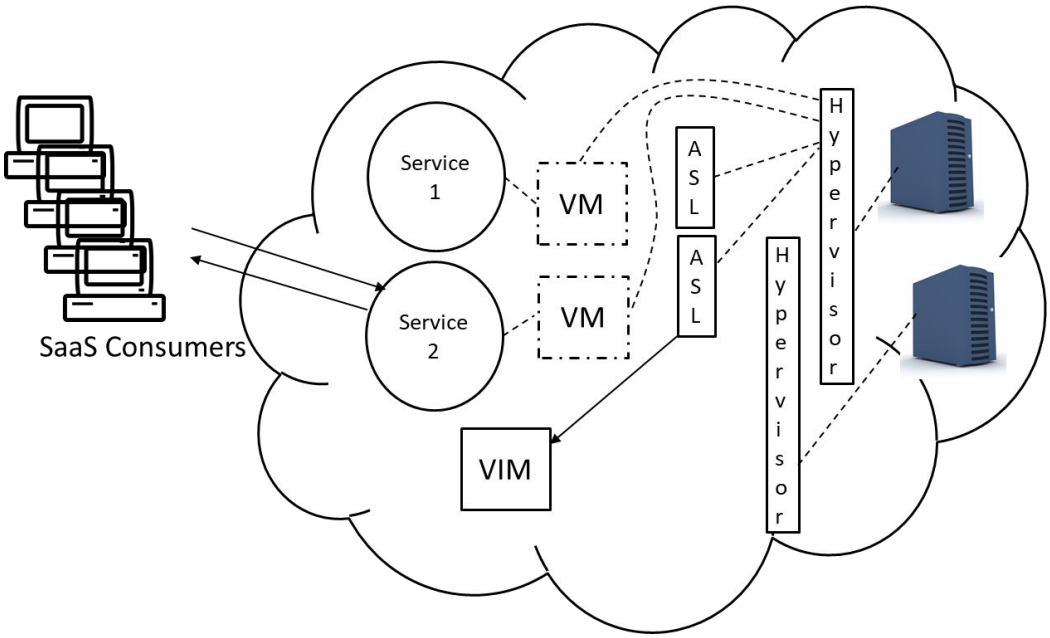
- It is the software module (service agent) which monitors and tracks the communication between Cloud service and the service consumer for dynamic scaling purpose.
  - Can indicate the need for scaling to cloud consumer.
  - Indicates to cloud manager for scaling in/out (if configured to auto scaling by the consumer).



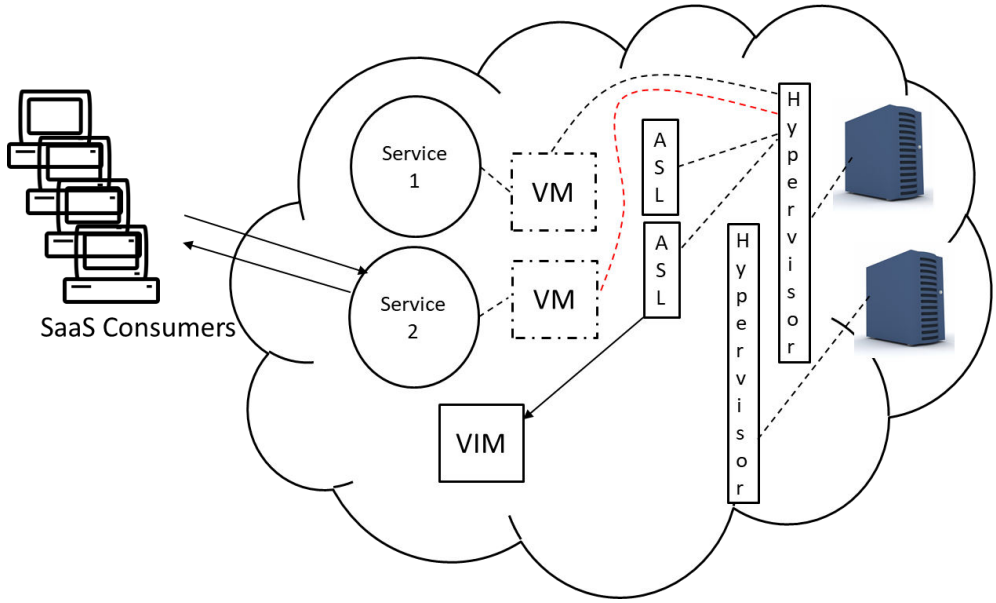

The startup setup with one consumer, two service instances and two ASL modules.



