

Cloud in a box

Seminar 181

Amit Merchant

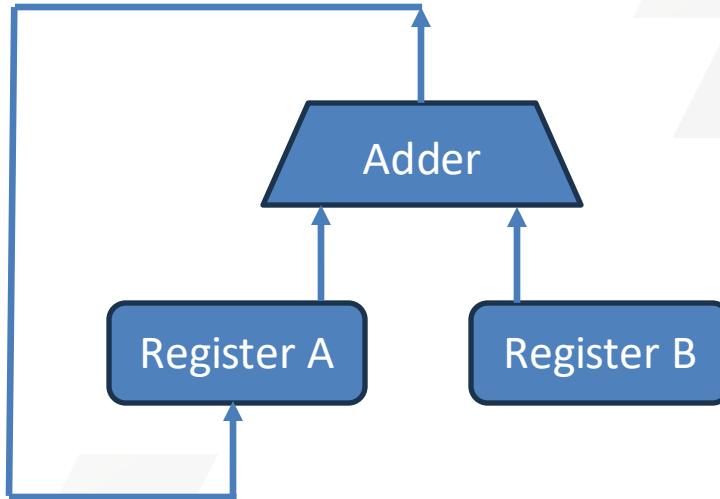
Sep 2025



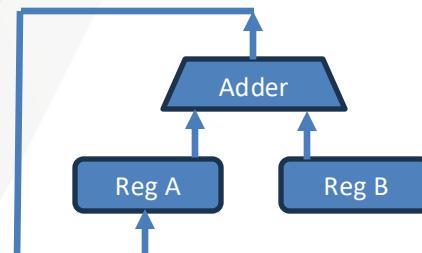
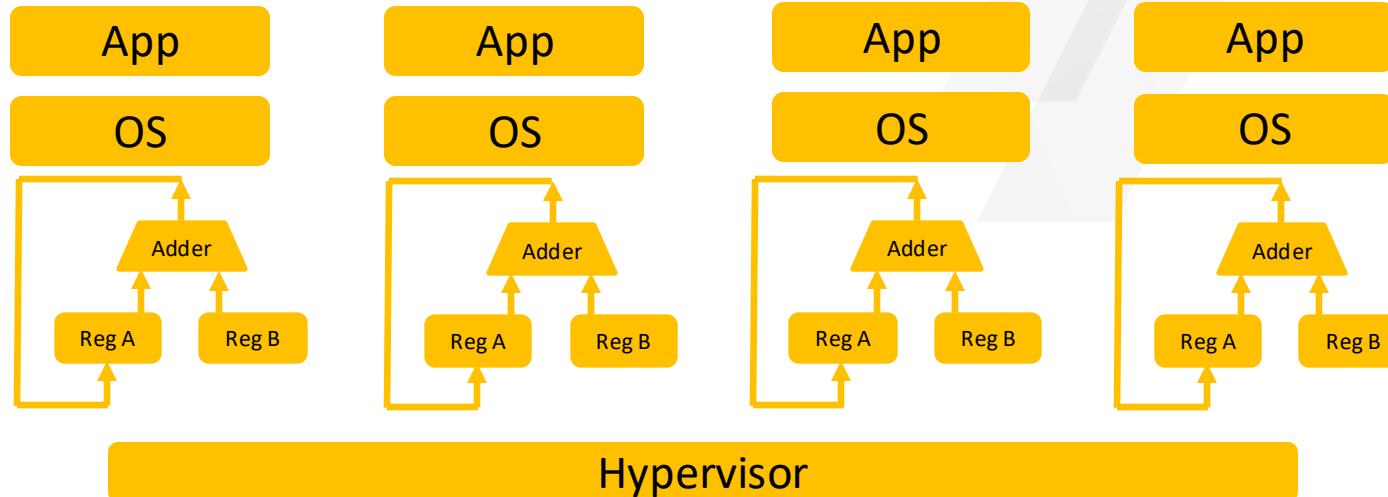


