Asignment-1: Example of Project Summary

Project Summary	
Team Members	List your team members (including SID) here and their contact email
Goal	State the goal for your project here – short statement
Methodology	State the systematic approach you are taking to complete the project – your methods – for example how you will gather, analyze and process the information for your project. What software are you using to manage the project.
Identified Challenges	Dot point the challenges that your project will overcome
Time Line	Enter the key dates for project. eg. Start date, end date, any important check point dates.
Key research articles	List the key reference articles (top three references only)

Cloud Computing Technologies

Lecture 4

Cloud Mechanisms

Lecture 4 Outline

- Cloud Mechanisms
  - Chapter 7 Cloud Infrastructure Mechanisms
  - Chapter 8 Specialised Cloud Mechanisms
  - Chapter 9 Cloud Management Mechanisms

Cloud Infrastructure Mechanisms

1. Logical Network Perimeter
2. Virtual Server
3. Cloud Storage Device
4. Cloud Usage Monitor
5. Resource Replication
6. Ready-Made Environment

Logical Network Perimeter

- Includes and isolates a group of related cloud-based IT resources.
  - Isolate IT resources in a cloud from non-authorized users, non-users and cloud consumers
  - Control the bandwidth that is available to isolated IT resources

Logical Network Perimeter

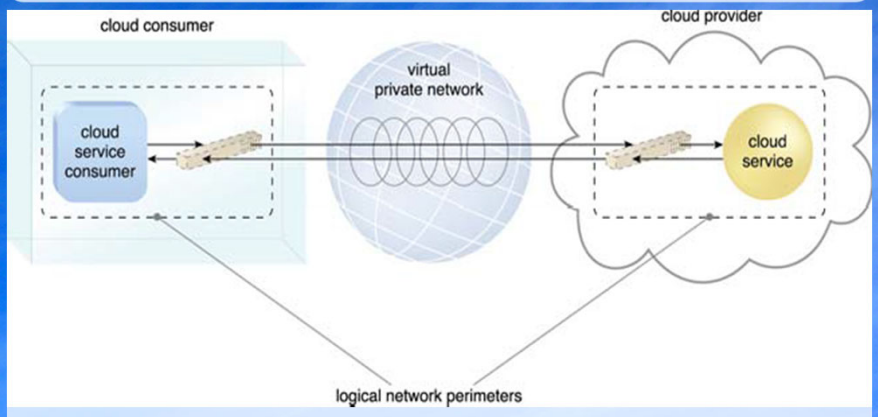


Figure 7.3 Two logical network perimeters surround the cloud consumer and cloud provider environments.

Logical Network Perimeter

- Established with the support of network devices:
  - Virtual Firewall: filters network traffic to and from the isolated network while controlling its interactions with the Internet
  - Virtual Network: acquired through VLANs, this IT resource isolates the network environment within the data centre infrastructure

Logical Network Perimeter

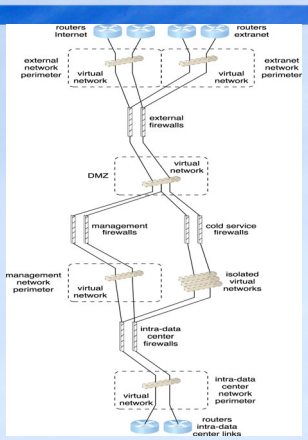


Figure 7.4 A logical network layout is established through a set of logical network perimeters using various firewalls and virtual networks

Virtual Server

- A virtual server is a form of virtualization software that emulates a physical server
  - It shares the same physical server with multiple cloud consumers (separate instances)
  - Foundation building block of cloud environments that can host numerous IT resources, cloud-based solutions, and other cloud computing mechanisms
  - Cloud consumers can customise their environments independently

Virtual Server

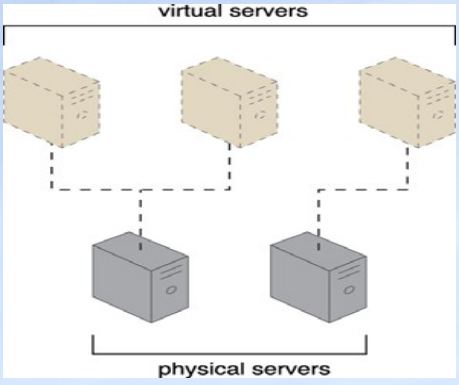


Figure 7.5 The first physical server hosts two virtual servers, while the second physical server hosts one virtual server.

Virtual Server

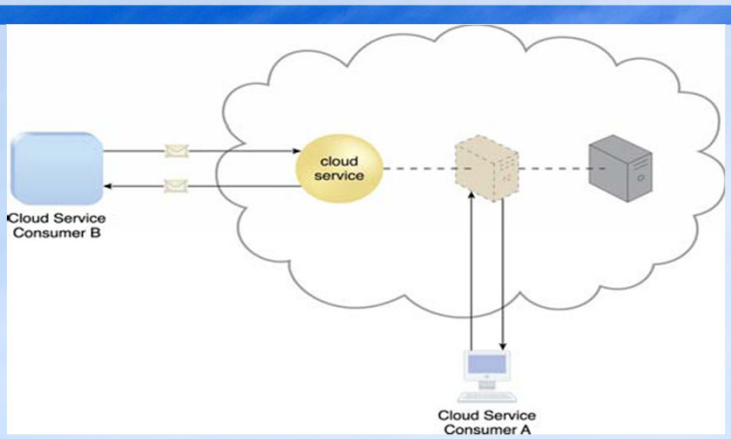


Figure 7.6 A virtual server hosts an active cloud service and is further accessed by a cloud consumer for administrative purposes.

Virtual Server

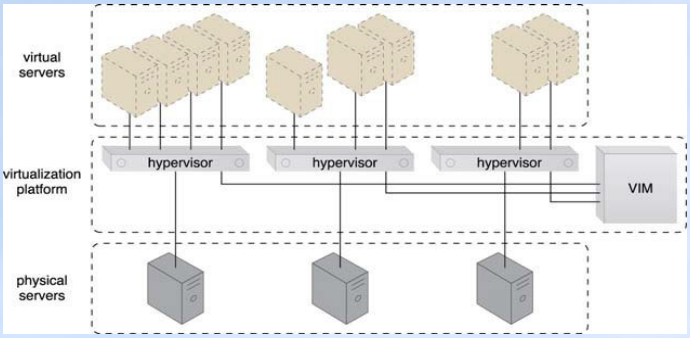


Figure 7.7 Virtual servers are created via the physical servers' hypervisors and a central VIM.