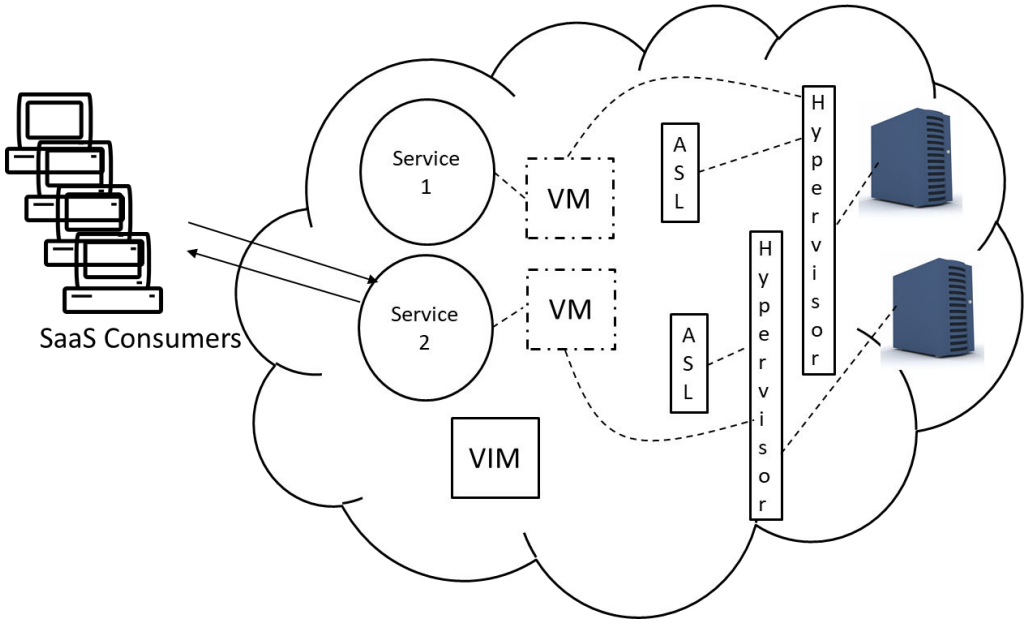

The number of Cloud service 2 consumers are increasing.



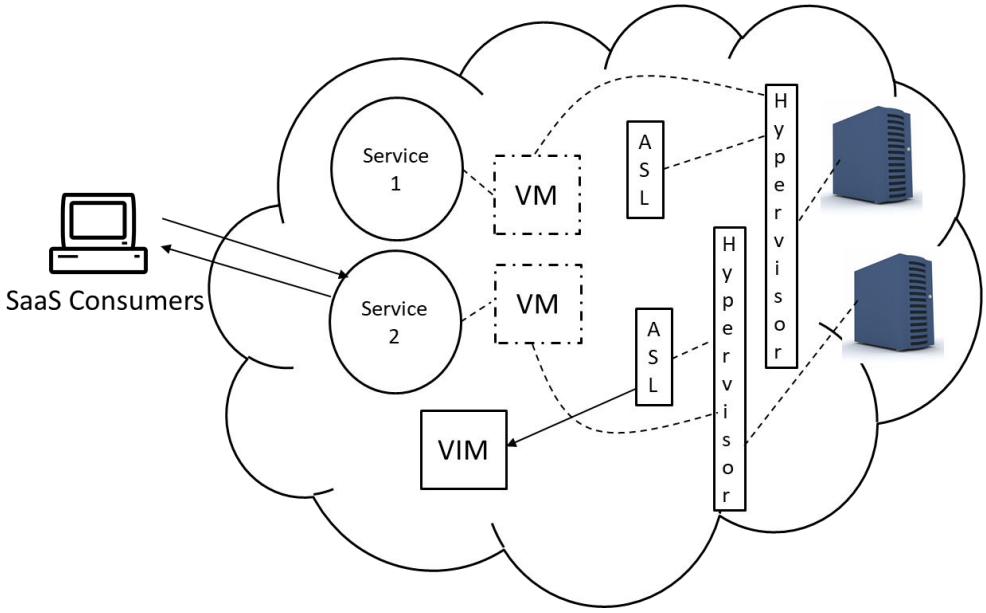

The ASL indicates the Virtual Infrastructure Manager (VIM) for increased load and lack of resources for VM hosting Service 2 on current host/server.



