UNIT IV RESOURCE MANAGEMENT AND SECURITY IN CLOUD 10

Inter Cloud Resource Management – Resource Provisioning and Resource Provisioning Methods – Global Exchange of Cloud Resources – Security Overview – Cloud Security Challenges – Software-as-a-Service Security – Security Governance – Virtual Machine Security – IAM – Security Standards.

4.1 INTER-CLOUD RESOURCE MANAGEMENT

Cloud of Clouds (Inter cloud)

- Inter cloud or 'cloud of clouds'-refer to a theoretical model for cloud computing services.
- Combining many different individual clouds into one seamless mass in terms of on-demand operations.
- The inter cloud would simply make sure that a cloud could use resources beyond its reach.
- Taking advantage of pre-existing contracts with other cloud providers.
- Each single cloud does not have infinite physical resources or ubiquitous geographic footprint.
- A cloud may be saturated to the computational and storage resources of its infrastructure.
- It would still be able satisfy such requests for service allocations sent from its clients.
- A single cloud cannot always fulfill the requests or provide required services.
- When two or more clouds have to communicate with each other, or another intermediary comes into play and federates the resources of two or more clouds.
- In inter cloud, intermediary is known as "cloud broker" or simply "broker."
- Broker is the entity which introduces the cloud service customer (CSC) to the cloud service provider (CSP)

Inter-Cloud Resource Management Consists of

- Extended Cloud Computing Services
- Resource Provisioning and Platform Management
- Virtual Machine Creation and Management
- Global Exchange of Cloud Resources

4.1.1 Extended Cloud Computing Services

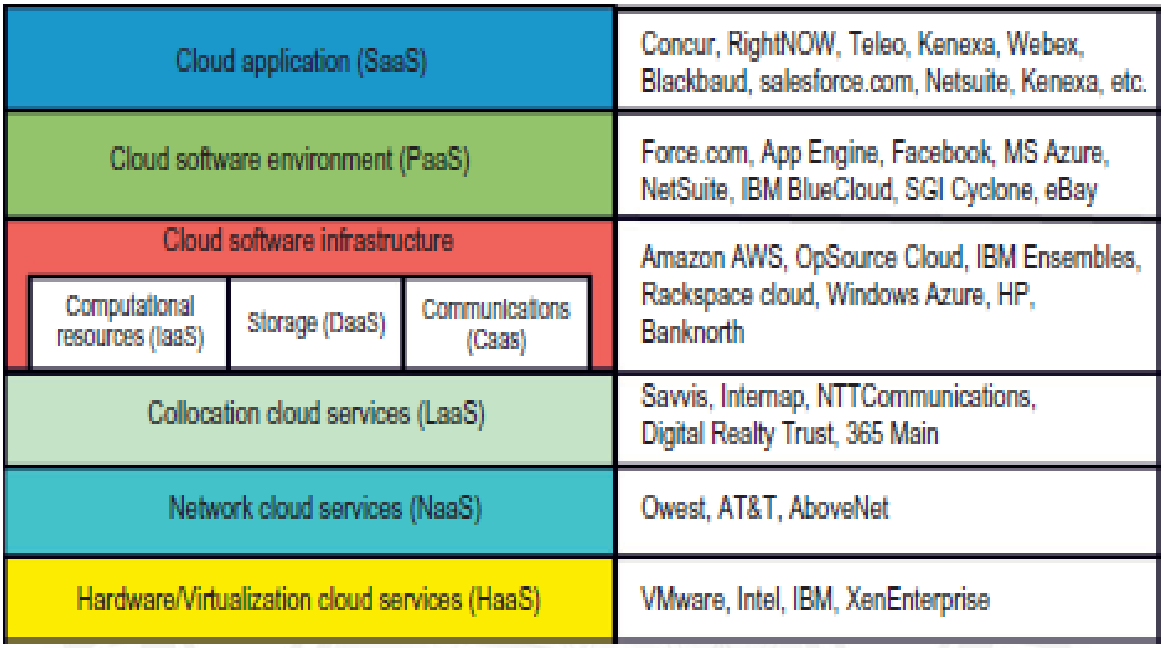


Fig: Six layers of cloud services and their providers

Six layers of cloud services

- Software as a Service(SaaS)
- Platform as a Service(PaaS)
- Infrastructure as a Service(IaaS)
- Hardware / Virtualization Cloud Services(HaaS)
- Network Cloud Services (NaaS)
- Collocation Cloud Services(LaaS)