The power of a Cloud rests on one
very important technology:
Virtualization

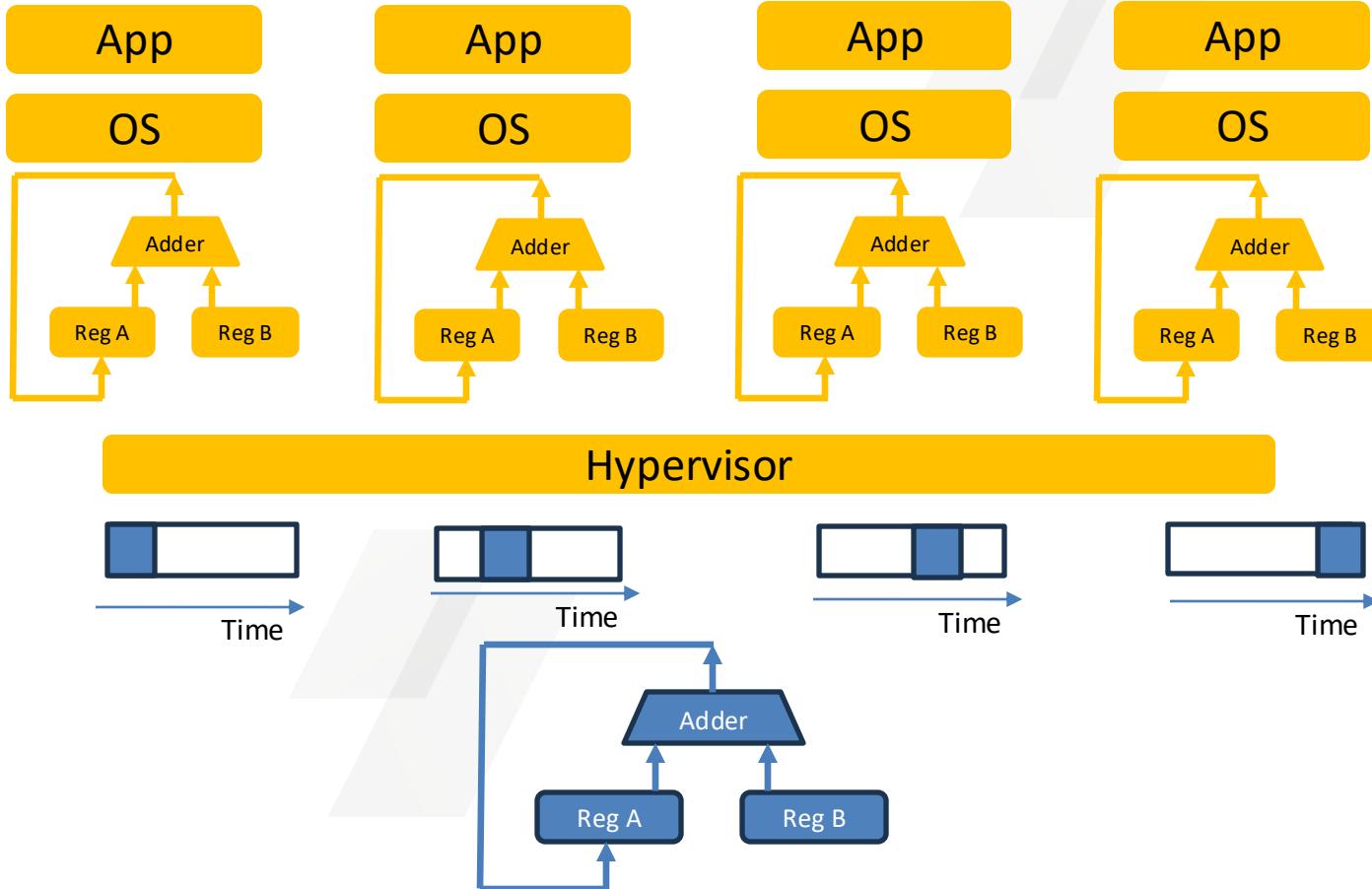
Virtualization: Simple model for a system