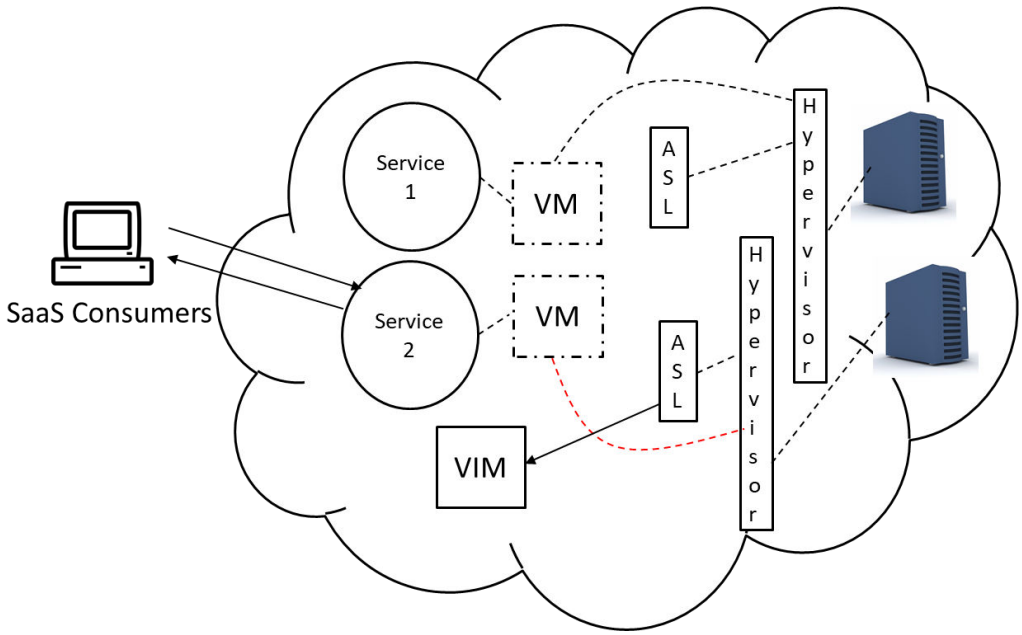

The VIM initiates the migration of VM hosting service 2 to new host for resource availability.



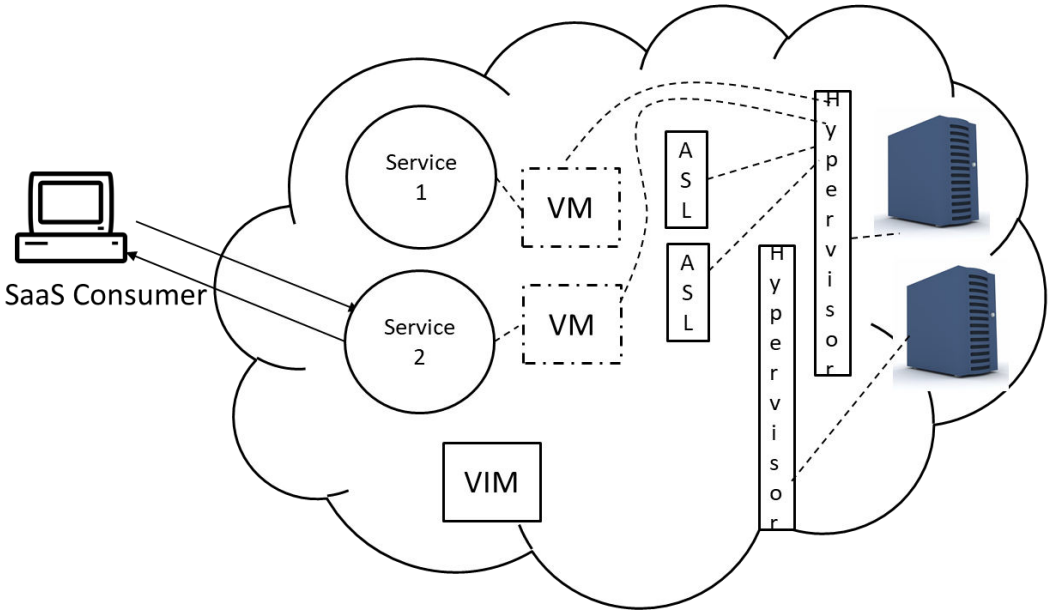

The VM hosting service 2 is migrated to the new host with more resources.




The number of service consumers of Service 2 have decreased. The ASL indicates this to VIM



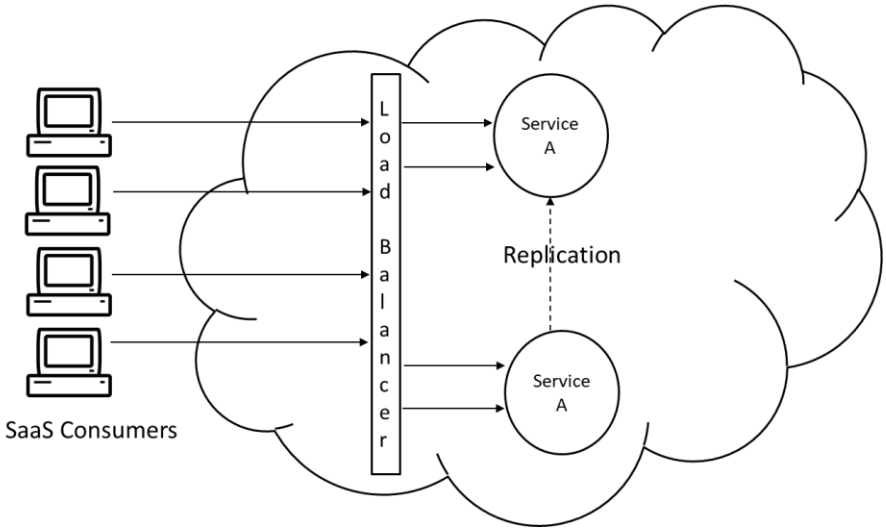

The VIM prepares for migration of service 2 hosting VM for server consolidation.




The VM is migrated to the (previous) host/server.