- The top layer offers SaaS which provides cloud application.
- PaaS sits on top of IaaS infrastructure.
- The bottom three layers are more related to physical requirements.
- The bottommost layer provides Hardware as a Service (HaaS).
- NaaS is used for interconnecting all the hardware components.

- Location as a Service (LaaS), provides security to all the physical hardware and network resources. This service is also called as Security as a Service.
- The cloud infrastructure layer can be further subdivided as
  - Data as a Service (DaaS)
  - Communication as a Service (CaaS)
  - Infrastructure as a Service(IaaS)
- Cloud players are divided into three classes:
  - Cloud service providers and IT administrators
  - Software developers or vendors
  - End users or business users.

Cloud Players	IaaS	PaaS	SaaS
IT administrators/ Cloud Providers	Monitor SLAs	Monitor SLAs and enable service platforms	Monitor SLAs and deploy software
Software developers (Vendors)	To deploy and store data	Enabling platforms	Develop and deploy software
End users or business users	To deploy and store data	To develop and test software	Use business software

Table: Cloud Differences in Perspective of Providers, Vendors, and Users

4.1.1 Cloud Service Tasks and Trends

- SaaS is mostly used for Business Applications
- Eg: CRM (Customer Relationship Management) used for business promotion, direct sales, and marketing services
- PaaS is provided by Google, Salesforce.com, and Facebook etc.
- IaaS is provided by Amazon, Windows Azure, and RackRack etc.
- Collocation services Provides security to lower layers.
- Network cloud services provide communications.

4.1.2 Software Stack for Cloud Computing

- The software stack structure of cloud computing software can be viewed as layers.
- Each layer has its own purpose and provides the interface for the upper layers.
- The lower layers are not completely transparent to the upper layers.

4.1.3 Runtime Support Services

- Runtime support refers to software needed in applications.
- The SaaS provides the software applications as a service, rather than allowing users purchase the software.
- On the customer side, there is no upfront investment in servers.

4.1.2 Resource Provisioning (Providing) and Platform Deployment

There are techniques to provision computer resources or VMs. Parallelism is exploited at the cluster node level.

4.1.2.1 Provisioning of Compute Resources (VMs)

- Providers supply cloud services by signing SLAs with end users.
- The SLAs must specify resources such as
  - CPU
  - Memory
  - Bandwidth

Users can use these for a preset (fixed) period.

- Under provisioning of resources will lead to broken SLAs and penalties.
- Over provisioning of resources will lead to resource underutilization, and consequently, a decrease in revenue for the provider.
- Provisioning of resources to users is a challenging problem. The difficulty comes from the following
  - Unpredictability of consumer demand
  - Software and hardware failures
  - Heterogeneity of services

- Power management
- Conflict in signed SLAs between consumers and service providers.

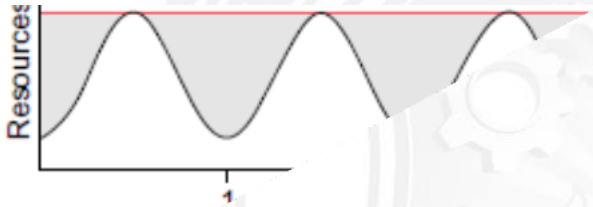
4.1.2.2 Provisioning Methods

Three cases of static cloud resource provisioning policies are considered.

