ITCS443 Parallel and Distributed Systems

Cloud Computing and Virtualization

Sudsanguan Ngamsuriyaroj

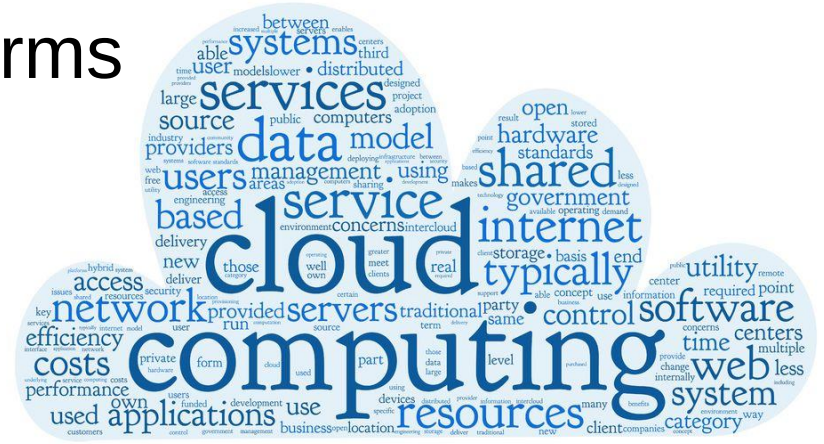
Ekasit Kijsipongse

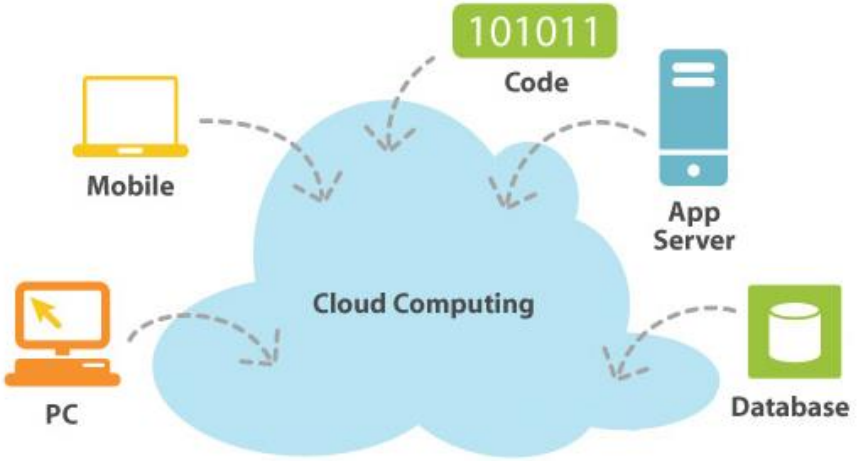
Ittipon Rassameeroj

Semester 1/2022

Outline

- Evolution of Computing Platforms
- NIST Cloud Framework
- Cloud Service Models
- Cloud Deployment Models
- Scaling
- Multitenant Technology
- Roles in Cloud
- Cloud Adoption Model
- Virtualization Concepts