Virtualization: Each app sees the same resources available

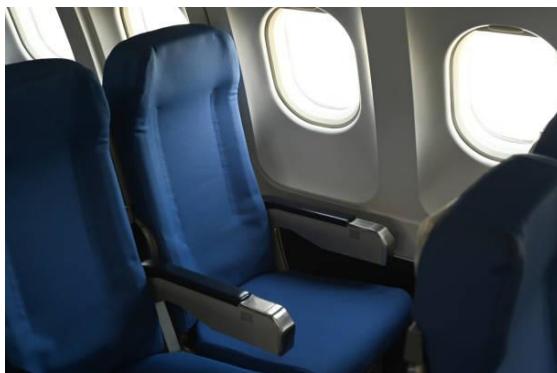


Hypervisor time shares on the actual hardware



Virtualization

- Why does time (sharing) multiplexing work?
 - On an average applications occupy CPUs for 10-30% of the time.
 - Rest: App is idling or waiting on IO.
 - Hypervisor manages CPU resources for each app, time slice for each application
 - Analogy is seat in an aircraft



One seat (CPU) is occupied by several people (Apps) daily, weekly, yearly.
The seat doesn't change, merely the occupants.
Each occupant use the seat for a brief time to complete his/her journey (Workload).

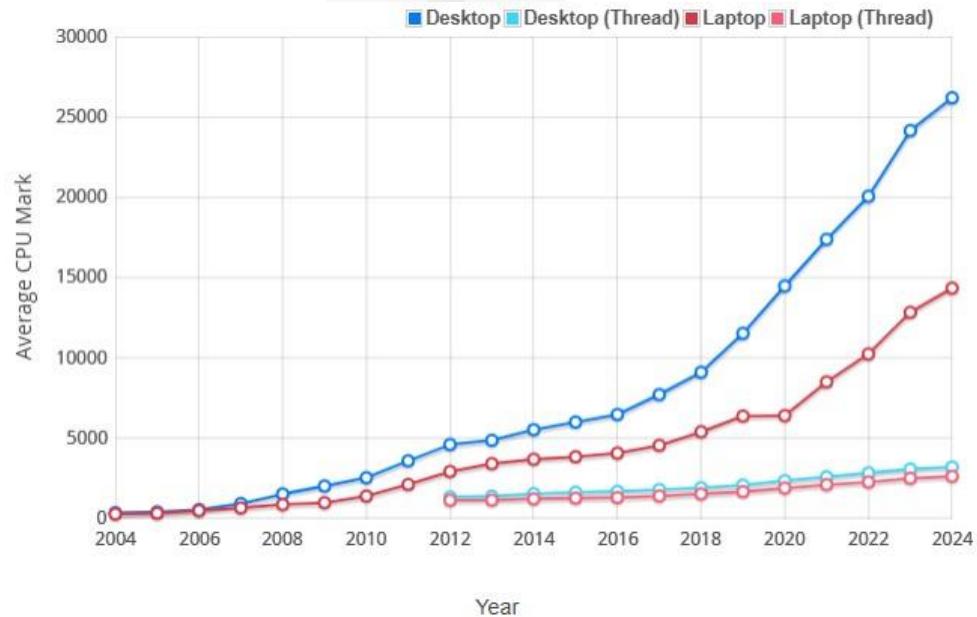
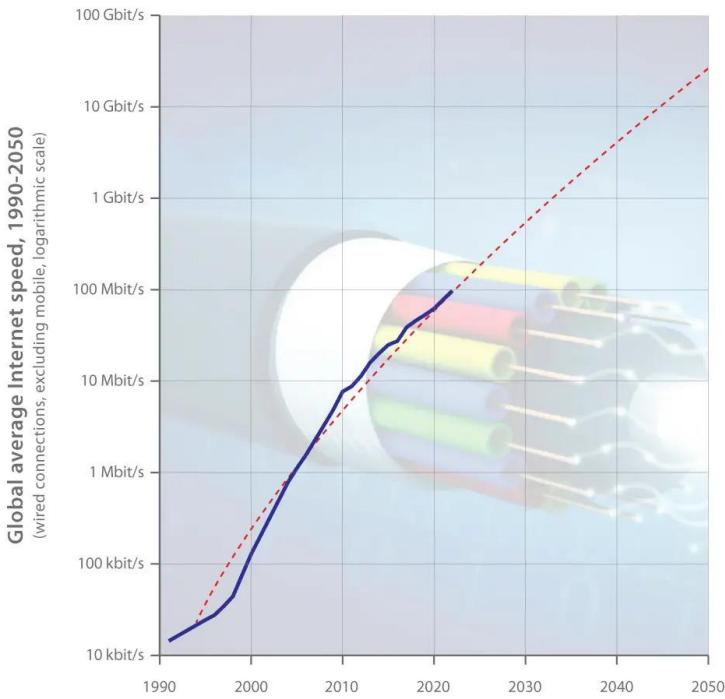
Brief History of Virtualization

