DALHOUSIE UNIVERSITY

Inspiring Minds

CLOUD ARCHITECTURE

CSCI 5408: Data Management, Warehousing, and Analytics

Prepared By: Suhaib Qaiser (suhaibqaiser@dal.ca)

Cloud Architecture Basics

Grid Computing
Distributed Computing
Parallel Computing
Cloud Computing

Web Platform hosted on Clouds

Application Servers
HTML5 and JavaScript Dynamics
CORS
Deploying application in Platform VS Infrastructure
Log Entries

Managed Services on the Cloud

Web Services
REST Services
Service Agents

Recap from last lecture …

Q1. Name 2 unique Cloud Service Providers for each:

a) SaaS
b) IaaS
c) PaaS

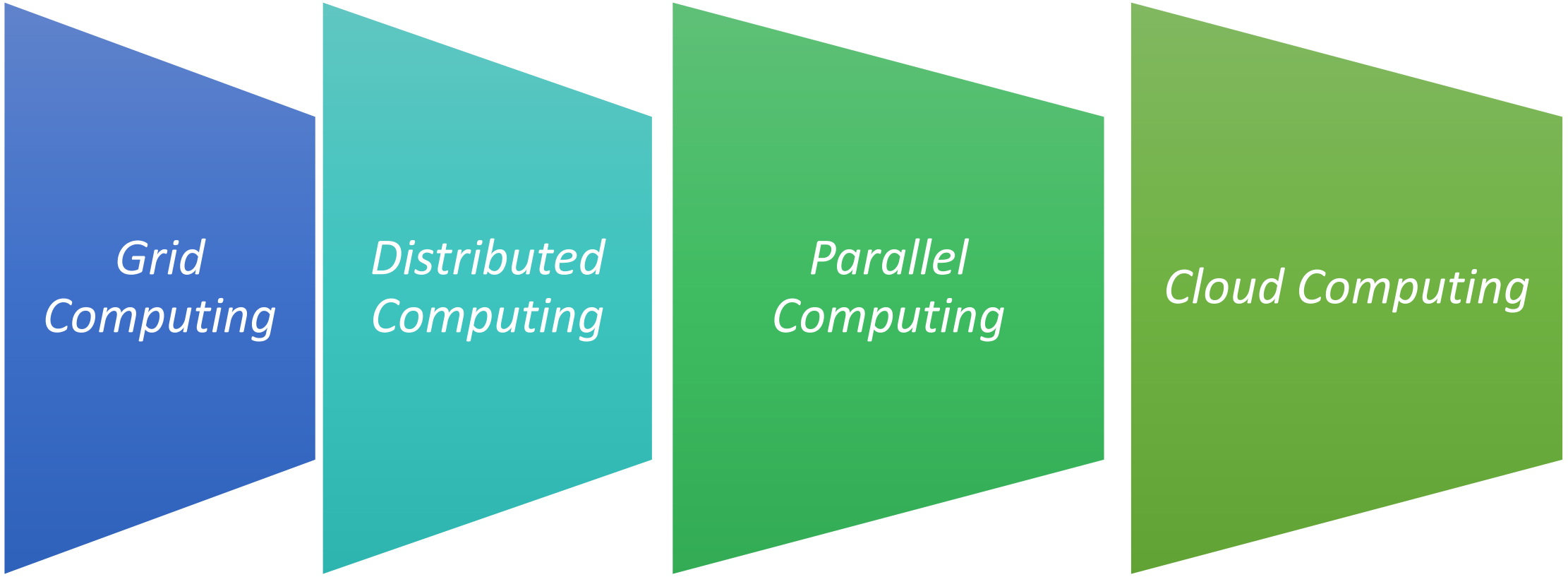
Q2. What are the limitations of multi-tenant cloud database system?

Q3. What is the purpose of Elastic Load Balancer?

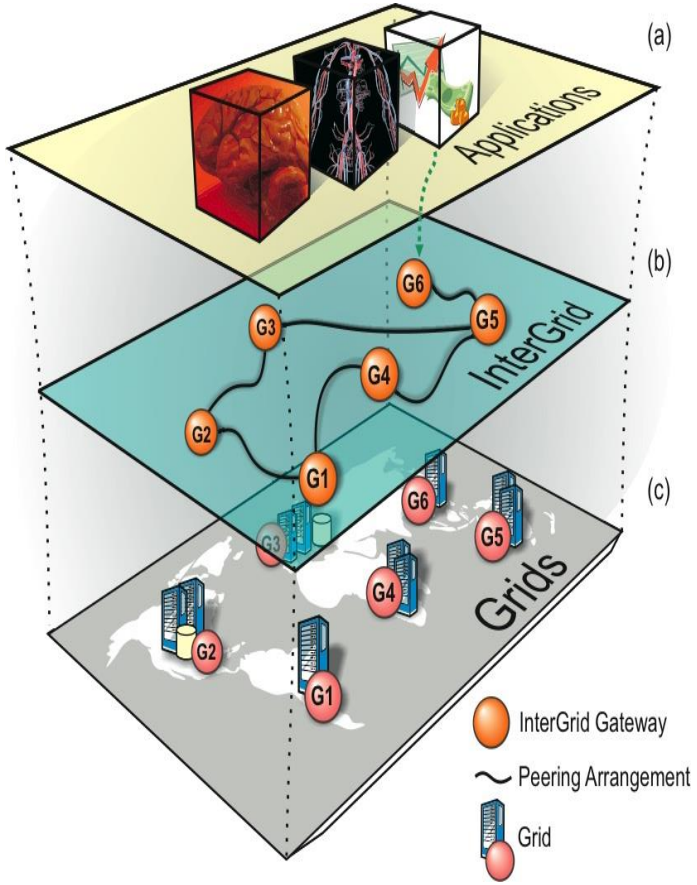
Q4. If I chose Platform over Infrastructure what unique feature to I get for free?

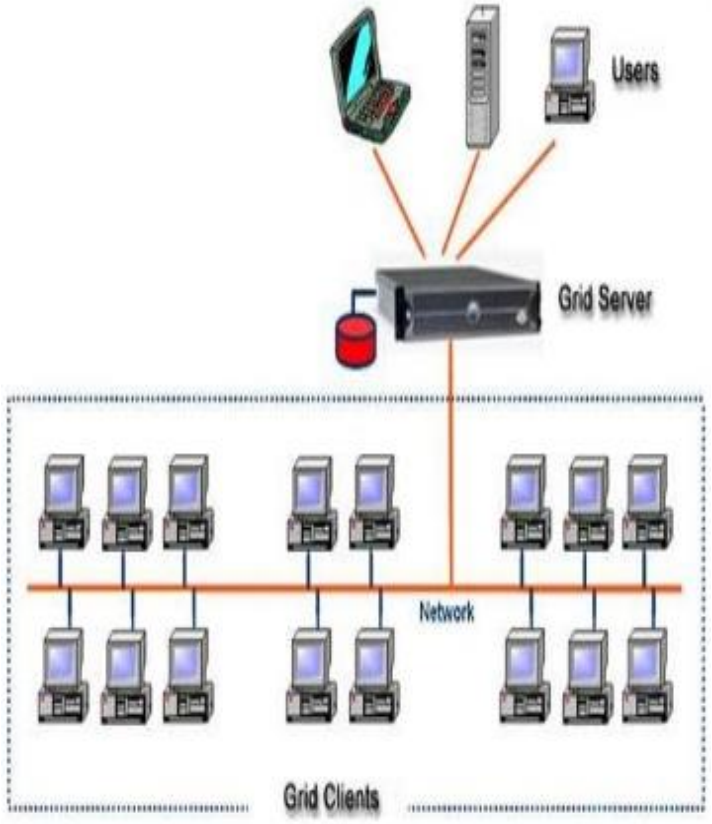
Q5. Why are public clouds better than private clouds?

Cloud Computing



Grid Computing





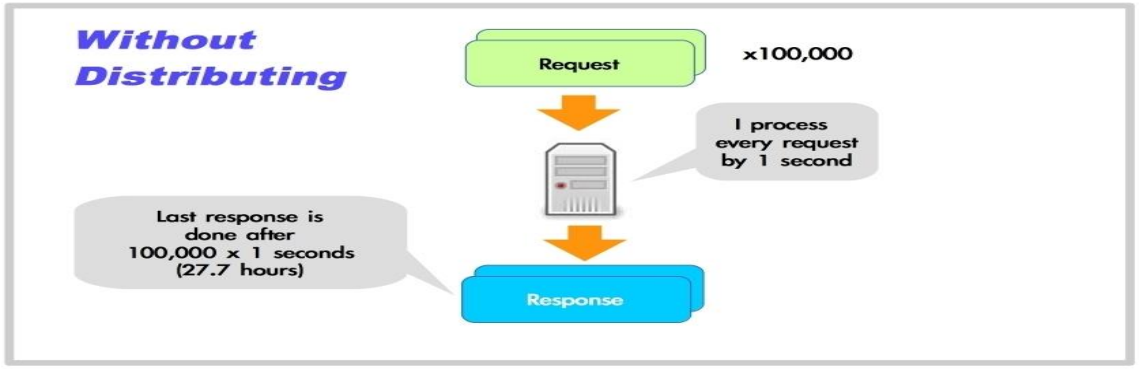
Grid computing is a computer network in which each computer's resources are shared with every other computer in the system. Processing power, memory and data storage are all community resources that authorized users can tap into and leverage for specific or grid computing can be seen as a special type of parallel computing that relies on complete computers (with onboard CPUs, storage, power supplies, network interfaces, etc.) connected to a computer network (private or public) by a conventional network interface, such as Ethernet.