Static cloud resource provisioning

case (a)

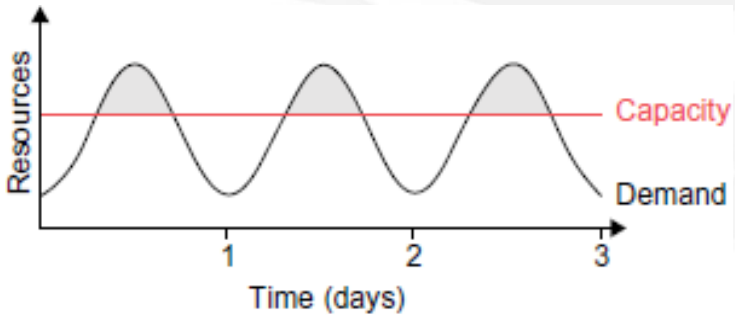
- over provisioning(Providing) with the peak load causes heavy resource waste (shaded area).



..

case (b)

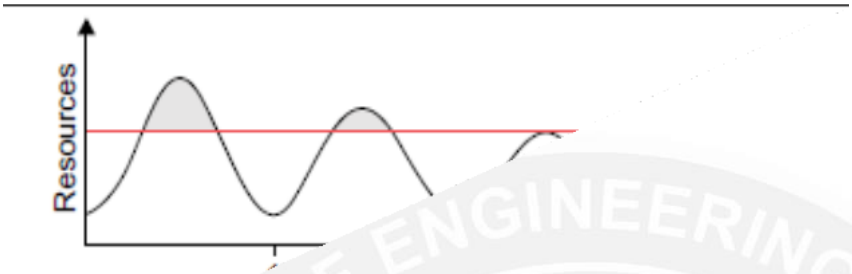
Under provisioning of resources results in losses by both user and provider. Users have paid for the demand (the shaded area above the capacity) is not used by users.



(b) Underprovisioning 1

case (c)

Declining in user demand results in worse resource waste.



Constant provisioning

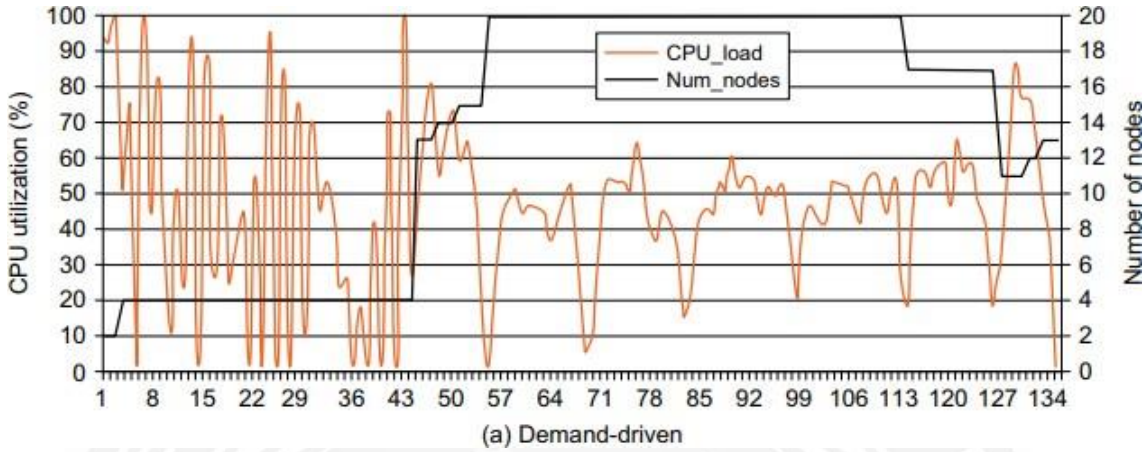
- Fixed capacity to a declining user demand could result in even worse resource waste.
- The user may give up the service by canceling the demand, resulting in reduced revenue for the provider.
- Both the user and provider may be losers in resource provisioning without elasticity.

