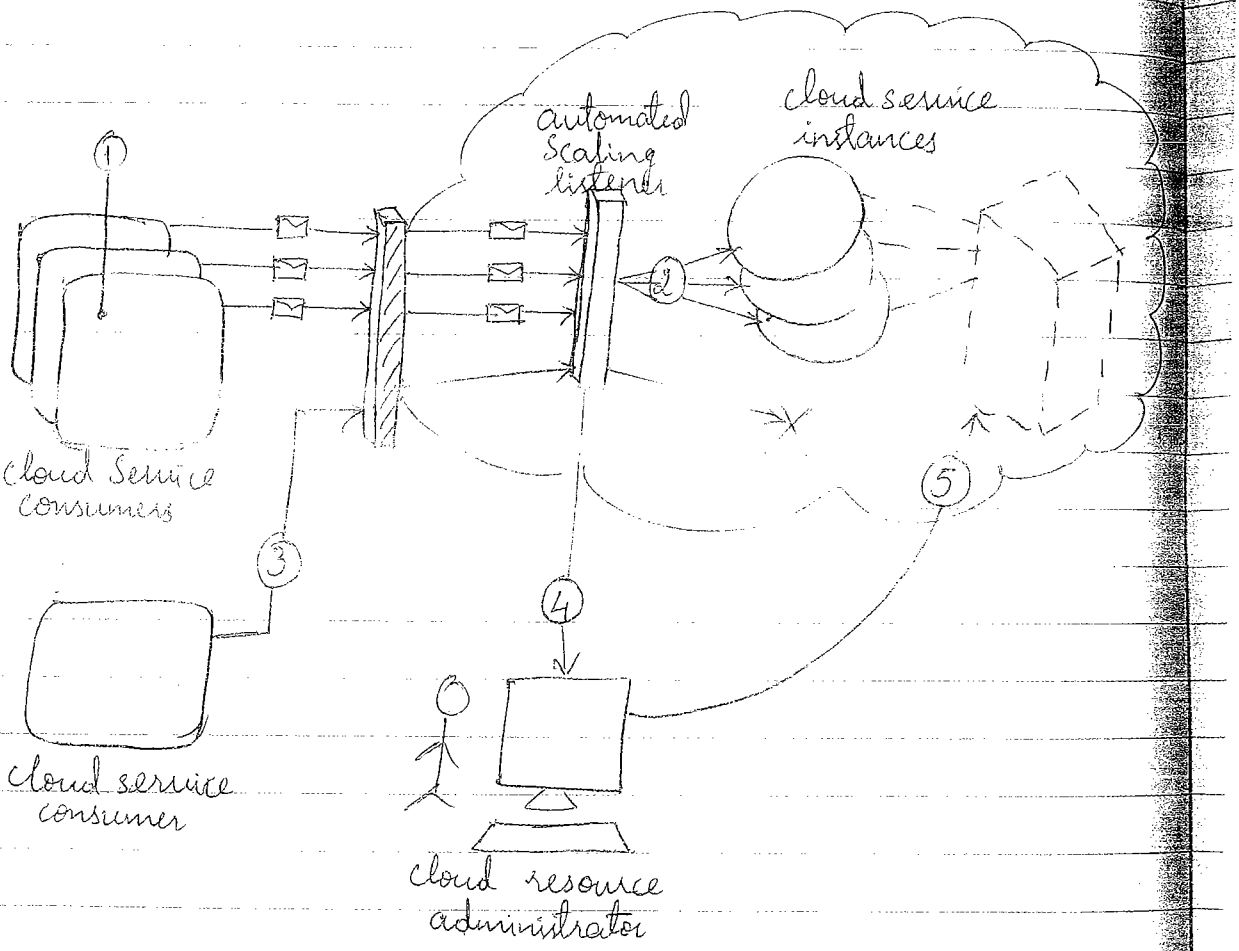
UNIT - 5

Specialized Cloud Mechanisms

Automated Scaling Listener

The automated scaling listner mechanism is a service agent that monitors & tracks communication between cloud service consumers & cloud services for dynamic scaling purposes.



Automated scaling listners can provide different types of responses to workload fluctuation conditions such as :

- Automatically scaling IT resources out or in based on parameters previously defined by the cloud consumer

- Automatic notification of the cloud consumer when workload exceed current thresholds or fall below allocated resources

Load Balancer

The load balancer mechanism is a runtime agent with fundamental logic of balancing a workload across two or more IT resources to increase performance & capacity beyond what a single IT resource can provide.

Specialized runtime workload distribution functs include.

- Asymmetric Distribution - larger workloads are issued to IT resources with higher processing capacity
- Workload Prioritization - workloads are scheduled, queued, discarded & distributed workloads according to their priority levels.
- Content-Aware Distribution - requests are distributed to different IT resources as dictated by the request content.

