Introduction

What is cloud computing?

1. On-Demand availability of computing system resources, especially like storage and computing power.
2. Option to use some else's computer via the internet.

Use a thin computer (Laptop) to connect to a remote computer which is powerful. (Client - servers)
But any failure on the powerful computer will bring down the client computers.
Hence a pool of computers is created and hosted and shared on an on-demand basis. The shared resource will be virtualized and delivered. Most of the maintenance of hardware is done by the cloud provider.

Impacts:

AWS -> 400Billion
IBM - > 100Billion
AWS is less than 20yrs old while IBM is more than 50yrs old.
The market believes cloud computing has to offer more value and hence the growth is exponential and future growth is expected to be high.

Azure added much value to Microsoft.
Successful apps today like Whatsapp, Salasforce, Snowflake etc were harnessing the power of cloud in their startup phase to grow exponentially. So the impact is by far most valued to business growth as well faster delivery.

Why such Impact?:

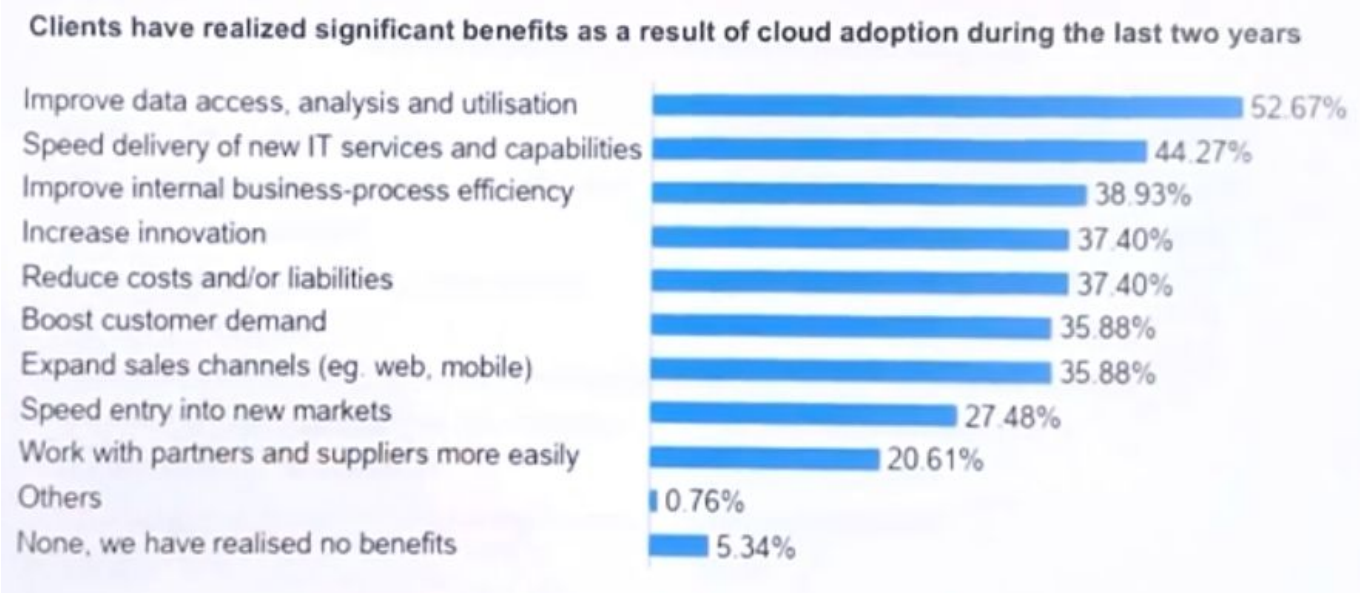
Division of labour! (Specialization) - Gives better efficiency since the expert in one field sticks to it and invests time improving it. The IT infra handling part separated from the core businesses allowing them to focus on the business value rather than the infra setup and maintenance.

Examples:Uber - Lets people hire - doesn't own a taxi
Facebook - Doesn't create content.
But both these examples provide the service of making the provider and the customer meet.

Business case for Cloud Computing

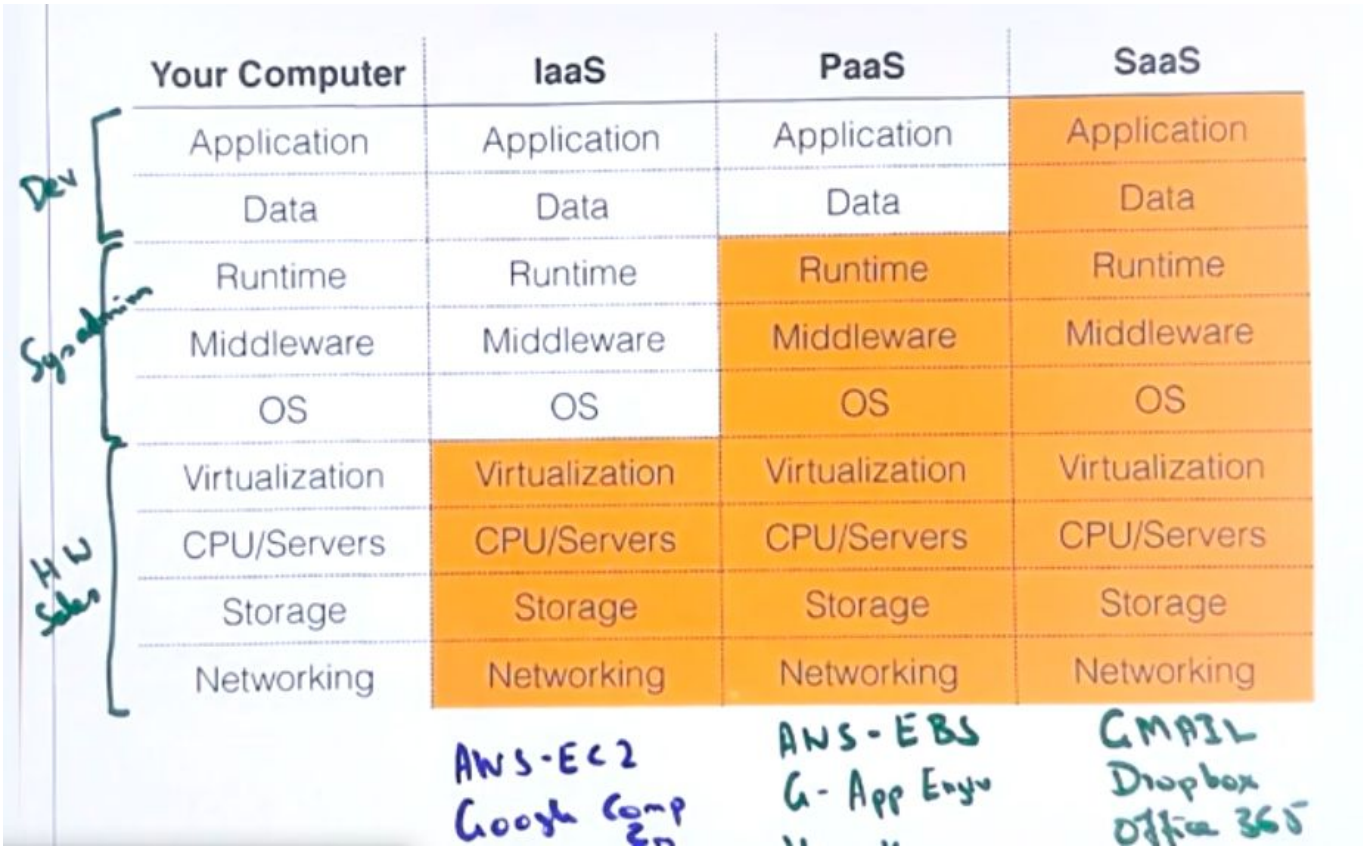
Business Case:

Division of labour, allows to decrease cost and increase efficiency.
Allows your technical team to get to market quickly, with low cost, while allowing them to exclusively to focus on the core business applications.



greatlearning
Power Ahead

Different types of Computer/Cloud



Your Computer	IaaS	PaaS	SaaS
Application	Application	Application	Application
Data	Data	Data	Data
Runtime	Runtime	Runtime	Runtime
Middleware	Middleware	Middleware	Middleware
OS	OS	OS	OS
Virtualization	Virtualization	Virtualization	Virtualization
CPU/Servers	CPU/Servers	CPU/Servers	CPU/Servers
Storage	Storage	Storage	Storage
Networking	Networking	Networking	Networking

SWOT Analysis

Strengths	Weaknesses
Low upfront cost	Managing spend
Low upfront time(time to market)	Governance and control
No Maintenance/upgrades	Data resides in remote
Low to no IT staff	Might be a shared resource
Allow risk sharing (unused resource handling)	Limited customization(any option outside the cloud providers scope)
Mirroring, High availability and security	Long run ROI unsure
Scalable and flexible	Multiple cloud integration
Option to abandon infra without cost leverage	Evolving internal systems and requirement for change
CAPEX vs OPEX (CAPEX is always high & Avoidable)	

SWOT

Opportunities	Threats
Cloud integration	Compliance and Regulation
All enterprise system can be on the cloud	Denial of service attacks - cybersecurity
Analytics built on the cloud	One point failure - data loss
Exciting new technologies and paradigms	Weak identity - access management
	Account hijacking

Some instances where the cloud is secure (physical security). Other instances where local is secure depends on the consumer.

