

Serverless Data Processing (CSCI 5410)

Dr. Saurabh Dey

Images and contents used in these slides are from web search and/or text books, and are used for academic purposes only

This content is protected and may not be shared, uploaded, or
distributed

Outline

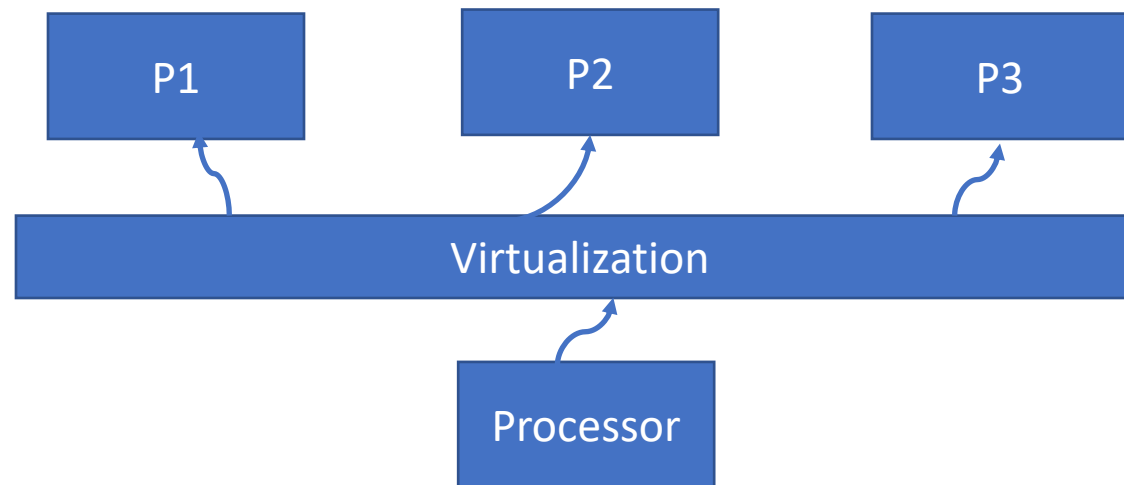
Cloud Computing Architecture,
Hardware and Software



Virtualization

- ❑ Virtualization is a critical element of the cloud infrastructure.
- ❑ Virtualization simulates the interface to a physical object by:
- ❑ **Multiplexing:** creates multiple virtual objects from one instance of a physical object. Example - a processor is multiplexed among a number of processes or threads.

Abstraction



Physical Layer

Citation: Dan C. Marinescu. (2018) *Cloud Computing Theory and Practice*, Second Ed

This content is protected and may not be shared, uploaded, or distributed

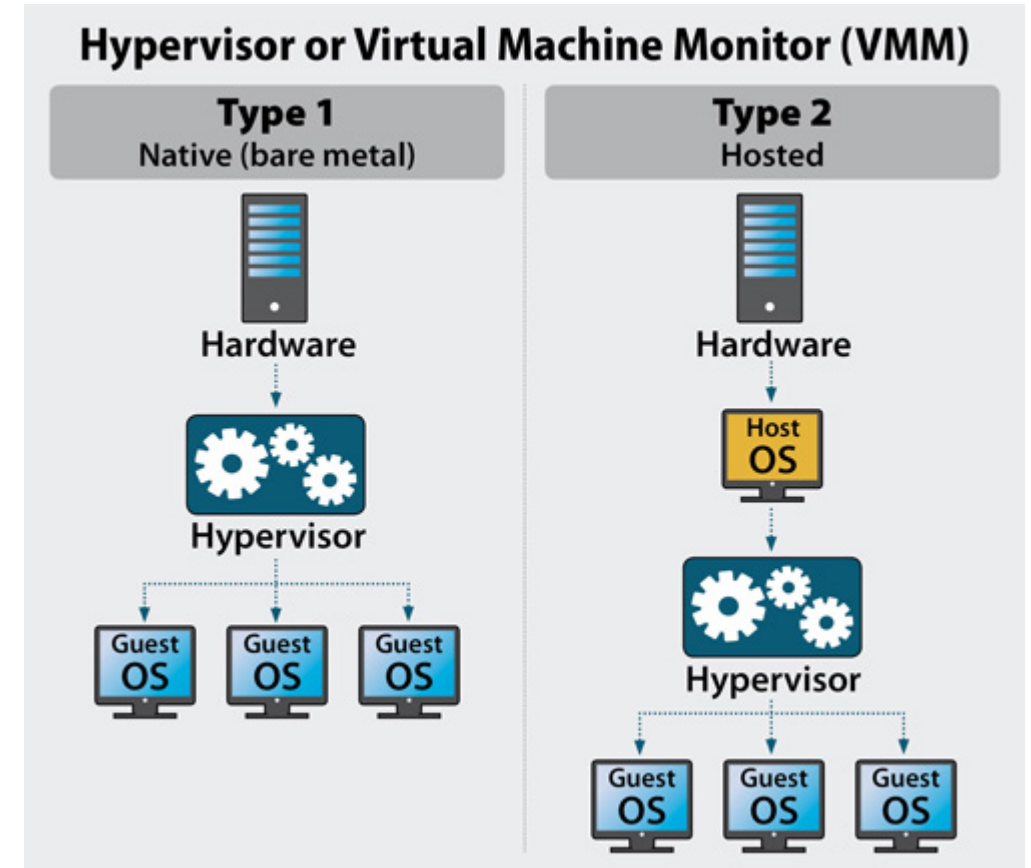
Virtual Machines (Cloud Computing uses VMs)