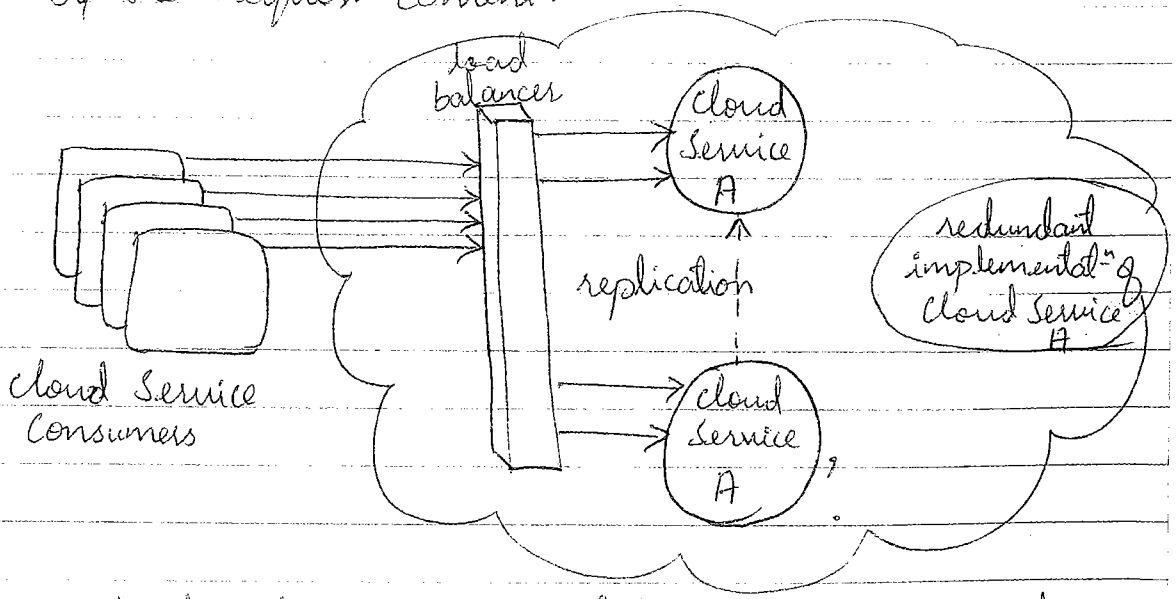


Fig: A load balancer implemented as a service agent transparently distributes incoming workload request message across two redundant cloud service implementations

Date
Page

Resource Cluster

- The resource cluster mechanism is used to group multiple IT resources instances so that they can be operated as a single IT resource.
- This increases the combined computing capacity, load balancing, & availability of the clustered IT resources.

Resource cluster architectures rely on high-speed dedicated network connections, or cluster nodes, between IT resources instances to communicate about workload distribution, task scheduling, data sharing, & system synchronization.

Common resource cluster types includes

- Server cluster

Physical or virtual servers are clustered to increase performance & availability. In Server cluster, physical servers to have access to shared storage, virtual servers are able to live-migrate from one to another. In this process, the virtualization platform suspends the execution of given virtual server at one physical server & resumes it on another physical server.

Database Cluster

Designed to improve data availability, this high-availability resource cluster has a synchronization feature that maintains the consistency of data being stored at different storage devices used in cluster.