Site Reliability Engineering

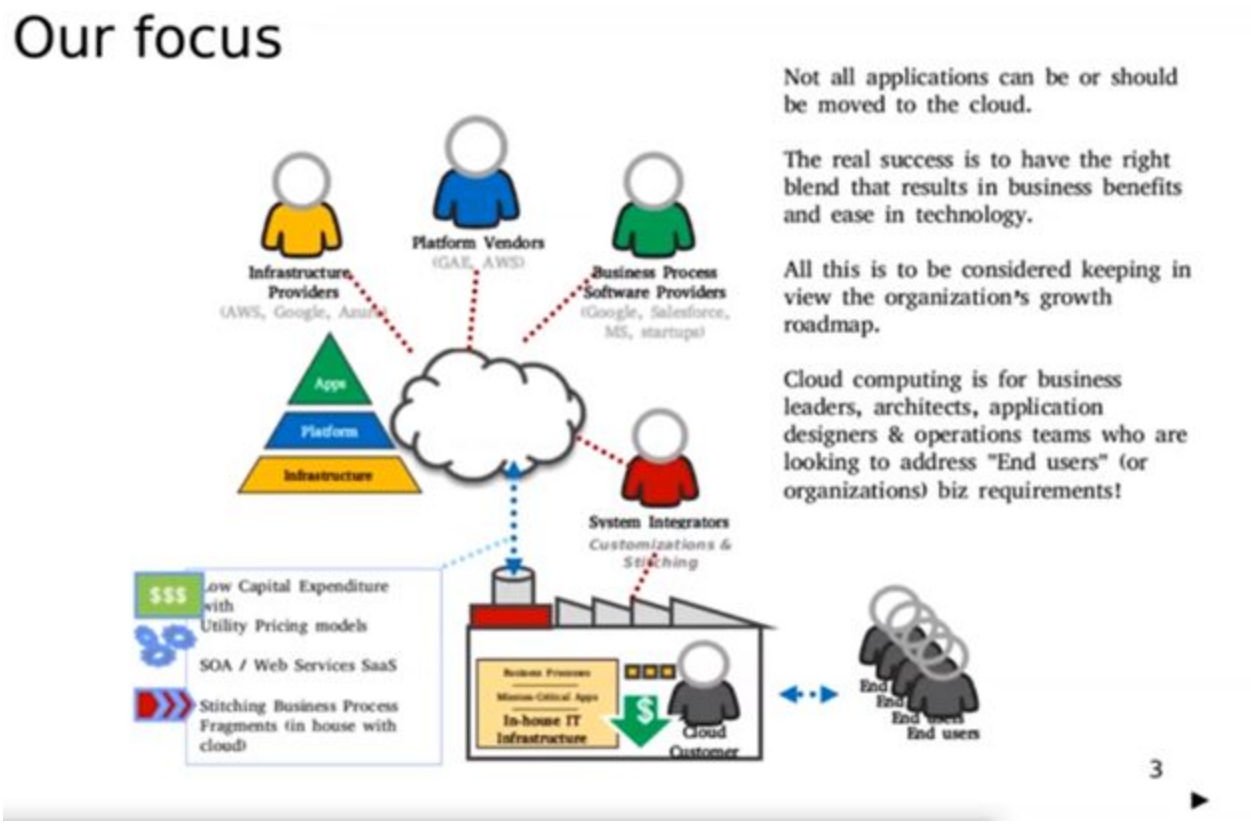
The developer keeps on improving the apps while sysAdmins tend to hold a state of the application without change.
So while both work together the conflict is always inevitable.

If we make a developer work with the ops team to enable changes seamlessly and add automations for deployment and maintenance.
Then the generalised term of ‘devops engineer’ arises. While google termed it as site reliability engineer.

Capacity Planning

The server needs for a specific application/business can be approximated and configured. But when the utilization of the server is less than the max capacity and when the utilization of the server is reaching above 100% (outage) then the capacity planning fails.
To avoid this cloud offer 2most important factors -> ‘Scale’ - increase capacity with the demand and ‘flexibility’ - reduce capacity when demand falls.

The customer is devoid of the all over capacity of the resources and only uses what is required for them. In the meantime, cloud providers have a separate use for the available resources on the cloud which is not utilized. It is provided as a low cost computation service to other customers.



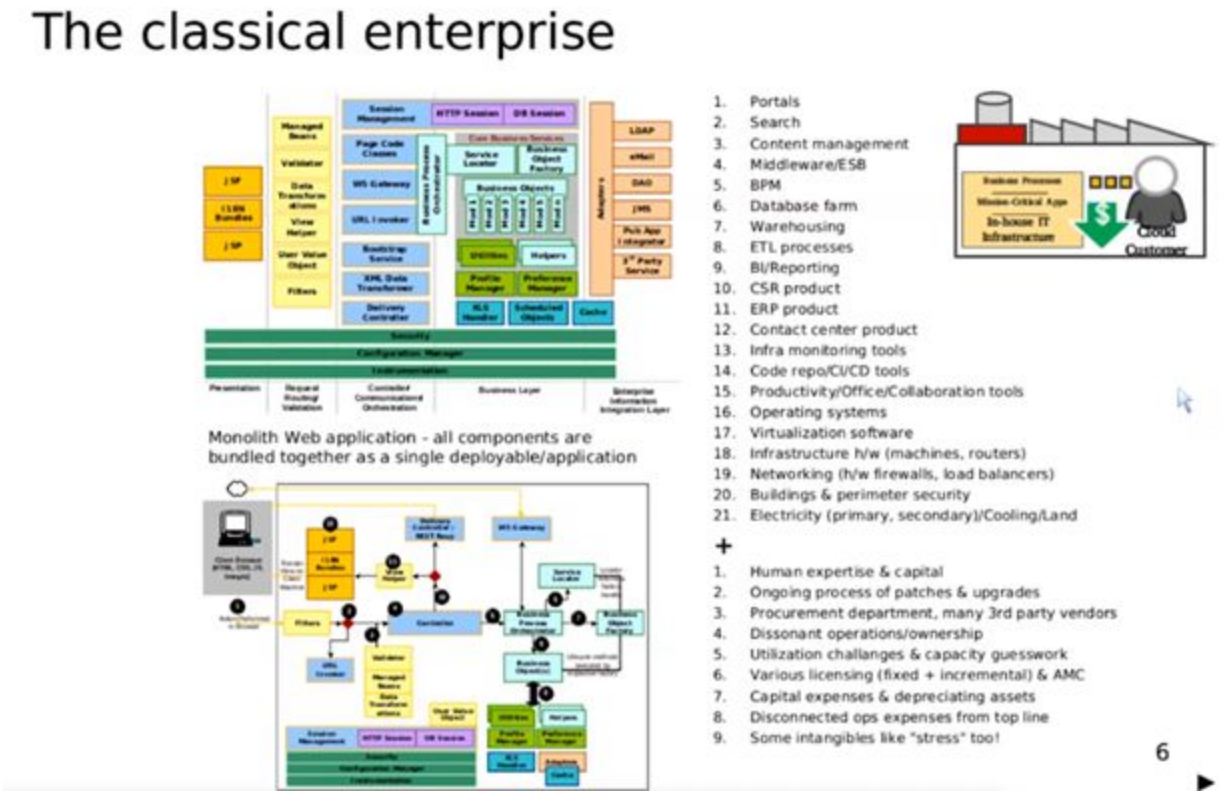
How is cloud computing needed to be used?
When it should be opted for.
How it would help in the organisation.
Which professional uses it and how it is used.

Business problems

Businesses would have to consider a lot of points pros and cons before taking the decision to implement cloud computing and the problems they face while using on-premise or data center are lists here:

a. Infrastructure as a service is required
b. Expertise of managing the data center
c. Business focus should not shift to handling the infra rather it should be within the business priority actions.
d. Availability at uncertain loads
e. Provisional latency
f. Elasticity
g. Redundancy(Failure tolerance)
h. Infra refresh(life cycle to hardware)
i. Long term/Lock in
j. Confidentiality
k. Compliance
l. Audit trail

The classical enterprise structure and requirements



All the requirements to handle a complete IT infrastructure is being shown. The list reminds of both the physical and virtual(software) requirements.