Resource-provisioning methods are

- Demand-driven method - Provides static resources and has been used in grid computing
- Event-driven method - Based on predicted workload by time.
- Popularity-Driven Resource Provisioning – Based on Internet traffic monitored

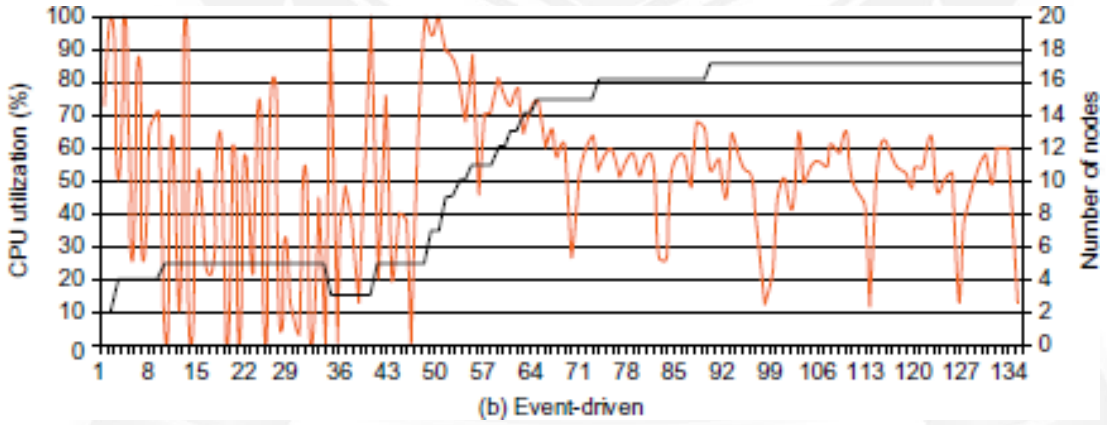
4.1.2.3 Demand Driven Methods

- Provides Static resources
- This method adds or removes nodes (VM) based on the current utilization(Use) level of the allocated resources.
- When a resource has surpassed (exceeded) a threshold (Upperlimit) for a certain amount of time, the scheme increases the resource (nodes) based on demand.
- When a resource is below a threshold for a certain amount of time, then resources could be decreased accordingly.
- This method is easy to implement.
- The scheme does not work out properly if the workload changes abruptly.



4.1.2.4 Event-Driven Resource Provisioning

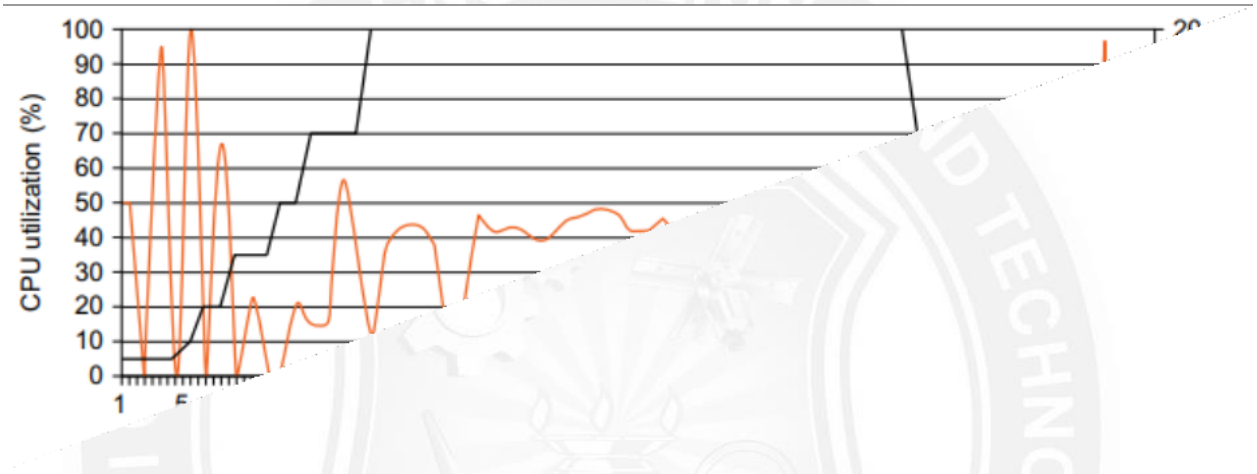
- This scheme adds or removes machine instances based on a specific time event.
- The scheme works better for seasonal or predicted events such as Christmastime in the West and the Lunar New Year in the East.
- During these events, the number of users grows before the event period and then decreases during the event period. This scheme anticipates peak traffic before it happens.
- The method results in a minimal loss of QoS, if the event is predicted correctly



4.1.2.5 Popularity-Driven Resource Provisioning

- Internet searches for popularity of certain applications and allocates resources by popularity demand.