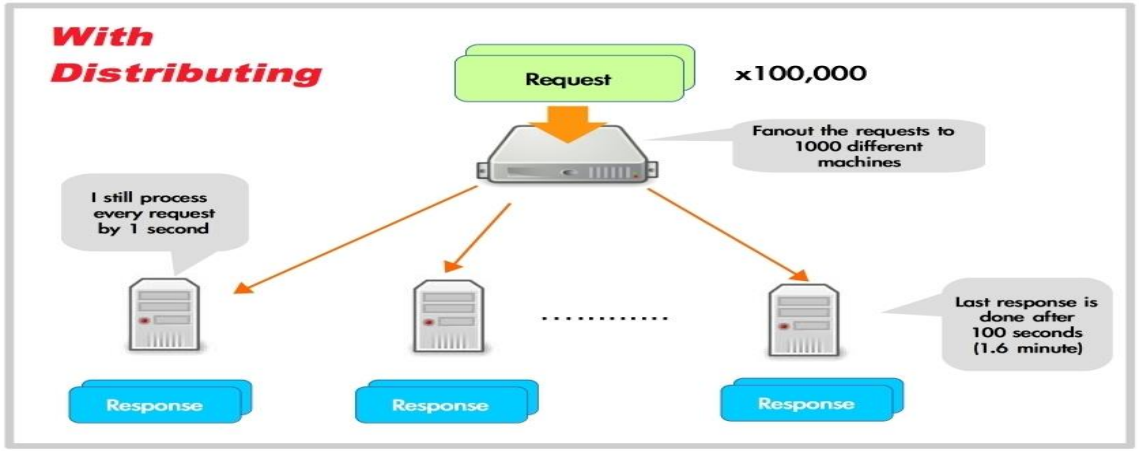
http://www.cloudbus.org/intergrid/

http://www.ijetae.com/files/Volume3Issue2/IJETAE_0213_102.pdf

Distributed Computing

Why need Distributed computing ?

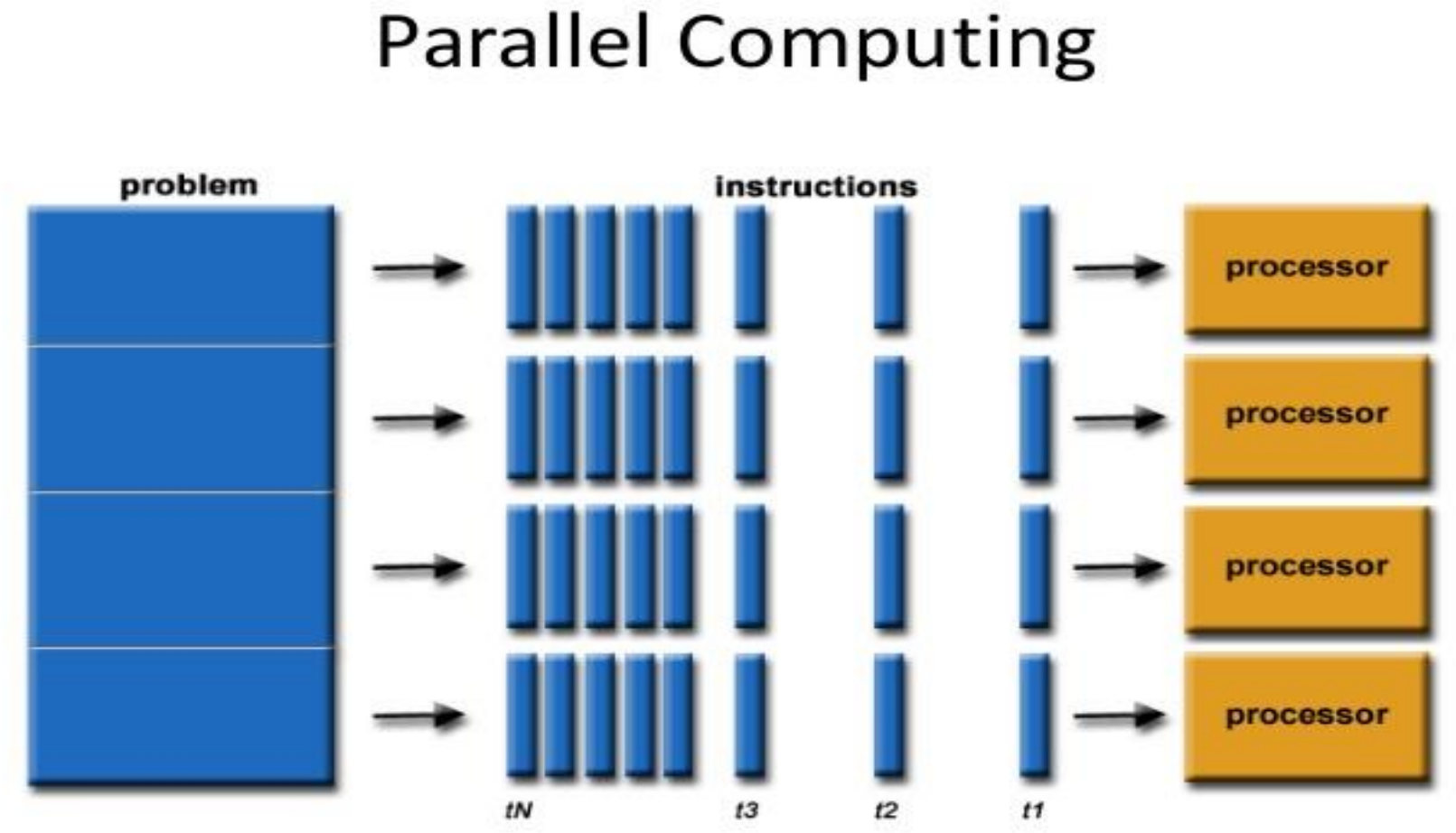




Distributed Computing is a field of computer science that studies distributed systems. A distributed system is a model in which components located on networked computers communicate and coordinate their actions by passing messages. The components interact with each other in order to achieve a common goal.