- ❑ Running multiple VMs on the same server allows applications to better share the server resources and lead to higher processor utilization.
- ❑ Virtualization also provides more freedom for the system resource management because VMs can be easily migrated.
- ❑ **VM migration steps:** a VM is stopped, its state is saved as a file, the file is transported to another server, and the VM is restarted

Citation: Dan C. Marinescu. (2018) *Cloud Computing Theory and Practice*, Second Ed

Hypervisors

- ❑ Hypervisor is a system software manages sharing of a physical processor among a number of virtual machines.
- ❑ A hypervisor is generally limited to one physical server and can therefore only create virtual images of that server.
- ❑ Two types
 - Native
 - Hosted



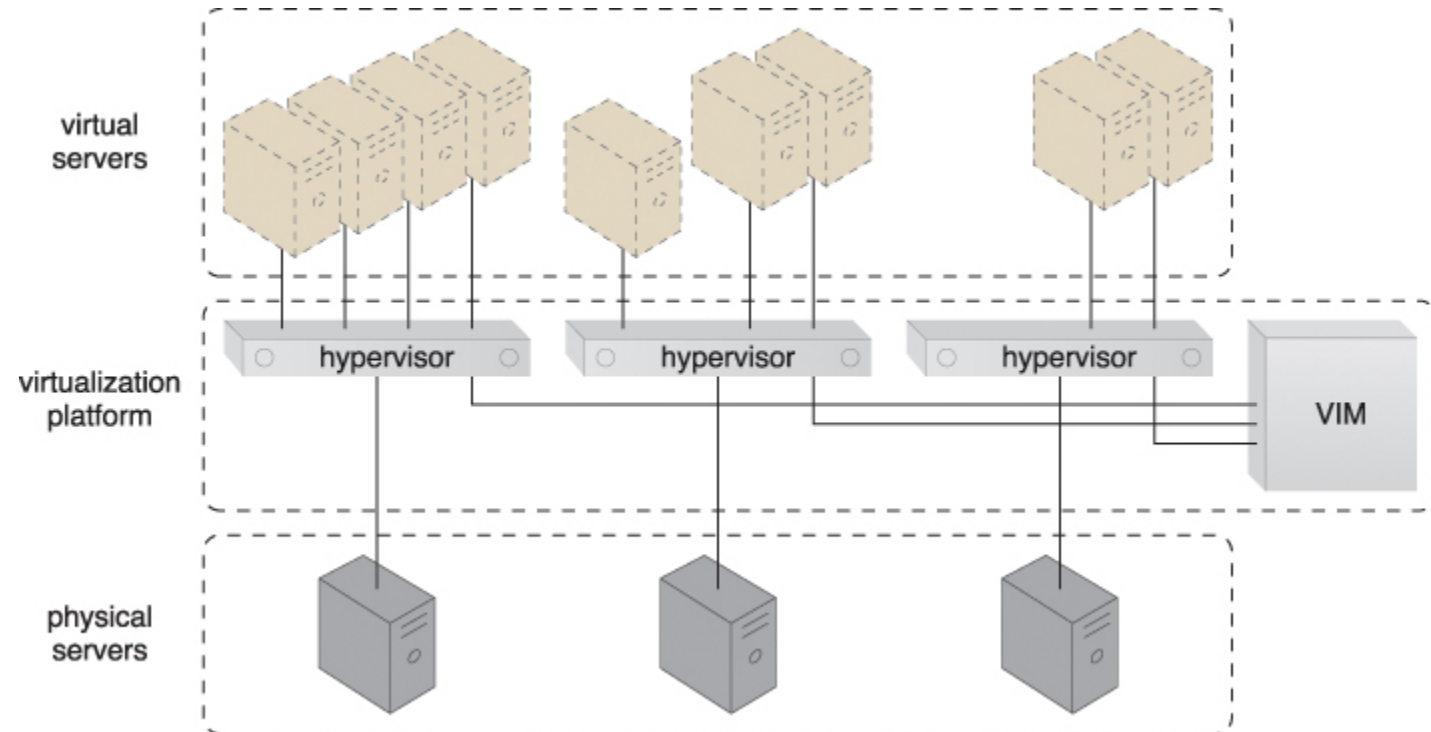
Source: <https://www.serverwatch.com/server-trends/guide-to-hypervisors.html> (Author: Christine Taylor, Nov 2018)

Citation: Dan C. Marinescu. (2018) *Cloud Computing Theory and Practice*, Second Ed