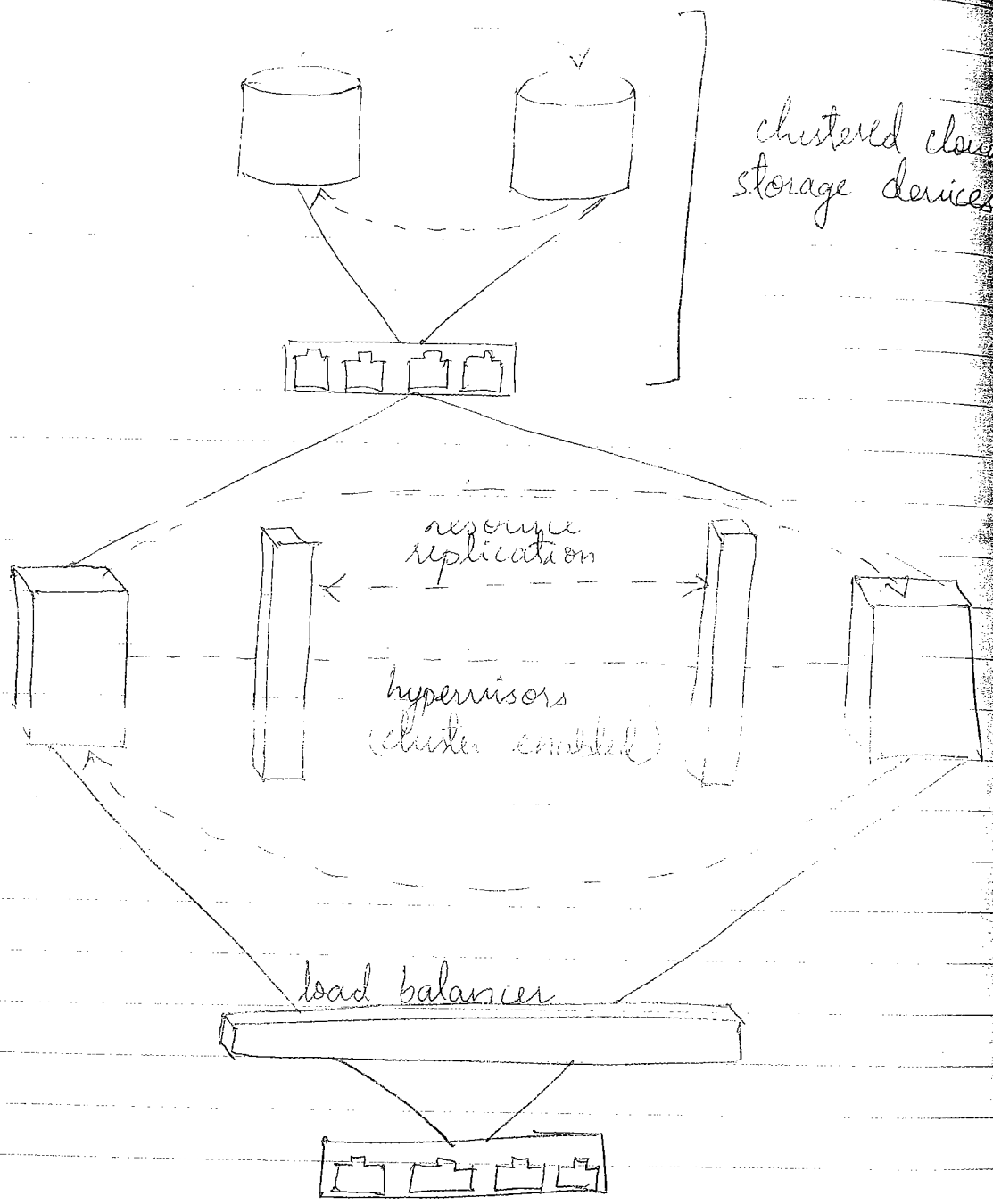


Fig: Load balancing & resource replication are implemented through a cluster-enabled hypervisor. A dedicated storage area network is used to connect the clustered storage & clustered servers which are able to share common disk storage devices.

Large Dataset Cluster - Data partitioning & distribution are implemented so that the target datasets can be efficiently partitioned without compromising data integrity or consistency