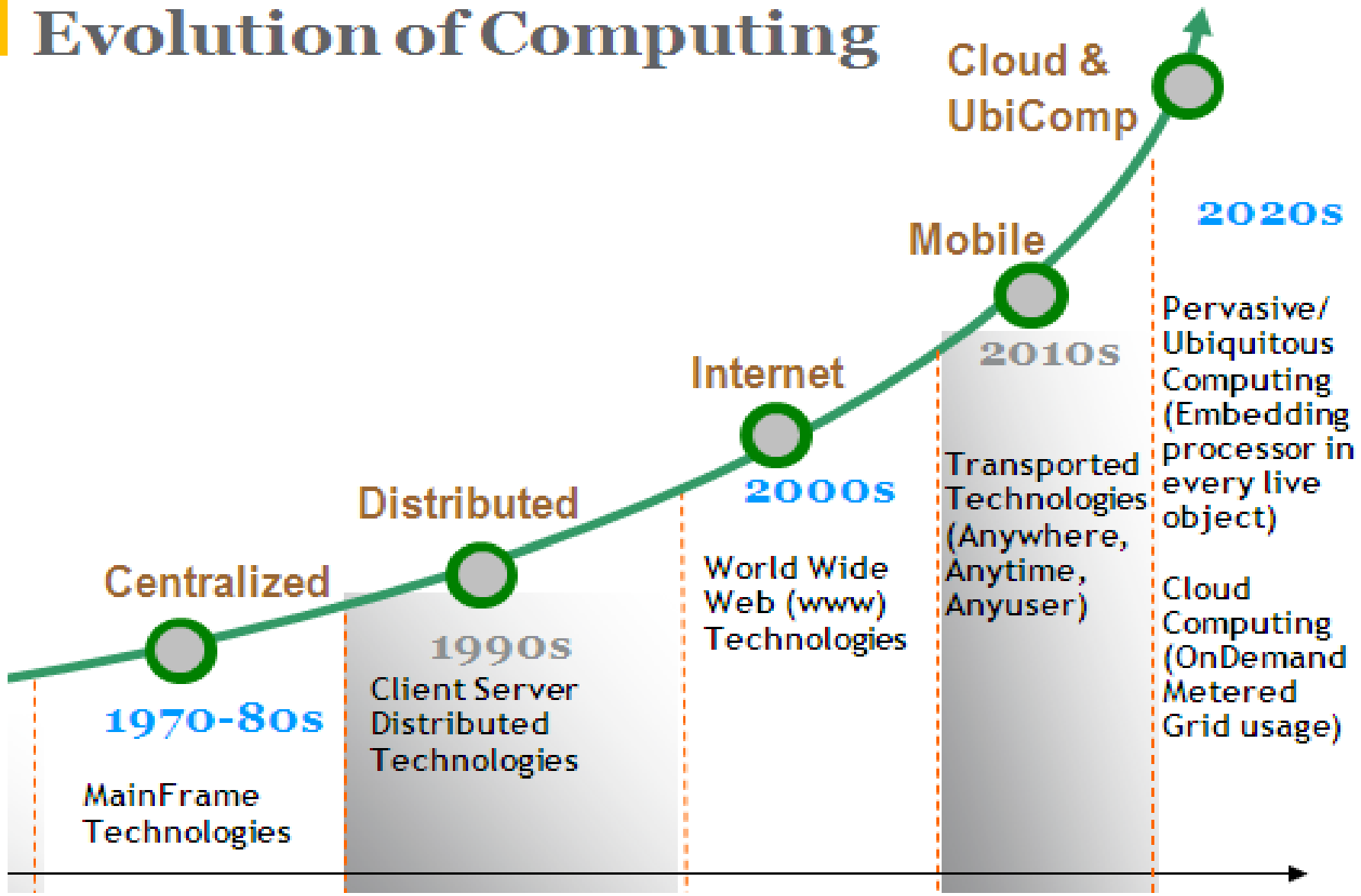
Evolution of Computing Platforms

70's – 80's
Mainframe Era

90's-2000's
Client Server Era

2010-beyond
Cloud Era

Evolution of Computing Platforms

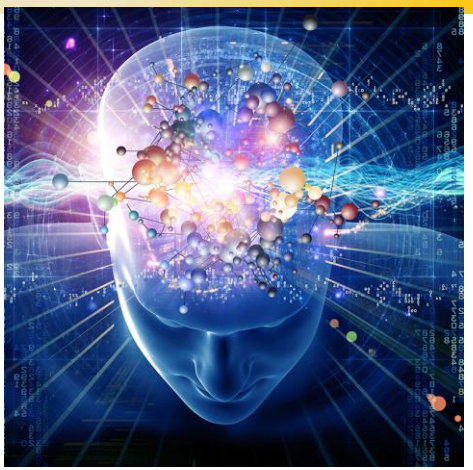


Latest Evolution

Cloud Computing

Big Data

Machine Intelligence



Internet of Things

Mobile Computing

Social Networks

Internet of Things

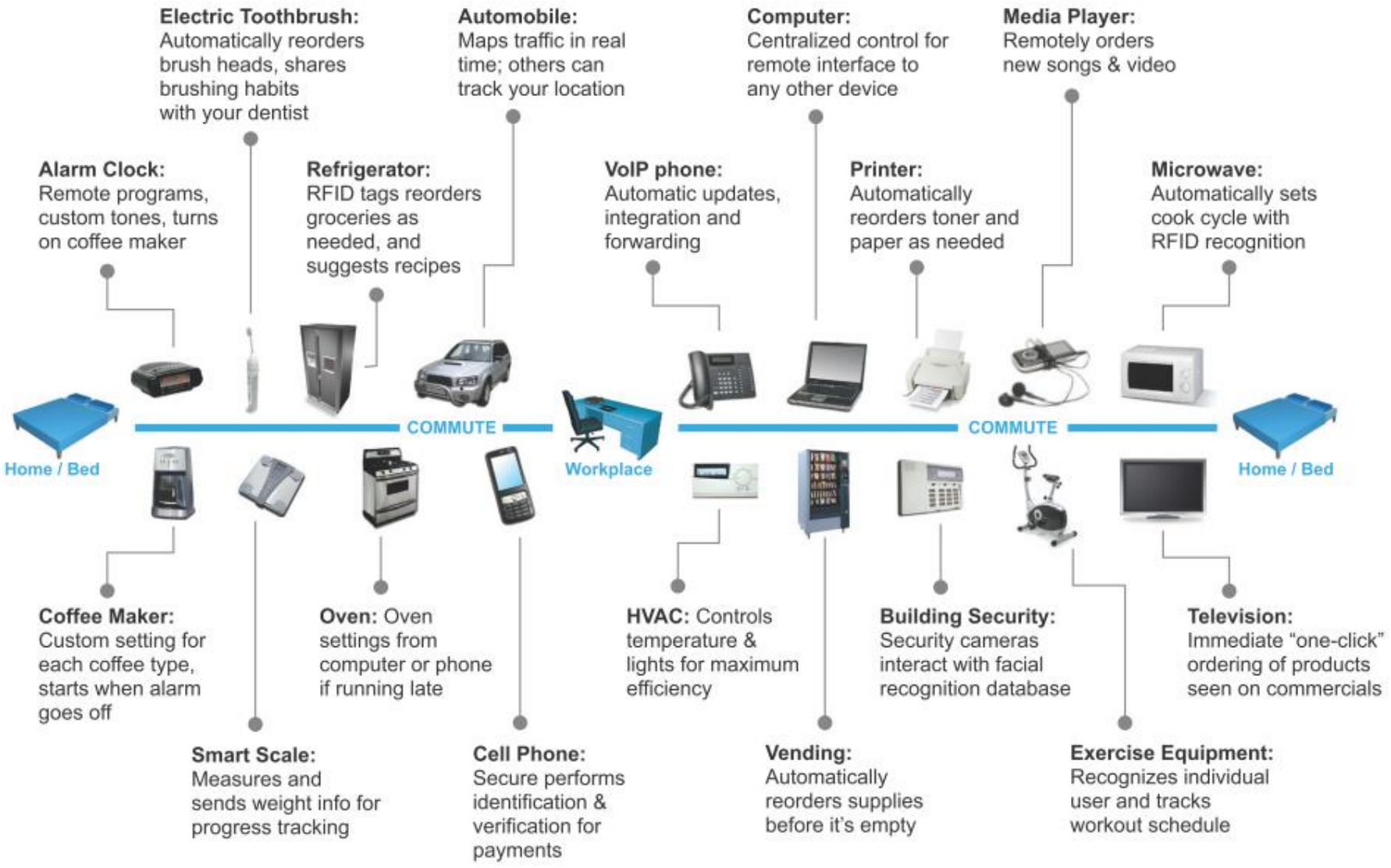
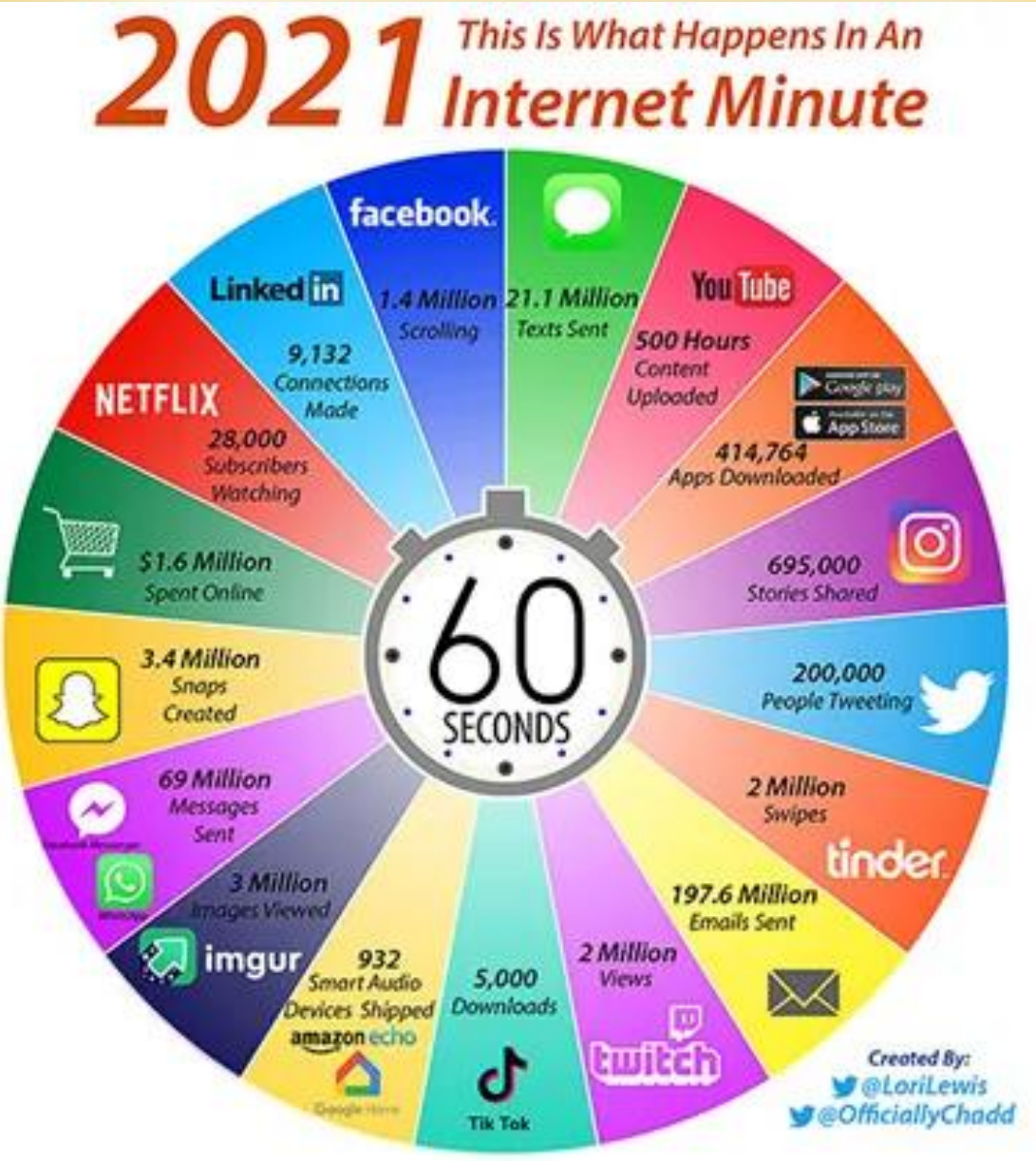


Figure 3. The Internet of Things

Big Data: From Biology to Astronomy

Data are created every minute!



Top Technology Trends 2022

Top Strategic Technology Trends for 2022

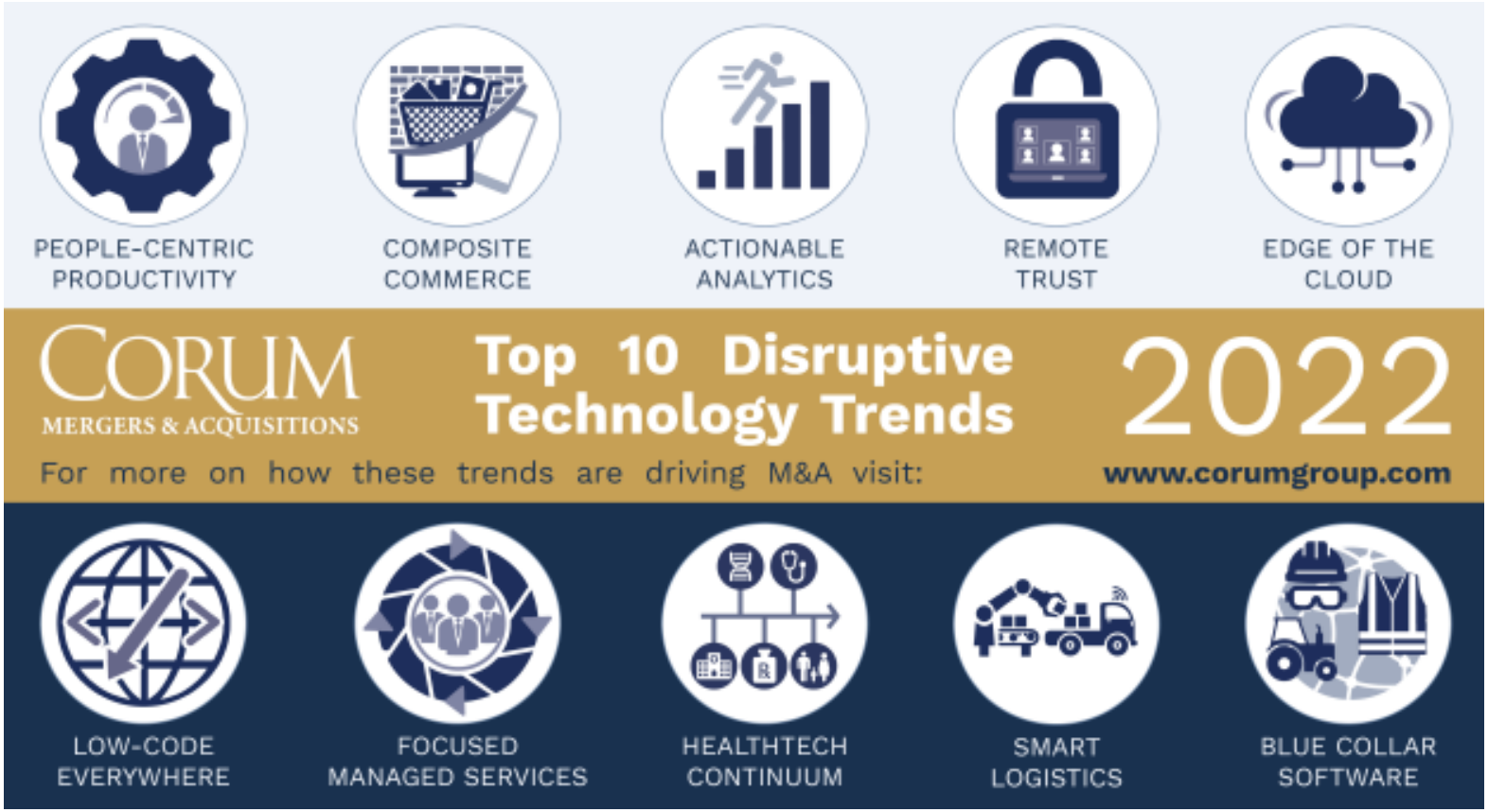
Accelerating Growth	Sculpting Change	Engineering Trust
• Generative AI	• AI Engineering	• Cloud-Native Platforms
• Autonomic Systems	• Hyperautomation	• Privacy-Enhancing Computation
• Total Experience	• Decision Intelligence	• Cybersecurity Mesh
• Distributed Enterprise	• Composable Applications	• Data Fabric