https://coderwall.com/p/klzs5a/hello-distributed-computing

Parallel Computing

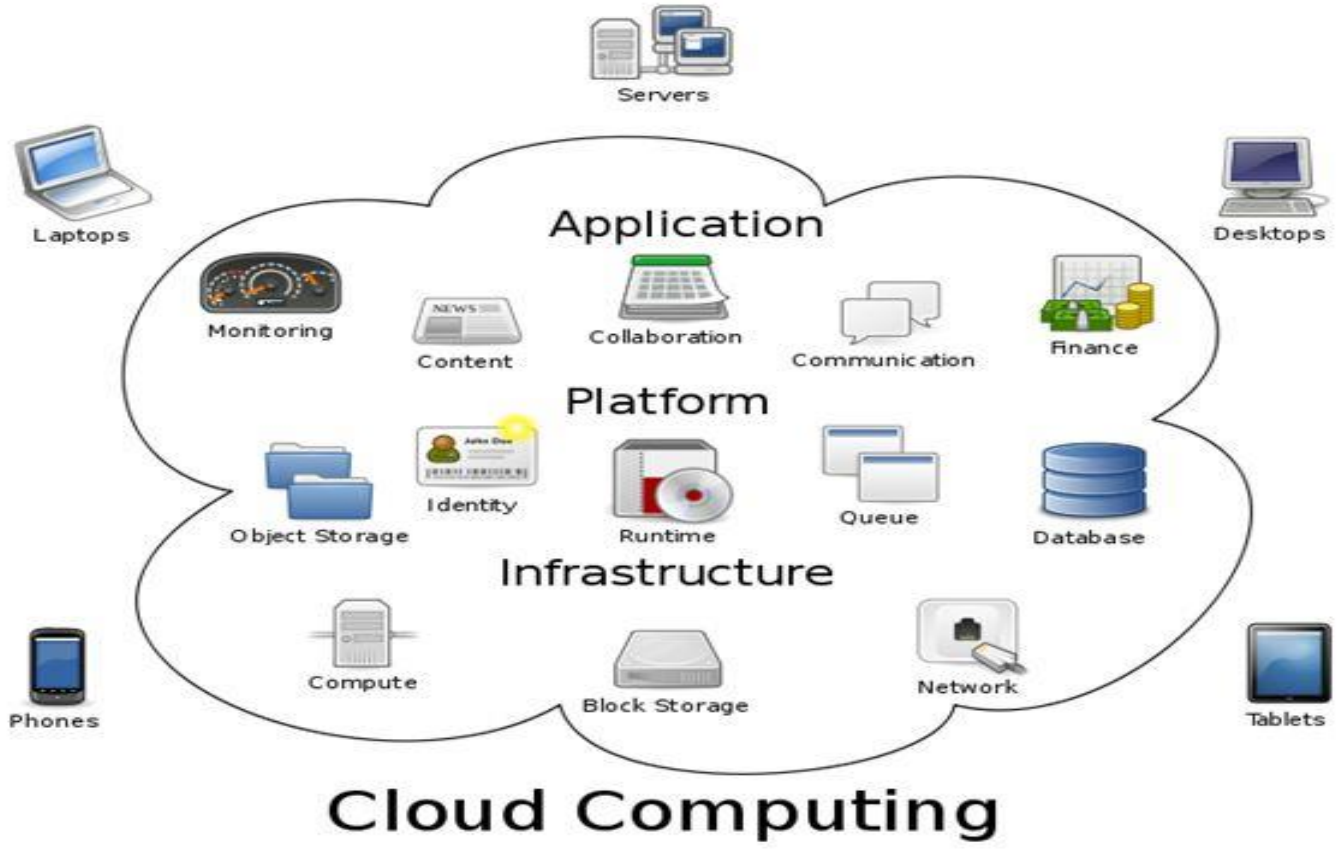


Parallel Computing is a type of computation in which many calculations or the execution of processes are carried out simultaneously.

Large problems can often be divided into smaller ones, which can then be solved at the same time.

Source: http://www.slideshare.net/EngrZardariSaddam/parallel-computing-47793374

Cloud Computing

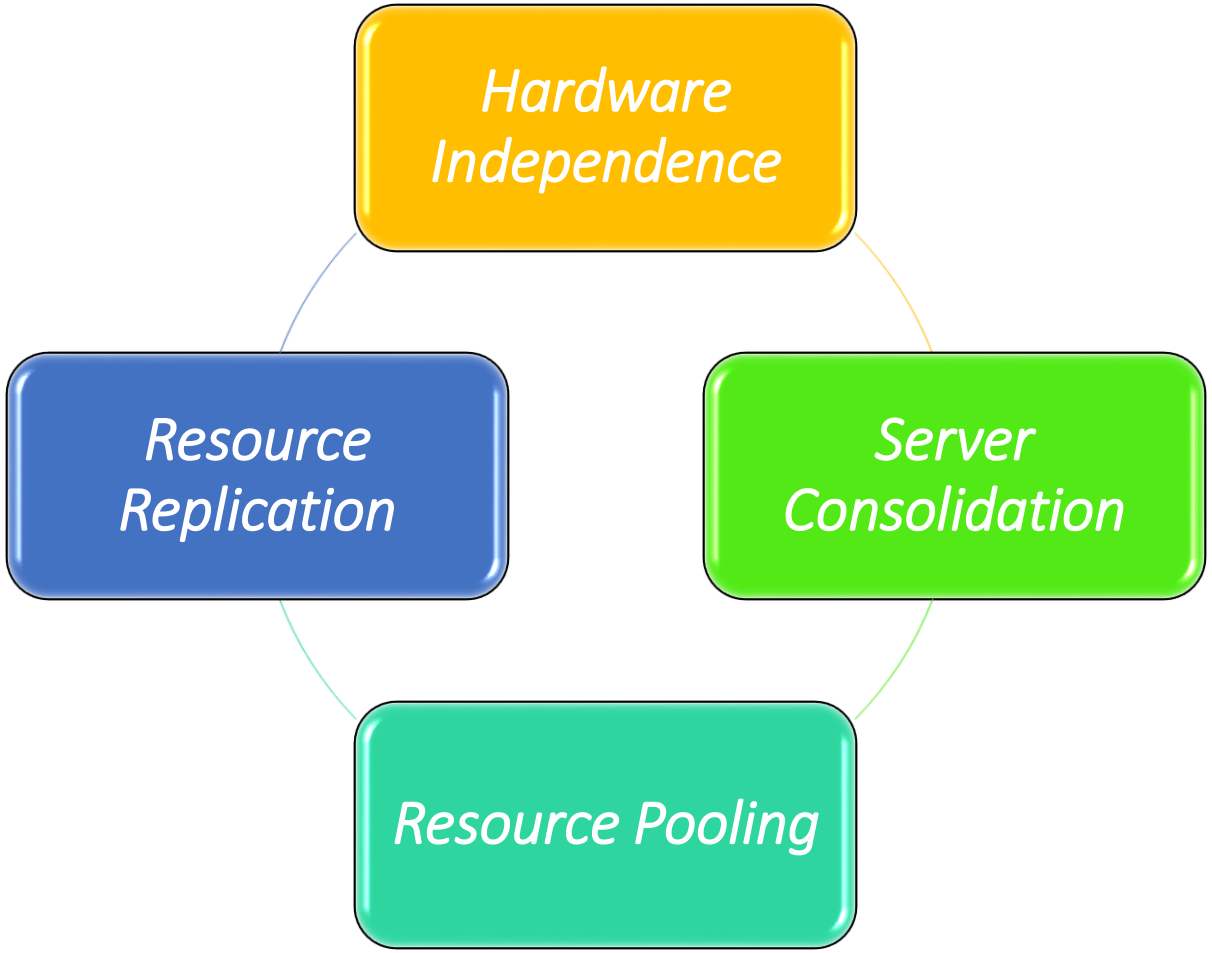


Cloud Computing is a type of Internet-based computing that provides shared computer processing resources and data to computers and other devices on demand.

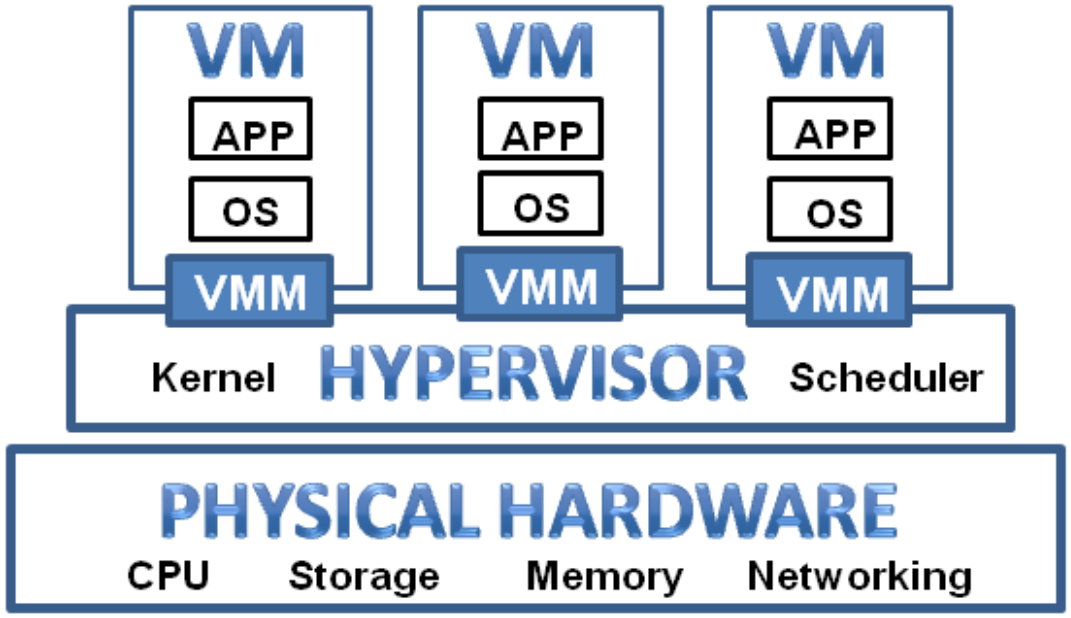
It is a model for enabling ubiquitous, on-demand access to a shared pool of configurable computing resources (e.g., computer networks, servers, storage, applications and services)

http://www.brighthub.com/environment/green-computing/articles/127086.aspx

Virtualization



Hardware Independence



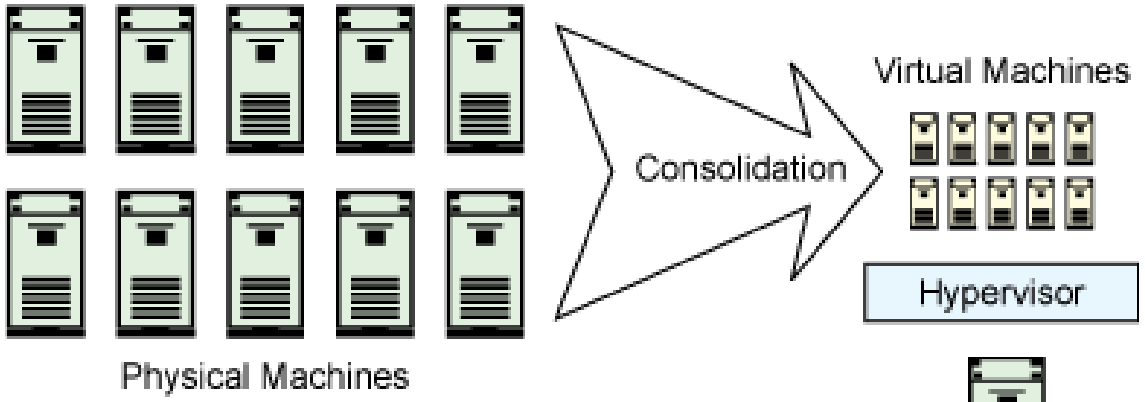
The availability of server operating system and ability to switch it between different servers without the need of hardware upgrade is called hardware independence