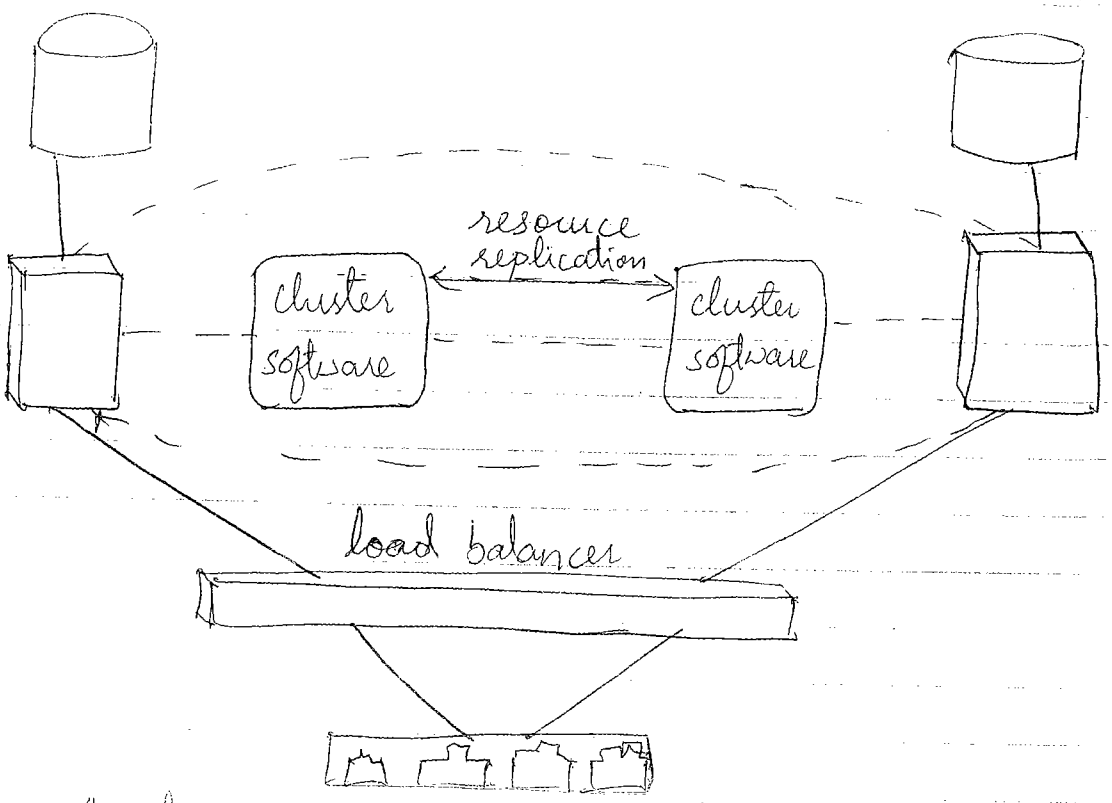


Fig 1) loosely coupled server cluster that incorporates a load balancer. There is no shared storage. Resource replication is used to replicate cloud storage devices through the network by cluster s/w.

There are two basic types of resource clusters.

1) Load Balanced Cluster.

2) HA Cluster - A high-availability cluster.

Multi-Device Broker

The multi-device broker mechanism is used to facilitate runtime data transformation so as to make a cloud service accessible to a wider range of cloud service consumer programs & devices.

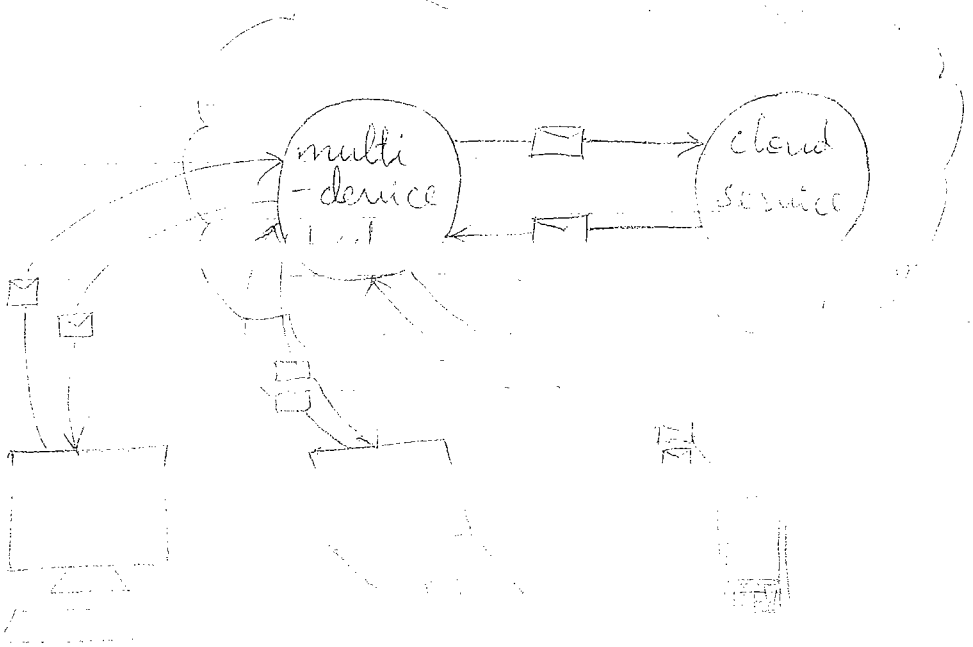


Fig. A multi-device broker contains the mapping logic necessary to transform data exchanges between a cloud service & different types of cloud service consumer devices. This scenario depicts the multi-device broker as a cloud service with its own API.

Multi-device brokers commonly exist as gateways or incorporate gateway components, such as:

- XML gateway – [illegible]
- Cloud Storage [illegible] to
& encodes storage devices to [illegible]
- Mobile Device [illegible] protocols used by mobile devices into [illegible]

The level at which transformation logic can be created include:

- transport protocols
- messaging protocols
- storage device protocols
- data schemas / data models

State Management Database

A state management database is a storage device that is used to temporarily persist state data for software programs.

- As an alternative to caching state data in memory, software programs can off-load state data to the database in order to reduce the amount of runtime memory they consume.