Source: Gartner
757234_C

Gartner

https://www.gartner.com/en/information-technology/insights/top-technology-trends

Corum Top Ten Disruptive Technology Trends



Data Centers

Building a Data Center

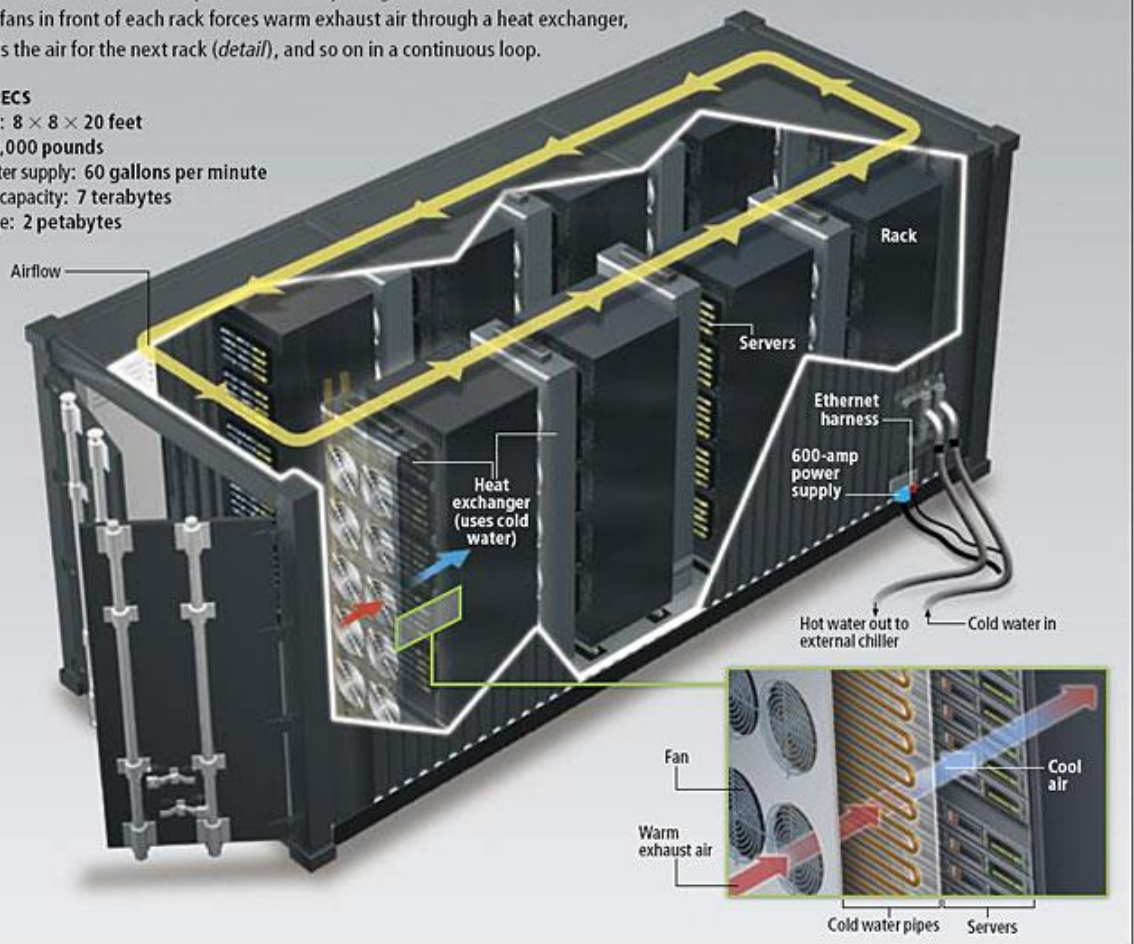
Sun's Modular Data Center: Project Blackbox

[HOW IT WORKS]

KEEPING COMPUTERS COOL

Inside Project Blackbox, racks of up to 38 servers apiece generate tremendous heat. A panel of fans in front of each rack forces warm exhaust air through a heat exchanger, which cools the air for the next rack (*detail*), and so on in a continuous loop.

DESIGN SPECS
Dimensions: 8 × 8 × 20 feet
Weight: 20,000 pounds
Cooling water supply: 60 gallons per minute
Computing capacity: 7 terabytes
Data storage: 2 petabytes



How Large of a Data Center?

Facts and Stats of World's largest data centers

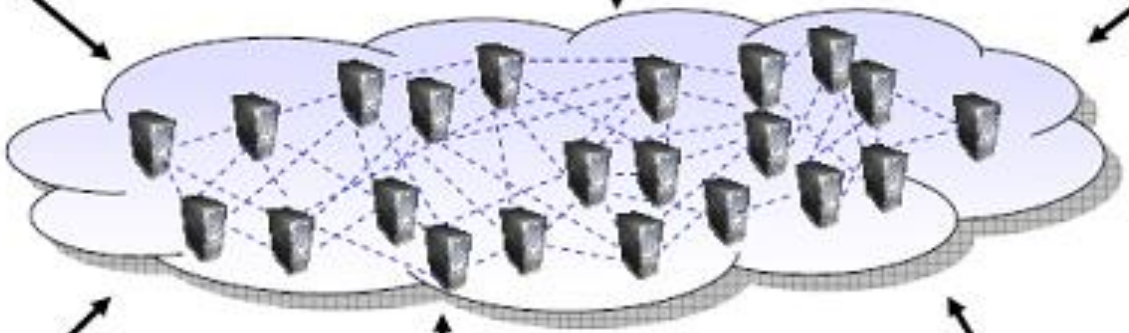
- Google 16 data centers: 9 in U.S., 1 in South America, 4 in Europe and 2 in Asia with about 2.5 million servers
- Facebook:
  - 1.13 billion daily active users (as of September 2016)
  - 60,000 servers (2010)
  - Own about $3.63 billion in "network equipment" (end of 2015)
- Amazon:
  - Amazon has 11 cloud regions
  - **1.5 million servers** in 7 locations
- Microsoft:
  - 1 billion users and 1,000,000 servers
  - Microsoft has 17 regions.
- Rackspace Hosting has 100,000 servers across 6 data centers.

Forces Driving Cloud Computing

Data-Intensive Applications:
- Explosion of applications and user-generated content:
- Exabyte in 2006
- Zettabyte in 2010

Datacenter Pressures:
Growing operational complexity and cost from infrastructure and application sprawls

Increased network capacity and availability



Innovation and Collaboration

Rising Energy Cost & Green compliance

Shared Services Across Lines of Business

Cloud Services

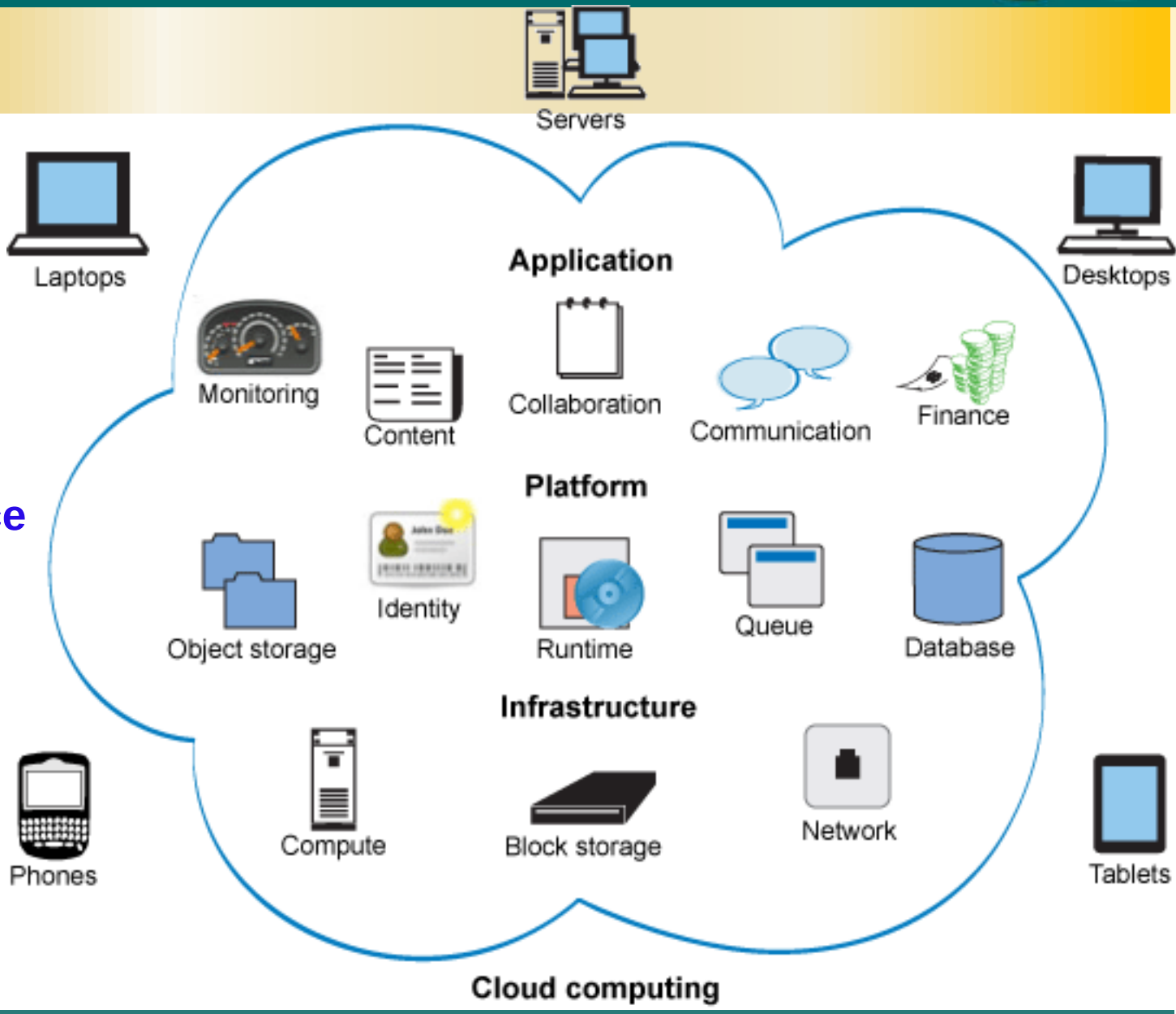
Service Catalogs

Self-service provisioning

A Big Picture of Cloud

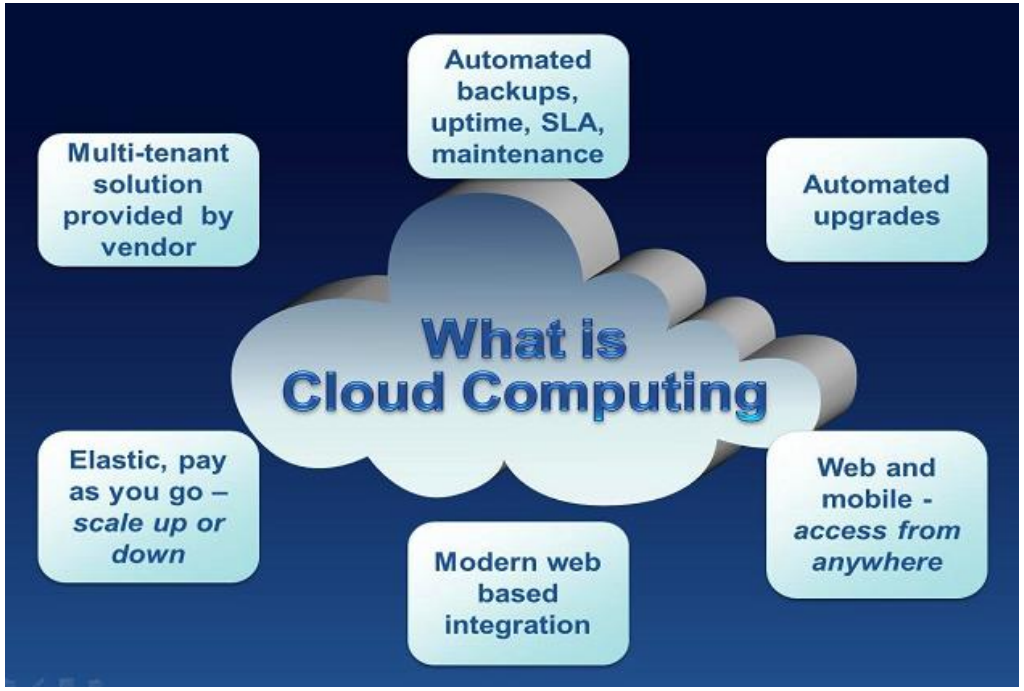
Goals

- **Always online**
- **Anywhere, any time, any device**
- **Unlimited resources**