Why cloud?

Utilization factor when compared to Datacenter is better in Cloud (Elasticity) and time incurred in deploying the hardware is reduced

History

From the 1950's the programming logic remained the same but the architecture of the infra has been improving and helping us speed up the processing and delivery.

Current Era - The Fifth Inflection Point – Virtualization and Cloud computing

Definitions

Style of computing in which massively scalable IT related capabilities are provided “as a service” using internet technologies to multiple “external customers”

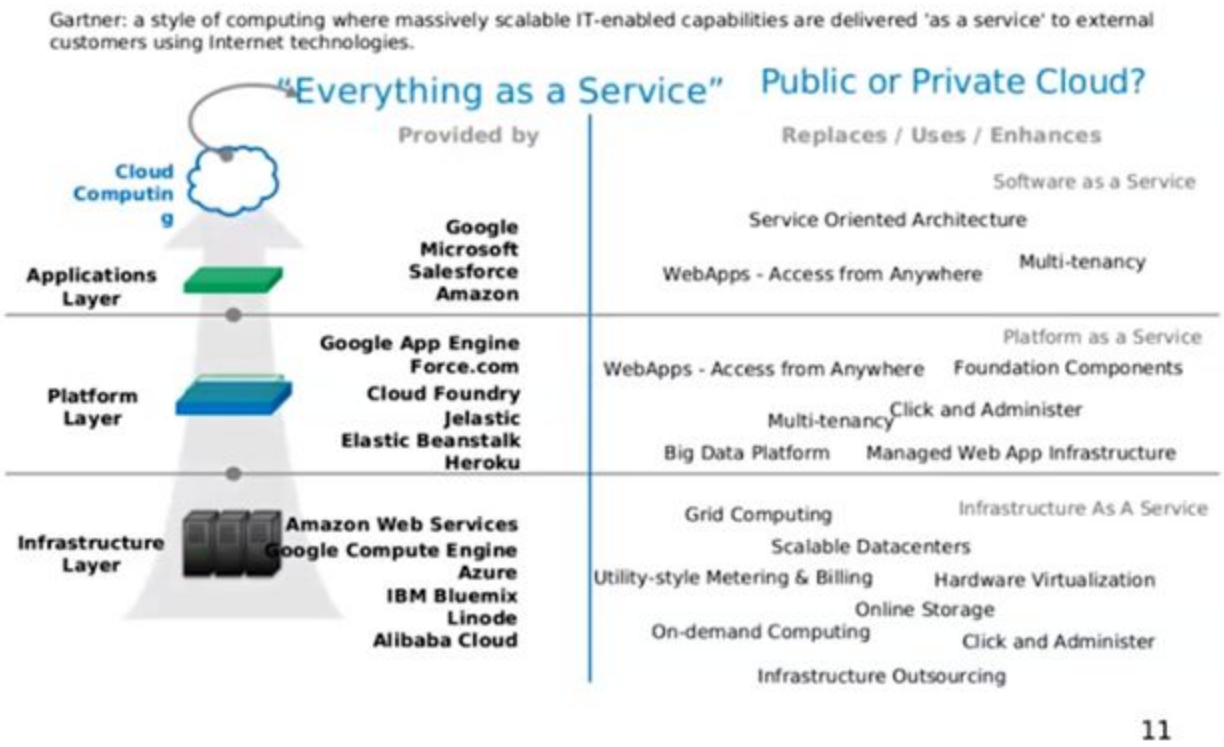
Pool of abstracted, highly scalable, and managed compute infrastructure capable of hosting end-customer applications and billed by consumption.

Myths of cloud computing

1. There’s one single “Cloud”
2. All you need is a credit card
3. The cloud always saves you money
4. The cloud always reduces your workload
5. Integration
    a. Seamlessly blend cloud and your on-premises
    b. Not suitable to blend public and private cloud
6. Cloud provider can guarantee security
7. Using virtualization and cloud computing is same
8. Cloud computing is only about technology

There are myths only but also would come true if implemented without much expertise or knowledge in it.

Service delivery models



Different models of service delivery have been explained here.