- 1967: IBM CP-40 first allowed **simultaneous multiple user** access to a mainframe
- 1972: IBM added virtual memory support in System/370
- 2001: VMware releases ESX Server 1.0, an enterprise class virtualization software
- 2006: Intel and AMD separately announce virtualization support in their processors
- 2006: AWS launches its first services (S3, EC2)
- 2010: Amazon migrates the entire Amazon marketplace to AWS
- 2012: Netflix announces migration of its entire infrastructure to AWS in re:Invent
- 2015: AWS announce first profitable quarter
- 2014: On Oct 31st Amazon announced that its AWS business generated \$27.4b of revenue, operating income of \$10b!

Cloud Computing

- Virtualization is the technological basis for cloud computing at gigantic scale.
 - Whereas Virtualization was first applied at enterprise's datacenter (mostly VMware)
 - Now it is being applied at gigantic scale by Cloud Providers
- Cloud is, “Highly accessible and economical model for computing, storage and software.”
- Cloud’s economic model is founded on:
 - Virtualization technology
 - High speed networking
 - Very highly automated
 - Shared services model
- In the last decade, cloud has become more secure, reliable, highly available, resilient and extremely user friendly.
- Cloud is often compared to a Utility model such as electricity.

What propelled cloud?



Exponential growths in CPU and network performance and storage capacity have been instrumental in enabling cloud business

Key Technical attributes

- Virtualization (a new software layer called Hypervisor)
- High Speed Networking
- Isolation for Workloads (Apps)
- Security
- High degree of automation
- Standardization

Key Economic attributes

- Pay as you use
- Pay less when you use more
- Pay less when you reserve capacity
- Pay less when you are flexible in your needs

Airline Analogy



Cloud	Airlines
Virtual CPU	Seat
Memory	Bin above the seat
Datacenter	Aircraft
Cloud Management SW	Staff
Highly occupied – More Profitable	Highly occupied – More Profitable

One seat (CPU) is occupied by several people (Apps) daily, weekly, yearly.
The seat doesn't change, merely the occupants.
Each occupant uses the seat for a brief time to complete his/her journey (Workload).