NIST Cloud Definition

Cloud computing is a *pay-per-use* model for enabling *ubiquitous, convenient, on-demand access* to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction



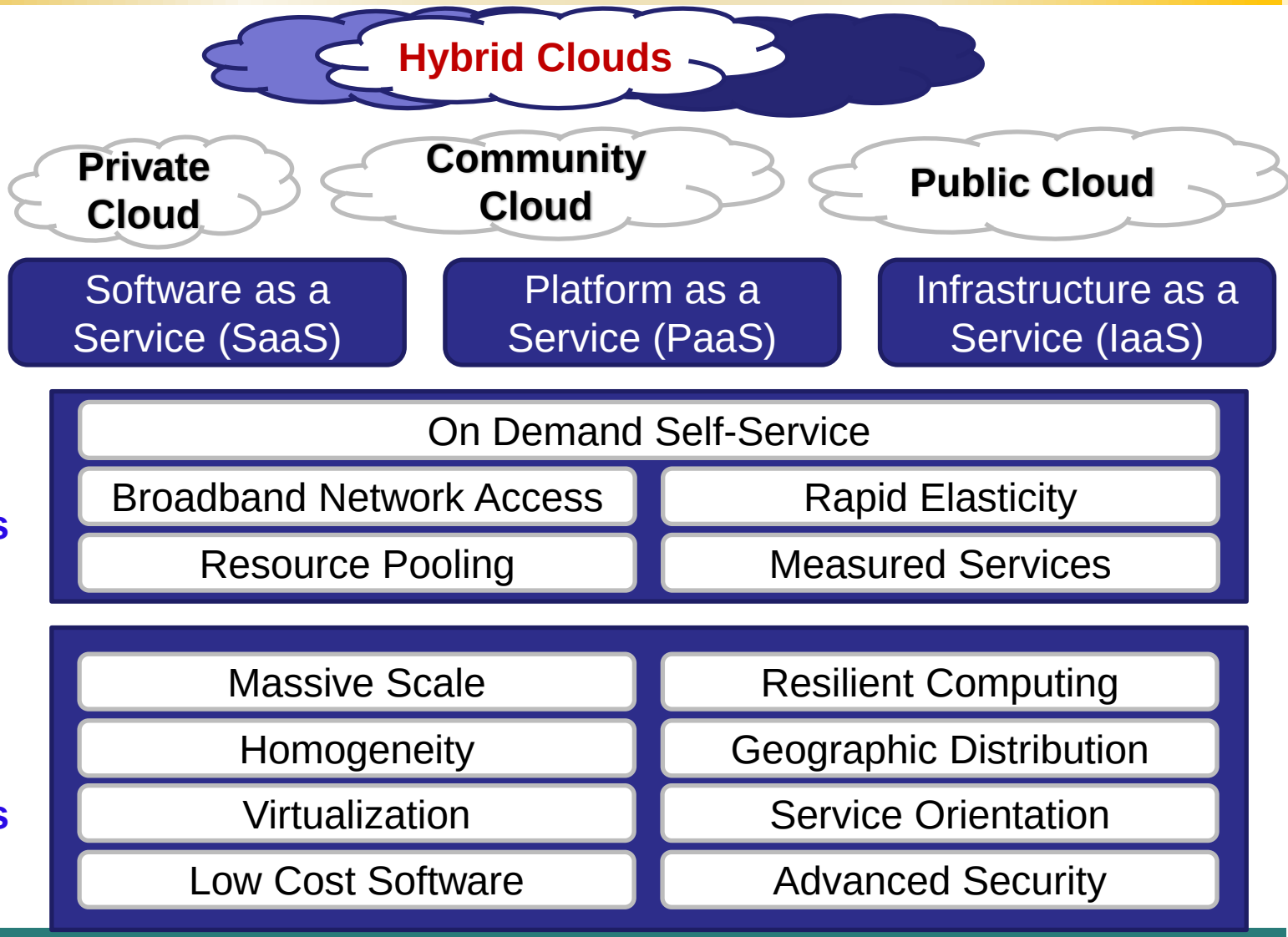
5 essential characteristics

3 service models

4 deployment models

National Institute of Standards and Technology (NIST)

NIST Cloud Framework



NIST Cloud Framework

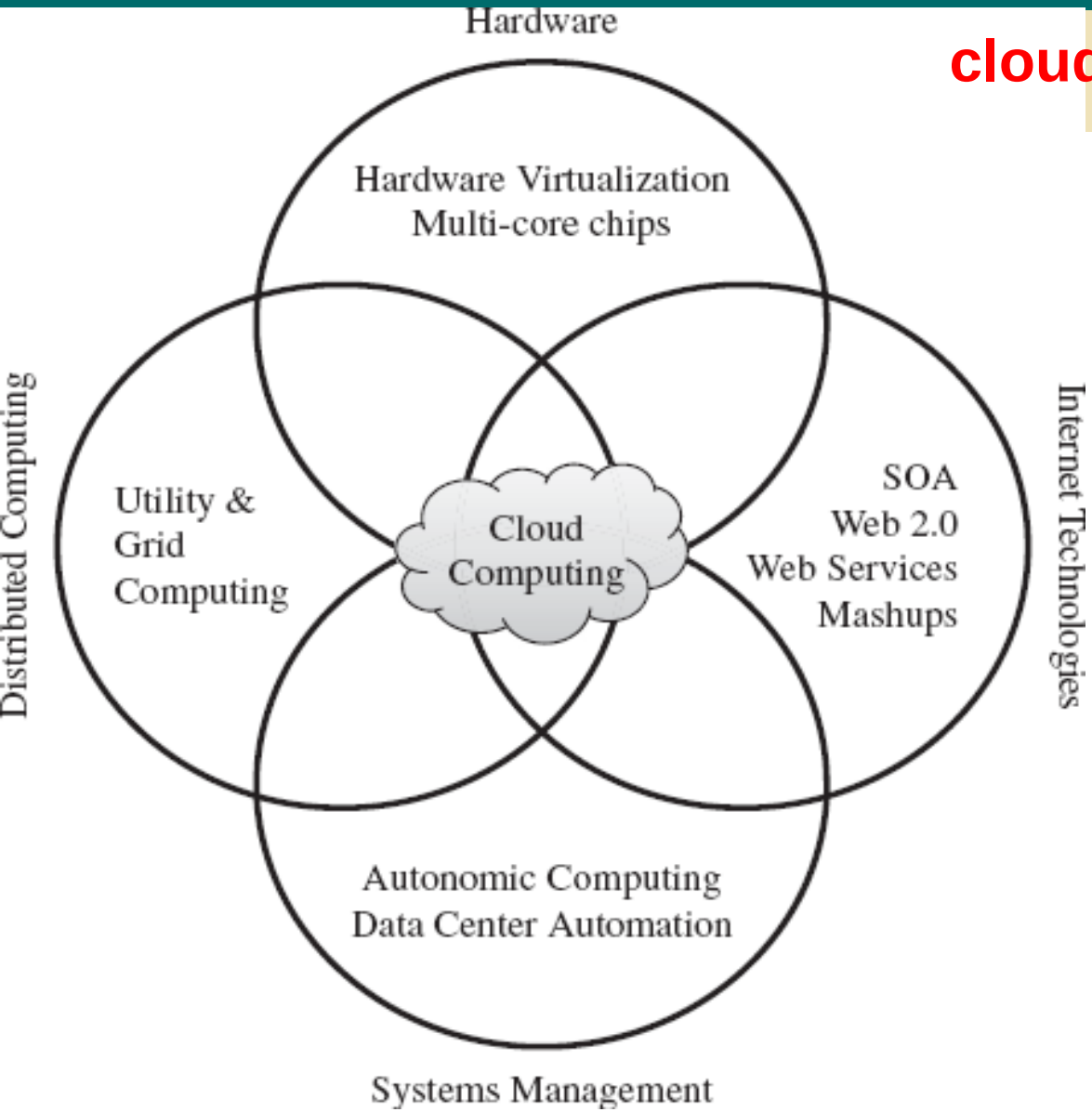
5 Essential Cloud Characteristics	
On-demand self-services	Users can provision servers and networks with little human intervention
Broadband Network access	Any computing capabilities are available over the network, and can be accessed thru many different devices
Resource pooling	Multiple users can access clouds that serve other consumers according to demand. Location independence.
Elasticity	Automated ability of a cloud to transparently scale IT resources. Provisioning is rapid and scales out or based on needs
Measured services	Ability to keep track of the usage of its IT resources Just like utilities paid by hours or transactions

NIST Cloud Framework

Additional Cloud Characteristics	
Multi-tenancy	• Enable an instance of the program to serve different consumers (tenants) whereby each is isolated from the others • Resource pooling allows cloud providers to pool large-scale IT resources to serve multiple cloud consumers
Resiliency	A form of failover that distributes redundant implementations of IT resources across physical locations

Clouds optimize resource usage and control it according to SLA (Service Level Agreement)