This content is protected and may not be shared, uploaded, or distributed

Hypervisor

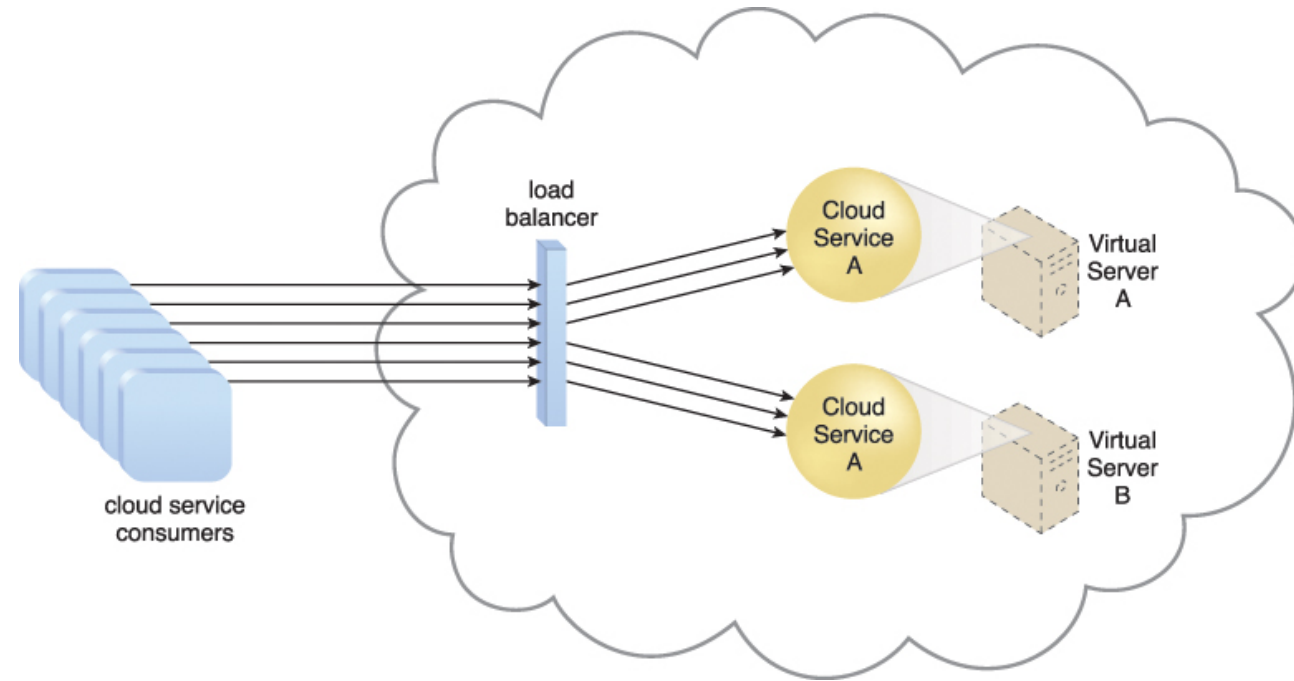
- ❑ A hypervisor has limited virtual server management features, such as increasing the virtual server's capacity or shutting it down.
- ❑ The VIM provides a range of features for administering multiple hypervisors across physical servers.



Citation: Erl, Thomas, Ricardo Puttini, Zaigham Mahmood. Cloud Computing.. [VitalSource Bookshelf].

Fundamentals of Cloud Architecture

- ❑ In **(1) Workload Distribution Architecture** IT resources can be horizontally scaled via
 - Addition of one or more identical IT resources, and
 - A load balancer that provides runtime logic capable of evenly distributing the workload



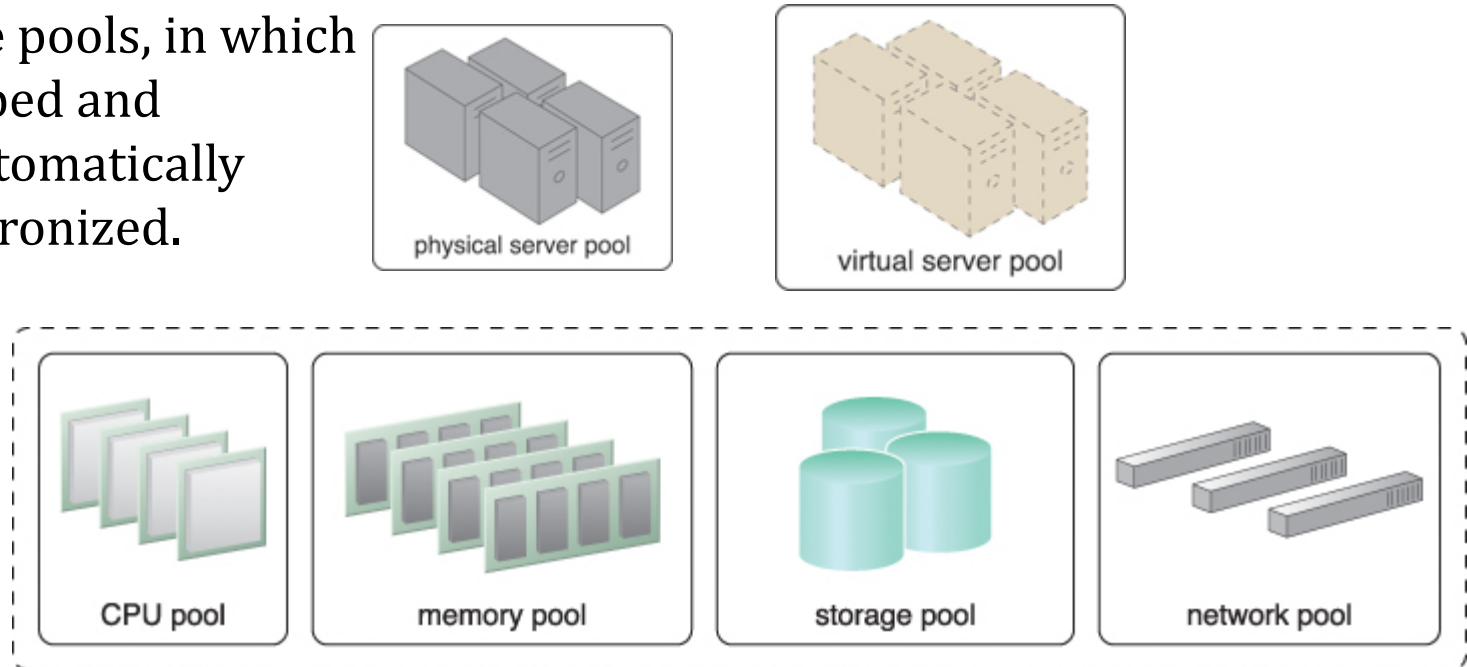
Workload Distribution Architecture

Citation: Erl, Thomas, Ricardo Puttini, Zaigham Mahmood. Cloud Computing.. [VitalSource Bookshelf].