SPIDERS

SaaS

PaaS

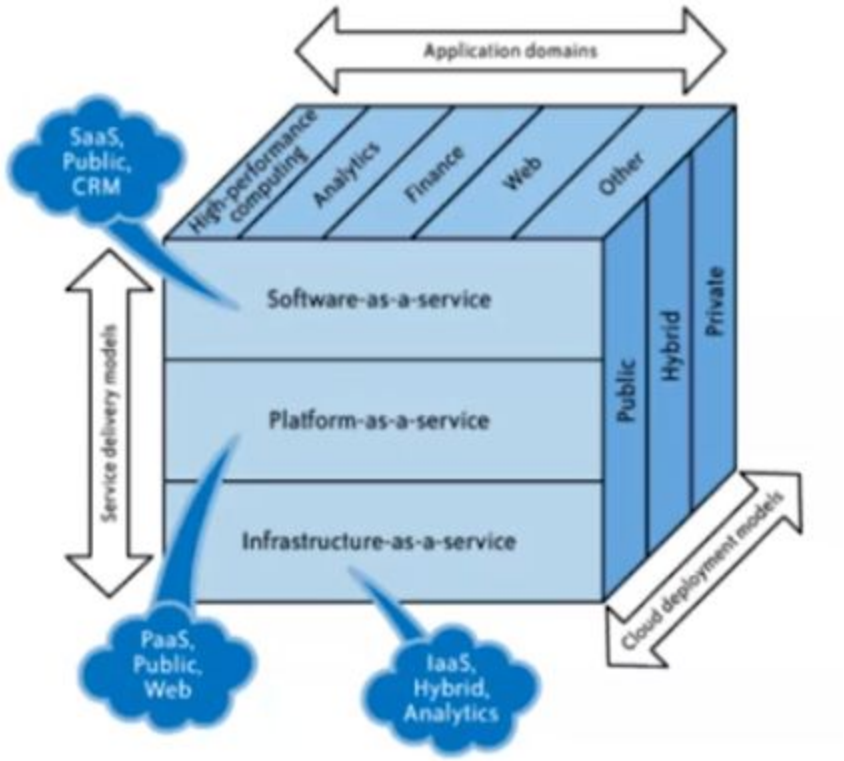
IaaS

bigData

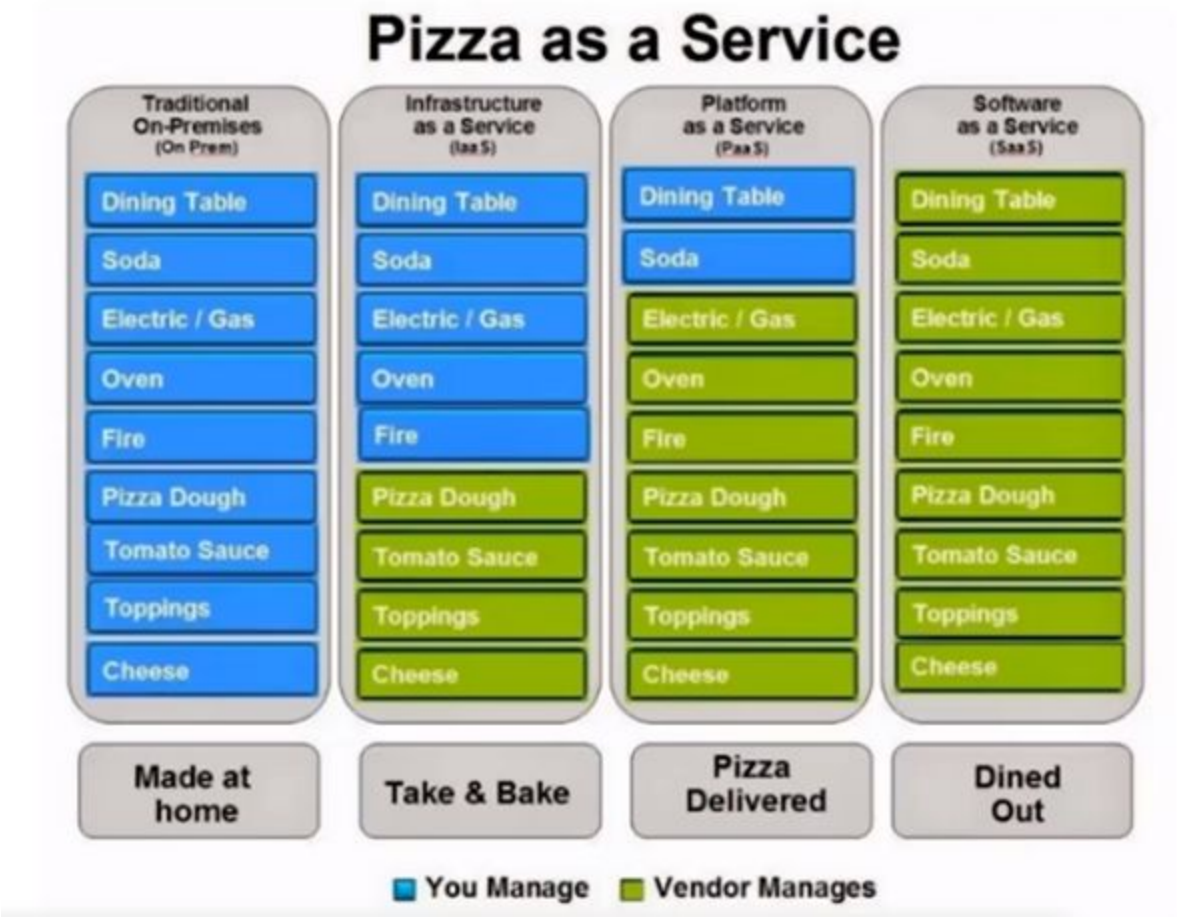
Elastic

Resilient

Subscription



Prospective pizza example

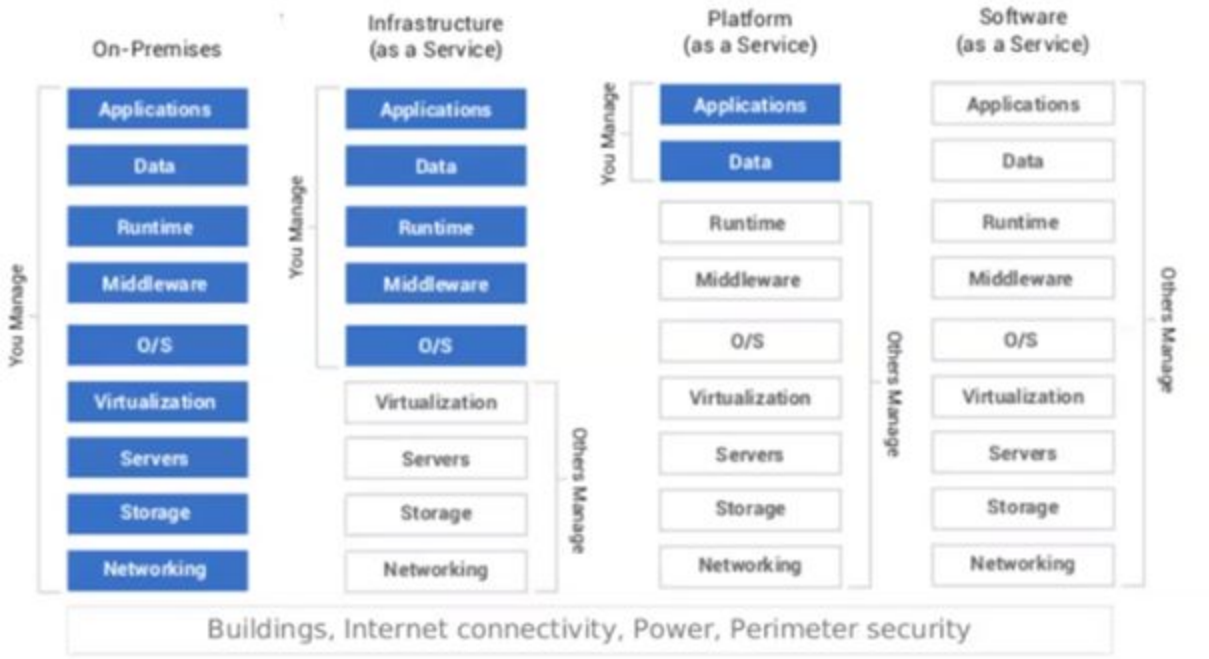


Example for different service models with a pizza example.

Cloud providers comparison

Source: Gartner (July 2019)

Degree of abstraction - app view



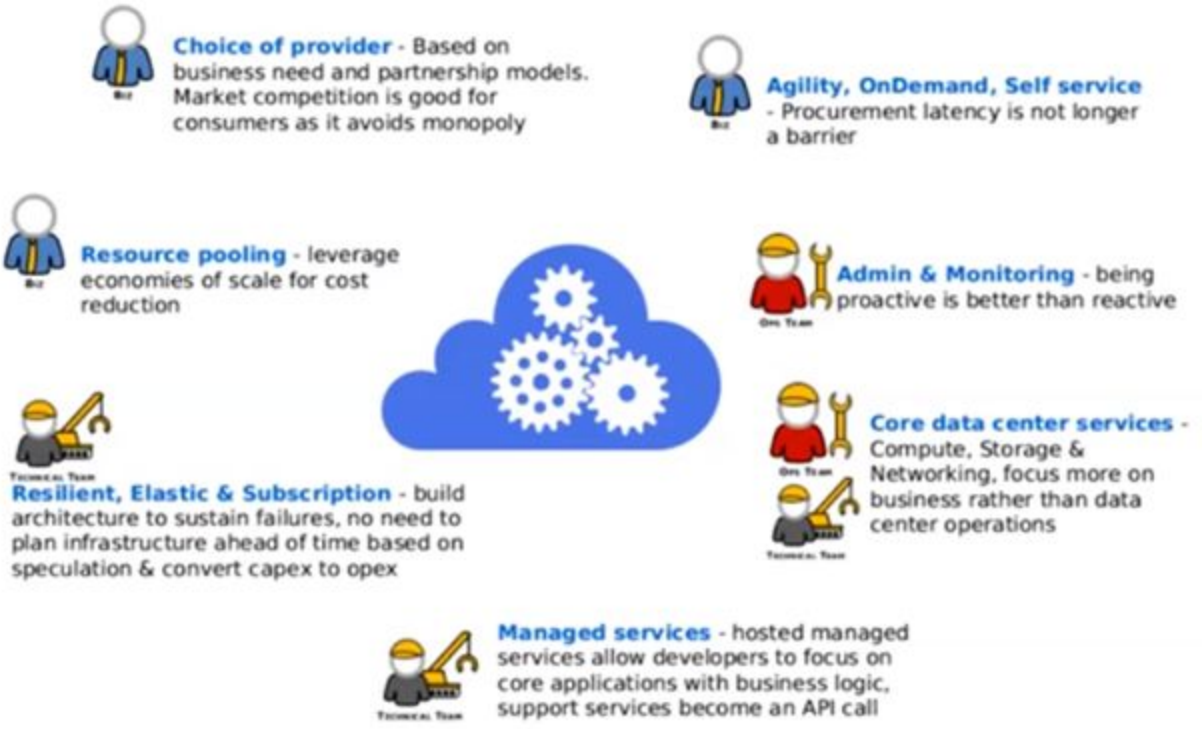
Same example over and application service.

Actual Datacenter view

Check out google datacenter:

https://www.youtube.com/watch?v=XZmGGAbHqa0

Cloud Computing attributes

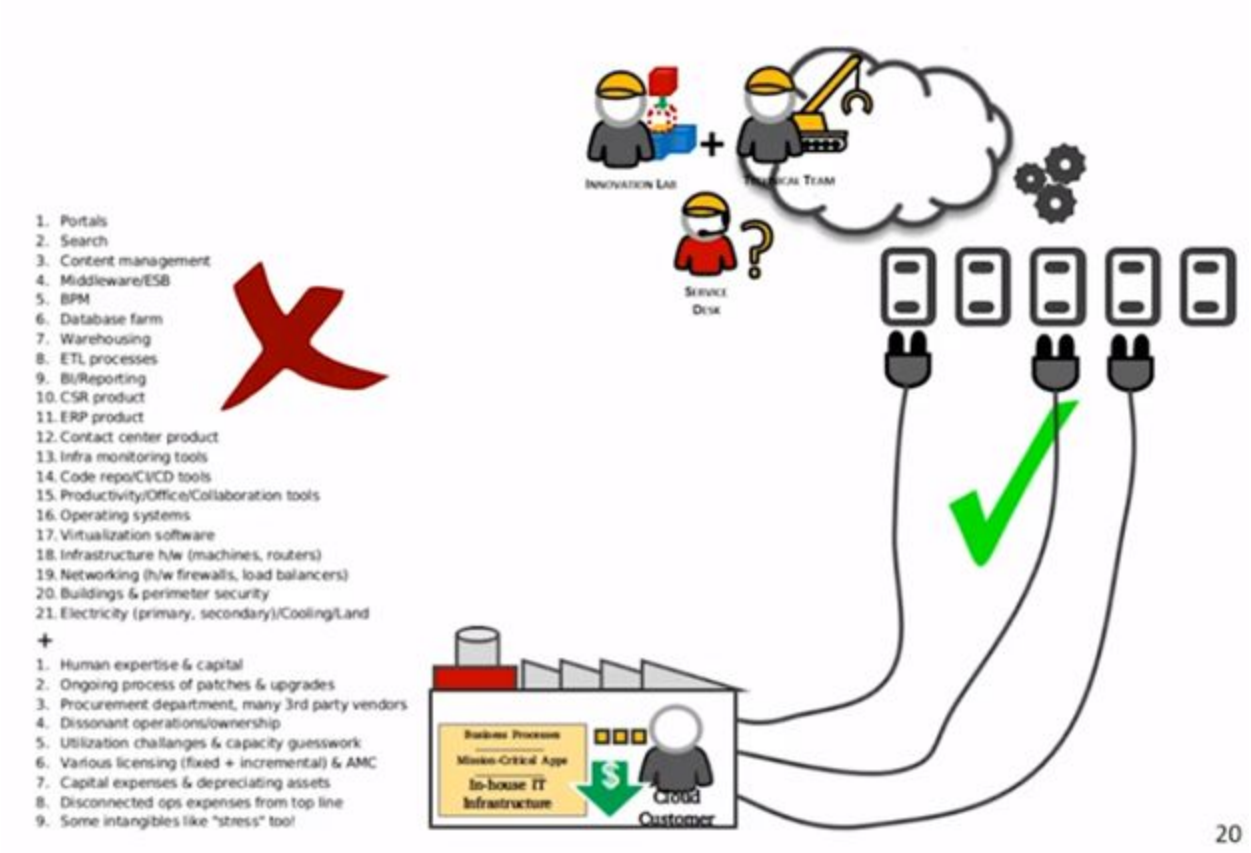


What powers up cloud computing and makes organisation, developers and administrators be inclined towards it. The pointers mentioned in the above picture add those values to the cloud.

Cloud Offerings

Since the type of tools used inside each cloud has its own name here are the generalized names of the tools that each cloud provides. For example, elastic infrastructure is the same technology but used as EC2 instances in AWS , virtual machines in Azure and GCP.

Hosted Managed services



The actions items of classical enterprise are now packaged as a cloud provider and hence all those services become managed services. The cloud customer just needs those services as plug and play and pay only for those services.

Cloud Deployment models

Public:

A model where a service provider makes resources, such as applications and storage, available to the general public over the internet.

The hosting environment is shared between many customers possibly reducing the costs for an individual customer.

Leveraging economies of scale enables a dynamic use of resources, because workload peaks of some customers occur during times of low workload of other customers

It is hosted and managed by a 3rd party from one or more data centers

Private:

Cloud computing properties are enabled in a company-internal datacenter

Alternatively, the private cloud may be hosted exclusively in the data center of an external provider, then referred to as an outsourced private cloud.

Public Cloud providers also offer means to create an isolated portion of their cloud made accessible to only one customer: a Virtual Private cloud which is the default behavior for many public cloud providers.

Community:

IT resources required by all collaborating partners are offered in a controlled environment accessible only by the community of companies that generally trust each other.

Carving out a dedicated area exclusively for a company could be possible(private)

A similar model to that of Asp, completely managed by a 3rd party.

Hybrid:

Any combination of Public, private and Community

Eg any cloud along with any static in-house data center are integrated.

Applications can choose the right environment leveraging the best from each option.

Enabled in “Cloud bursting”

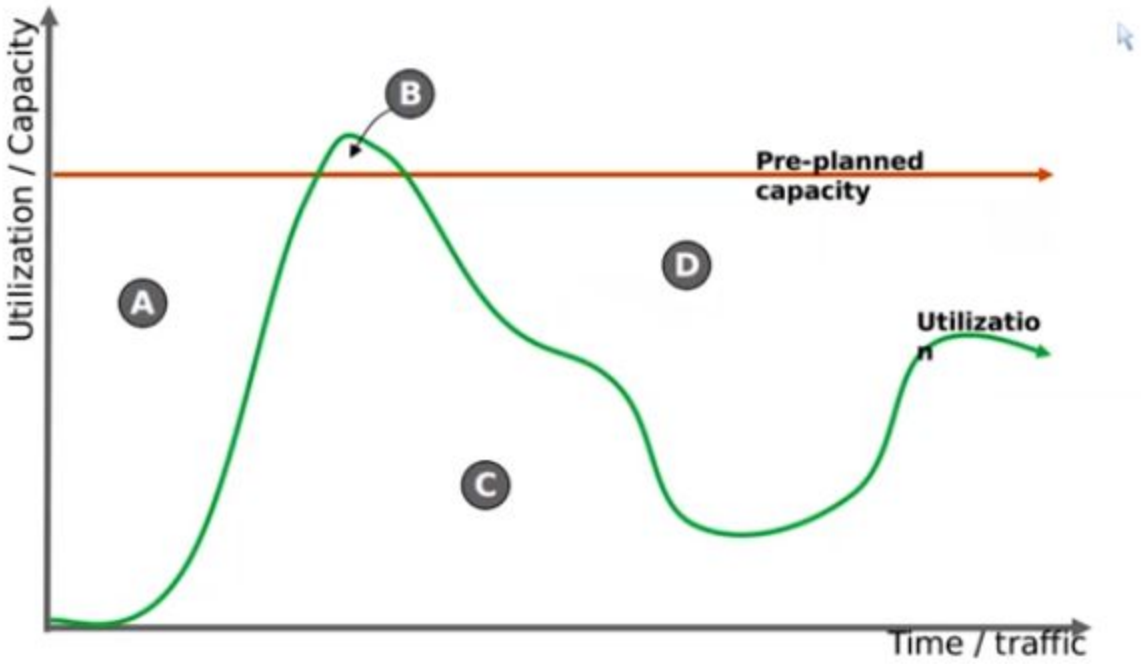
Interconnecting usually happens via VPN

Private connect – no connection to actual internet.

Subscription model

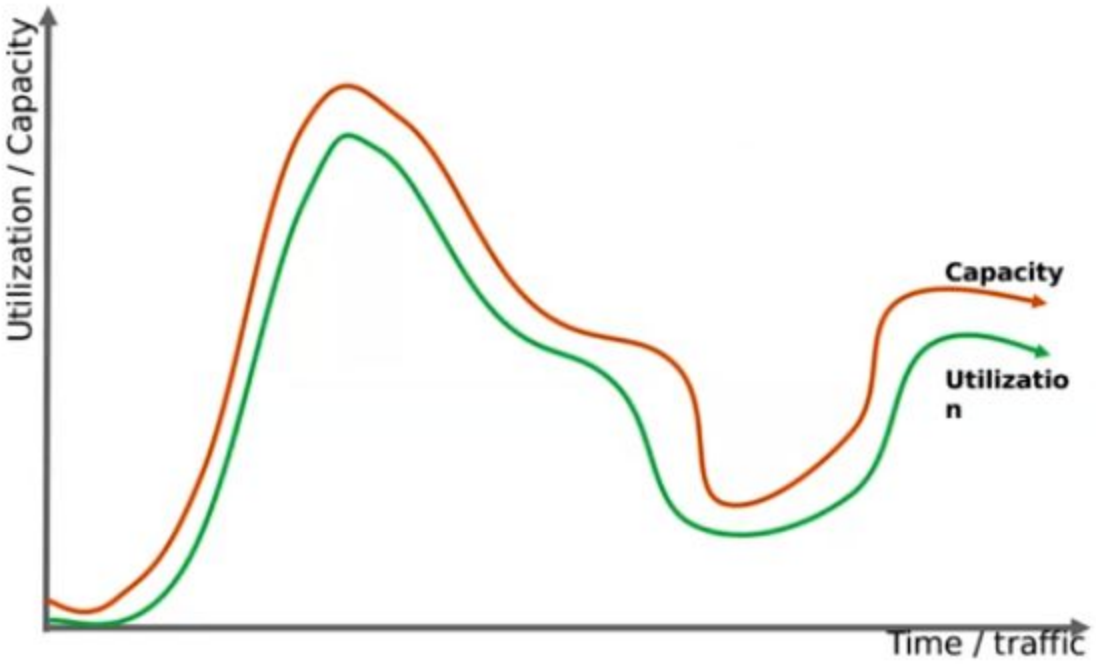
1. Customers avoid large upfront capital expense
2. Pay as an ongoing operational expense
3. Easily and quickly scale up or down based on business demand and only pay for what is needed.
4. Better matches today's financial drivers
5. No buying, it's always 'Renting'

Cost economics – Classical mode



What's happening in areas A, B, C & D? What about OPEX? 31

Cost economics – Cloud model



Delta T calculation for triggering the ‘add provision’ or ‘remove provision’ needs to be set by the user with respect to the application needs and demand monitoring.