- This scheme has a minimal loss of QoS, if the predicted popularity is correct.
- Resources may be wasted if traffic does not occur as expected.
- Again, the scheme has a minimal loss of QoS, if the predicted popularity is correct.
- Resources may be wasted if traffic does not occur as expected.



4.1.2.6 Dynamic Resource Deployment

- The cloud uses VMs as building blocks to create an execution environment across multiple resource sites.
- Dynamic resource deployment can be implemented to achieve scalability in performance.
- Peering arrangements established between gateways enable the allocation of resources from multiple grids to establish the execution environment.
- Dynamic resource deployment can be implemented to achieve scalability in performance.
- InterGrid is used for interconnecting distributed computing infrastructures.
- InterGrid provides an execution environment on top of the interconnected infrastructures.
- IGG(InterGridGateway) allocates resources from an
  - Organization's local cluster (Or)
  - Cloud provider.
- Under peak demands, IGG interacts with another IGG that can allocate resources from a cloud computing provider.
- Component called the DVE manager performs resource allocation and management.
- Intergrid gateway (IGG) allocates resources from a local cluster three steps:

(1) Requesting the VMs(Resources)

(2) Enacting (Validate) the leases

(3) Deploying (install) the VMs as requested.

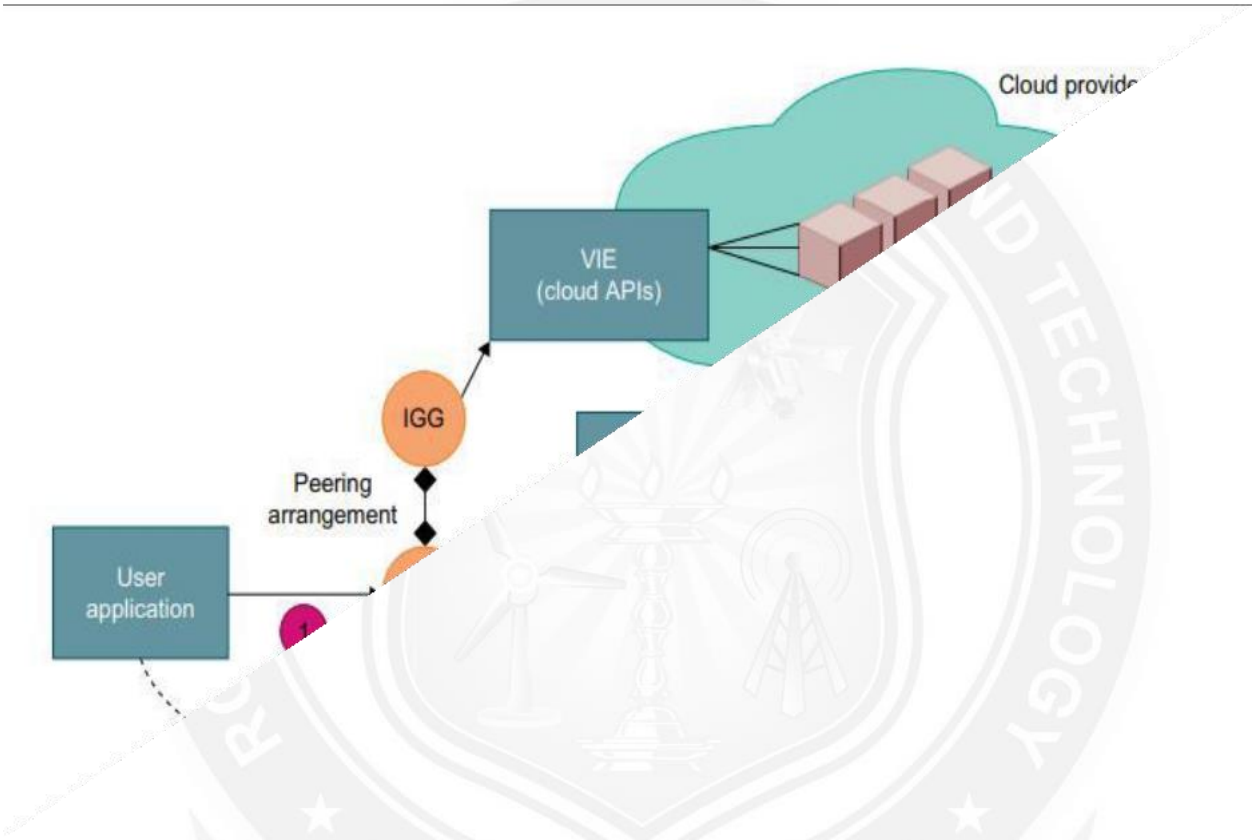


Fig: Cloud resource deployment using an IGG (intergrid gateway) to allocate the VMs from a Local cluster to interact with the IGG of a public cloud provider.

- Under peak demand, this IGG interacts with another IGG that can allocate resources from a cloud computing provider.
- A grid has predefined peering arrangements with other grids, which the IGG manages.
- Through multiple IGGs, the system coordinates the use of InterGrid resources.
- An IGG is aware of the peering terms with other grids, selects suitable grids that can provide the required resources, and replies to requests from other IGGs.