**Module No – 113:**

- **Load Balancer:** It is the service agent which distributes workload among multiple processing resources such as multiple service instances. Workload is distributed on the basis of:
  - Processing capacity of the IT resource
  - Workload prioritization
  - Content-Aware distribution




Follow the video lecture to understand fully

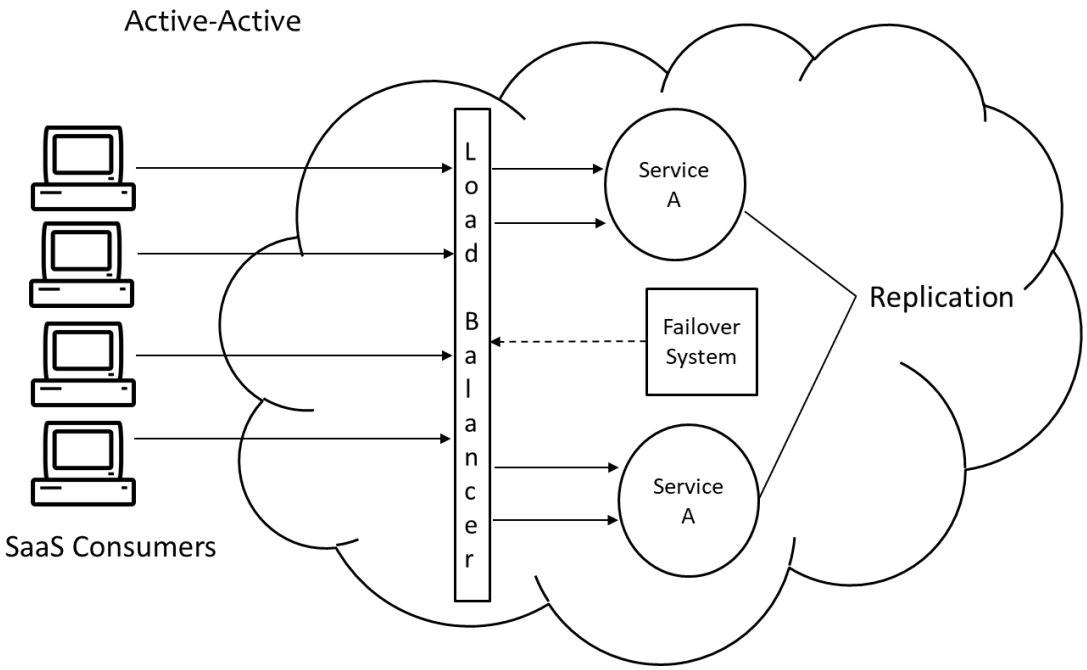
- **SLA Monitor:** Works by pinging (for example) to a service instance to record the "down" status with time.
    - The statistics are used to evaluate the extent of SLA violation.
    - Uses a polling agent ~~(studied before)~~.
- **Pay-per-use Monitor:** It is based upon a monitoring agent ~~(studied before).~~
    - It collects the resource usage by intercepting the messages sent to a Cloud service by the consumer.
    - Collected data (such as transmitted data volume, bandwidth consumption etc.) is used for billing purpose.
    -

## Module No – 114: Failover System

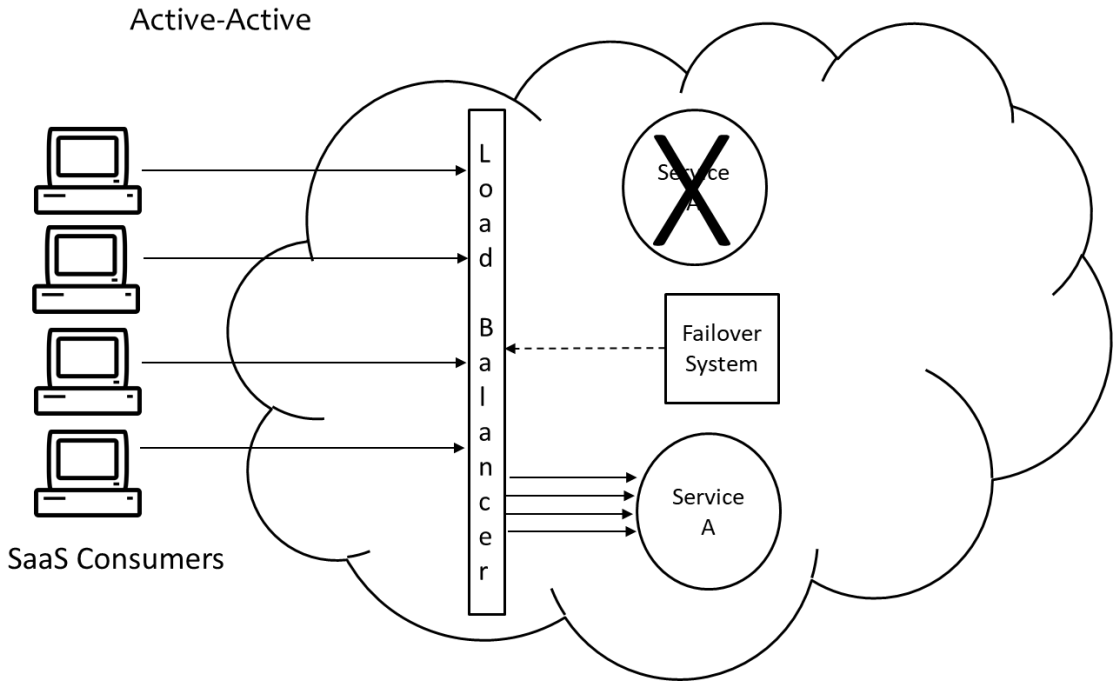
- This mechanism is used to increase the reliability and availability of IT resources by using redundant implementations (for example of Cloud services).
- Used for:
    - Mission critical programs
    - Cloud (supporting) services which can cause a single point of failure.
- The redundant implementations are actively monitored for error detection and unavailability of resources.
- Configurations:
    - Active-Active: The redundant implementation is actively processing the workload. Load balancer implementation is required. The failover system detects the resource failure and directs the load balancer to allocate workload only to active (redundant) implementation. When the failed instance is recovered or replicated, the failover system directs the load balancer to start allocating the workload to all (including replicated) instances.
    - Active-Passive: The redundant instance is passive till the active instance fails. The failover system when detects a failure, it activates a redundant instance and redirects the workload towards the newly activated instance. Upon recovery or replication of failed instance, the failover system puts it to stand-by state while the previously activated instance continues to serve as the active instance.