This content is protected and may not be shared, uploaded, or distributed

Fundamentals of Cloud Architecture

- ❑ A **(2) Resource Pooling Architecture** is based on the use of one or more resource pools, in which identical IT resources are grouped and maintained by a system that automatically ensures that they remain synchronized.

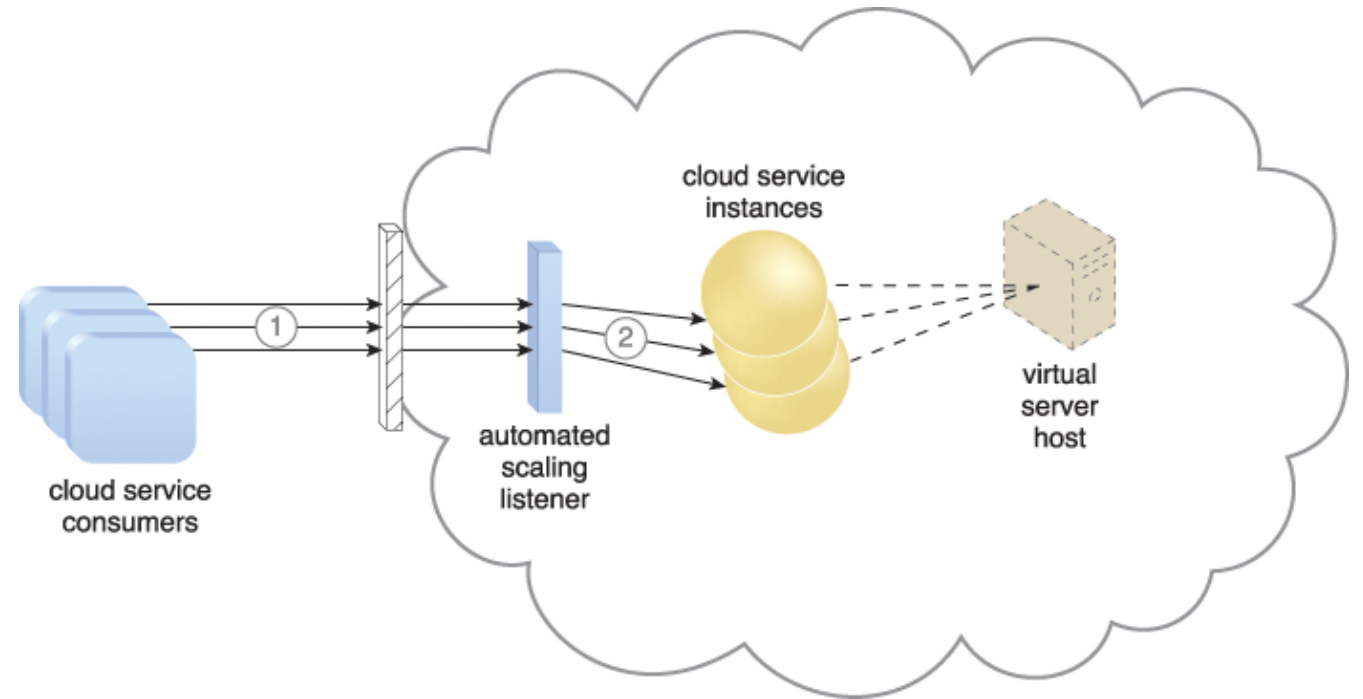


Citation: Erl, Thomas, Ricardo Puttini, Zaigham Mahmood. Cloud Computing.. [VitalSource Bookshelf].

Resource Pooling Architecture

Fundamentals of Cloud Architecture

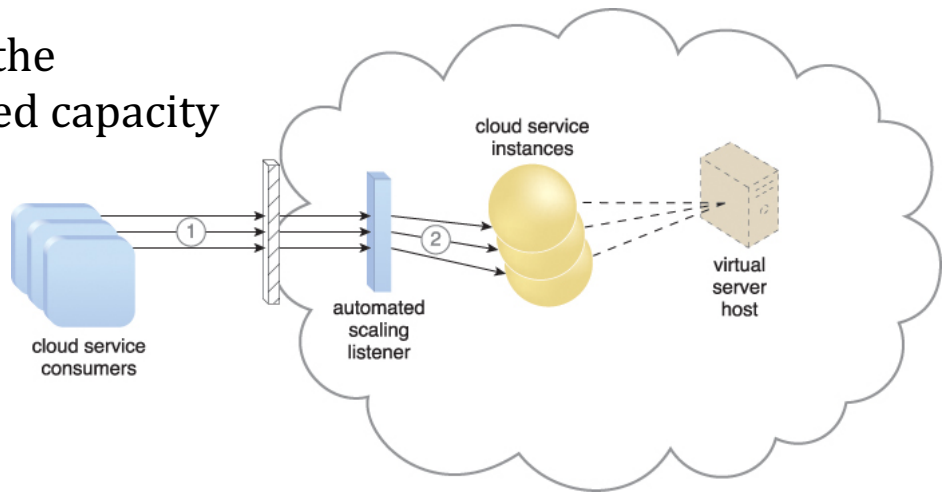
- ❑ A **(3) Dynamic Scalability Architecture** is an architectural model based on a system of predefined scaling conditions that trigger the dynamic allocation of IT resources from resource pools