Cloud Storage Device

- Storage devices designed specifically for cloud-based provisioning
  - It can be virtualized similar to virtual servers
  - Provides fixed-increment capacity allocation
  - Visible for remote access via cloud storage services
  - Primary concern is the security, integrity, and confidentiality of data

Cloud Storage Device

- Cloud Storage Levels:
  - Files: collections of data grouped into files located in folders
  - Blocks: lowest level of storage and closest to hardware
  - Datasets: data organised into table-based, delimited, or record format
  - Objects: data and its associated metadata organised as web-based resources

Cloud Storage Device

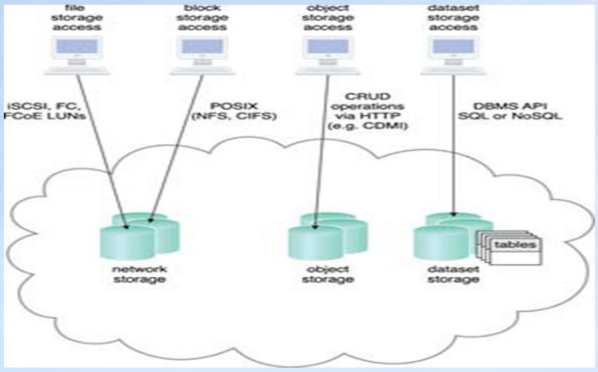


Figure 7.9 Different cloud service consumers utilize different technologies to interface with virtualized cloud storage devices. (Adapted from the CDMI Cloud Storage Reference Model.)

Cloud Storage Device

- Storage Interfaces
  - Network Storage Interfaces: including SCSI for storage blocks, server message block (SMB), common Internet file service (CIFS), and Network File System (NFS)
  - Object Storage Interfaces: typically accessed via REST or web service-based cloud services, also Storage Network Industry Association's (SNIA's) Cloud Data Management Interface (CDMI)

Cloud Storage Device

- Database Storage Interfaces: support a query language
  - Relational Data: usually SQL-based, e.g., DB2, Oracle, MSSQL, MySQL
  - Non-relational Data (or NoSQL databases): uses a "looser" structure (no data relations) to avoid complexity and processing overhead

Cloud Usage Monitor

- A lightweight autonomous software program that collects and processes IT resource usage data
  - Monitoring: intermediary, event-driven program existing as a service agent that transparently monitors and analyses data flows
    - Collects network traffic and message metrics

Cloud Usage Monitor

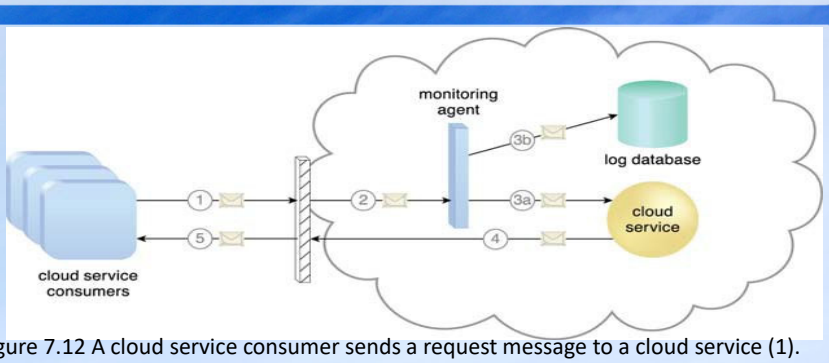


Figure 7.12 A cloud service consumer sends a request message to a cloud service (1). The monitoring agent intercepts the message to collect relevant usage data (2) before allowing it to continue to the cloud service (3a). The monitoring agent stores the collected usage data in a log database (3b). The cloud service replies with a response message (4) that is sent back to the cloud service consumer without being intercepted by the monitoring agent (5).

Cloud Usage Monitor

- Resource Agent: processing module collecting usage data by having event-driven interactions with specialised resource software
  - Collects observable resource events, e.g., initiating, suspending, resuming, vertical scaling
- Polling Agent: processing module that collects cloud service usage data
  - Monitor IT resource status, e.g., uptime vs downtime

Cloud Usage Monitor

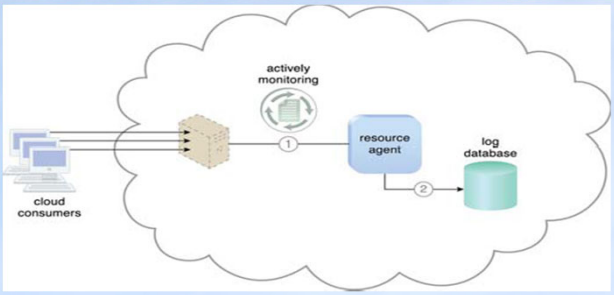


Figure 7.13:

(1)The resource agent is actively monitoring a virtual server and detects an increase in usage.

(2) The resource agent receives a notification from the underlying resource management program that the virtual server is being scaled up and stores the collected usage data in a log database, as per its monitoring metrics.

Cloud Usage Monitor

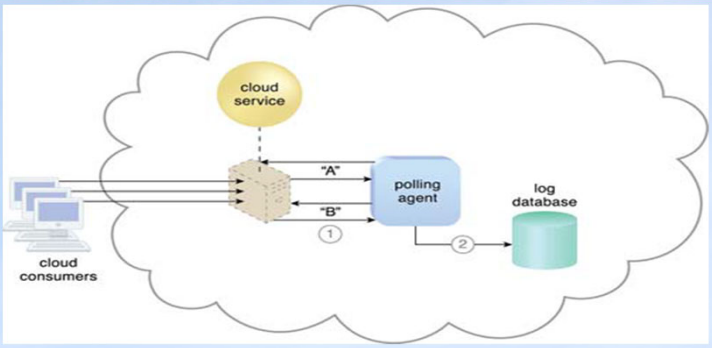
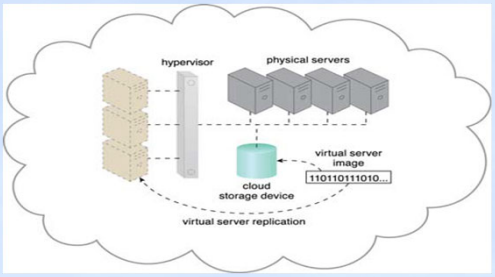


Figure 7.14 A polling agent monitors the status of a cloud service hosted by a virtual server by sending periodic polling request messages and receiving polling response messages that report usage status "A" after a number of polling cycles, until it receives a usage status of "B" (1), upon which the polling agent records the new usage status in the log database (2).