- Request redirection policies determine which peering grid InterGrid selects to process a request and a price for which that grid will perform the task.
- An IGG can also allocate resources from a cloud provider.
- The InterGrid allocates and provides a distributed virtual environment (DVE).

- This is a virtual cluster of VMs that runs isolated from other virtual clusters.
- A component called the DVE manager performs resource allocation and management on behalf of specific user applications.
- The core component of the IGG is a scheduler for implementing provisioning policies and peering with other gateways.
- The communication component provides an asynchronous message-passing mechanism.

4.1.2.7 Provisioning of Storage Resources

- Storage layer is built on top of the physical or virtual servers.
- Data is stored in the clusters of the cloud provider.
- The service can be accessed anywhere in the world.

- Eg:
  - E-mail system might have millions of users and each user can have thousands of e-mails and consume multiple gigabytes of disk space.
  - Web searching application.
  - To store huge amount of information solid-state drives are used instead of hard disk drives

In storage technologies, hard disk drives may be augmented (increased) with solid-state drives in the future.

Table 4.8 Storage Services in Three Cloud Computing Systems

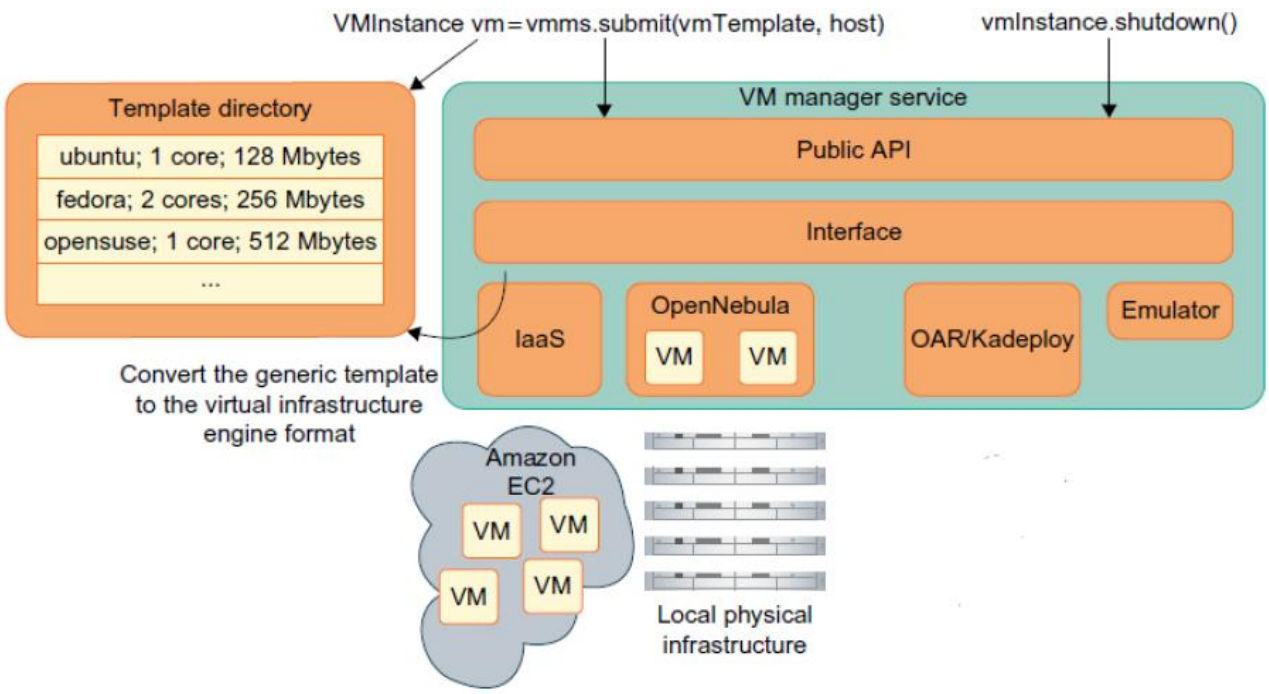
Storage System	Features
GFS: Google File System	Very large sustainable reading and writing bandwidth, mostly continuous accessing instead of random accessing. The programming interface is similar to that of the POSIX file system accessing interface.
HDFS: Hadoop Distributed File System	The open source clone of GFS. Written in Java. The programming interfaces are similar to POSIX but not identical.
Amazon S3 and EBS	S3 is used for retrieving and storing data from/to remote servers. EBS is built on top of S3 for using virtual disks in running EC2 instances.