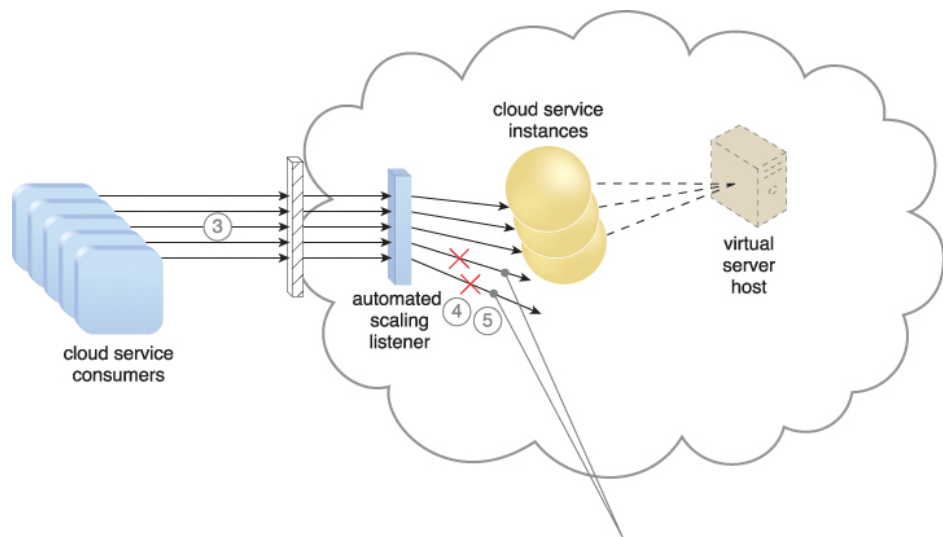
Citation: Erl, Thomas, Ricardo Puttini, Zaigham Mahmood. Cloud Computing.. [VitalSource Bookshelf].

Dynamic Scalability Architecture

Listener monitors the traffic for predefined capacity

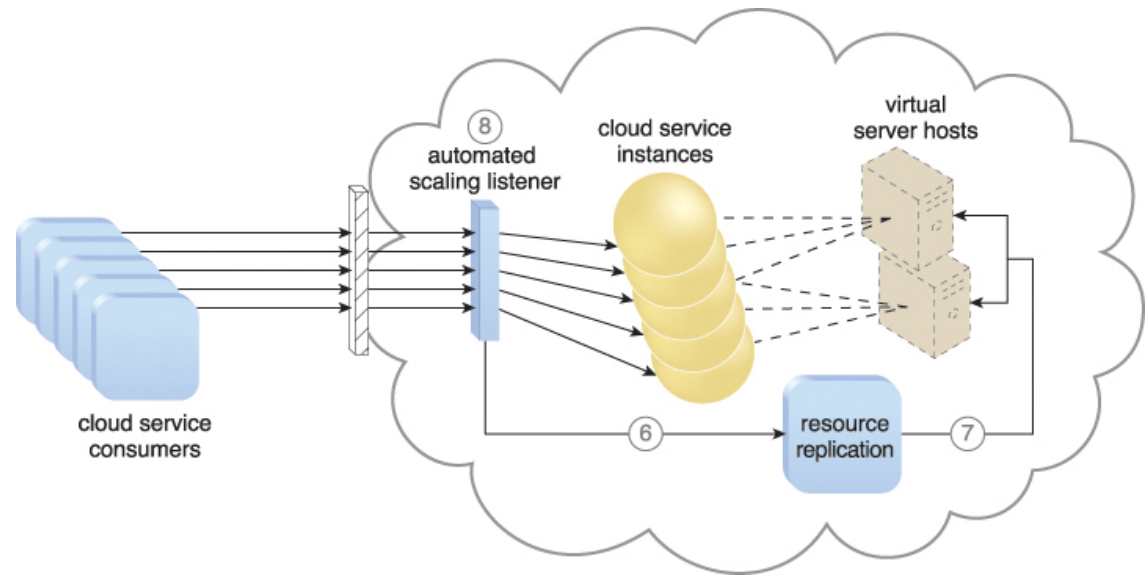


Automated Scaling is done
Contract has provisions to increase capacity



Additional requests.
Overloading system

automated scaling listener
notices cloud service instances
overloaded with requests

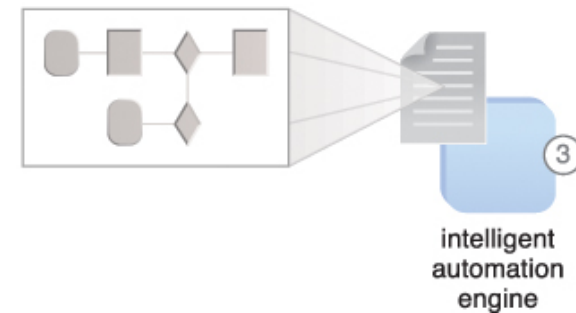
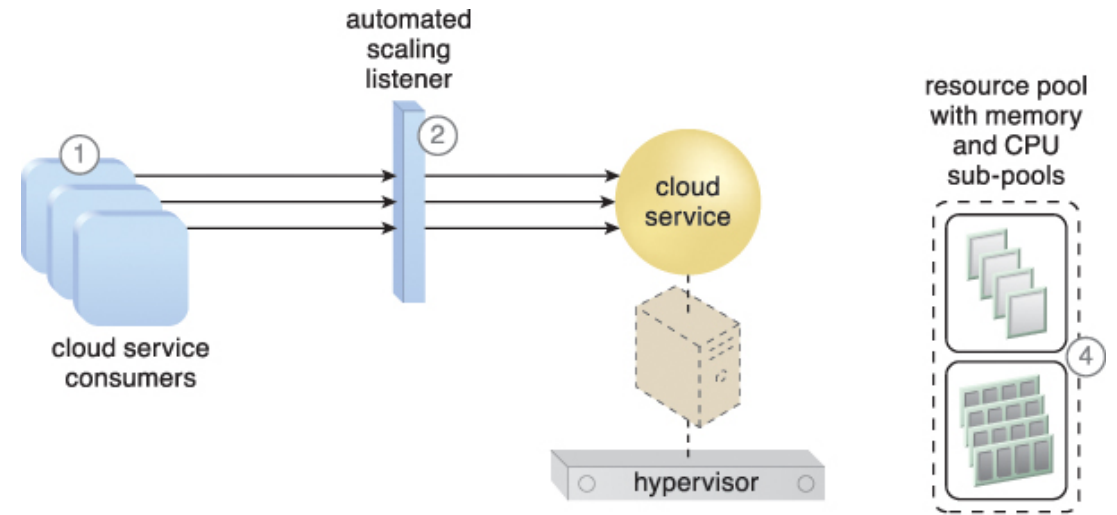