Virtualization is a conversion process that translates unique IT hardware into emulated and standardized software-based copies

Cloning and manipulating of virtual IT resources is much easier than duplicating physical hardware

Source: http://www.cse.wustl.edu/~jain/cse567-11/ftp/hypervsr/fig1.png

Server Consolidation



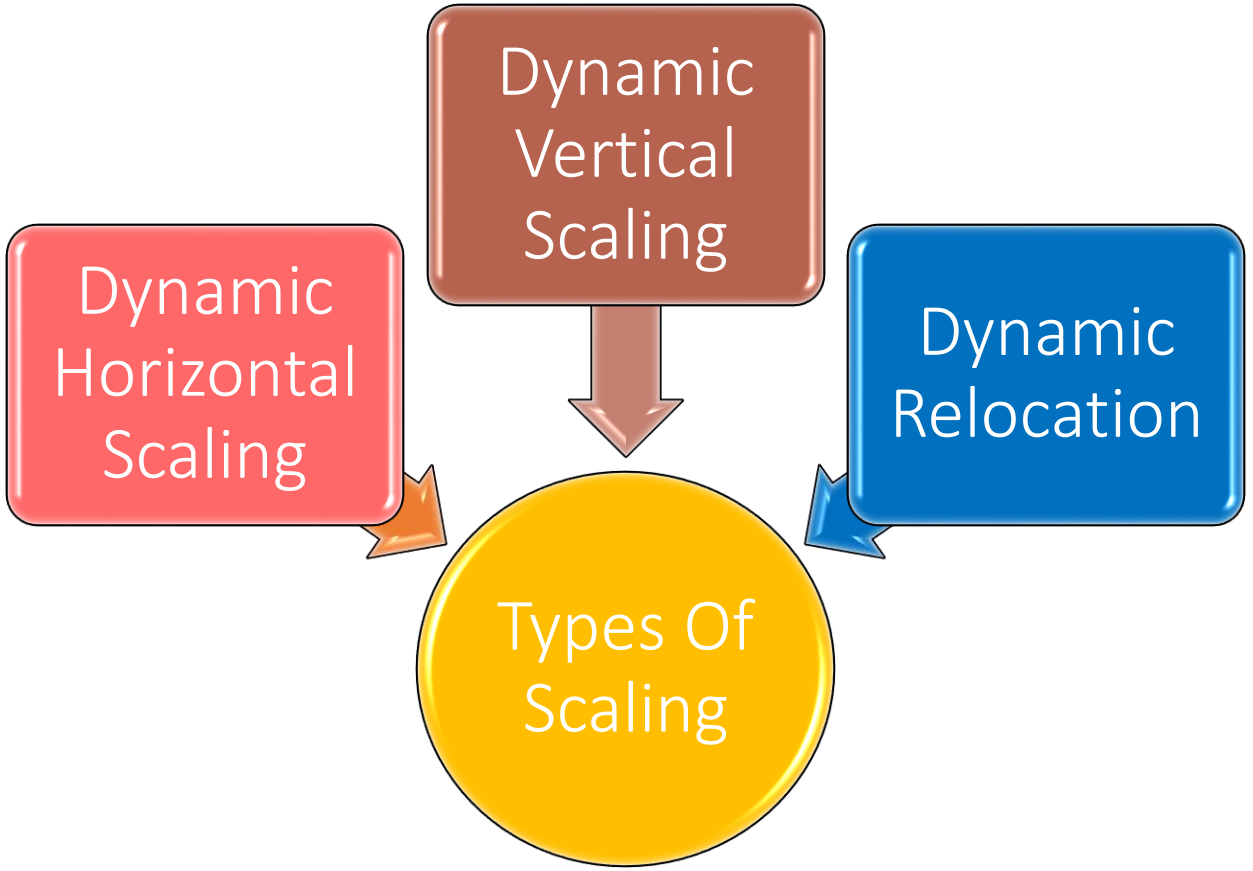
The coordination function provided by the virtualization software allows multiple virtual servers to be created by the same host

This fundamental capability directly supports common cloud features such as on-demand usage, resource pooling, elasticity, scalability and resiliency

Provides “Dynamic Scalability Architecture”

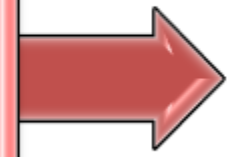
Source: https://www.ibm.com/developerworks/library/l-kernel-shared-memory/figure1.gif

Types of Scaling



Types of Scaling

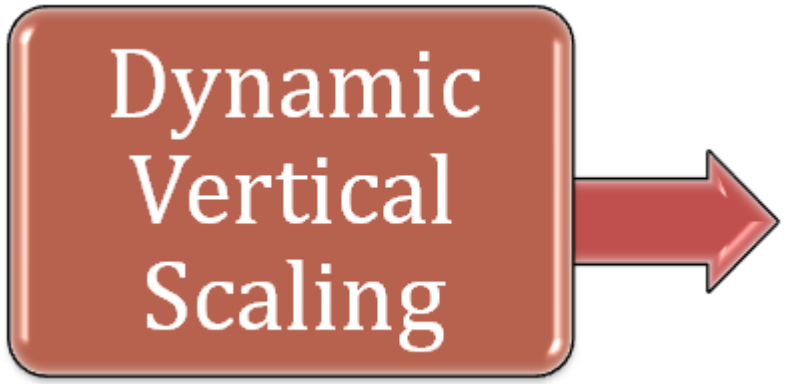
Dynamic Horizontal Scaling



- **IT resources are scaled out and in to handle fluctuating workloads. Horizontal scaling always duplicate resources by using replication as a mode. Requests would be redirected to new server nodes to handle and service response**

A common example would be increasing a virtual machine to handle load of incoming requests

Types of Scaling



IT resource instances are scaled up and down when there is a need to adjust the processing capacity of single IT resource

An example would be addition of processing core or memory on a single IT instance or a virtual machine.

A more common example is adding nodes to a platform service cloud on runtime when demand increases

Types of Scaling

Dynamic Relocation

This type of scaling moves the entire architecture to a new host with better specifications. Such as moving database to another with more capacity. Or shifting networks that have more bandwidth or response times

Sometime infrastructure can be shifted to a bigger and stronger infrastructure to handle load in a better and efficient manner