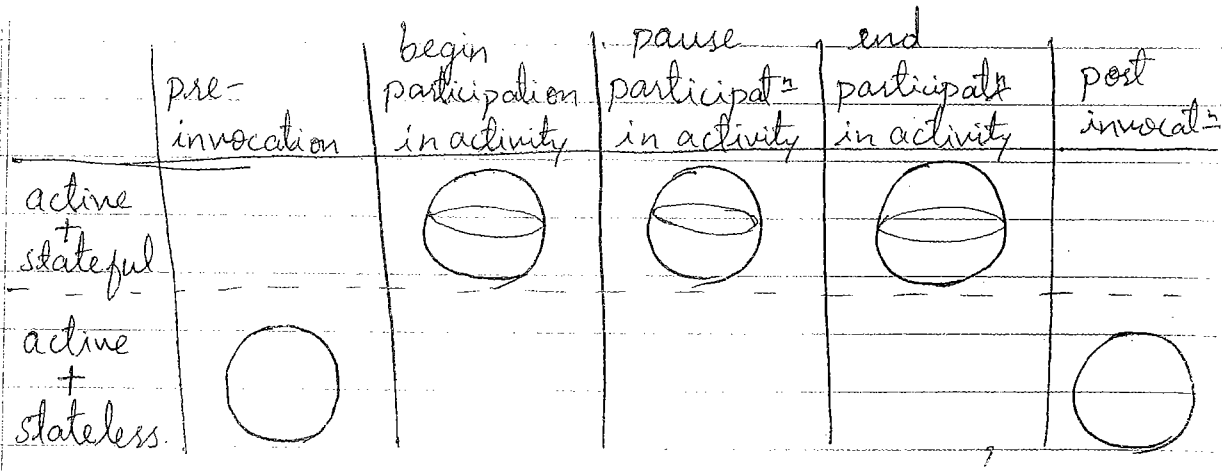


Fig: During the lifespan of a cloud service instance it may be required to remain stateful & keep state data cached in memory, even when idle.

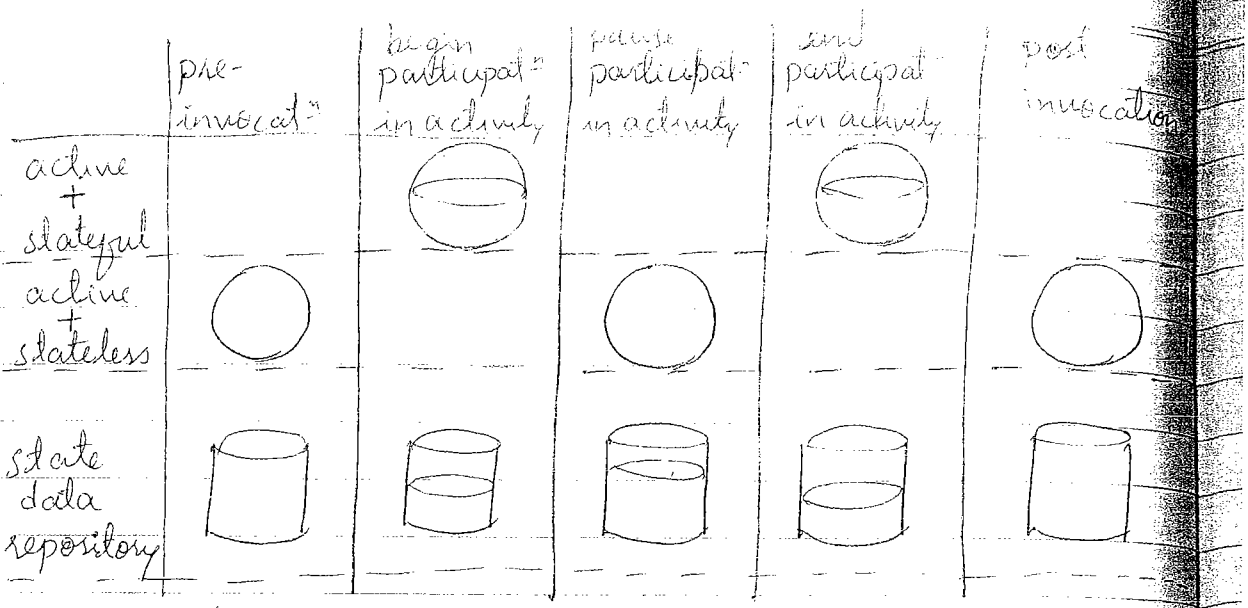


Fig: By deserting state data to a state repository, the "cloud service" is able to transition to a stateless condition, thereby temporarily freeing system resources.

Cloud Management Mechanisms

Remote Administration System

The remote administration system mechanism provides tools & user-interfaces for external cloud resource administrators to configure & administer cloud-based IT resources.