Advantages

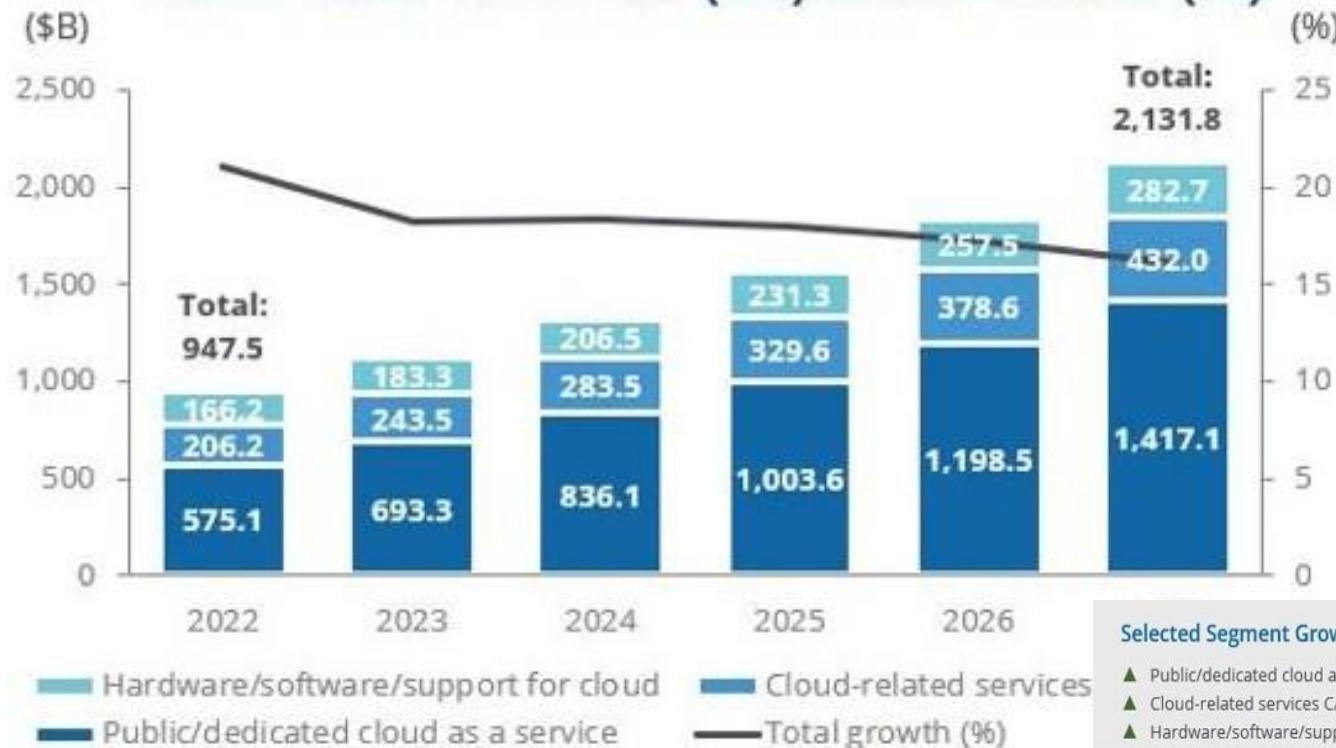
- Cost reduction
- Scaling (you can go from serving 1 customer to 100s of customers instantly)
- Ease of use
- Lower up front capital investment – propelled startup culture
- Focus on business and purpose rather than IT and means

Disadvantages and Challenges

- Security remains a top concern, especially for financial institutions
- Shadow IT (Businesses start using cloud on the side and leave sensitive data vulnerable)
- New skills required – opportunity for engineers as they graduate from CS and CE
- Many enterprise agreements involve buying long term licenses for fixed or dedicated CPUs (example Oracle)
- VM sprawl (Finops challenges) – wasted money
- Migration of workloads from traditional datacenters
- Cloud is not the reason, but cloud has enabled complex integrations – environments are continuously getting more challenging to manage

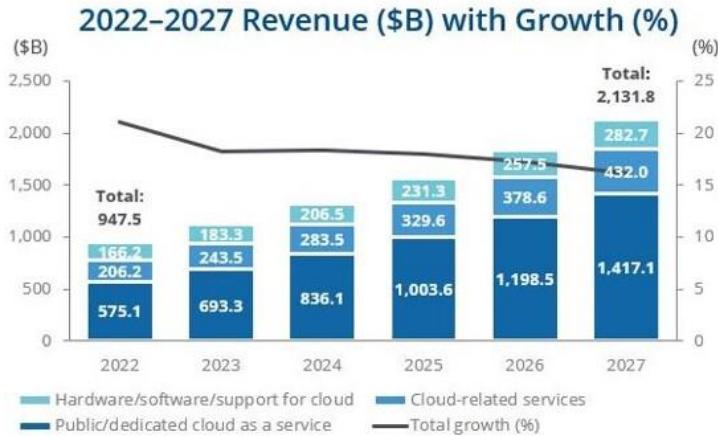
Worldwide Whole Cloud Revenue: IDC

2022–2027 Revenue (\$B) with Growth (%)