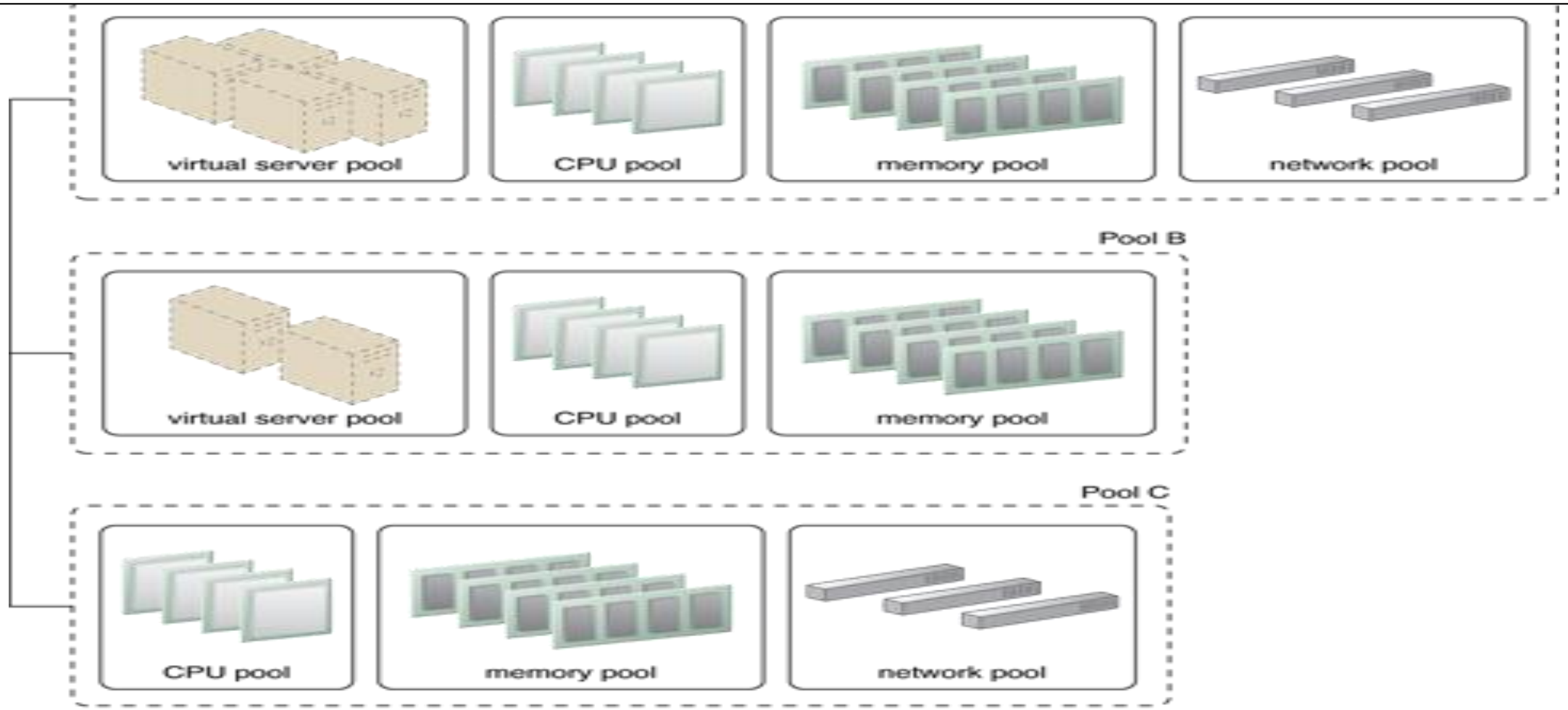
Resource Pooling

A resource pooling architecture is based on the use of one or more resource pools, in which identical IT resources are grouped together and maintained by a system to automatically ensure that they are synchronized.

Types of Pools:

- Physical Server Pools – network servers with operating systems for immediate use
- Virtual Server Pools – servers provisioned using templates and chosen by cloud consumer
- Storage Pools – storage devices file based or block based available for use
- Network Pools – different preconfigured network connectivity devices such as load balancers, VPN, firewalls, etc
- CPU Pools – processing units available for virtual servers broken into individual cores
- Memory Pools – Pools of physical RAM used to vertically scale virtual servers

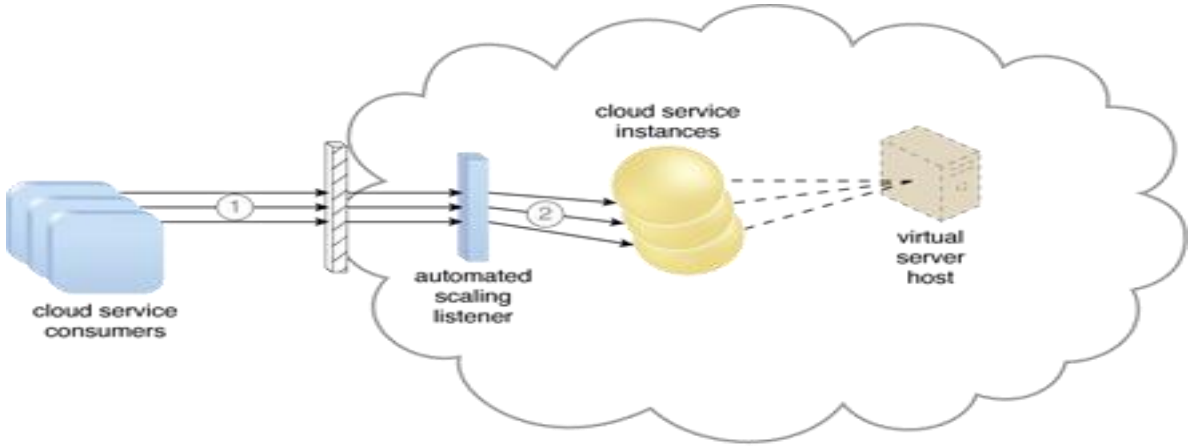
Resource Pooling

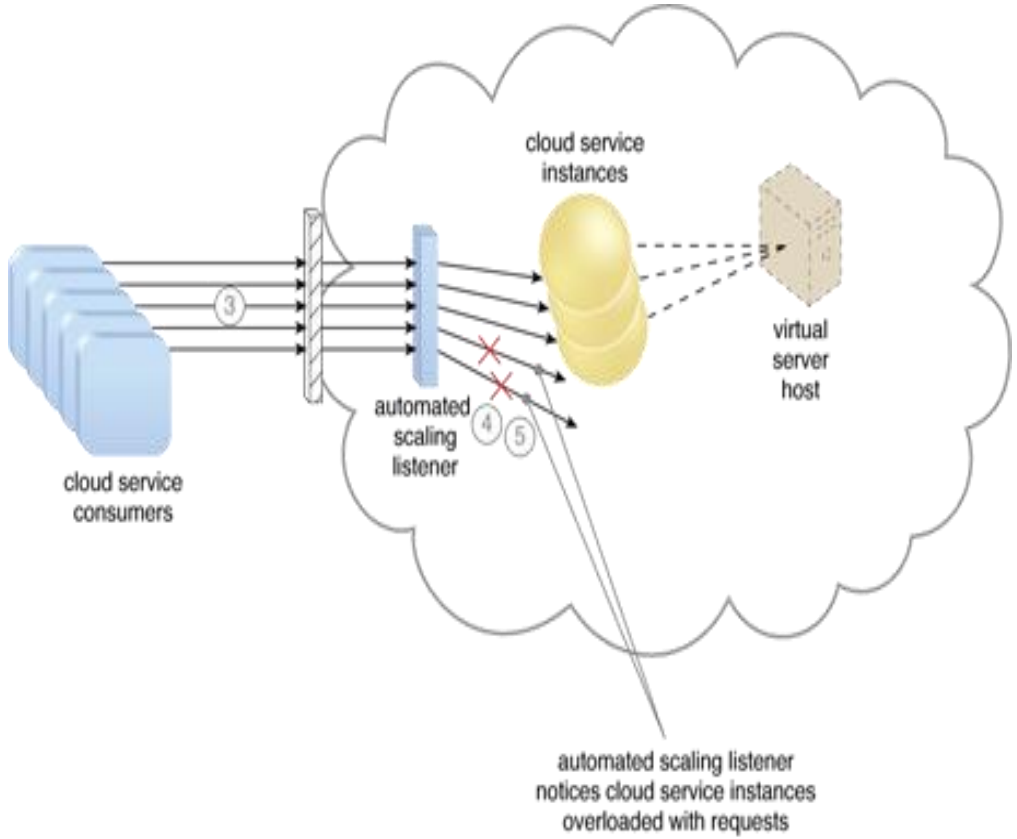


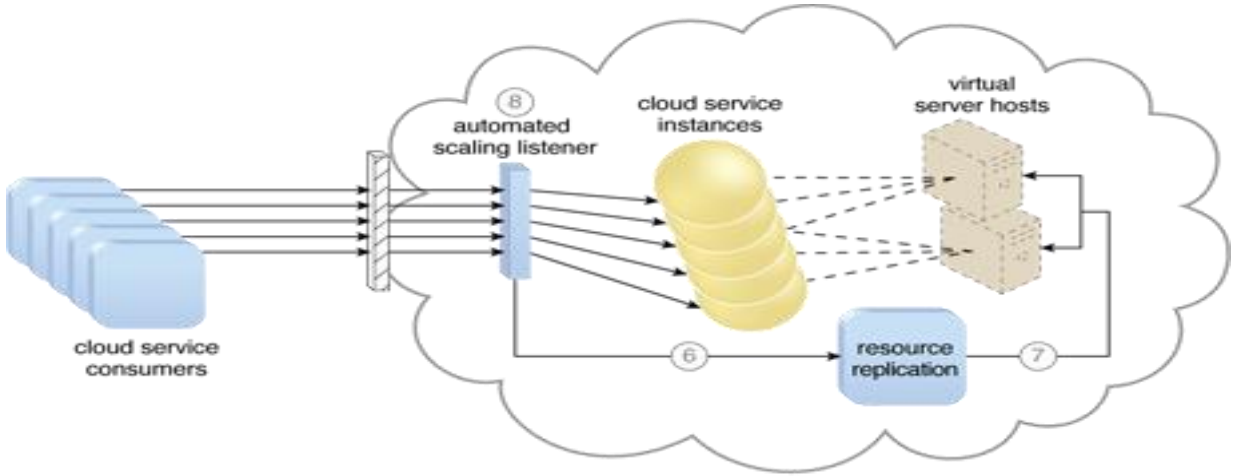
Resource Replication

- Virtual servers are created as virtual disk images that contain binary file copies of hard disk content
- These copies are available to replicate virtual servers and migrate to other sources if required
- Resource replication provides following benefits:
  - Physical Server Pools – network servers with operating systems for immediate use
  - It is easy for instantaneous deployment of virtual resources on-demand
  - Servers can be backed up instantaneously since all the files are stored (memory snapshots, operating system files, etc) as disk images
  - Easy to scale up and down
  - Provides more fail over capability and increase reliability and efficiency of IT resources