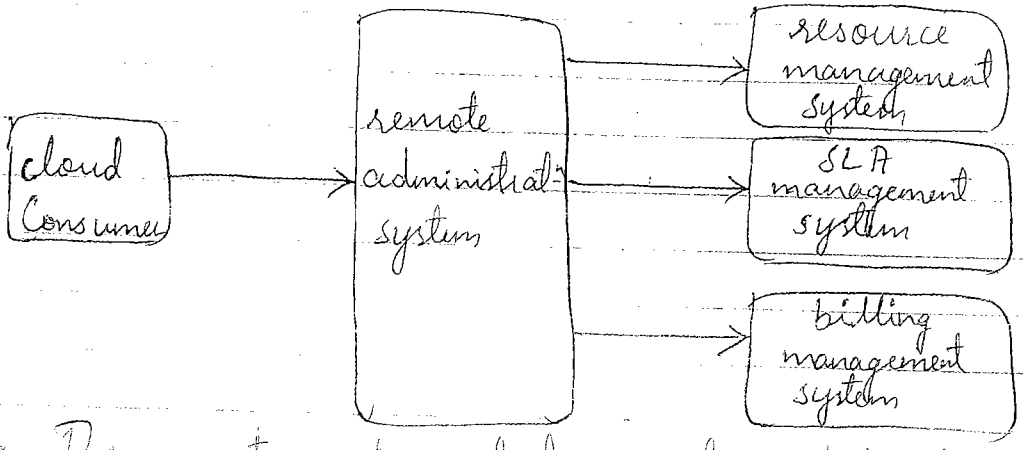
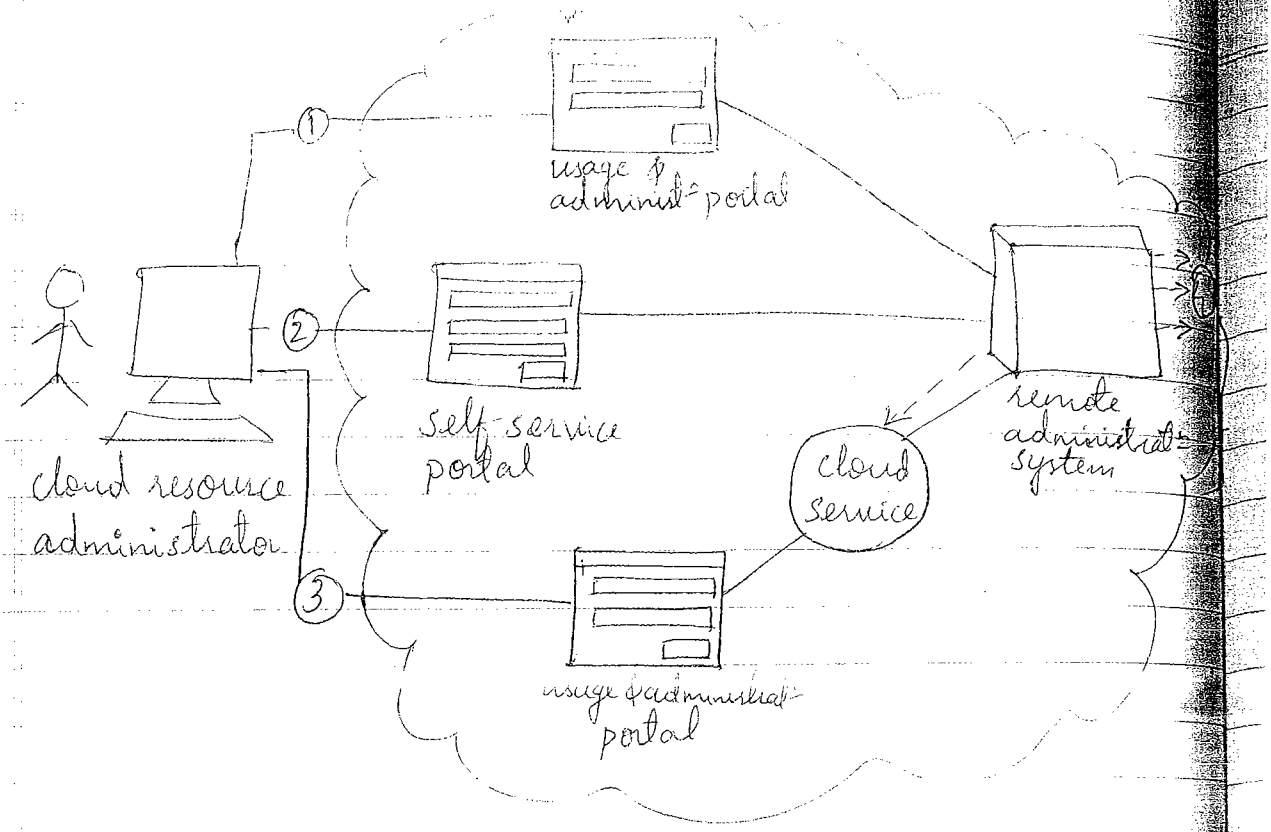


Fig. The remote administration system abstracts underlying management system to expose & centralize administrat control to external cloud resource administration.