Dynamic Scalability Architecture

This content is protected and may not be shared, uploaded, or distributed

Fundamentals of Cloud Architecture

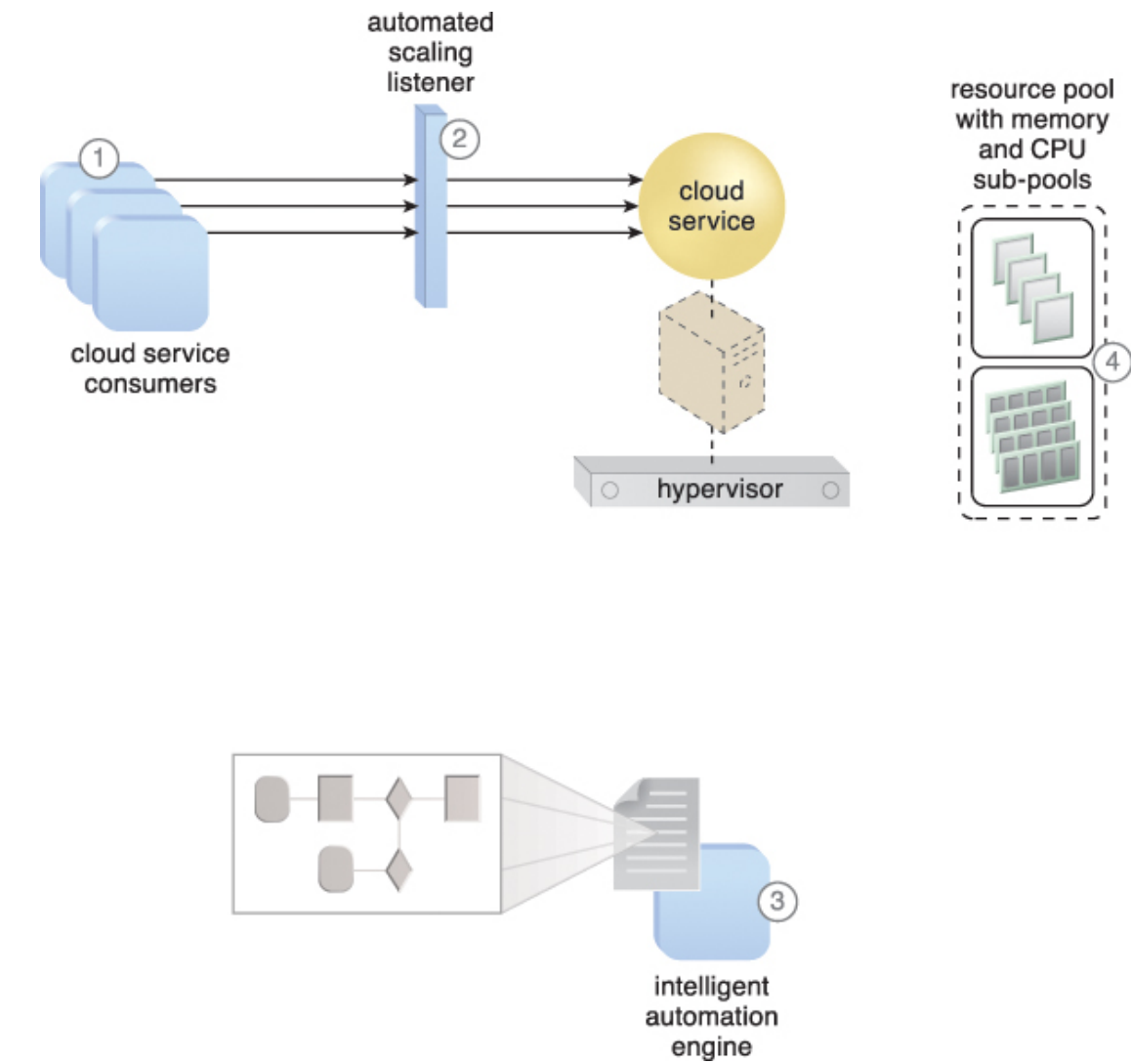
- ❑ An **(4) Elastic Resource Capacity Architecture** is primarily related to the dynamic provisioning of virtual servers, using a system that allocates and reclaims CPUs and RAM in immediate response to the fluctuating processing requirements of hosted IT resources