Source: IDC, Worldwide Whole Cloud Forecast, 2023–2027: Cloud Is the Critical Support Enabling the AI Everywhere Era, #US50435323

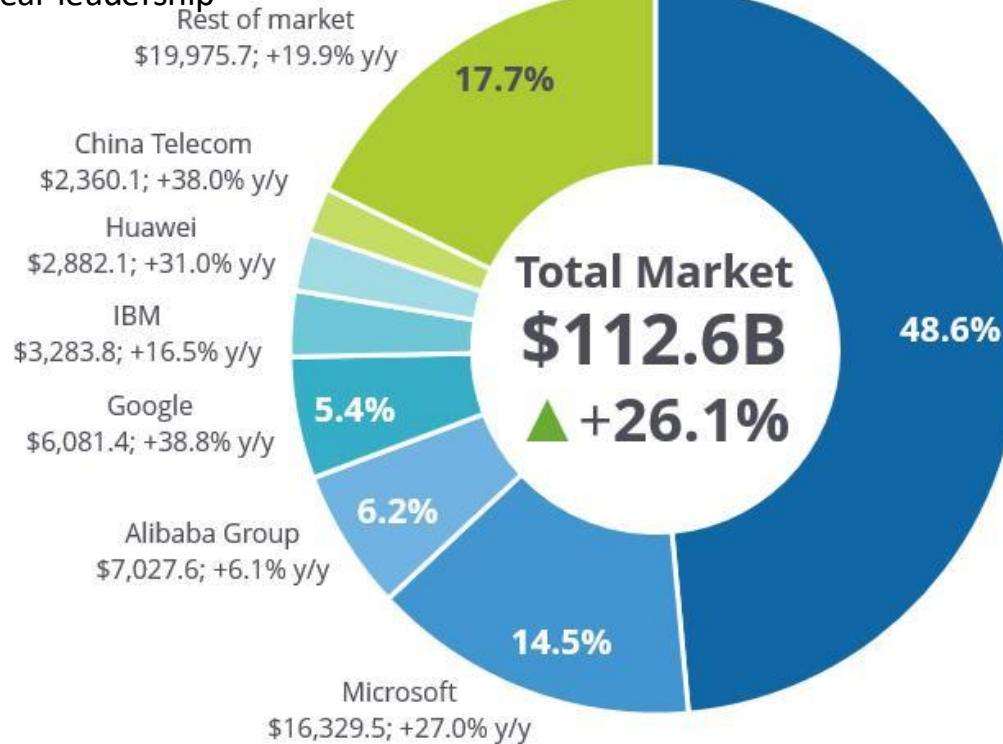
Worldwide Whole Cloud Revenue: IDC



- Cloud is a \$1T market.
- It's still growing at double digit annually (17.6%)
- But has slowed down from the dizzying 20-30% growth in previous years
- Major enterprises around the world have embraced cloud and moving workloads to cloud aggressively
- Cloud adoption is a secular trend (long term trend not affected by short term factors) with double digit growth for foreseeable future.

Market Share of Major Cloud Service Providers: IDC

AWS has clear leadership



Amazon Web Services
\$54,674.6; +30.1% y/y

Update: Based on latest results, AWS is headed for a \$110B annual revenue

- Source: IDC, Worldwide Public Cloud Infrastructure as a Service Market Shares, 2022: Investments in High-Performance Services Create Differentiation, #US51035622

IaaS, PaaS, SaaS

- Infrastructure as a Service (IaaS) is a type of cloud computing service in which the provider offers virtualized computing resources (such as servers, storage, and networking) over the internet. These resources can be used to build, deploy, and manage their applications and services.
- Platform as a Service (PaaS) is a type of cloud computing service in which the user gets access to a platform for developing, deploying, and managing applications. PaaS providers typically offer tools and services that help developers build, test, and deploy applications and also can manage the underlying infrastructure. PaaS has less management overhead than IaaS.
- Software as a Service (SaaS) is a type of cloud computing service in which the provider hosts software applications and makes them available to customers over the internet. SaaS applications are usually accessed through a web browser and users have to pay a subscription charge.