## Module No – 115: Failover System: Case study

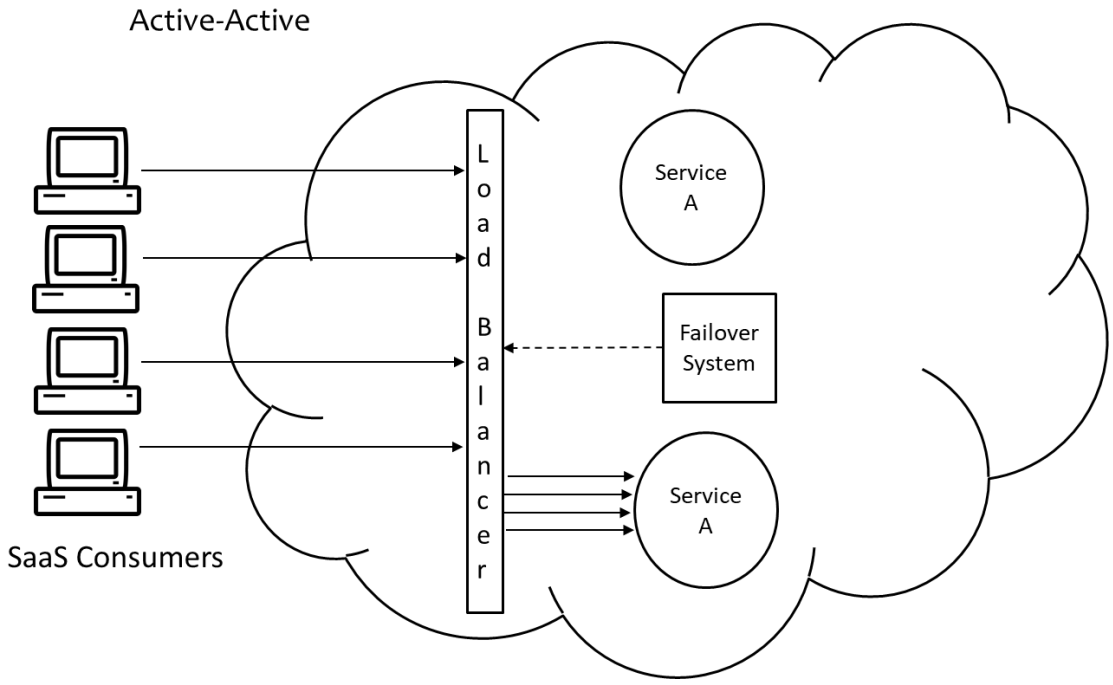
- Let us see the implementations of Failover System:



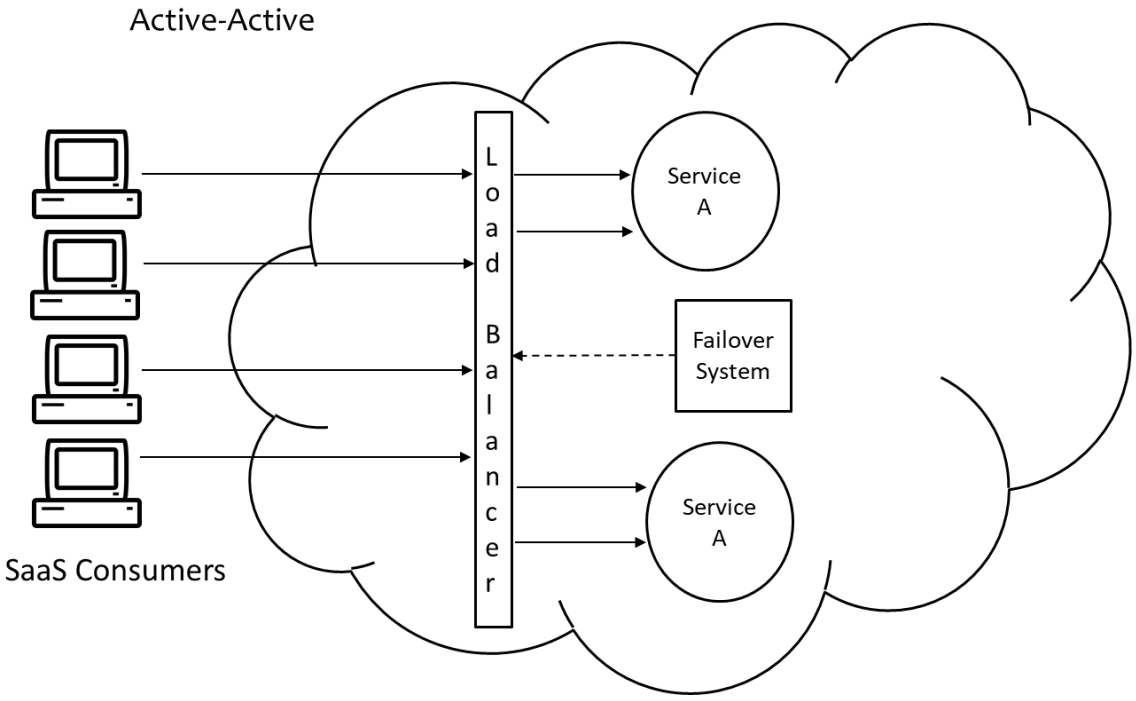

Follow the video lecture to understand fully



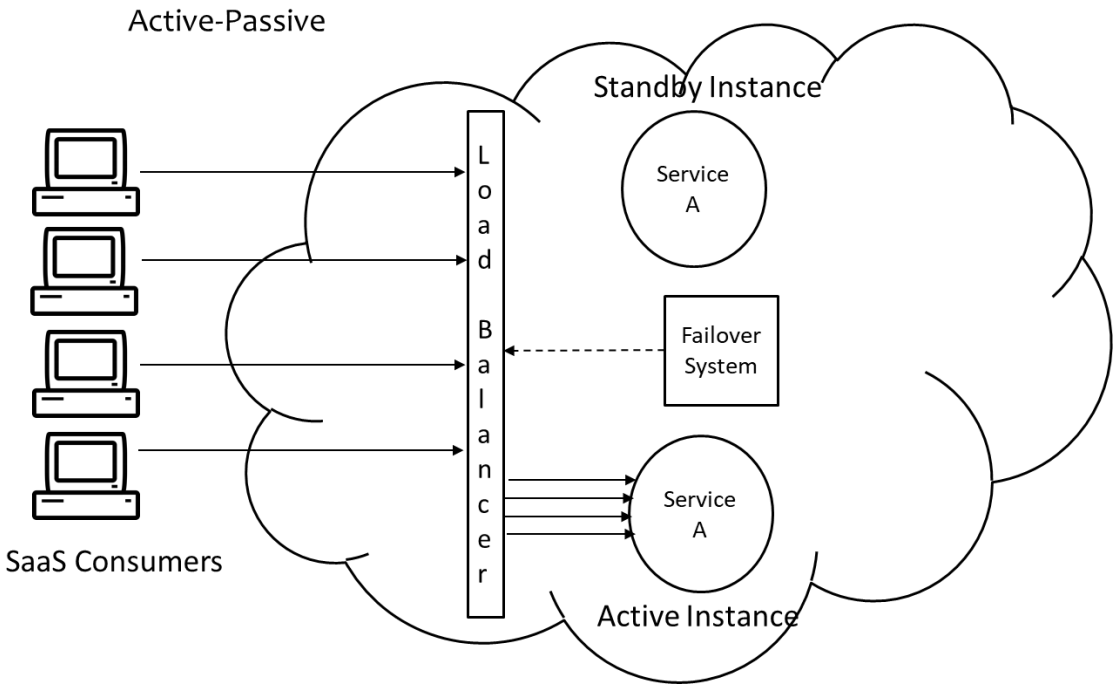

Follow the video lecture to understand fully



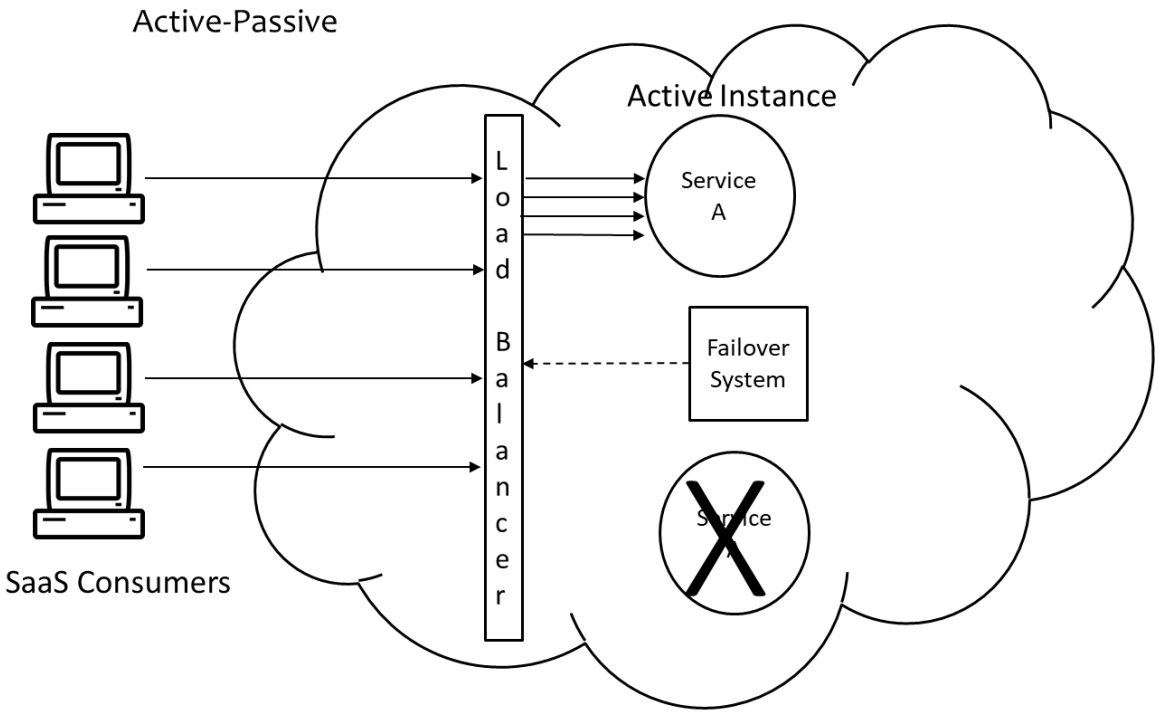

Follow the video lecture to understand fully



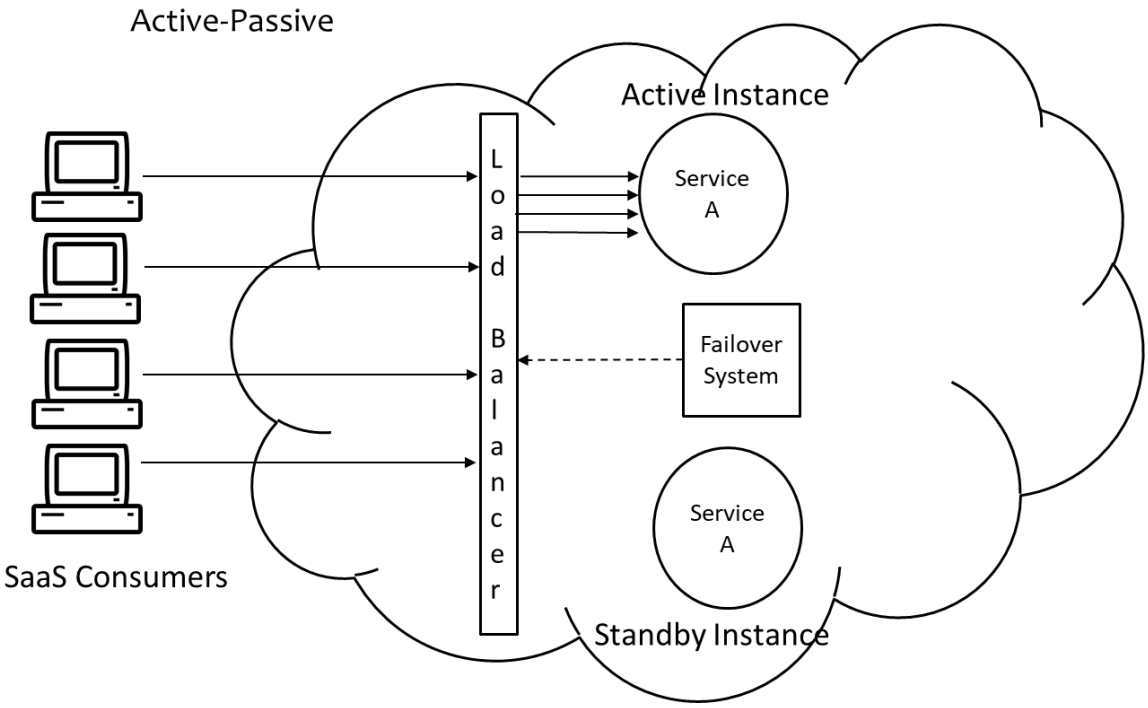

Follow the video lecture to understand fully




Follow the video lecture to understand fully




Follow the video lecture to understand fully




Follow the video lecture to understand fully

**Module No – 116: Resource Cluster Mechanism:**

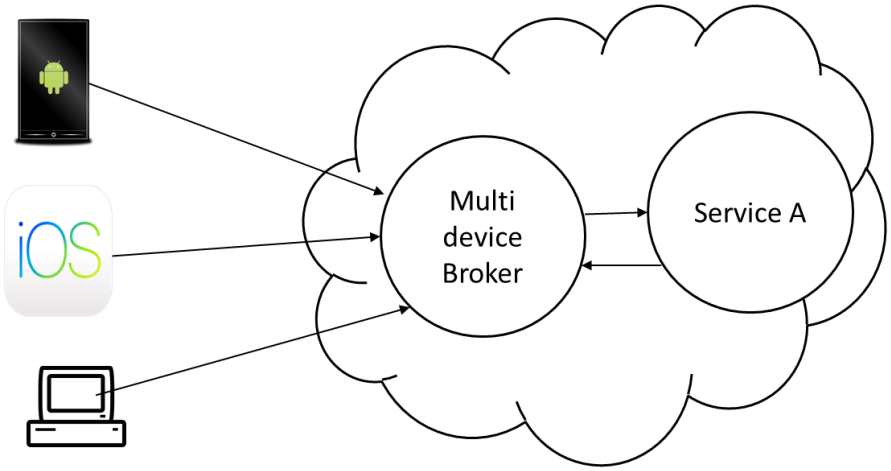
- The Cloud promises virtually unlimited IT resources.
- These IT resources are (although virtualized) but can not be provided through a single physical server.
- It is obvious that the Cloud IT resources are provisioned from multiple physical servers located in a single or multiple data center/s.
- The resource cluster mechanism is used to group multiple IT resources so that they can be used as a single IT resource.
- This increases the computing capacity, load balancing capacity and availability of the clustered IT resources.
- High speed communication links are used to connect the clustered IT resources for:
  - Workload distribution
  - Task scheduling
  - Data sharing
  - System synchronization
- Server clusters may or may not have a shared storage.
- Common types:
  - *Server Cluster*: Consisting of physical or virtual servers. The virtualized clusters support the migration of VMs for scaling and load balancing.

- *Database Cluster*: Is used to keep redundant implementation of databases. It has features to synchronize the data across all the redundant instances.