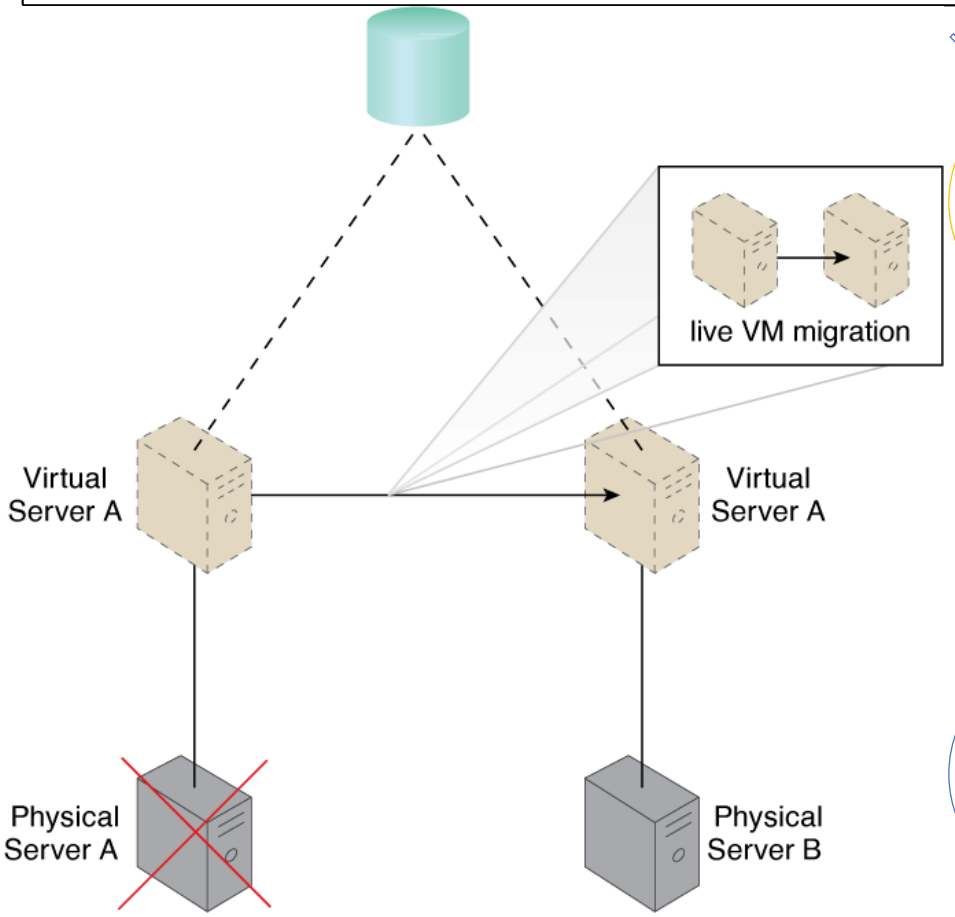
Resource Replication







Zero Downtime Architecture

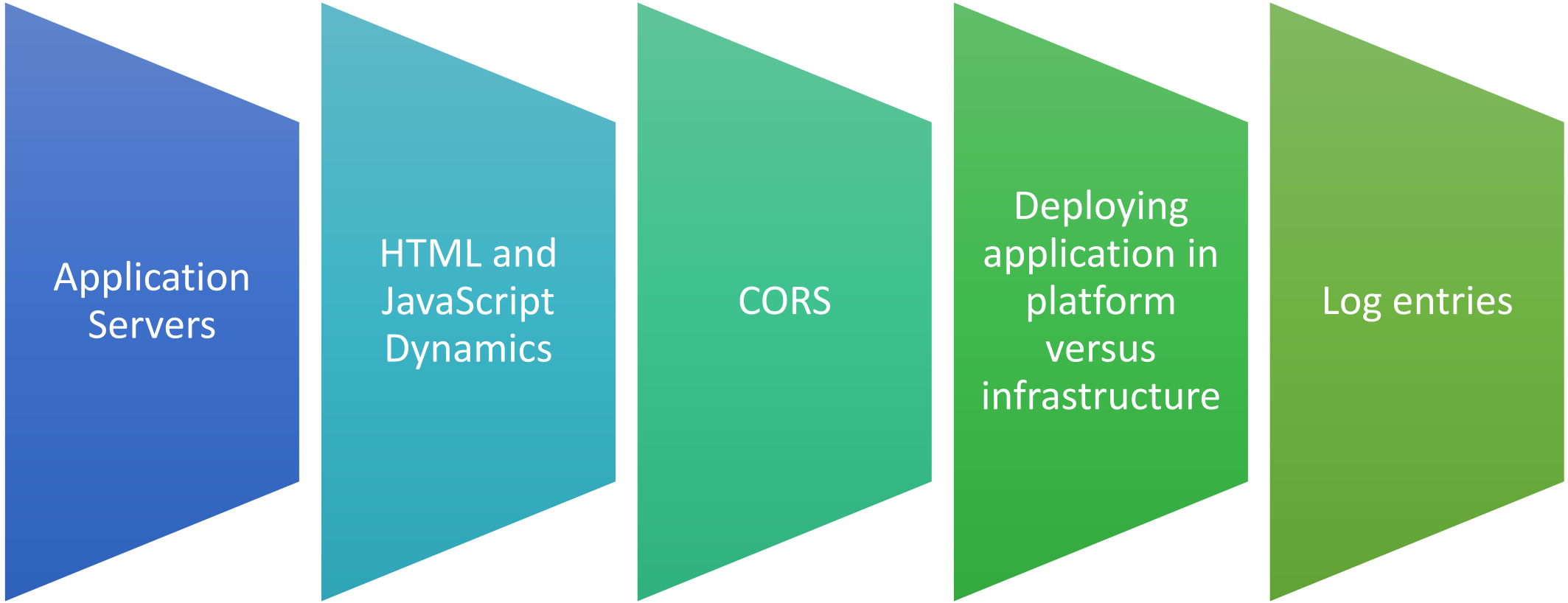


The zero downtime architecture establishes a sophisticated failover system that allows virtual servers to be dynamically moved to different physical server hosts in an event of failure

Multiple physical servers are assembled into a group that is controlled by a fault tolerant system capable of switching activity from one physical server to another

The VM migration is typically a core part of high availability in cloud architecture

Web Platforms Hosted on Cloud



Application Servers

Application servers are responsible for running web applications on the server

Based on the platform and language there are different options for using application servers

Web server clients sends requests to application server which application server responds with appropriate content

Web resources can be either generated through static content or dynamic content

Most cloud architectures as well as application servers makes a clear distribution between static and dynamic content when serving to request