Market leaders

- AWS clear leader as Cloud Services Provider. AWS is also the clear leader from a technical standpoint introducing new services every month.
 - Microsoft is clear SaaS leader in market share terms, but mostly because of Office 365.
 - Smaller SaaS companies producing far more innovative products than big 3 (Amazon, Microsoft, Google)
 - Google holds enormous advantage because of its undisputed leadership in search, no one even close to Google on search.
- Managed services is a highly fragmented market. Leader Accenture only holds 11.8% market share.
 - Professional cloud managed services likely underestimated because of fragmentation
- SaaS growing faster than IaaS.

What are professional cloud managed services?

- At the most basic level this is professional services for:
 - Provisioning a service
 - Deprovisioning the service
 - Life cycle management of the service
- What is life cycle management of the service?
 - Managing all changes, problems and issue relating to the service.
 - Changes can be user driven, e.g. attaching a larger storage volume, or scaling the compute engine.
 - Changes can also be software driven, e.g. new software patch requires OS update.
 - Problems and issues arise during execution. E.g. Performance slowness or OS/System crash.

What has Cloud done for Software?

- Eliminated need for sending disks or downloading and installing software on the client (client = your laptop).
- Software can be accessed via web browsers and setting up an account.
- Freeing up tech companies from marketing and distribution costs and long timelines, has enabled a huge revolution in software.
- Really when a layman speaks of cloud, it usually means storage more than compute. E.g. “My photos are in the cloud.”
- An entire generation has now grown up that does not know how to store songs on their device. Why ipods are no longer popular.

Technical trends on Cloud

- Containerization: lightweight virtual machines for hosting applications. Open Source.
- Kubernetes: Container orchestration. Open source. (What is orchestration? It means how to monitor and manage cloud resources)
- Serverless: Milliseconds of computing, often at the edge. No infra to manage and pay on the cloud. Ideal for microservices. But we are still trying to harness the power of serverless.
- Infrastructure as code: Define the end state of your resources and tooling will produce the required deployed resources. Terraform is the key player.
- Specialized security products as hackers become more sophisticated.

Technical trends on Cloud

- Microservices: An approach to software development that breaks up a large application into smaller independent software services that communicate with each other through APIs. (API is Application Programming Interface. It is a set of rules and protocols that software must use to communicate with other software services).
- FinOps: Cloud sprawl brings major cost challenges. Enterprises are not seeing the financial benefits of moving to cloud because of cloud sprawl – decentralized provisioning and management.
- Data Engineering. Newer and more sophisticated tools to manage large streams of data, perform real time analytics, and create real time actions. E.g. video surveillance in sensitive areas.
- Advent of no-code, low code services, citizen developer services. AWS abandoned Honeycode unexpectedly in 2023. But keep watching this space for more.

When to remain on the datacenter vs cloud?

- Clients always ask this question
- Intensive, mission critical and predictable should remain in virtualized datacenters (example core banking)
- Workloads that need to scale or delivered at the edge should move to cloud
- Workloads that are constantly changing should move to cloud
- Development should be done on cloud

What is a workload?

It is the demand placed in IT systems and resources such as CPU and networking. It is interchangeably used with applications or a set of applications that fulfills certain enterprise functions, example a database that holds inventory for the enterprise, or an application that provides HR data to the entire company.

Useful links

- The Challenge of the Computer Utility, Douglas F Parkhill, 1966
- Wikipedia: https://en.wikipedia.org/wiki/Cloud_computing
- Demystifying Cloud Computing: [LINK](#)
- Above the Clouds: A Berkeley View of Cloud Computing: [LINK](#)
- Cloud Tutorial: [LINK](#)
- Cloud Computing Misc Presentations: [LINK](#)
- Google Cloud: <https://cloud.google.com/>
- Cloud Computing Primer for 2016: [LINK](#)

Next steps

- Cloud Computing Class 412 covers the basics of cloud services and gets hands-on