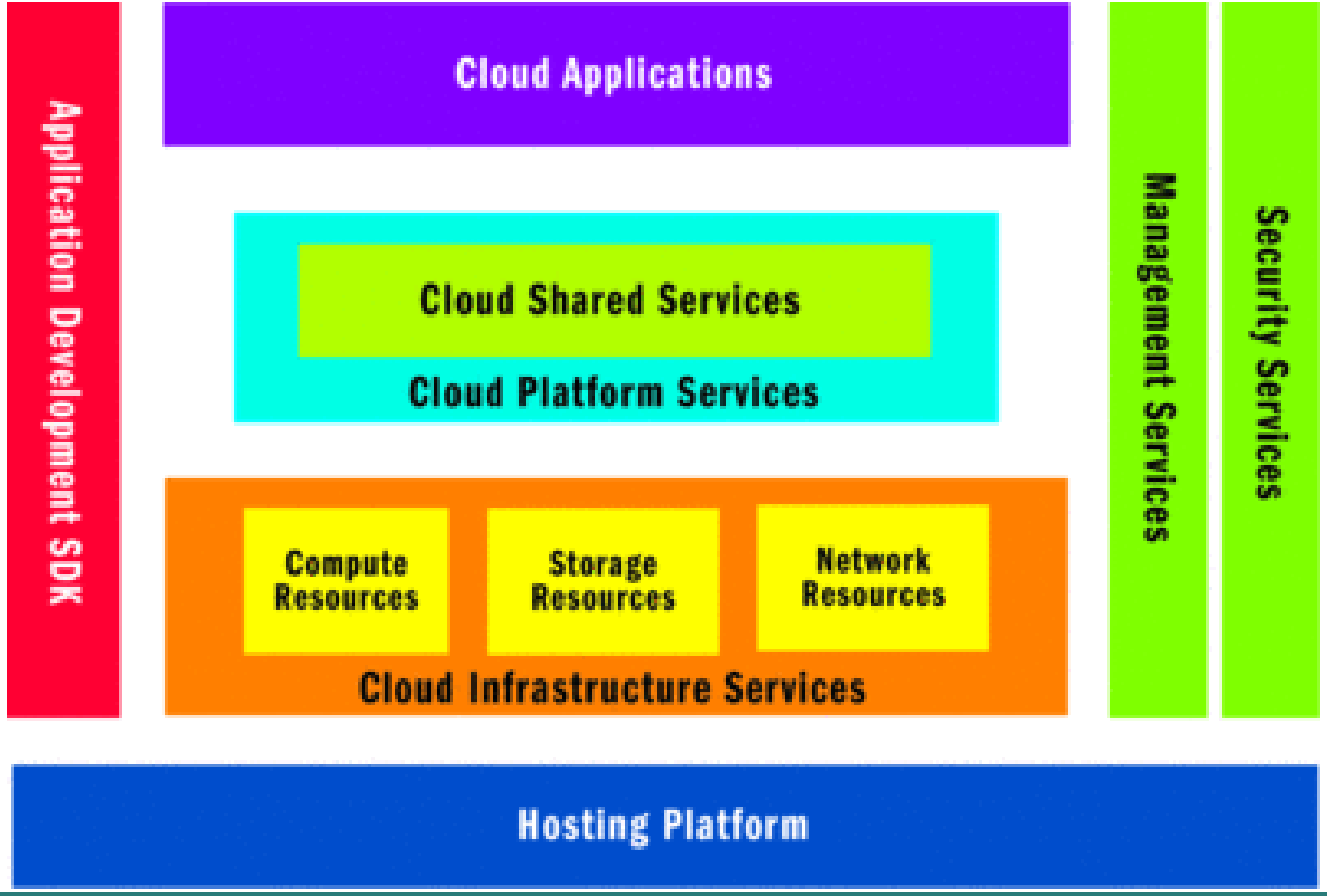
What Technologies Leads to Clouds



cloud-enabling technologies

- **Broadband Networks**
- **Internet Architecture**
- **Data Center Technology**
- **Virtualization Technology**
- **Web Technology**
- **Multitenant Technology**
- **Service Technology**

Cloud Computing Platform



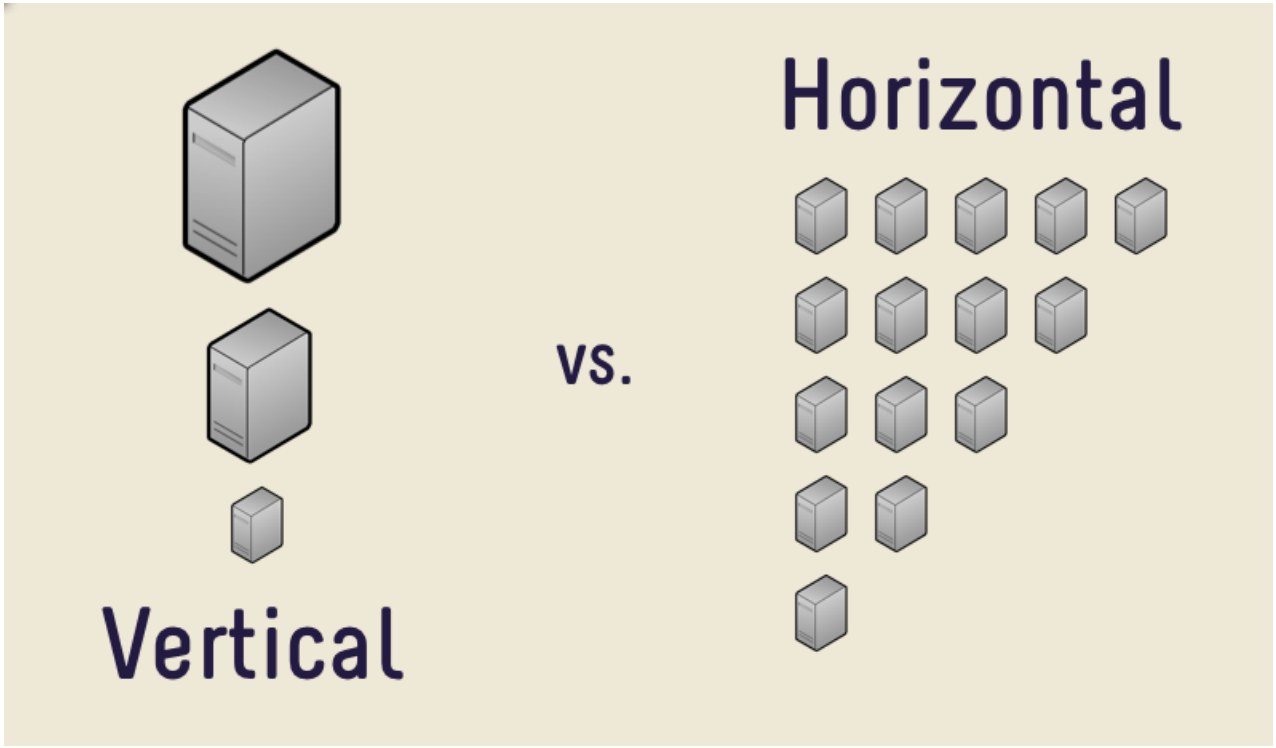
Vertical vs. Horizontal Scaling

Scale up/down vs. Scale out/in

Ability of IT resources to handle increased / decreased demands

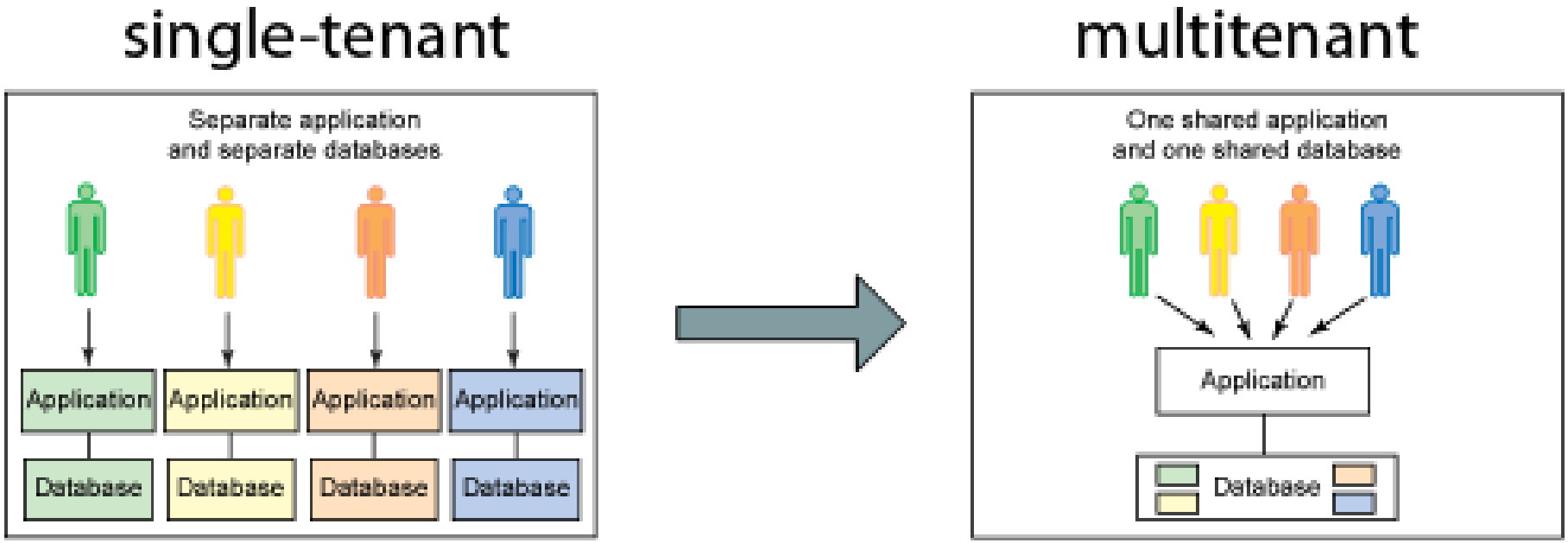
Horizontal Scaling: allocating or releasing IT resources of the same type