Note: Delta T is the time taken for a new resource to get added into the infra and act as added capacity to the infra. Without proper calculation of which the capacity graph will look like steep steps up and step down.

Vertical scaling (Specialized) vs. Horizontal scaling (Commodity)

Large Operation expense	Less Operating expense
Wasted/Idle resource	Maximum resource utilization
Failure takes out a large chunk	Failure takes out a small chunk
Expensive redundancy model	Inexpensive redundancy
One shoe fitting all model	Specific h/w for specific task
Too much co-existence	Very less to no co-existence

Introduction to Virtualization

Option to utilize a resource to fullest. Divide and rule formula, rather than having one big machine.

a. Virtualization of the computing resources, including servers, network and storage, allows dynamic flexibility.

b. Capacity can be more efficiently utilized.

c. Quickly add new servers without delay due to procurement or installation.

d. Easy to turn on or off virtual servers to handle scalability.

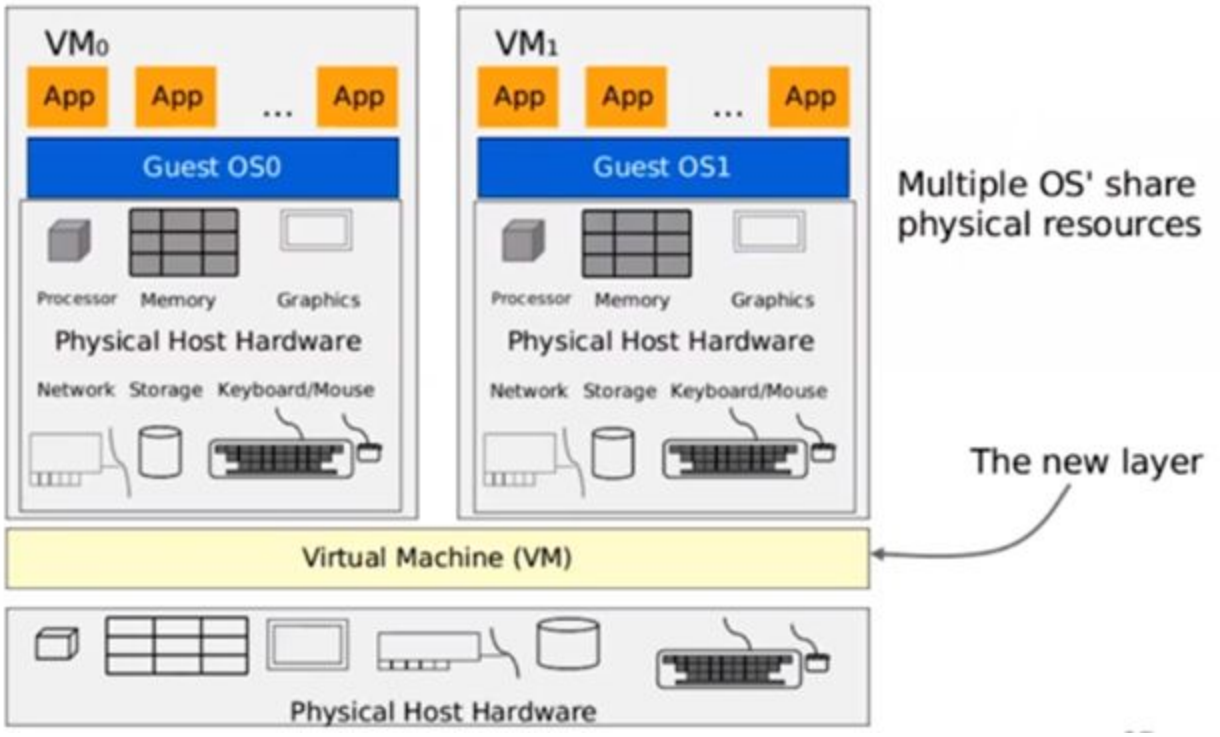
e. Physical connectivity is done up front and configuration is done in software at provisional time.

f. Networking equipment and storage is virtualized as well.

A typical application stack

The single OS bears all the underlying bare metal to the application level.

Virtualized Stack



Multiple OS on the bare metal and application runs within an OS on the same machine but on different virtual machines

Characteristics of VM in Cloud

Business Perspective:

1. Option of more computing resources available on demand
2. Elimination of upfront monetary commitment
3. Ability to pay for computing resources on a short term basis as needed.
4. There are options to acquire compute resources by ‘auction’
5. A service centric approach with self-service and is self-managed

Technology Perspective:

1. Partitioning: In virtualization, many applications and operating systems are supported in a single physical system by partitioning the available resources.
2. Isolation: Each virtual machine is isolated from its host physical system and other virtualized machines. Because of the isolation, if one virtual-instance crashes, it doesn't affect the other virtual machines. In addition, data isn't shared between one virtual container and another.
3. Encapsulation: A virtual machine can be represented as a single file, so you can identify it based on the service it provides. In essence, the encapsulated process could be a business service.
4. Flexible: Should be able to configure and reconfigure to meet the growing and changing needs.

Virtualization Drawbacks

Amplified physical failures

Skills required to set up perfectly

Complex root cause analysis

Estimating the number of VMs per physical hardware

Performance reduces with added layers

Some application do not play well in a virtualized environment

Licensing costs

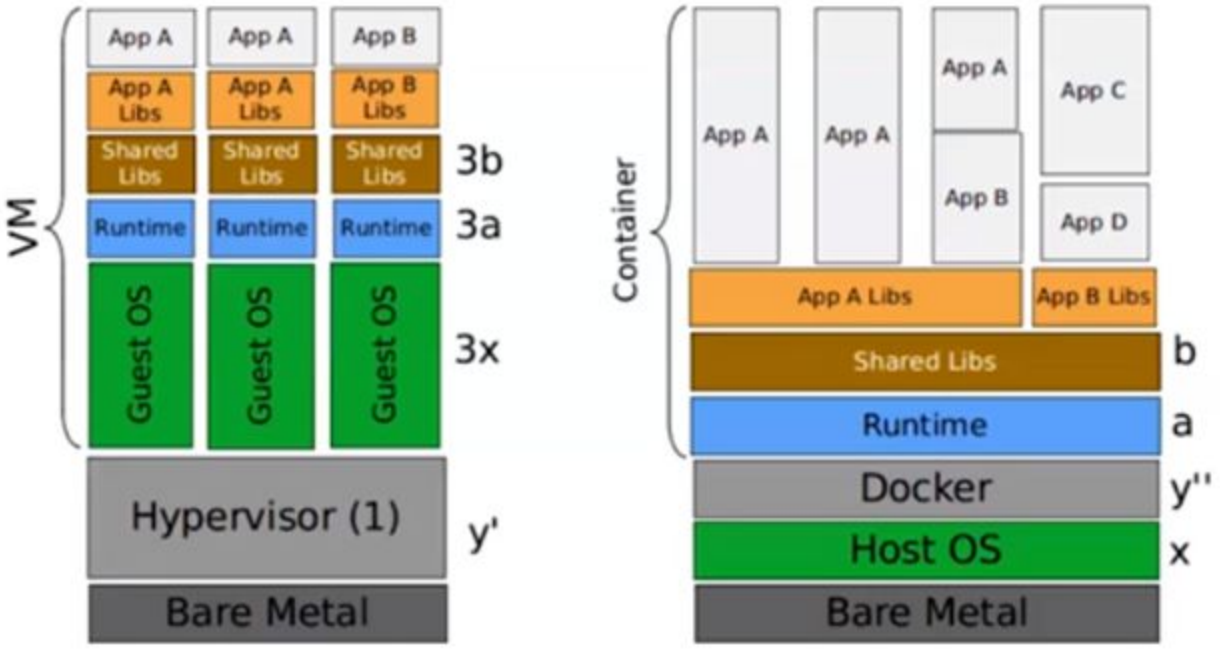
Next gen Virtualization

1. Dynamic swapping in virtualization is a need to avoid amplified failures
2. Virtualized machines come with fixed sizes inside the fixed machines. i.e. max cores cannot exceed the hardware machines limits.
3. How to fix the number of VMs per bare metal machine?