Application Servers

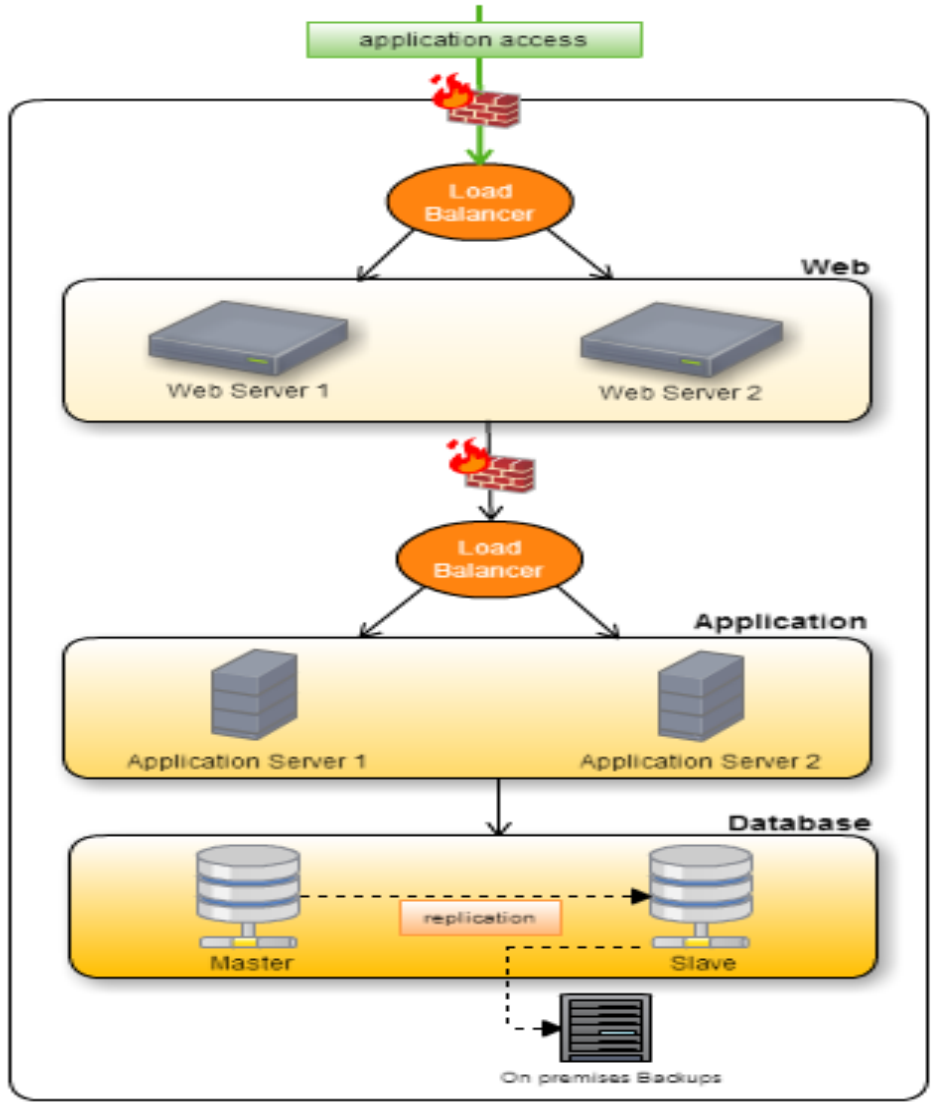
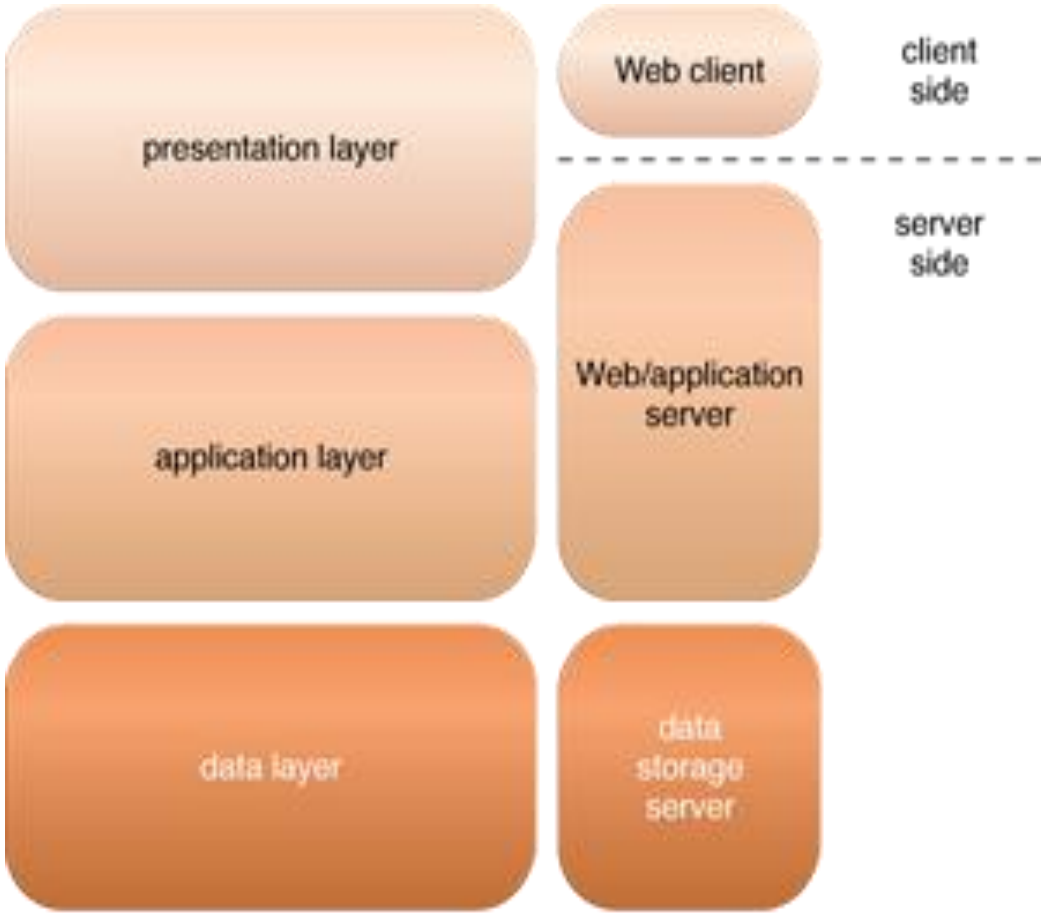


Figure 1: Traditional Software Architecture



Source: https://www.captechconsulting.com/blogs/library/145DFC74A28240298DCCE5889D141027.ashx

Use of Putty

Putty is an open source Telnet / SSH client used to open a terminal with virtual host

Normally it requires a host url and a port that can accept SSH connection

Putty is used with Windows base machines. Mac OS already has built in SSH client that can be used

Web resources can be either generated through static content or dynamic content

Most Virtual machines are protected by either a password or a key file. Putty KeyGen can be used to generate a private key file

HTML and JavaScript Dynamics

Web Client plays an important role in Cloud Computing

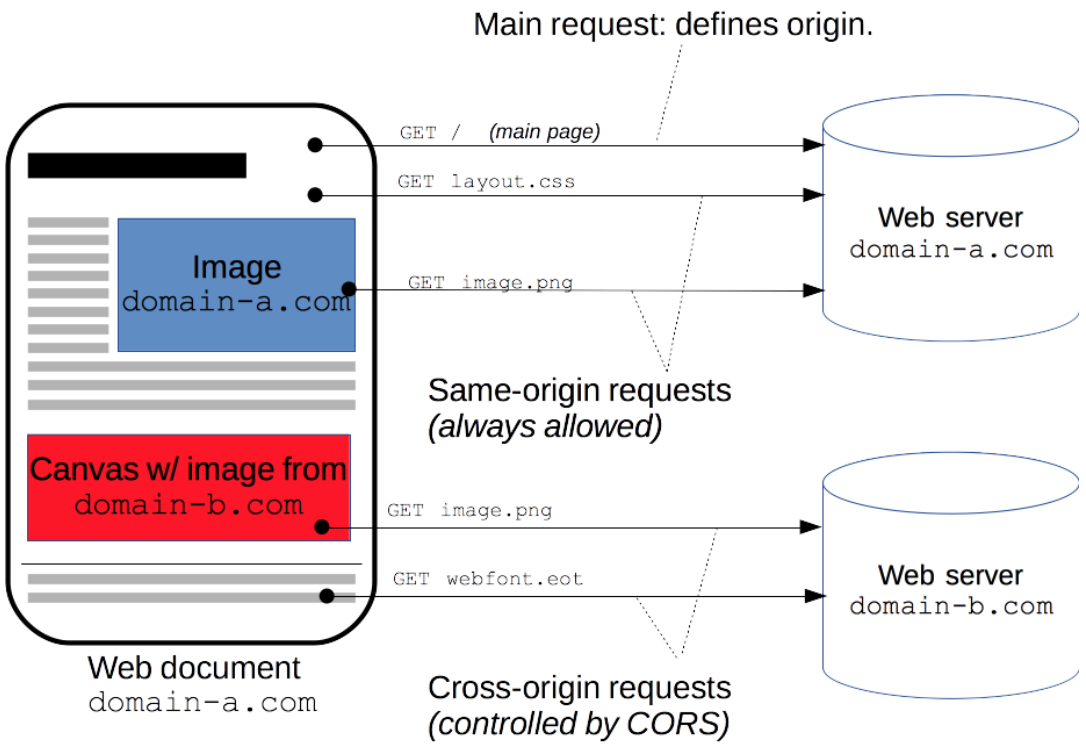
Industry is moving towards thick client – thin server concept where JavaScript plays an important role on the web clients

HTML5 have introduced new UI components that need very less interaction from server

Web Application Servers provides data and business logic whereas all the view related functionality is performed by the browser through JavaScript

Many JavaScript frameworks have become popular like AngularJS, Node.JS, Knockout.JS, React.JS, etc

CORS – Cross Origin Resource Sharing



A resource makes a cross-origin HTTP request when it requests a resource from a different domain than the one which the first resource itself serves

The Cross-Origin Resource Sharing standard works by adding new HTTP headers that allow servers to describe the set of origins that are permitted to read that information using a web browser

Simple Request: If the server responds with a Access-Control-Allow-Origin: *
This means that the resource can be accessed by any domain in a cross-site manner

Preflighted Requests: first send an HTTP request by the OPTIONS method to the resource on the other domain, in order to determine whether the actual request is safe to send

Source: https://developer.mozilla.org/en-US/docs/Web/HTTP/Access_control_CORS

Deploying application in platform VS Infrastructure

Application Servers can be maintained either on using Infrastructure services or Platform services but there is a key difference

Services deployed on infrastructure can only have high level of monitoring on the Cloud. Measured usage can only be performed on overall infrastructure and not individual web service. You can only scale horizontally

Deploying services on a platform gives ability to consumers to scale resources as per need horizontally and also vertically

Consumers can also have low level metrics and measured usage stats of each individual service

Platform Service are often more expensive than infrastructure but they are usage dependant. Prices are determined based on usage from users

Platform services can scale vertically as per need

Log Entries

Logging is the most critical and important feature in a Cloud system

Logs provide effective means of troubleshooting the problem in a Cloud system

Logging is also available as a cloud service or an add-on in many cloud offerings

Error reporting and fail over benchmarks can be determined by effective logging of the system

Administrative operations can be performed through logging critical information such as errors and alerts from the system

Managed Services on the Cloud

Managed Services on the Cloud

The field of service technology is a key stone foundation of cloud computing and also formed the basis of “as-a-service”.