Vertical Scaling: replacing IT resources with higher or lower capacity

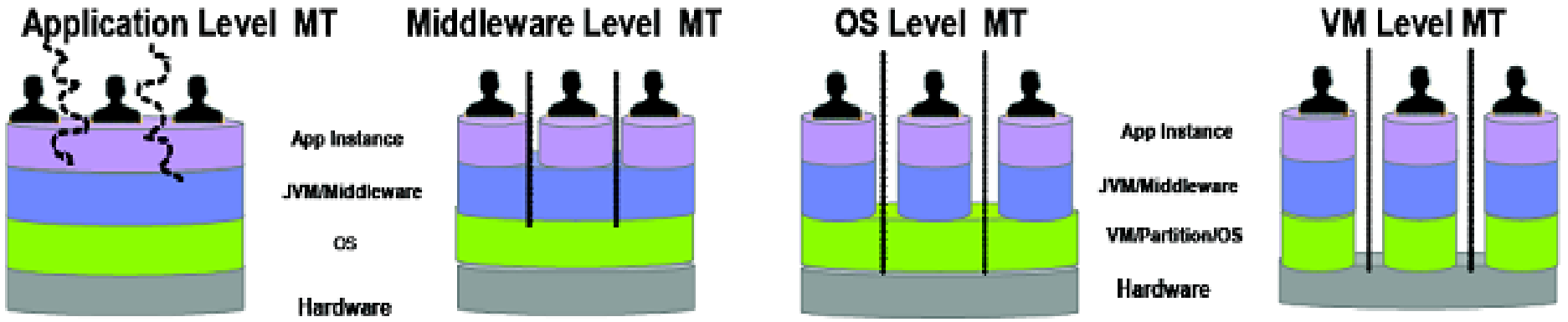


Multitenant Technology

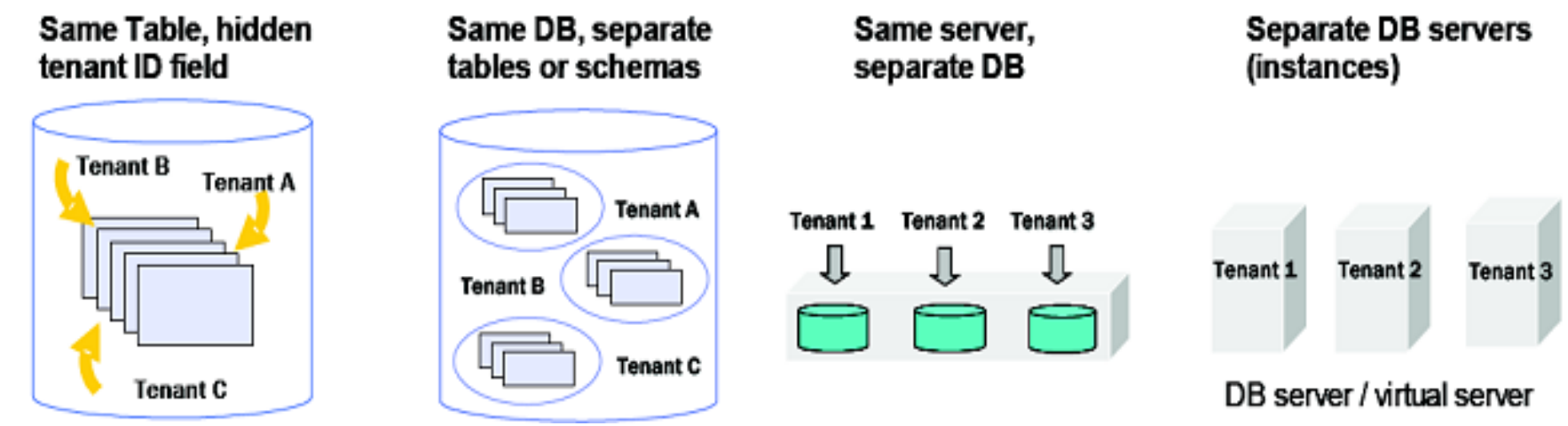
- **Single-tenancy**: each customer has their own software instance
- **Multi-tenancy**: multiple customers (tenants) share the **same application**, running on the same OS, same hardware, with same data-storage mechanism
- **Benefits: isolation, cost saving, improved security**



Multi-tenancy at Different Levels



Example: Database Multi-tenancy



Cloud Service Models

SaaS

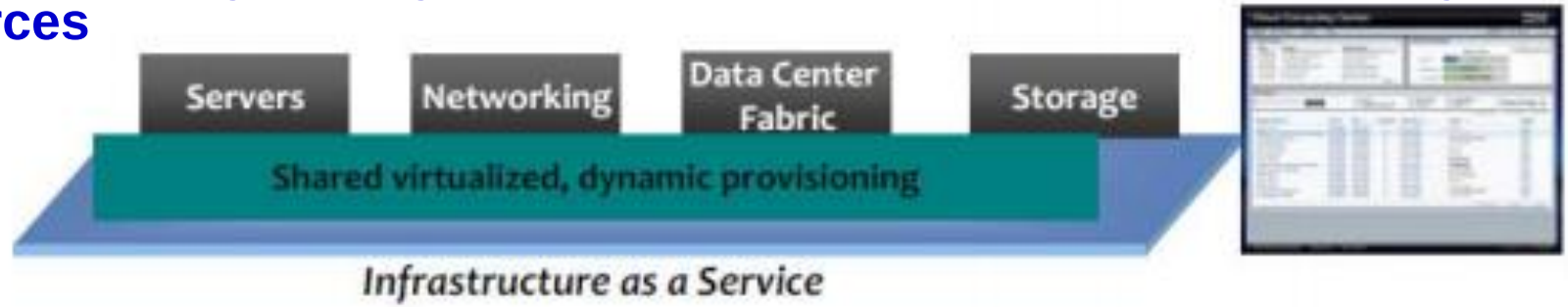
Use provider's applications over a network