Containers :- (LinuX Containers – LXC)

1. Application focused bundling
2. Common aspects(OS, library etc.) are shared
3. Smaller footprints
4. Quick load time

Containerization vs Virtualization



Container vs Virtualization

All apps need to be deployed on the same OS image (good choice for PaaS)	VM can have a choice of OS(good option for Iaas)
Containers have less isolation among them	VM allows for higher isolation
Containers will usually take milliseconds to start	VM may take minutes

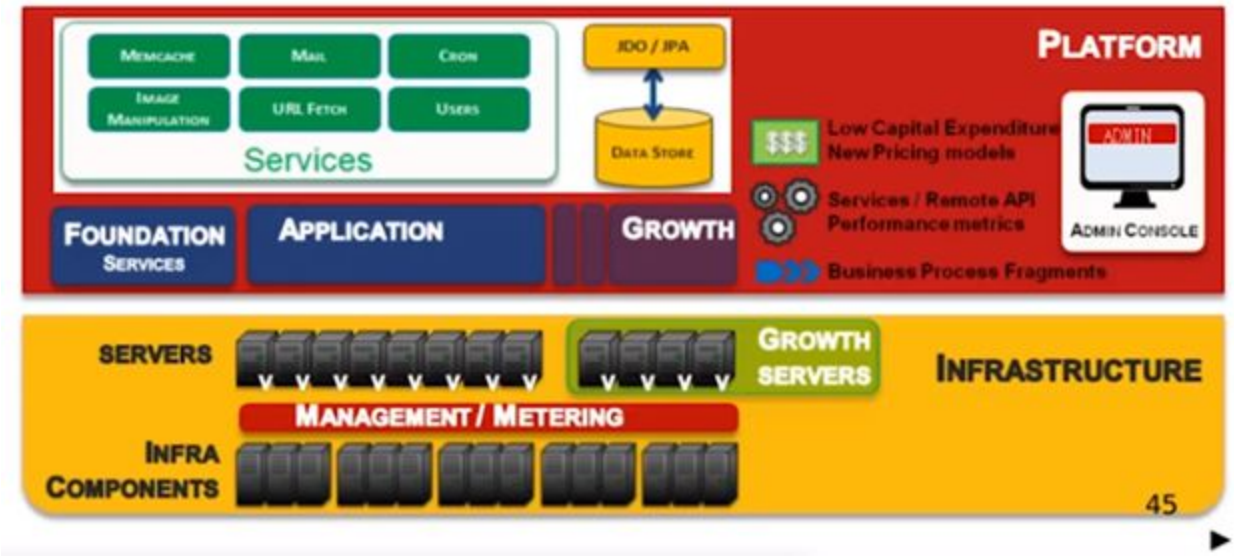
Use Container or virtualization?

Depends on the business requirement totally

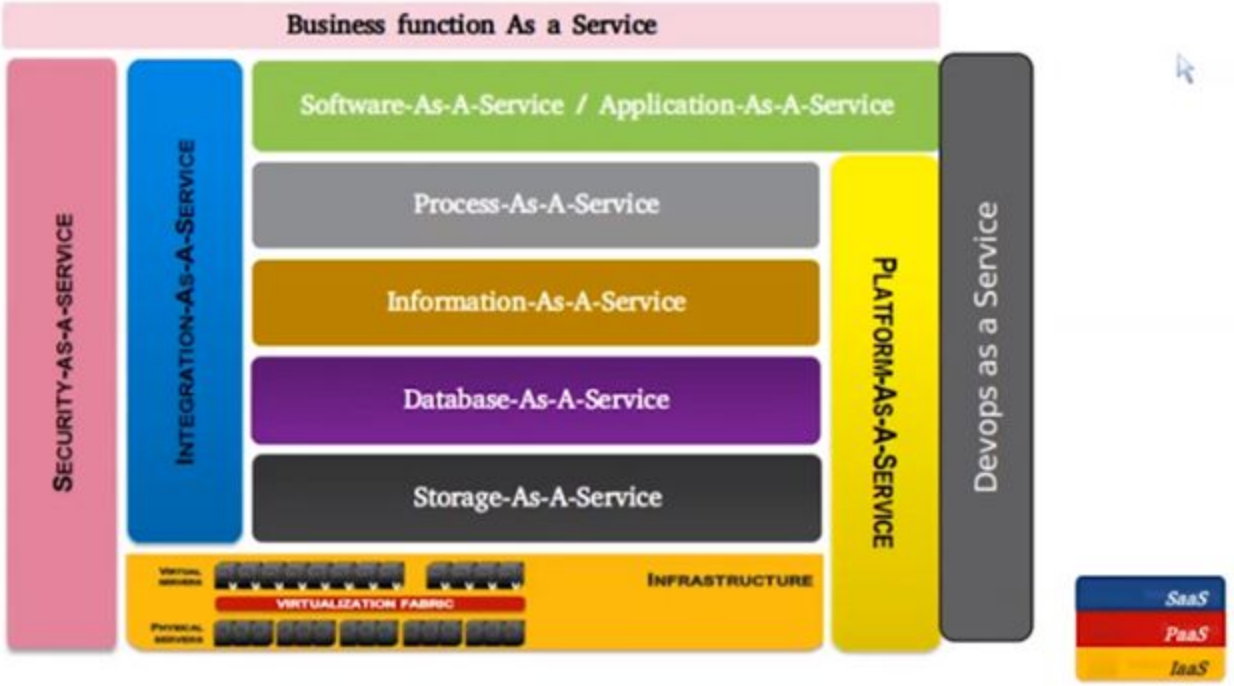
Use cases where both options are combined is also a valid scenario.

PaaS overview

Abstracts the infra layer and enables us to monitor both infra and application layers. Helps in managing the underlying layers.



Cloud service Taxonomy



The main models being the primary IaaS, PaaS and SaaS, any higher level service models can be created out of these and above and combination goes on.

Price economics & Decision making

Profit = Revenue – Costs (Since everything is pay-as-you-go)

Cloud allows you to control and monitor the ops costs

Cost Based:

- Cost of building product
- Cost of maintaining the product
- Providing support for the product
- Other factors that may be specific to the product as well as the organization
- Charging for the services offered and/or enhanced
- Pass through costs can find a way in the product price

Value based:

- Customer perception about the product
- What values customer derives from the product
- Longevity of the product
- Recognizable name associated with luxury

And hence the cost in different regions will be different.

And hence you will need to analyze the cost aligned with multiple factors and decide/plan the strategies.

Data to Information xform velocity

The decision to make the cost optimized needs to be analyzed according to your requirements and then the cost should be calculated.

Cloud provides managed service which handles these cost optimization

Similarly many other such data collection is managed by cloud and provided as a managed service like monitoring and dashboards etc.

Challenges with Distributed computing

Heterogeneity (Apps, DB, Cache, Storage)

Fault handling

Consistency – Strict and Eventual

Global concurrency

Upgrades and maintenance

Local file systems

Application sessions – transient data (synchronization)

Designing application for the cloud

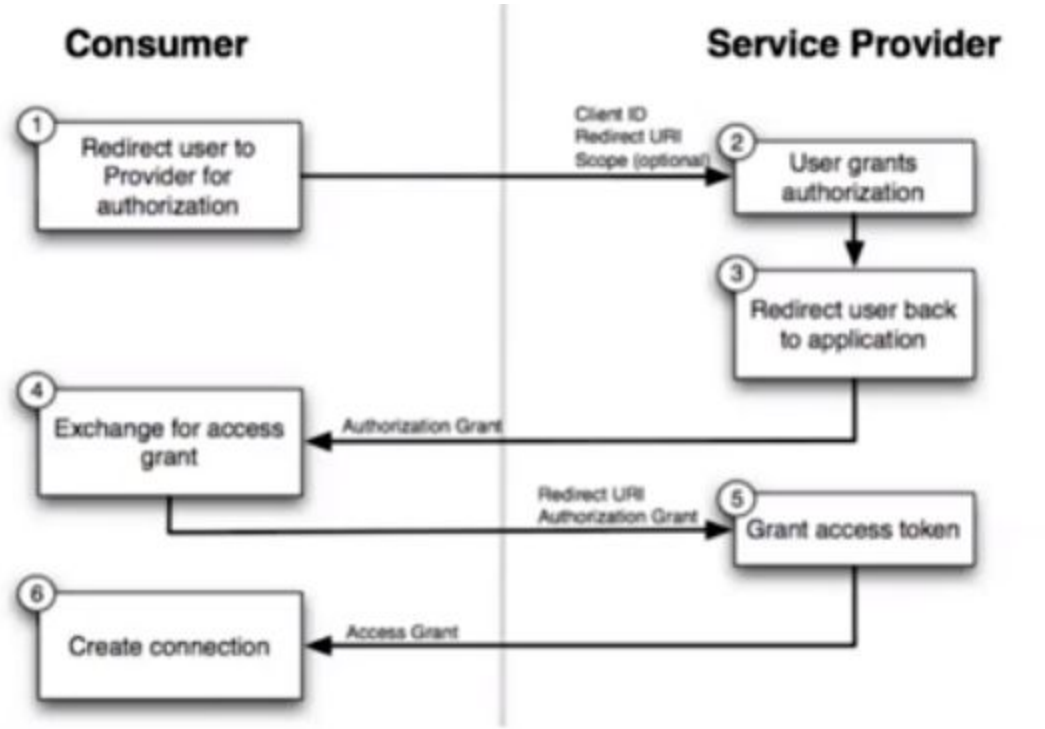
1. Impact analysis on architecture
2. Which cloud provider
3. Factoring defining the business process and integration
4. Security options
5. Foundation, stateless, future etc.
6. Micro service approach