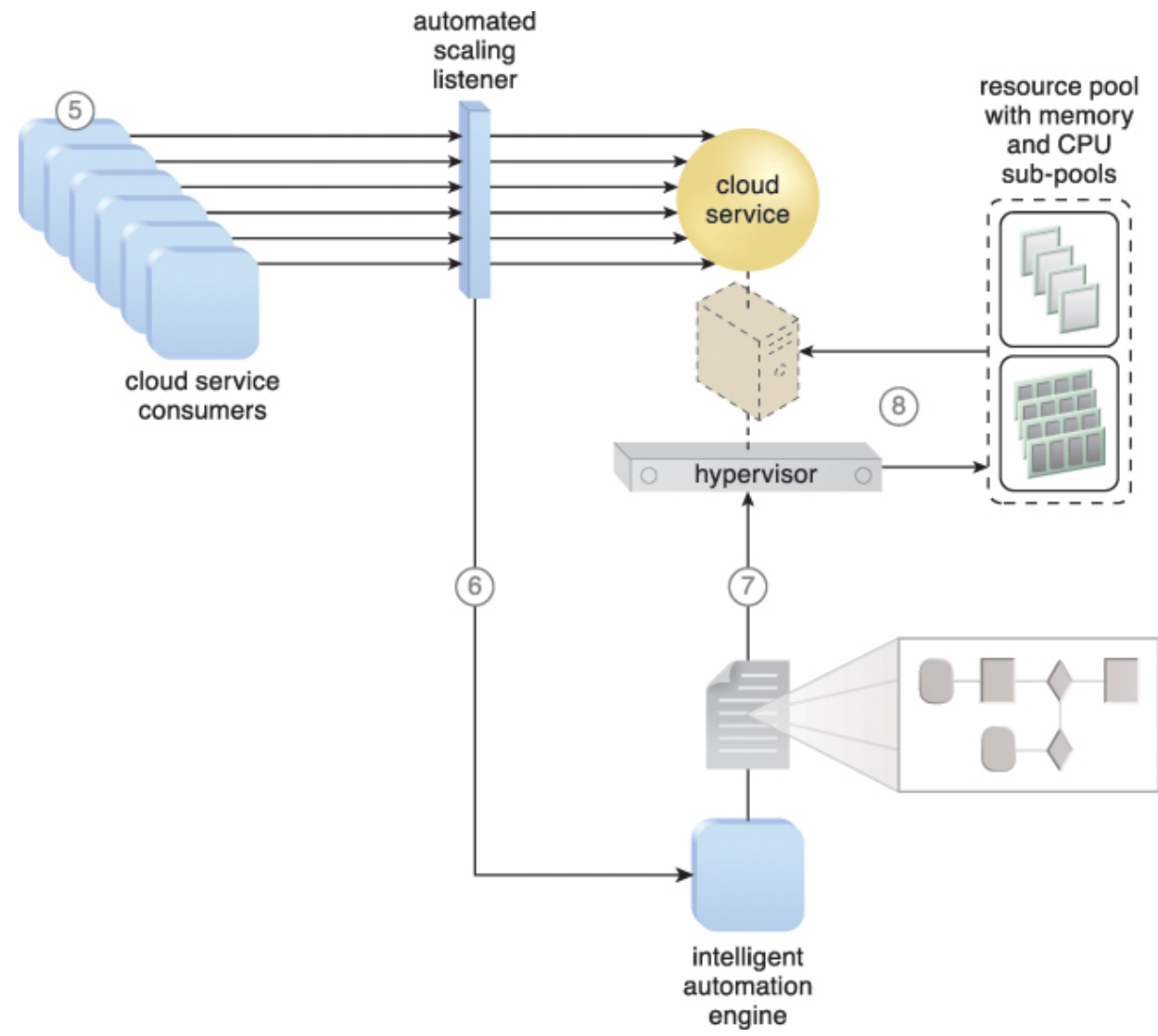
Elastic Resource Capacity Architecture

Citation: Erl, Thomas, Ricardo Puttini, Zaigham Mahmood. Cloud Computing.. [VitalSource Bookshelf].