PaaS

Deploy customer created applications to a cloud

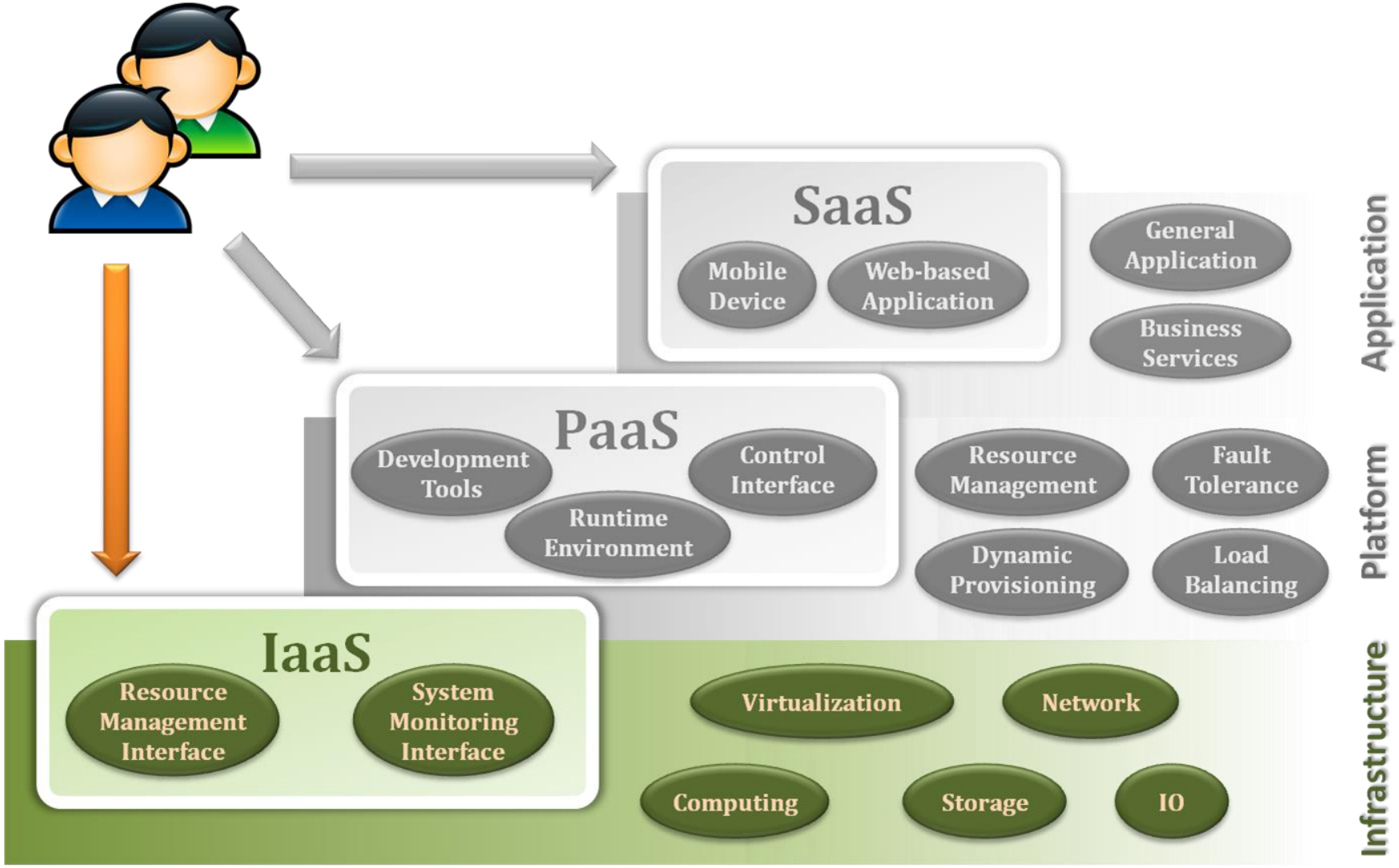
IaaS

Rent processing, storage, network capacity, and other computing resources



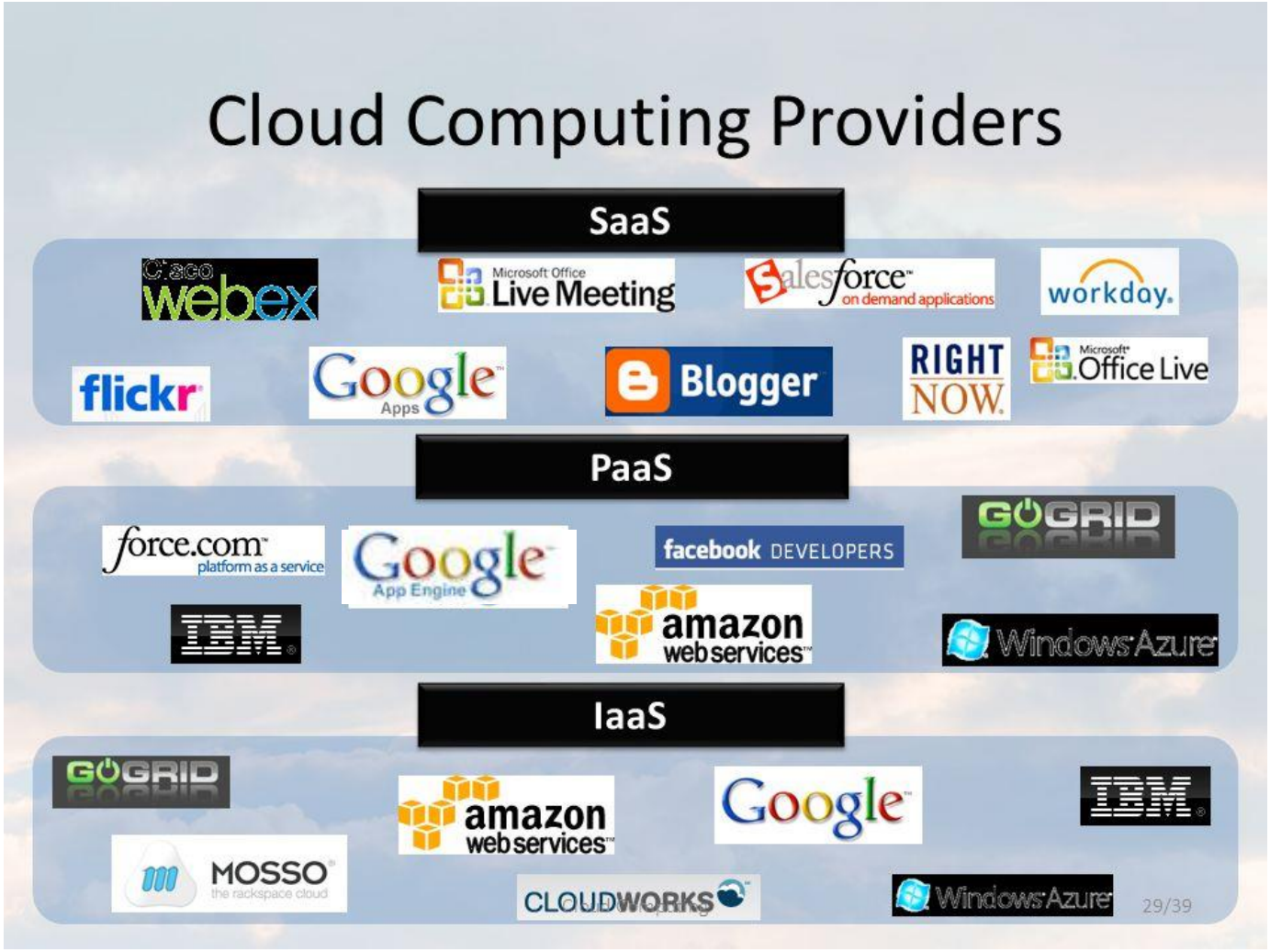
Cloud Service Models

Service Features



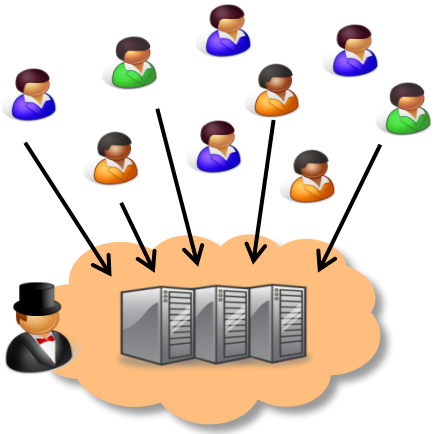
Cloud Service Models

Who are at which service?

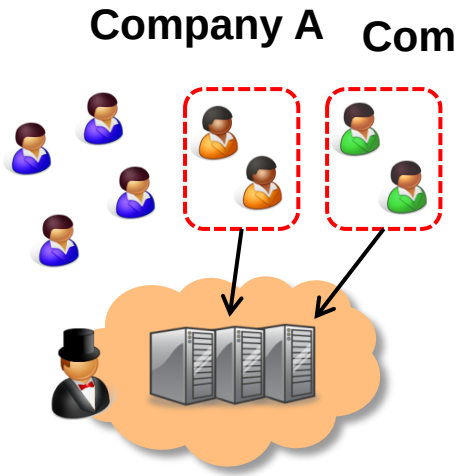


Cloud Deployment Models

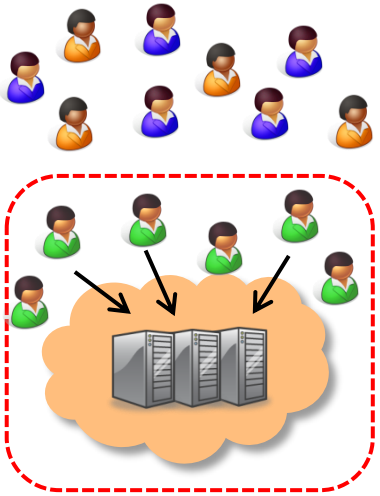
Who is a customer of the cloud?



Public



Community



Private

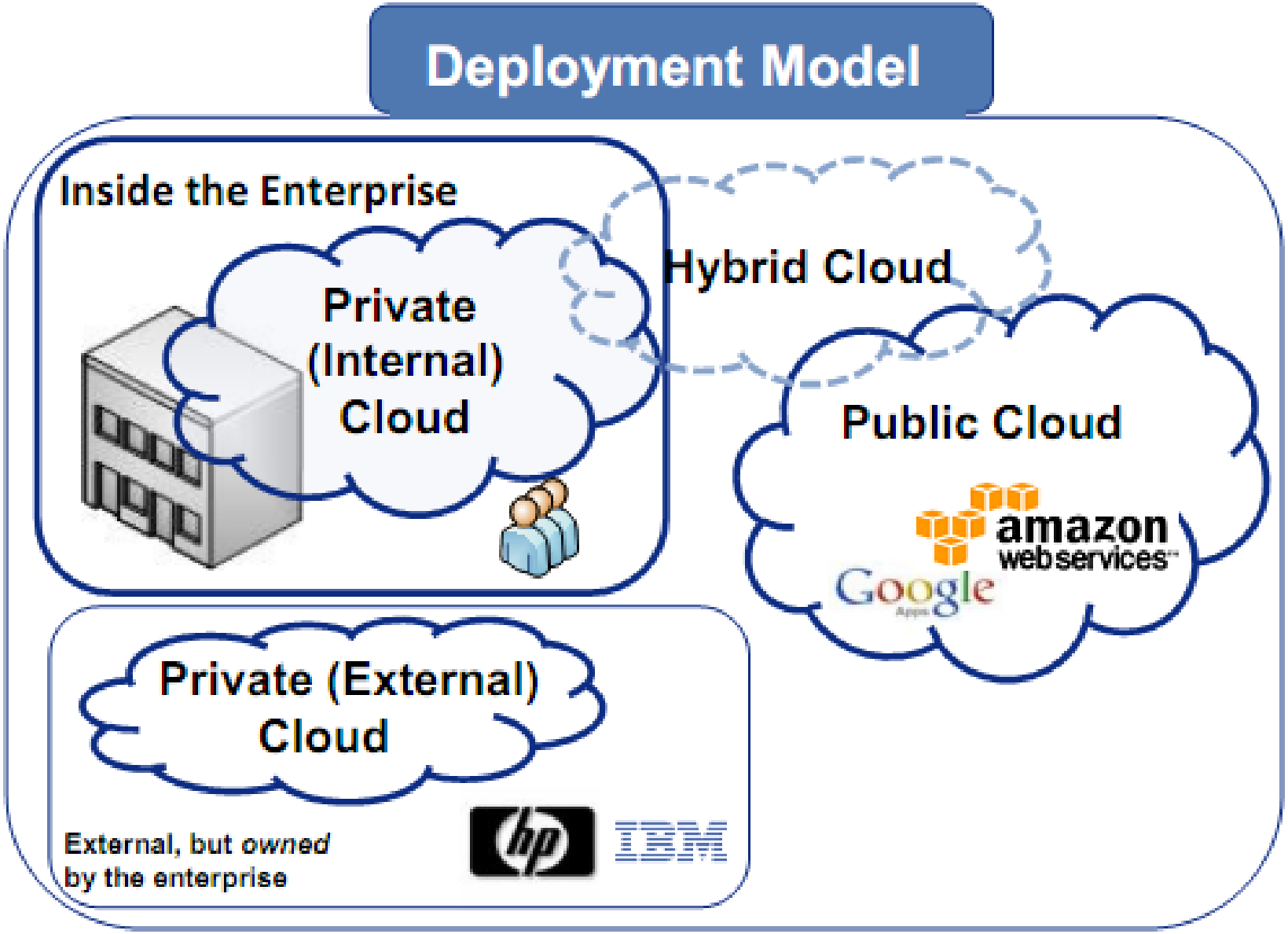
Public cloud: Amazon, Microsoft, Google, IBM