Distribution of Control between Service Models

1. Decentralized Administration – each service model retains administrative control over its resources
2. Secure Distributed Collaboration – SLA maintenance even when Cloud provider faces some service offline
3. Credential Federation – decentralized single-sign-on mechanism
4. Placement of functionality – huge impact if not placed properly
5. Federated Data Collaboration – use on-prem and cloud else bring everything inside cloud
6. Loose coupling – different evolution cycle for the different services and hence loose coupling will help in retaining the infra intact and still act with individuality.

Security of architecture with multiple providers.

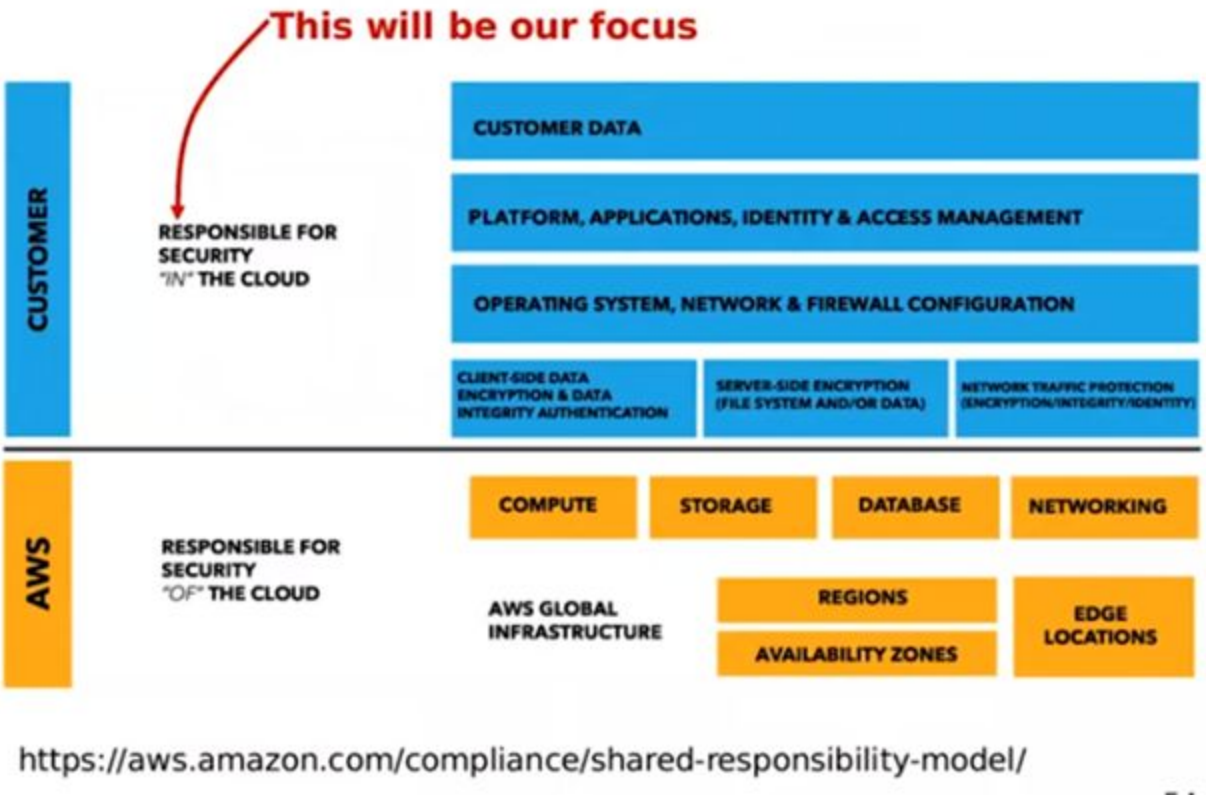
Authenticator like OAuth, facebook, Gmail can be used for user login/data.



Shared responsibility model

Security “of the cloud” – with cloud provider

Security of your application “in the cloud” – you are responsible



Diversity of programming languages

There are many languages and the cloud is compatible with anything you want until you fix the architecture in the cloud accordingly. Popularity index:

1. Python
2. Javascript
3. Java

Infrastructure automation

Infrastructure provisioning can be automated using machine-readable scripts, instead of performing manual provisioning. This leads to considerable increase in speed of infrastructural operations, and eliminates the human error.This is known an Infrastructure as Code (IaC)

Options:

a. Chef.io

b. Puppet.com

c. Ansible

d. CI using Jenkins

e. Cloud provider based automation tools. Each has their own.

CI CD models

The developer pushes their codes to version control (git)

The CI/CD automatically kicks in when the code is pushed to git and publishes the code in staging and starts testing/pushes to production, according to the pipeline flow design. Can also be used with a docker.

IoT

Physical objects to interact with digital and connected through the internet. Ex. Smart refrigerators, smart AC, smart water heaters, etc.

Google Brillo:-

https://www.youtube.com/watch?v=2rPkbyyviGI

Glossary

- **Cloud Computing:** The delivery of on-demand computing resources — everything from applications to data centres — over the internet on a pay-for-use basis.
- **Latency:** Time interval between the stimulation and response.
- **Elasticity:** Elasticity is the ability to grow or shrink infrastructure resources dynamically as needed to adapt to workload changes in an autonomic manner.
- **Redundancy:** Provision or existence of more than one means or resources to perform an activity or function.
- **Compliance:** A state of being in accordance with established guidelines or specifications, or the process of becoming so
- **Business Process Management:** The process of analysing and improving business processes to create a more efficient and effective organization.
- **ETL (Extract, Transform Load):** A process in data warehousing responsible for pulling data out of the source systems and placing it into a data warehouse.
- **Microservices:** An architectural style that structures an application as a collection of loosely coupled services, which implement business capabilities.
- **Resilient:** Strong and able to recover from damage quickly.
- **SPIDERS: S**aaS **P**aaS Iaa**S** big**D**ata **E**lastic **R**esilient **S**ubscription
- **IaaS (Infrastructure as a Service):** A standardized, highly automated offering, where compute resources, complemented by storage and networking capabilities are owned and hosted by a service provider and offered to customers on-demand.
- **PaaS (Platform as a Service):** A broad collection of application infrastructure (middleware) services (including application platform, integration, business process management and database services).
- **SaaS (Software as a Service):** A software that is owned, delivered and managed remotely by one or more providers.
- **Resource Pooling:** Resources pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand.
- **Managed Service:** An information technology (IT) task provided by a third-party contractor and delivered to a customer.
- **Public Cloud:** Computing services offered by third-party providers over the public Internet, making them available to anyone who wants to use or purchase them.
- **Private Cloud:** Computing services offered either over the Internet or a private internal network and only to select users instead of the general public.
- **Community Cloud:** A shared cloud computing service environment that is targeted to a limited set of organizations or employees.
- **Hybrid Cloud:** A hybrid cloud is an integrated cloud service utilising both private and public clouds to perform distinct functions within the same organisation.
- **Vertical Scalability:** Addition of resources to a single system node, such as a single computer or network station, which often results in additional CPUs or memory.

- **Horizontal Scalability:** Ability to connect multiple hardware or software entities, such as servers, so that they work as a single logical unit.
- **Encapsulation:** It is the mechanism that binds together code and the data it manipulates.
- **Containerization:** an OS-level virtualization method used to deploy and run distributed applications without launching an entire VM for each app.
- **Session Affinity:** Session affinity overrides the load-balancing algorithm by directing all requests in a session to a specific application server.
- **Stateless:** Class that does not have any class level attribute or an object that is instantiated, used and thrown away e.g. HTTP