Normal traffic



More requests

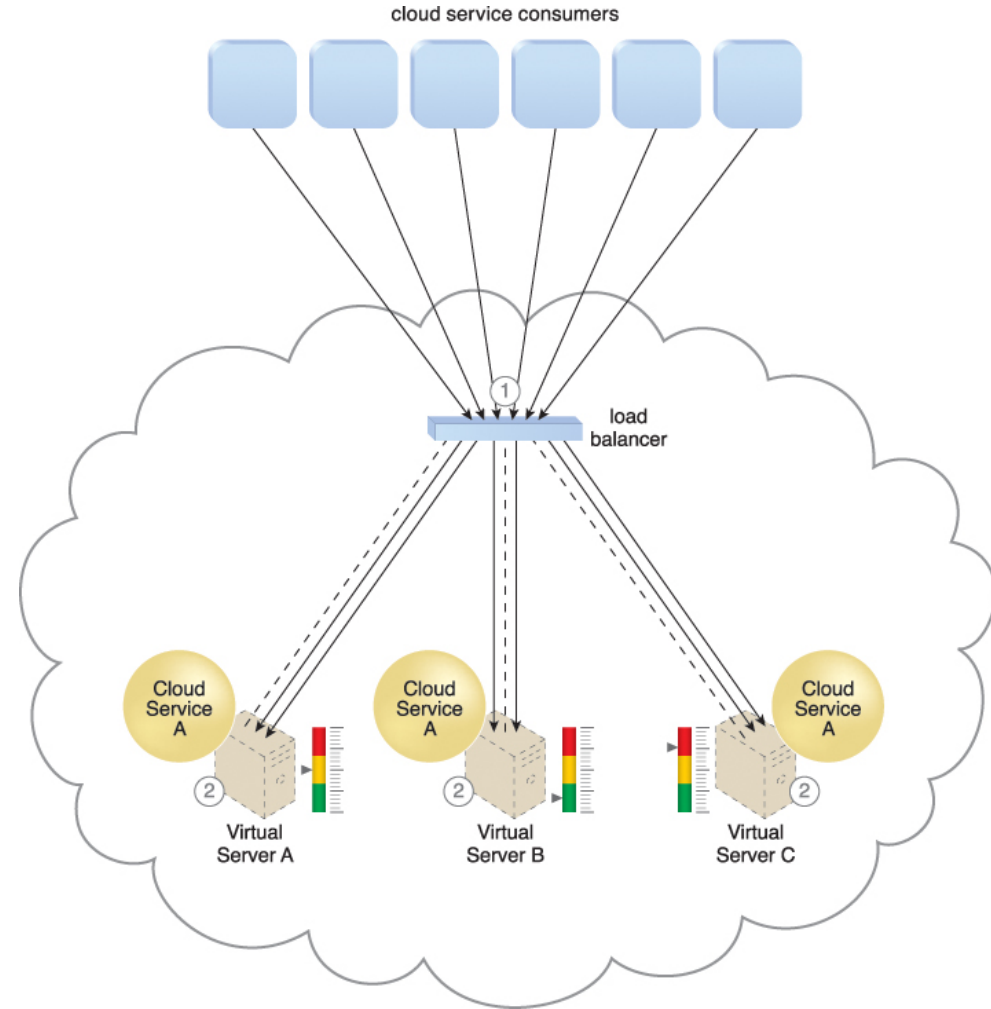


Elastic Resource Capacity Architecture

This content is protected and may not be shared, uploaded, or distributed

Fundamentals of Cloud Architecture

- ❑ An **(5) Service Load Balancing Architecture** can be considered a specialized variation of the workload distribution architecture that is geared specifically for scaling cloud service implementations. Redundant deployments of cloud services are created, with a load balancing system added to dynamically distribute workloads.

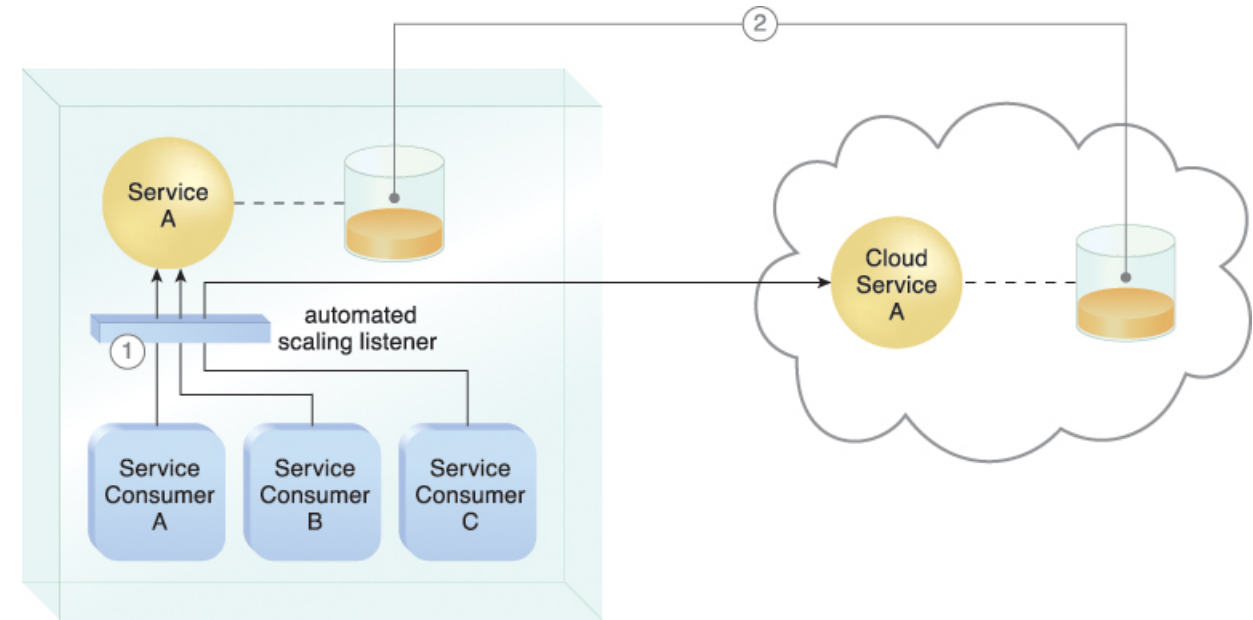


Service Load Balancing Architecture

Citation: Erl, Thomas, Ricardo Puttini, Zaigham Mahmood. Cloud Computing.. [VitalSource Bookshelf].

Fundamentals of Cloud Architecture

- ❑ A **(6) Cloud Bursting Architecture** establishes a form of dynamic scaling that scales or “bursts out” on-premise IT resources into a cloud whenever predefined capacity thresholds have been reached.
- ❑ IT resources remain inactive until cloud bursting occurs.
- ❑ If no longer required, the cloud-based IT resources are released and the architecture “bursts in” back to the on-premise environment.



Cloud Bursting Architecture

Citation: Erl, Thomas, Ricardo Puttini, Zaigham Mahmood. Cloud Computing.. [VitalSource Bookshelf].

Questions to Consider

- Do you need to study or consider cloud architecture for serverless application development?