4.5.3 Virtual Machine Creation and Management

The managers provide a public API for users to submit and control the VMs

.

Fig. Virtual Machine Creation and Management



Independent Service Management:

- Independent services request facilities to execute many unrelated tasks.
- Commonly, the APIs provided are some web services that the developer can use conveniently.

Running Third-Party Applications

- Cloud platforms have to provide support for building applications that are constructed by third-party
  application providers or programmers.

- The APIs are often in the form of services.
- Web service application engines are often used by programmers for building applications.
- The web browsers are the user interface for end users.

Virtual Machine Manager

The manager manage VMs deployed on a set of physical resources

- VIEs(Virtual Infrastructure Engine) can create and stop VMs on a physical cluster
- Users submit VMs on physical machines using different kinds of hypervisors

- To deploy a VM, the manager needs to use its template.
- Virtual Machine Templates contains a description for a VM with the following static information:
    - The number of cores or processors to be assigned to the VM
    - The amount of memory the VM requires
    - The kernel used to boot the VM's operating system.
    - The price per hour of using a VM
- OAR/Kadeploy is a deployment tool
- API(Application Programming Interface) - An API is a software intermediary that makes it possible for application programs to interact with each other and share data

Virtual Machine Templates

- A VM template is analogous to a computer's configuration and contains a description for a VM with the following static information:
- The number of cores or processors to be assigned to the VM
- The amount of memory the VM requires
- The kernel used to boot the VM's operating system
- The disk image containing the VM's file system
- The price per hour of using a VM

Distributed VM Management

- A distributed VM manager makes requests for VMs and queries their status.
- This manager requests VMs from the gateway on behalf of the user application.
- The manager obtains the list of requested VMs from the gateway.
- This list contains a tuple of public IP/private IP addresses for each VM with Secure Shell (SSH) tunnels.