Resource Replication

- Multiple instances of the same IT resource to enhance availability and performance
  - Implemented using virtualisation technology

Figure 7.16 The hypervisor replicates several instances of a virtual server, using a stored virtual server image



Ready-Made Environment

- A set of pre-installed IT resources, ready to be used and customised
  - Example: PaaS cloud delivery model
  - Includes:
    a. databases,
    b. middleware,
    c. development tools,
    d. governance tools,
    e. SDKs, runtime execution environments, etc.

Ready-Made Environment

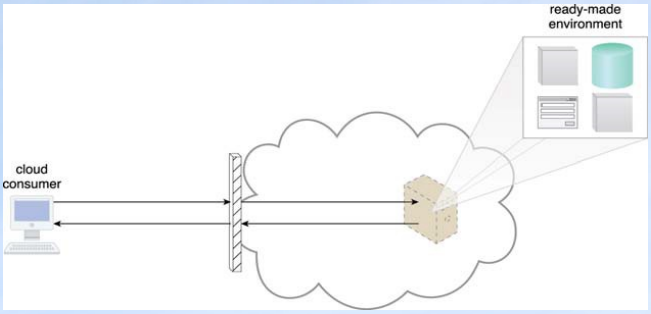


Figure 7.20 A cloud consumer accesses a ready-made environment hosted on a virtual server

Specialized Cloud Mechanisms

1. Automated Scaling Listener
2. Load Balancer
3. SLA Monitor
4. Pay-Per-Use Monitor
5. Audit Monitor
6. Failover System
7. Hypervisor
8. Resource Cluster
9. Multi-Device Broker
10. State Management Database

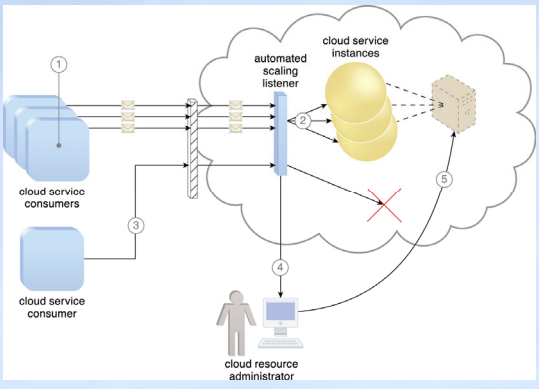
Specialised Cloud Mechanisms

- Automated Scaling Listener: service agent that monitors and tracks communications between cloud service consumers and cloud services for dynamic scaling purposes
  - Typically deployed in the cloud near the firewall to track workload by monitoring the volume of cloud consumer-generated requests
  - Responses include scaling IT resources out or in, notifications sent to the cloud consumer

Specialised Cloud Mechanisms

- Automated Scaling Listener:

Figure 8.1
(1).Three cloud service consumers attempt to access one cloud service simultaneously
(2). The automated scaling listener scales out and initiates the creation of three redundant instances of the service
(3). A fourth cloud service consumer attempts to use the cloud service
(4). Programmed to allow up to only three instances of the cloud service, the automated scaling listener rejects the fourth attempt and notifies the cloud consumer that the requested workload limit has been exceeded
(5). The cloud consumer's cloud resource administrator accesses the remote administration environment to adjust the provisioning setup and increase the redundant instance limit

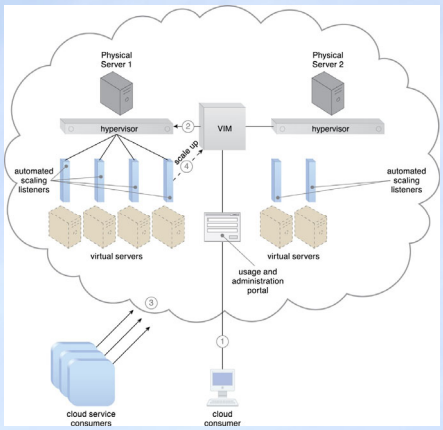


Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

Specialised Cloud Mechanisms

- Automated Scaling Listener:

Figure 8.2
(1) A cloud consumer creates and starts a virtual server with 8 virtual processor cores and 16 GB of virtual RAM.
(2) The VIM creates the virtual server at the cloud service consumer's request and allocates it to Physical Server 1 to join 3 other active virtual servers .
(3). Cloud consumer demand causes the virtual server usage to increase by over 80% of the CPU capacity for 60 consecutive seconds.
(4). The automated scaling listener running at the hypervisor detects the need to scale up and commands the VIM accordingly.



Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

Specialised Cloud Mechanisms

- **Load Balancer:** balances workload across two or more IT resources to increase performance and capacity beyond a single IT resource
  - Asymmetric distribution: larger workloads issued to IT resources with higher processing capacities
  - Workload prioritisation: workloads scheduled, queued, discarded, and distributed according to priority level
  - Content-aware distribution: requests distributed to different IT resources depending on request content

Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

Specialised Cloud Mechanisms

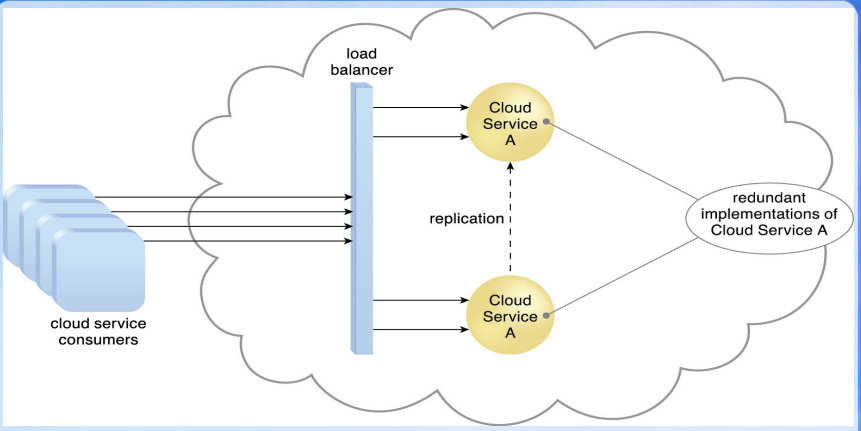


Figure 8.5 A **load balancer implemented** as a service agent transparently distributes incoming workload request messages across two redundant cloud service implementations, which in turn maximizes performance for the cloud service consumers.

Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

Specialised Cloud Mechanisms

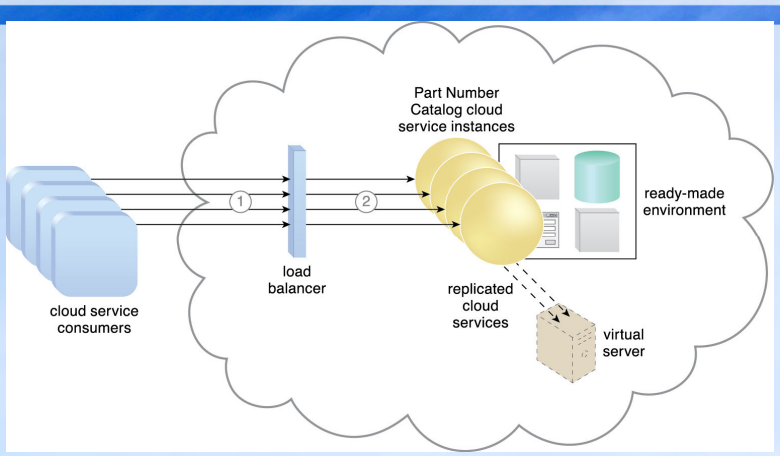


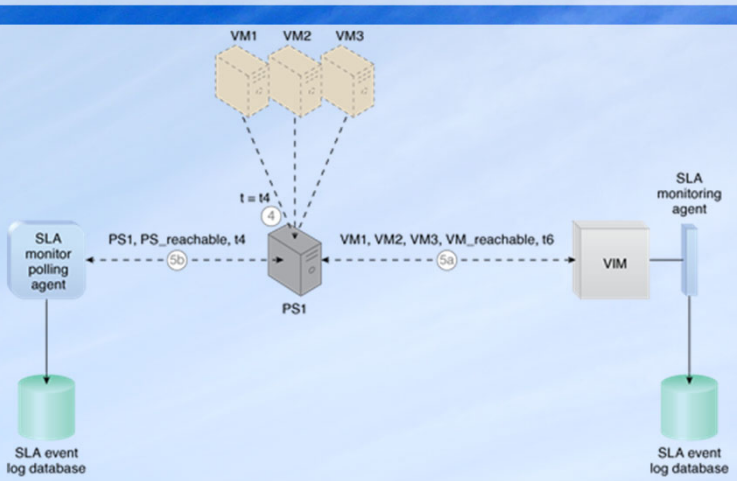
Figure 8.6 New instances of the cloud services are automatically created to meet increasing usage requests. The **load balancer** uses round-robin scheduling to ensure that the traffic is distributed evenly among the active cloud services.

Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

SLA Monitor

- **SLA Monitor:** observes runtime performance of cloud services to ensure they fulfil contractual QoS requirements as published in SLAs
  - Data collected is aggregated into reporting metrics
  - Can proactively repair or failover cloud services when exception conditions occur

SLA Monitor

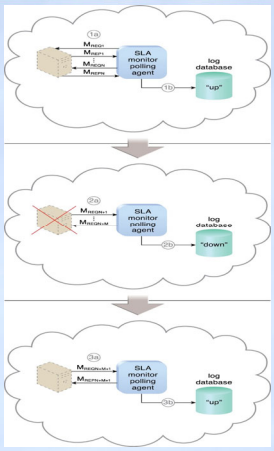


Specialised Cloud Mechanisms: SLA Monitor

Figure 8.7 The SLA monitor polls the cloud service by sending over polling request messages (M_{REQ1} to M_{REQN}). The monitor receives polling response messages (M_{REP1} to M_{REPN}) that report that the service was "**up**" at each polling cycle (1a). The SLA monitor stores the "up" time—time period of all polling cycles 1 to N—in the log database (1b).

The SLA monitor polls the cloud service that sends polling request messages (M_{REQN+1} to M_{REQN+M}). Polling response messages are not received (2a). The response messages continue to time out, so the SLA monitor stores the "**down**" time—time period of all polling cycles N+1 to N+M—in the log database (2b).

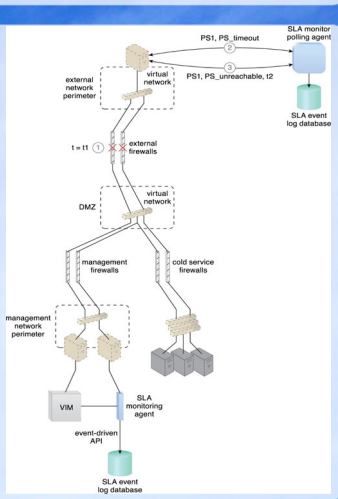
The SLA monitor sends a polling request message ($M_{REQN+M+1}$) and receives the polling response message ($M_{REPN+M+1}$) (3a). The SLA monitor stores the "**up**" time in the log database (3b).



Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

Specialised Cloud Mechanisms: SLA Monitor

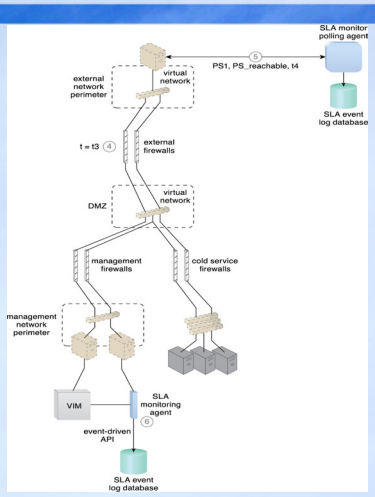
Figure 8.8 At timestamp = t1, a firewall cluster has failed and all of the IT resources in the data center become unavailable (1). The **SLA monitor** polling agent stops receiving responses from physical servers and starts to issue PS_timeout events (2). The **SLA monitor** polling agent starts issuing PS_unreachable events after three successive PS_timeout events. The timestamp is now t2 (3).



Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

SLA Monitor

Figure 8.9 The IT resource becomes operational at timestamp = t3 (4). The SLA monitor polling agent receives responses from the physical servers and issues PS_reachable events. The timestamp is now t4 (5). The SLA monitoring agent did not detect any unavailability since the communication between the VIM platform and physical servers was not affected by the failure (6).



Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

Pay-Per-Use Monitor

- measures cloud-based IT resource usage for fee calculations and billing purposes
- includes:
  - request/response message quantity
  - transmitted data volume
  - bandwidth (data transfer rate)

Specialised Cloud Mechanisms:
Pay-Per -Use Monitor

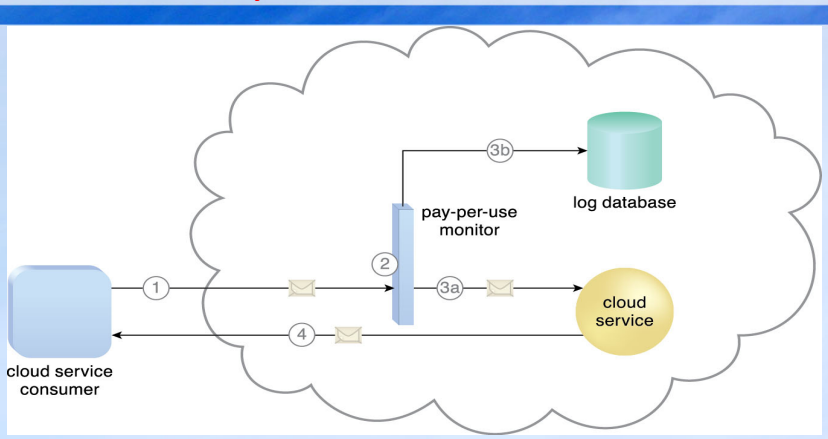


Figure 8.13 A cloud service consumer sends a request message to the cloud service (1). The pay-per-use monitor intercepts the message (2), forwards it to the cloud service (3a), and stores the usage information in accordance with its monitoring metrics (3b). The cloud service forwards the response messages back to the cloud service consumer to provide the requested service (4).

Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

Audit Monitor

- Collects data (audit tracking data) for networks and IT resources in support of regulatory and contractual obligations

Specialised Cloud Mechanisms:
Audit Monitor

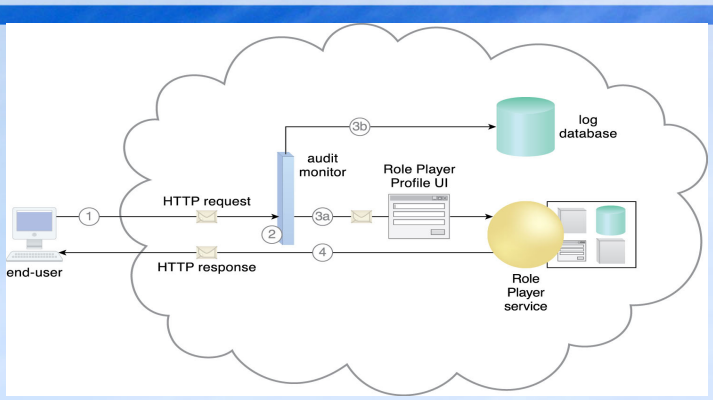


Figure 8.16 An end-user attempts access to the Role Player cloud service (1). An **audit monitor** transparently intercepts the HTTP request message and analyzes the message header to determine the geographical origin of the end-user (2). The audit monitoring agent determines that the end-user is from a region that Innovartus is not authorized to charge a fee for access to the application. The agent forwards the message to the cloud service (3a) and generates the audit track information for storage in the log database (3b). The cloud service receives the HTTP message and grants the end-user access at no charge (4).

Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

Failover System:

- increases reliability and availability of IT resources through redundancy
  - Automatically switches over to a redundant or standby IT resource instance when currently active IT resource becomes unavailable

Failover System:

- Approaches:
  - Active-Active: redundant IT resources actively serving workload synchronously using load balancing, remaining operational resources continue service after failure
  - Active-Passive: standby or inactive implementation activated upon failure of active IT resource

Specialised Cloud Mechanisms: Failover System

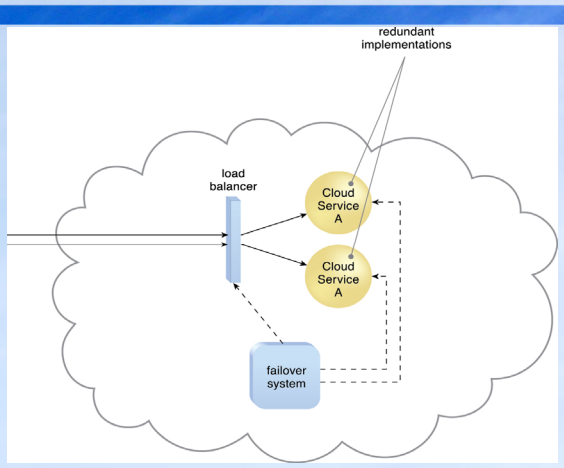


Figure 8.17 The failover system monitors the operational status of Cloud Service A.

Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

Specialised Cloud Mechanisms: Failover System

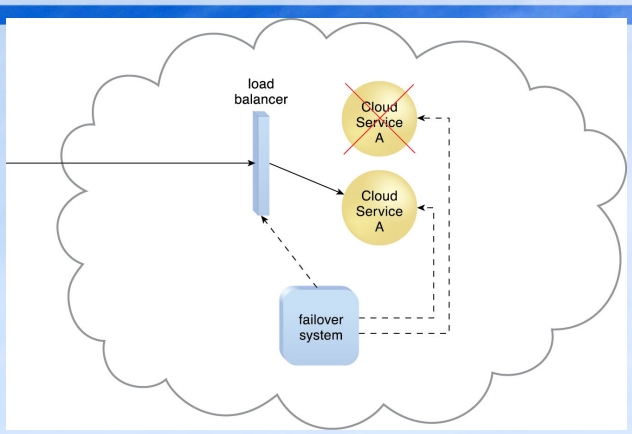


Figure 8.18 When a failure is detected in one Cloud Service A implementation, the **failover system** commands the load balancer to switch over the workload to the redundant Cloud Service A implementation.

Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

Specialised Cloud Mechanisms: Failover System

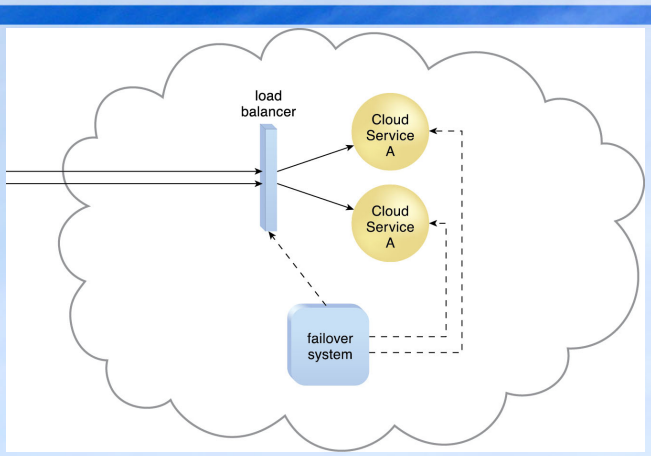


Figure 8.19 The failed Cloud Service A implementation is recovered or replicated into an operational cloud service. The **failover system** now commands the load balancer to distribute the workload again.

Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

Hypervisor:

- fundamental part of virtualization infrastructure used to generate virtual server instances

Specialised Cloud Mechanisms:
Hypervisor:

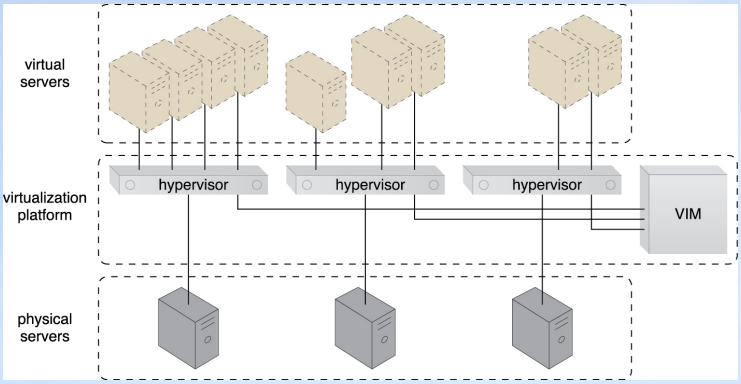


Figure 8.27 Virtual servers are created via individual hypervisor on individual physical servers. All three hypervisors are jointly controlled by the same VIM.

Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

Specialised Cloud Mechanisms:
Hypervisor:

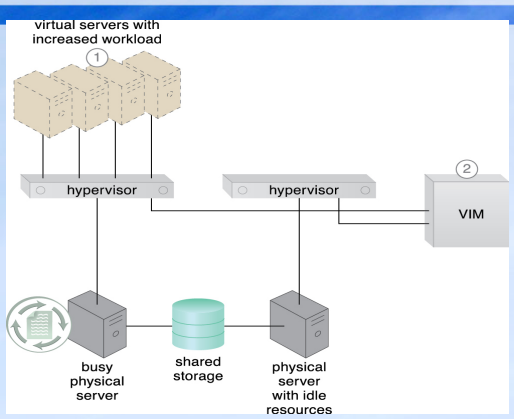


Figure 8.28 A virtual server capable of auto-scaling experiences an increase in its workload (1). The VIM decides that the virtual server cannot scale up because its underlying physical server host is being used by other virtual servers (2).

Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

Resource Cluster:

- geographically diverse cloud-based IT resources are combined into groups to improve their allocation and use, examples:
  - Server Cluster: physical or virtual servers clustered to increase performance and availability
  - Database Cluster: improve data availability by synchronising data stored in different devices
  - Large Dataset Cluster: implement data partitioning and distribution to efficiently divide large data sets without compromising data integrity or computing accuracy

Resource Cluster:

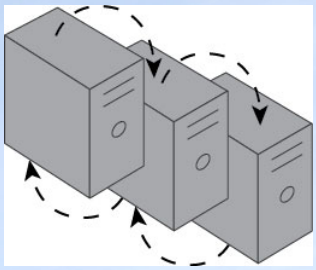


Figure 8.30 The curved dashed lines are used to indicate that IT resources are clustered

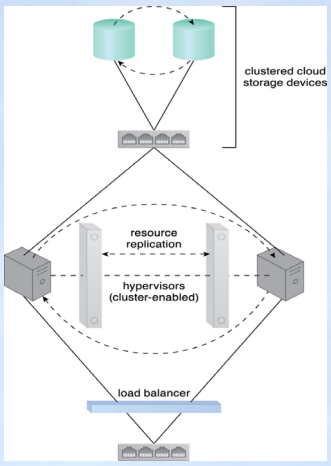
Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

SIT706 Week 4

51

Specialised Cloud Mechanisms:
Resource Cluster:

Figure 8.31 Load balancing and resource replication are implemented through a cluster-enabled hypervisor. A dedicated storage area network is used to connect the clustered storage and the clustered servers, which are able to share common cloud storage devices. This simplifies the storage replication process, which is independently carried out at the storage cluster. (See the *Hypervisor Clustering Architecture* section in Chapter 12 for a more detailed description.)



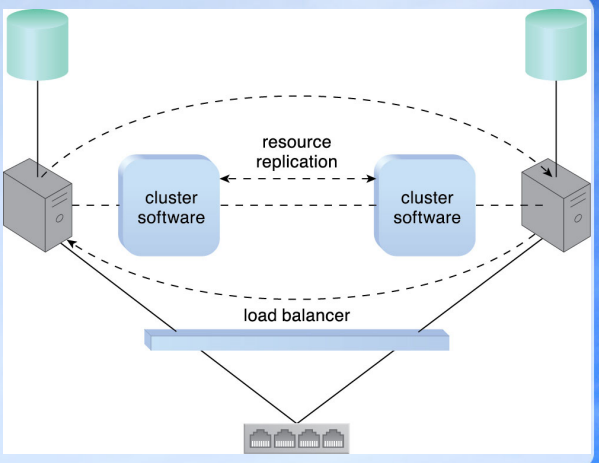
Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

SIT706 Week 4

52

Specialised Cloud Mechanisms:
Resource Cluster:

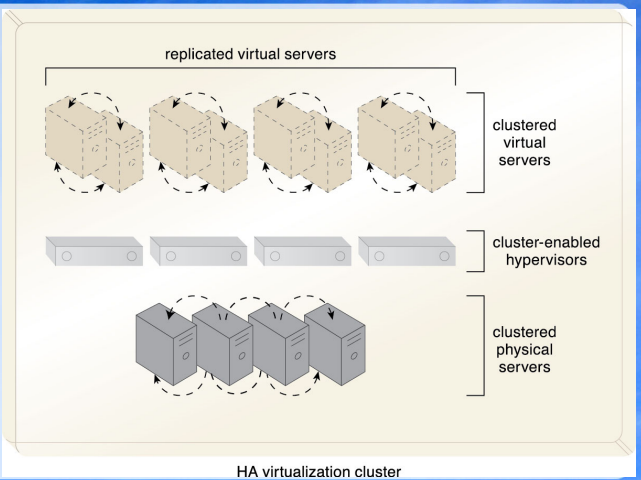
Figure 8.32 A loosely coupled server cluster that incorporates a load balancer. There is no shared storage. Resource replication is used to replicate cloud storage devices through the network by the cluster software.



Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

Specialised Cloud Mechanisms:
High Availability Virtualization Cluster:

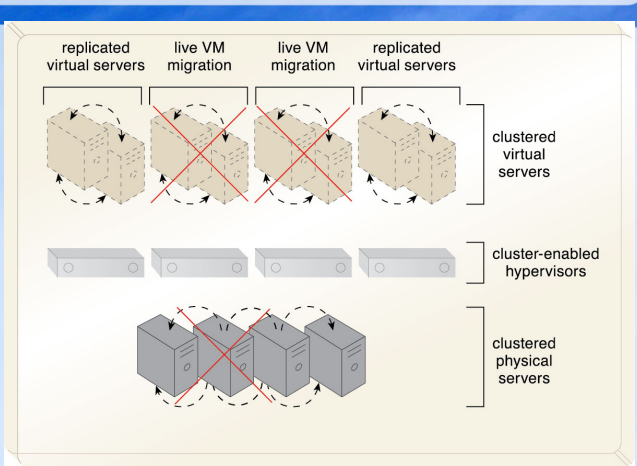
Figure 8.33 An HA Virtualization cluster of physical servers is deployed using a cluster-enabled hypervisor, which guarantees that the physical servers are constantly in sync. Every virtual server that is instantiated in the cluster is automatically replicated in at least two physical servers.



Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

Specialised Cloud Mechanisms:
High Availability Virtualization Cluster:

Figure 8.34 All of the virtual servers that are hosted on a physical server experiencing failure are automatically migrated to other physical servers.



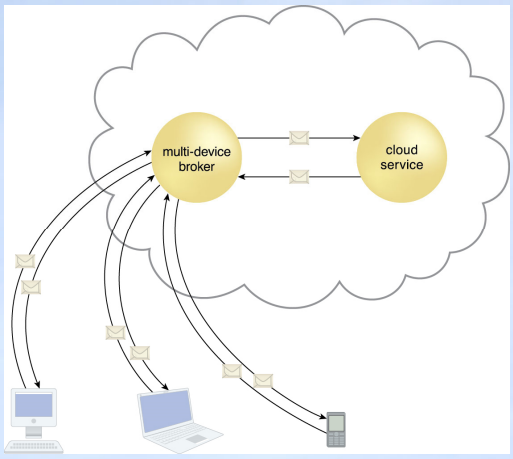
Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

Multi-Device Broker:

- runtime data transformation to allow cloud services access to a wider range of cloud service consumer programs and devices.

Multi-Device Broker:

Figure 8.35 A multi-device broker contains the mapping logic necessary to transform data exchanges between a cloud service and different types of cloud service consumer devices. This scenario depicts the multi-device broker as a cloud service with its own API. This mechanism can also be implemented as a service agent that intercepts messages at runtime to perform necessary transformations.
.



Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

SIT706 Week 4

57

Multi-Device Broker:

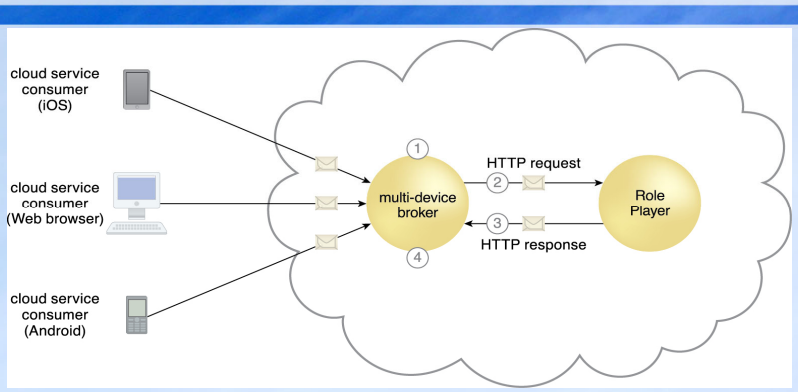


Figure 8.36 The multi-device broker intercepts incoming messages and detects the platform (Web browser, iOS, Android) of the source device (1). The multi-device broker transforms the message into the standard format required by the Innovates cloud service (2). The cloud service processes the request and responds using the same standard format (3). The multi-device broker transforms the response message into the format required by the source device and delivers the message (4).

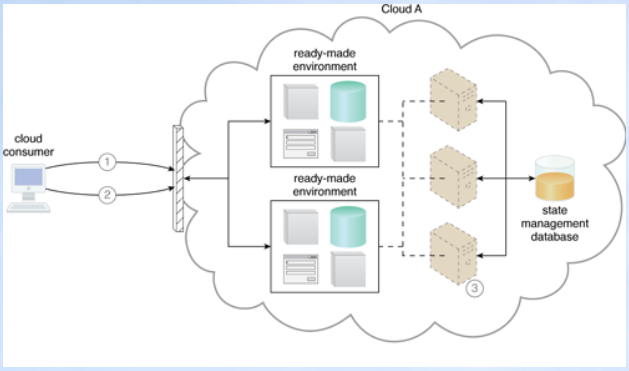
Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

SIT706 Week 4

58

State Management Database:

- storage device for temporarily persisting state information instead of caching state data in memory



Cloud Management Mechanisms

1. Remote Administration System
2. Resource Management System
3. SLA Management System
4. Billing Management System

Remote Administration System

- External configuration and administration of cloud-based IT resources
  - Primarily two types of portal:
    - Usage and Administration Portal: management controls for cloud-based IT resources
    - Self-Service Portal: effectively a "shopping portal" allowing cloud consumers to choose cloud-based IT resources for provisioning
  - Common tasks include:
    - Configuring cloud services, provisioning/releasing IT resources
    - Monitoring status, usage, and performance, QOS and SLA fulfilment
    - Managing leasing costs and usage fees, user accounts, security credentials, authorization, access control
    - Tracking internal and external access to leased services
    - Planning and assessing IT resource provisioning
    - Capacity planning

Remote Administration System

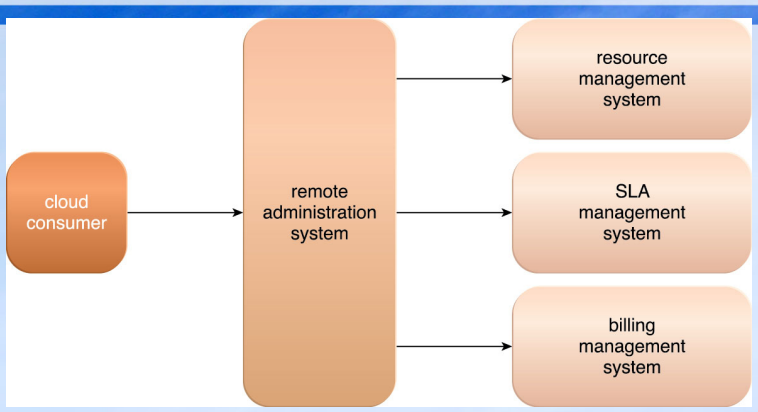


Figure 9.2 The remote administration system abstracts underlying management systems to expose and centralize administration controls to external cloud resource administrators. The system provides a customizable user console, while programmatically interfacing with underlying management systems via their APIs.

Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

Remote Administration System

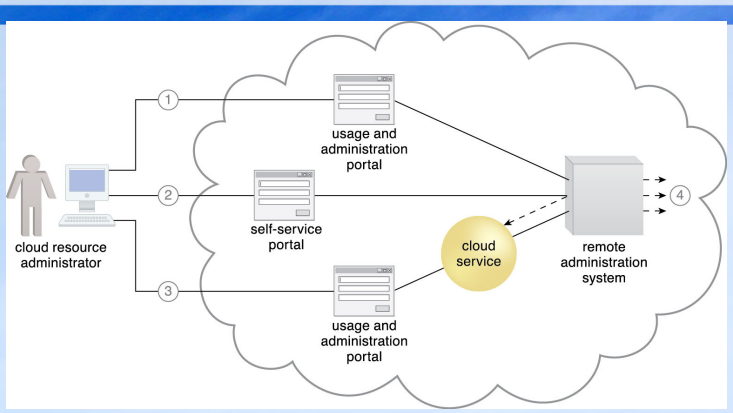


Figure 9.3 A cloud resource administrator uses the usage and administration portal to configure an already leased virtual server (not shown) to prepare it for hosting (1). The cloud resource administrator then uses the self-service portal to select and request the provisioning of a new cloud service (2). The cloud resource administrator then accesses the usage and administration portal again to configure the newly provisioned cloud service that is hosted on the virtual server (3). Throughout these steps, the remote administration system interacts with the necessary management systems to perform the requested actions (4).

Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

Resource Management System

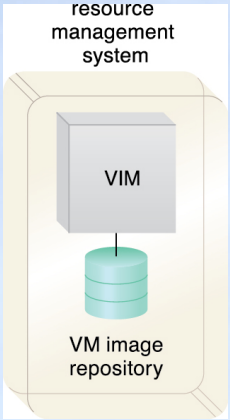
- Managing templates used to create instances e.g., virtual server images
- Allocating/releasing virtual IT resources in response to start/pause/resume/terminate of virtual IT resource instances

Resource Management System

- Coordinate IT resources for other mechanisms such as resource replication, load balancing, failover systems
- Enforcing usage and security policies
- Monitoring operational conditions of IT resources

Resource Management System

Figure 9.5 A resource management system encompassing a VIM platform and a virtual machine image repository. The VIM may have additional repositories, including one dedicated to storing operational data.



Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

Remote Administration System

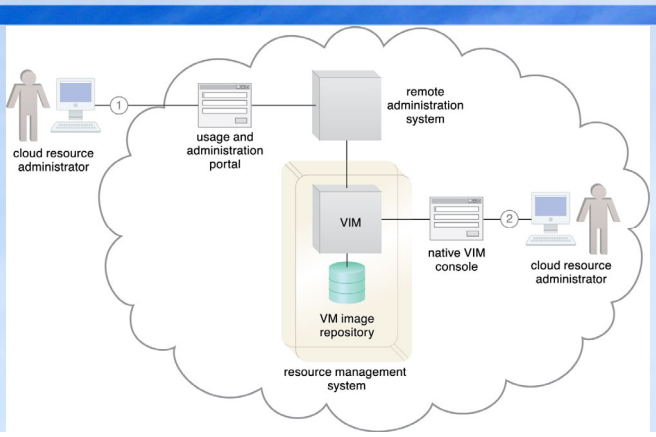


Figure 9.6 The cloud consumer's cloud resource administrator accesses a usage and administration portal externally to administer a leased IT resource (1). The cloud provider's cloud resource administrator uses the native user-interface provided by the VIM to perform internal resource management tasks (2).

Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

SLA Management System

- Includes repository for storing/retrieving SLA data
- one or more SLA monitoring mechanisms collect SLA data
- Metrics monitored, for individual cloud services, are aligned with SLA guarantees

SLA Management System

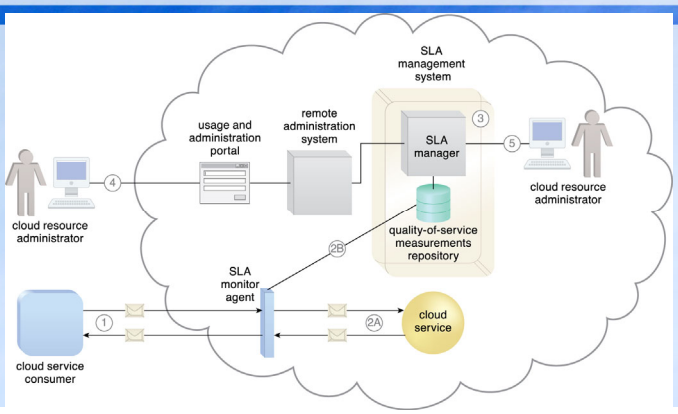


Figure 9.8 A cloud service consumer interacts with a cloud service (1). An SLA monitor intercepts the exchanged messages, evaluates the interaction, and collects relevant runtime data in relation to quality-of-service guarantees defined in the cloud service's SLA (2A). The data collected is stored in a repository (2B) that is part of the SLA management system (3). Queries can be issued and reports can be generated for an external cloud resource administrator via a usage and administration portal (4) or for an internal cloud resource administrator via the SLA management system's native user-interface (5).

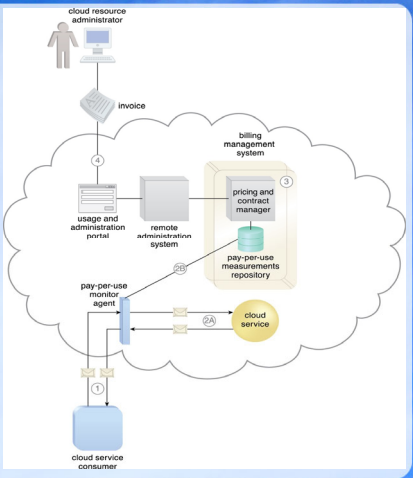
Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

Billing Management System

- Collection and processing of usage data relevant for accounting and billing

Billing Management System

Figure 9.10 A cloud service consumer exchanges messages with a cloud service (1). A pay-per-use monitor keeps track of the usage and collects data relevant to billing (2A), which is forwarded to a repository that is part of the billing management system (2B). The system periodically calculates the consolidated cloud service usage fees and generates an invoice for the cloud consumer (3). The invoice may be provided to the cloud consumer through the usage and administration portal (4).



Source: *Cloud Computing* by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini

Summary

- Cloud Infrastructure Mechanisms,
- Specialised Cloud Mechanisms and
- Cloud Management Mechanisms