Services are of different types but more generally SOAP and REST service are common in use

Web Services

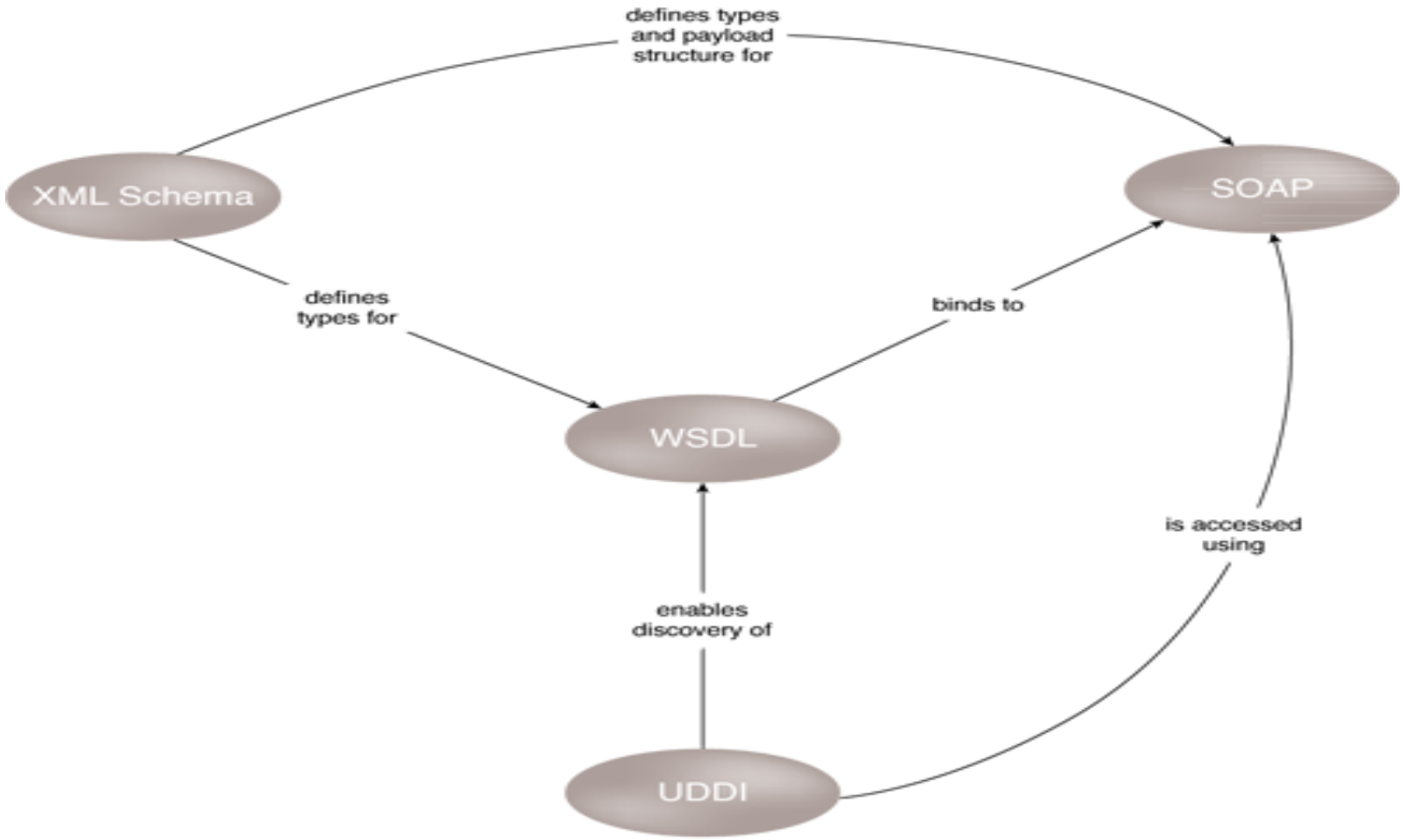
Also commonly known as SOAP services are served with XML based request and response

Web Service Description Language defines operations and request and response attributes

SOAP – Simple Object Access Protocol is a wrapper around HTTP

UDDI – Universal Description, Discovery and Integration regulates the services and identifies access points

SOAP Service Architecture

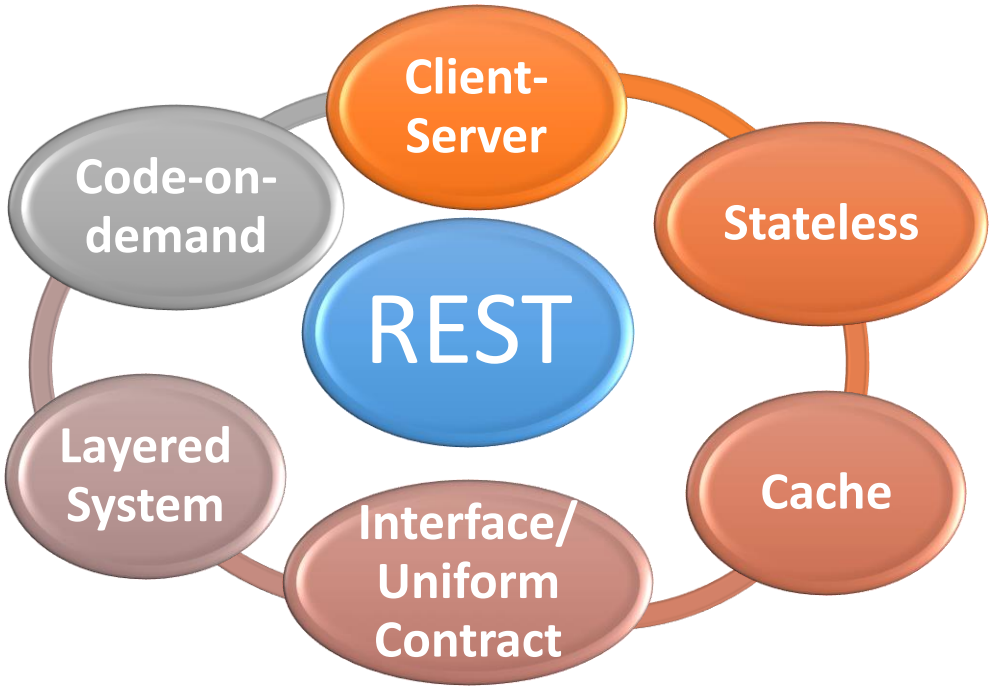


REST Services

REST services are easy to use service that runs on top of HTTP

REST – Representational State Transfer is designed for web clients to load data from server in a representational state

The six defining factors for REST are:



Services Agents

Service agents are event driven programs designed to intercept messages at runtime

Active service agents performs an action upon intercepting and reading the contents of the message

Cloud environment rely heavily on the use of system-level and custom service agents to perform monitoring and measured usage in a cloud service

Service agents can be used for monitoring, logging, reporting, pay-for use billing and other requirements

Consumers can also design a service agent on top of services they use on the clouds

Scenario: Intelligent Weather Forecast API

Problem: Samsung Electronics wants to build a unique calendar app for their customers which can predict weather and advice on user routine change in real time. Samsung is forecasting that millions of users would be using this service and requests can vary based on locality and time periods. Samsung is not worried about a high budget on this service but wants service to be secure. Samsung wants to utilize public cloud domain. Following are the requirements:

a) Samsung needs a high performing cloud system that is capable of returning real time weather forecast

b) Samsung needs a maximum capacity data store where it can store analytical information, weather forecast information, patterns and parameters related to prediction

c) Samsung needs failover and data replication

Please provide a complete design of back end services, choices of cloud providers, some comments on costing, security mechanics, infrastructure needs, database needs, development team licenses, etc.

Quiz

Q.1 What is the difference between Server Consolidation and Hardware Independence?

Q.2 What is the difference between Grid Computing and Parallel Computing?

Q.3 What is the difference between resource pooling and resource replication?

Q.4 Why is JavaScript more important in Cloud computing than any other language?

Q.5 If a company server crashed or we get complains from service center, as a production support what should I do first?

Q.6 Are REST services more efficient then SOAP services? If yes than WHY

Reading Material

Chapter 5: Cloud Computing: Concepts, Technology & Architecture
by Zaigham Mahmood, Thomas Erl, Ricardo Puttini

URL: https://www.safaribooksonline.com/library/view/cloud-computing-concepts/9780133387568/ch05.html

OPTIONAL READING:
Cloud Computing Tutorial

URL: https://www.tutorialspoint.com/cloud_computing/cloud_computing_tutorial.pdf


ANY
QUESTIONS
?