4.1.4 Global Exchange of Cloud Resources

- Cloud infrastructure providers (i.e., IaaS providers) have established data centers in multiple geographical locations to provide redundancy and ensure reliability in case of site failures.
- Amazon does not provide seamless/automatic mechanisms for scaling its hosted services across multiple geographically distributed data centers.
- This approach has many shortcomings
- First, it is difficult for cloud customers to determine in advance the best location for hosting their services as they may not know the origin of consumers of their services.
- Second, SaaS providers may not be able to meet the QoS expectations of their service consumers originating from multiple geographical locations.
- The figure the high-level components of the Melbourne group's proposed InterCloud architecture

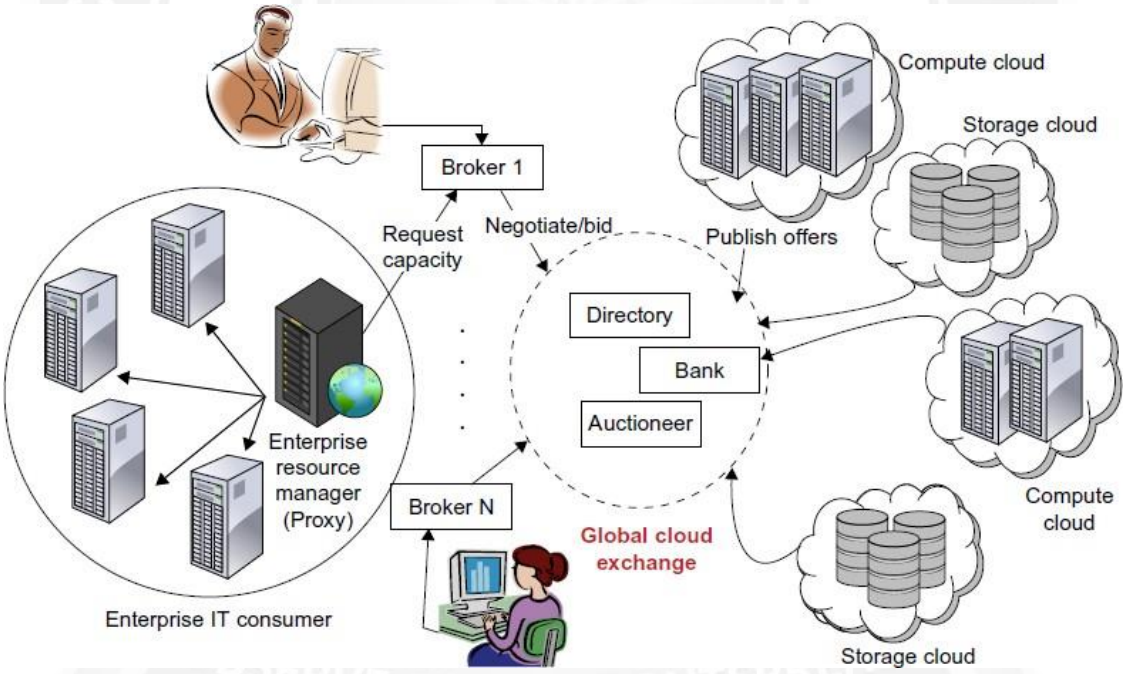


Fig: Inter-cloud exchange of cloud resources through brokering

- It is **not possible** for a cloud infrastructure provider to establish its **data centers at all possible locations** throughout the world.
- This results in **difficulty** in meeting the **QOS expectations** of their customers.

- Hence, services of **multiple cloud infrastructure** service providers are used.
- **Cloud coordinator** evaluates the available resources.
- The availability of a banking system ensures that financial transactions related to SLAs are carried out in a securely.
- By realizing InterCloud architectural principles in mechanisms in their offering, cloud providers will be able to dynamically expand or resize their provisioning capability based on sudden spikes in workload demands by leasing available computational and storage capabilities from other cloud.
- They consist of client brokering and coordinator services that support utility-driven federation of clouds:
    - application scheduling
    - resource allocation
    - migration of workloads.
- The architecture cohesively couples the administratively and topologically distributed storage and compute capabilities of clouds as part of a single resource leasing abstraction.
- The system will ease the crossdomain capability integration for on-demand, flexible, energy-efficient, and reliable access to the infrastructure based on virtualization technology
- The Cloud Exchange (CEx) acts as a market maker for bringing together service producers and consumers.
- It aggregates the infrastructure demands from application brokers and evaluates them against the available supply currently published by the cloud coordinators.
- It supports trading of cloud services based on competitive economic models such as commodity markets and auctions.
- CEx allows participants to locate providers and consumers with fitting offers.