  - Useful for active-active and active-passive failover systems.
- *Large Dataset Clusters*: This type of cluster is used to partition and distribute large datasets without affecting the data integrity or computing accuracy.
  - Each node processes workloads without any need to depend/communicate with other nodes.

- Additional types:
  - *Load Balanced Cluster*: Implements a load balancer mechanism (~~discussed before~~).
  - *HA Cluster*: Implements a failover system (~~discussed before~~).

**Module No – 117:**

- **Multi-Device Broker:** This mechanism is used to transform the messages (received from heterogenous devices of Cloud consumers ) into a standard format before conveying them to the Cloud service.
  - The response messages from Cloud service are intercepted and transformed back to the device specific format before conveying to the devices through the multi-device broker mechanism.






Follow the video lecture to understand fully

- **State Management Database:** It is a device used to temporarily store the state data of software programs.
  - State data can be (for example) the configuration and number of VMs being employed to support a user subscription to a PaaS instance.
  - In this way, the programs do not use the RAM for state-caching purposes and thus the amount of memory consumed is lowered.

- The services can then be in a "stateless" condition.
- For example, a PaaS instance (ready-made environment) requires three VMs. If user pauses activity, the state data is saved in state management software and the underlying infrastructure is scaled in to a single VM.
- When the user resumes the activity, the state is restored by scaling out on the basis of data retrieved from state management database.

**Lesson No. 25**

## CLOUD MANAGEMENT

### Module No – 118: Remote Administration System

- It is a Cloud mechanism which provides the APIs and tools to the providers to develop and used online portals.
- These portals also provide some administrative controls to the Cloud consumers as well.
- Usage and Administration Portal:
  - Management controlling of Cloud IT resources
  - IT resources usage reports
- Self-Service Portal:
  - The consumer can look at and choose various Cloud services
  - The chosen services/package is submitted to Cloud provider for automated provisioning
- The remote administration console can be used to:
  - Configure and setting cloud services
  - Provision and releasing IT resources for on-demand usage
  - Monitor cloud service status, usage and performance
  - QoS and SLA fulfillment monitoring
  - IT resource leasing cost and usage fee management
  - Managing user accounts, security credentials, authorization and access control
  - The remote administration console can be used to:
  - Capacity planning
- If allowed, a Cloud consumer can create its own front-end application using API calls of remote administration system.

### Module No – 119: Resources Management System

- Utilizes the virtual infrastructure manager (VIM) for creating and managing the virtual IT resources.
- Typical tasks include:
  - Managing the templates used to initialize the VMs
  - Allocating and releasing the virtual IT resources
  - Starting, pausing, resuming and termination of virtual IT resources in response to allocation/release of these resources