The two primary types of portals that are created with remote administration system:

1) Usage and Administration Portal.
A general purpose portal that centralizes managment controls to different cloud-based IT resources & can further provide IT resource usage reports.

2) Self-Service Portal
This is a shopping portal that allows cloud consumers to search up-to-date list of cloud services & IT resources that are available from a cloud provider.



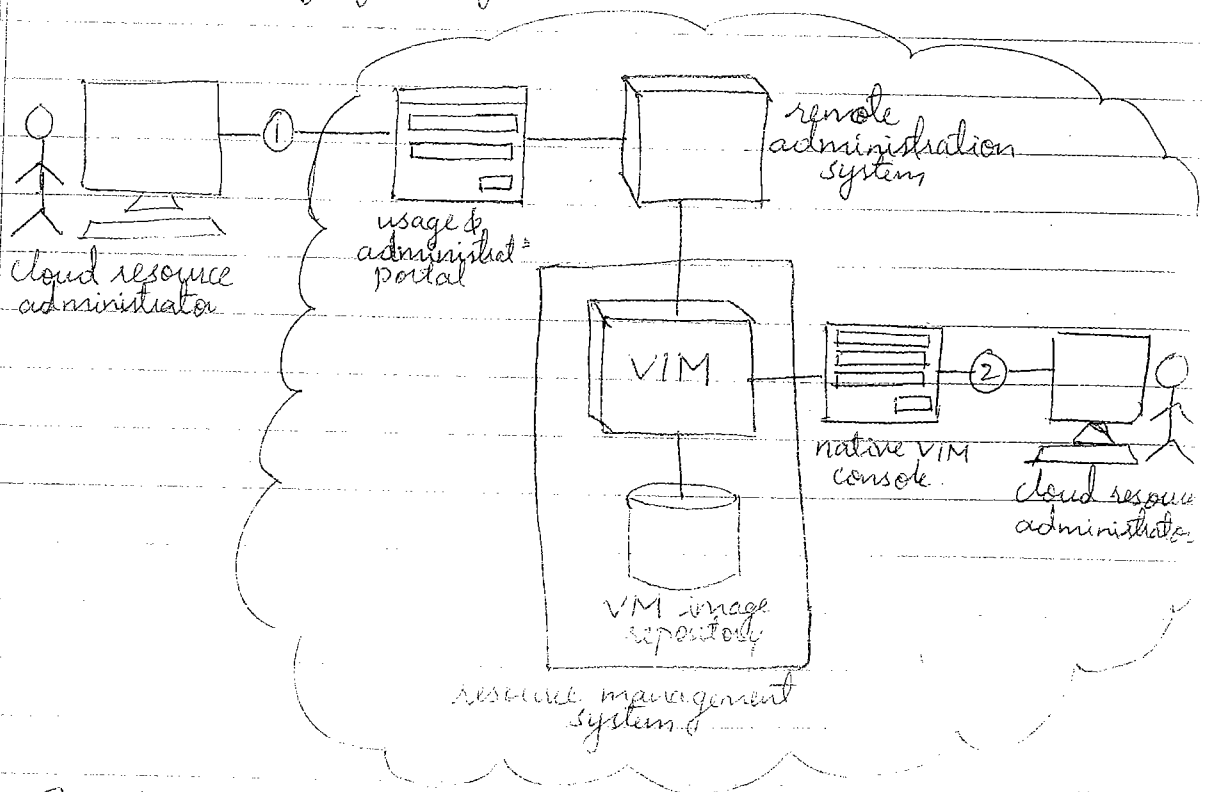
(1) The cloud resource administrator uses the usage & administ- portal to configure an already leased virtual server to prepare it for hosting.
(2) The cloud resource administrator then uses the self-service portal to select & request the provisioning of a new cloud service
(3)

Depending on i) cloud delivery model,
ii) the level of access control granted to cloud consumer,
iii) which underlying management systems the remote administration system interacts with,
tasks that can be performed via administration console include:
- Configuring & setting up cloud service
- provisioning & releasing IT resources for on-demand cloud services

- monitoring cloud services status, usage & performance
- monitoring QoS & SLA fulfillment
- managing leasing costs & usage fees
- managing user accounts, security credentials, authorizat= , & access control

Resource Management System

- The resource management system mechanism helps coordinate IT resources in response to management actions performed by both cloud consumers & cloud providers.
- Core to this system is the virtual infrastructure manage (VIM) that coordinates the server hardware so that virtual server instances can be created from underlying physical server.



① The cloud consumer's cloud resource administrator accesses a usage & administration portal externally to administer a leased IT resource.

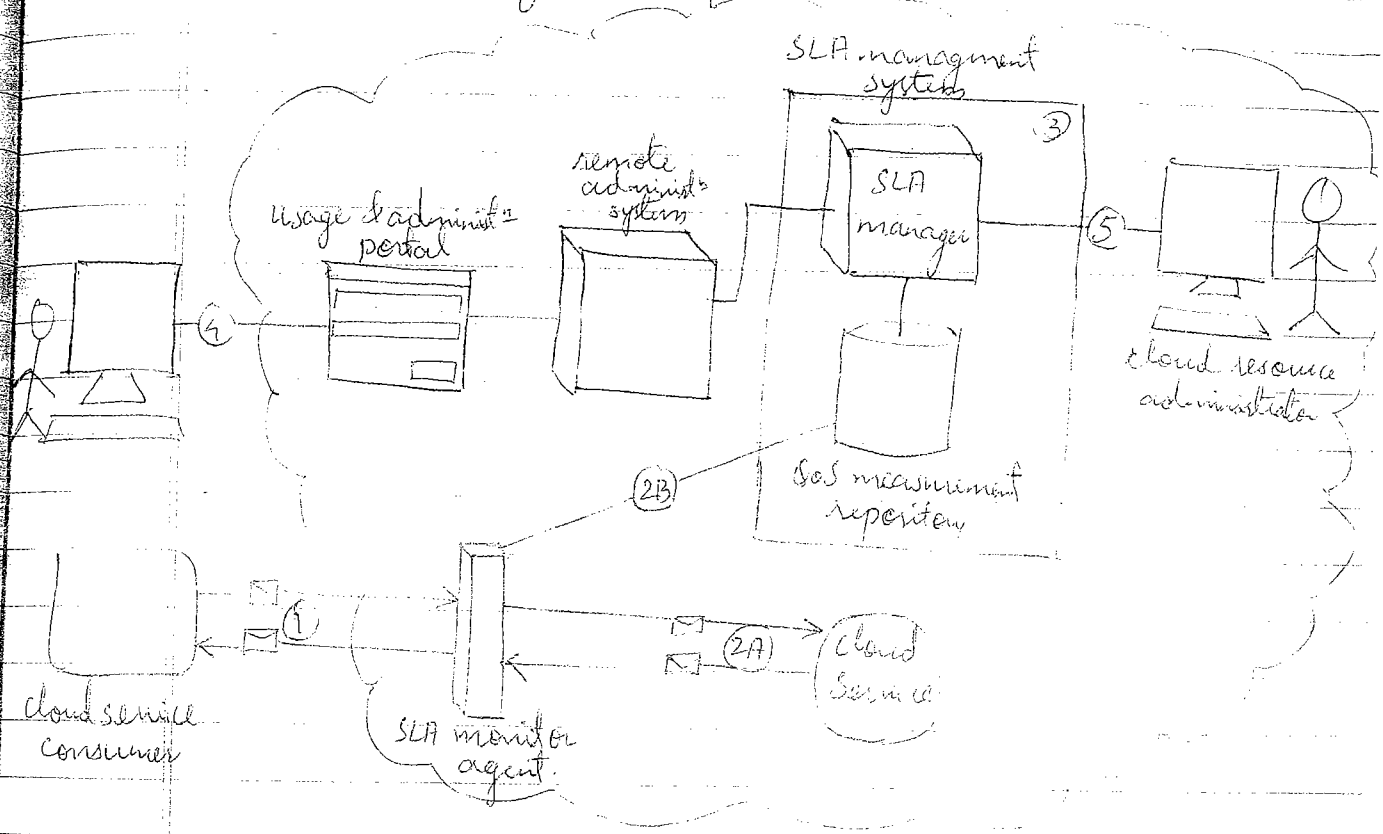
② The cloud provider's cloud resource administrator uses the native user interface provided by the VIM to perform internal resource management tasks.

Resource management system task include:

- managing virtual IT resource template that are used to create pre-built instances, such as virtual server image.
- allocating & releasing virtual IT resources into available physical infrastructure in response to starting, pausing, resuming & termination of virtual IT resource instances
- coordinating IT resources in relation to the involvement of other mechanisms, such as resource replication, load balancer, & failover system
- enforcing usage & security policies throughout the lifecycle of cloud service instances
- monitoring operational conditions of IT resources.

SLA Management System.

The SLA management system mechanism represents product that provides features pertaining to the administration, collection, storage, reporting, & runtime notification of SLA data.



Billing Management system

- The billing management system mechanism is dedicated to the collection & processing of usage data as it pertains to cloud provider accounting & cloud consumer billing.
- It relies on pay-per-use monitors to gather runtime usage data that is stored in repository & used for billing, reporting & invoicing purpose.