Community cloud: Shared by several similar organizations

Private cloud: Shared within a single organization

Hybrid cloud: Join between private and public cloud

Cloud Deployment Models

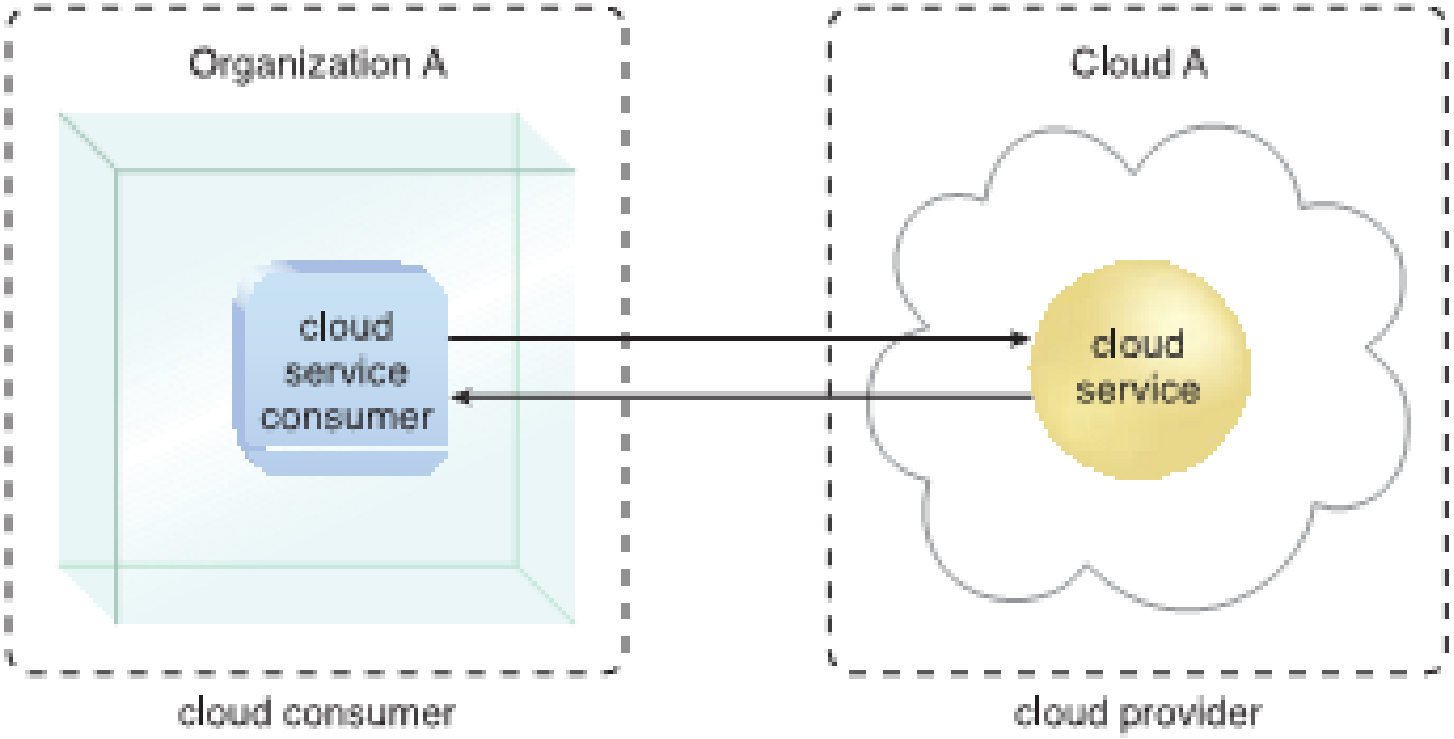


Roles in Cloud

Cloud Provider vs. Cloud Consumer

Cloud provider provides cloud-based IT resources

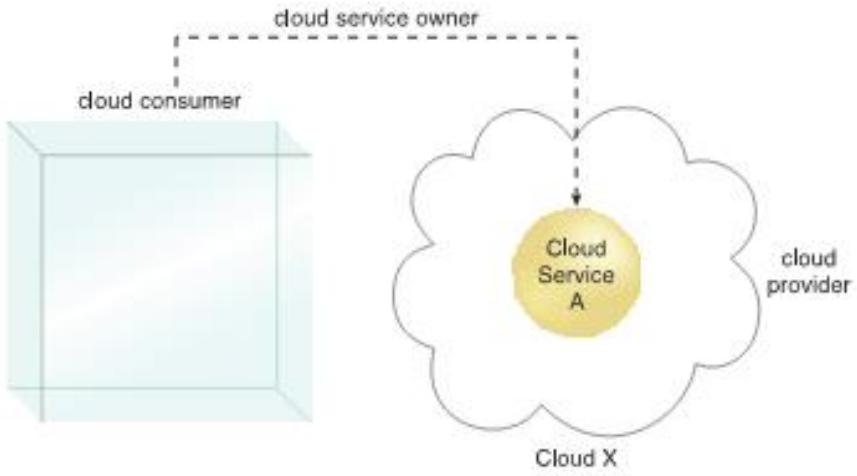
Cloud consumer uses IT resources made available by cloud providers

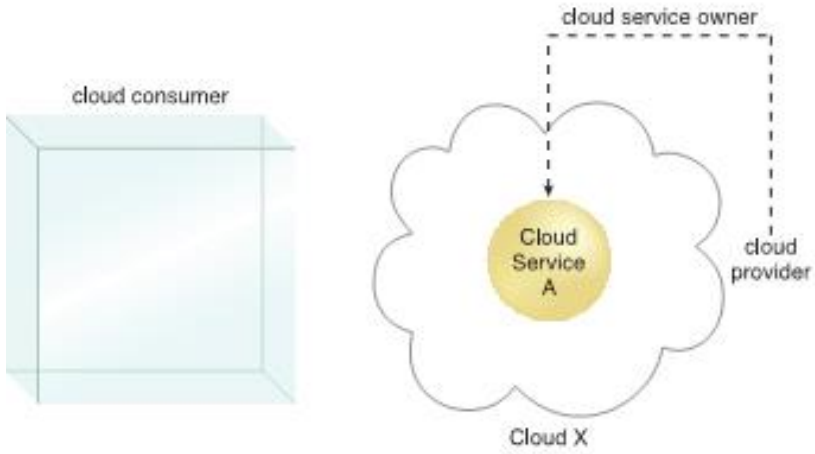


Roles in Cloud

Cloud Service Owner

A person/organization who legally owns a cloud service

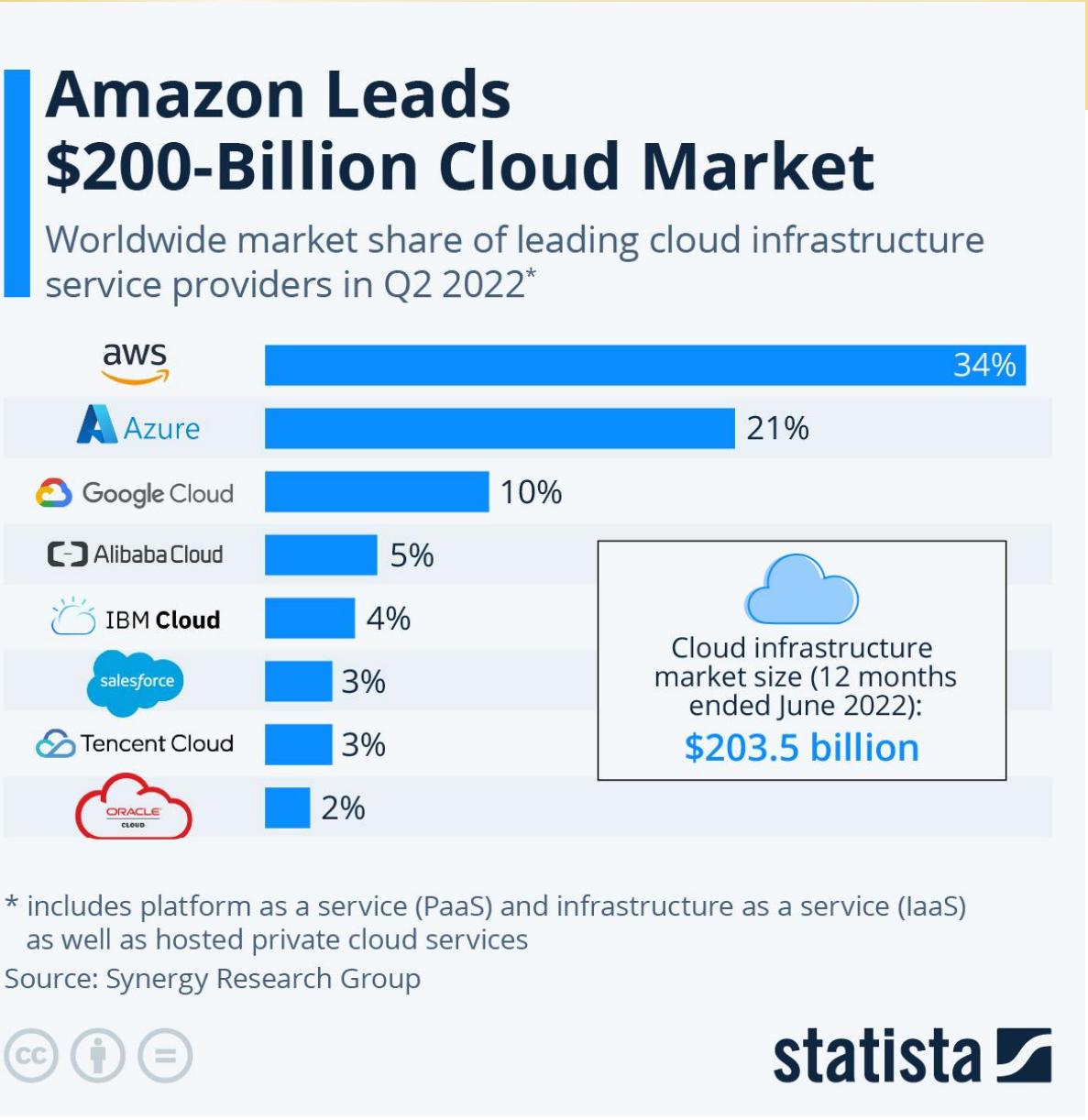




A cloud consumer can be a cloud service owner when it deploys its own service in a cloud

A cloud provider becomes a cloud service owner if it deploys its own cloud service, typically for other cloud consumers to use

Cloud Market Share

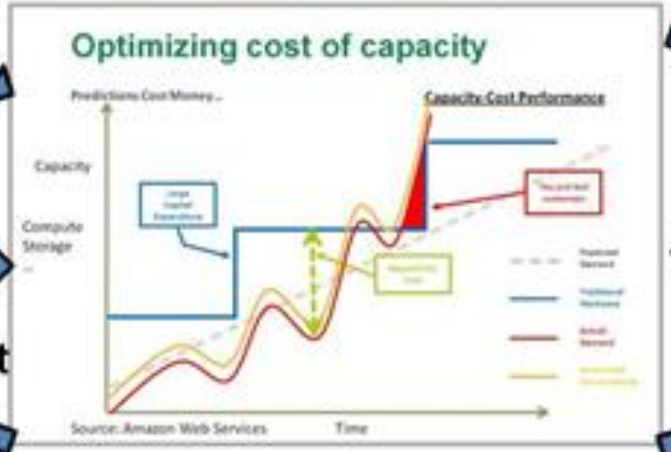


Cloud Capacity and Utilization

Motivation

Increase Margin
(Make more money)

Speed of Cost reduction
- Cost of adoption / de-adoption



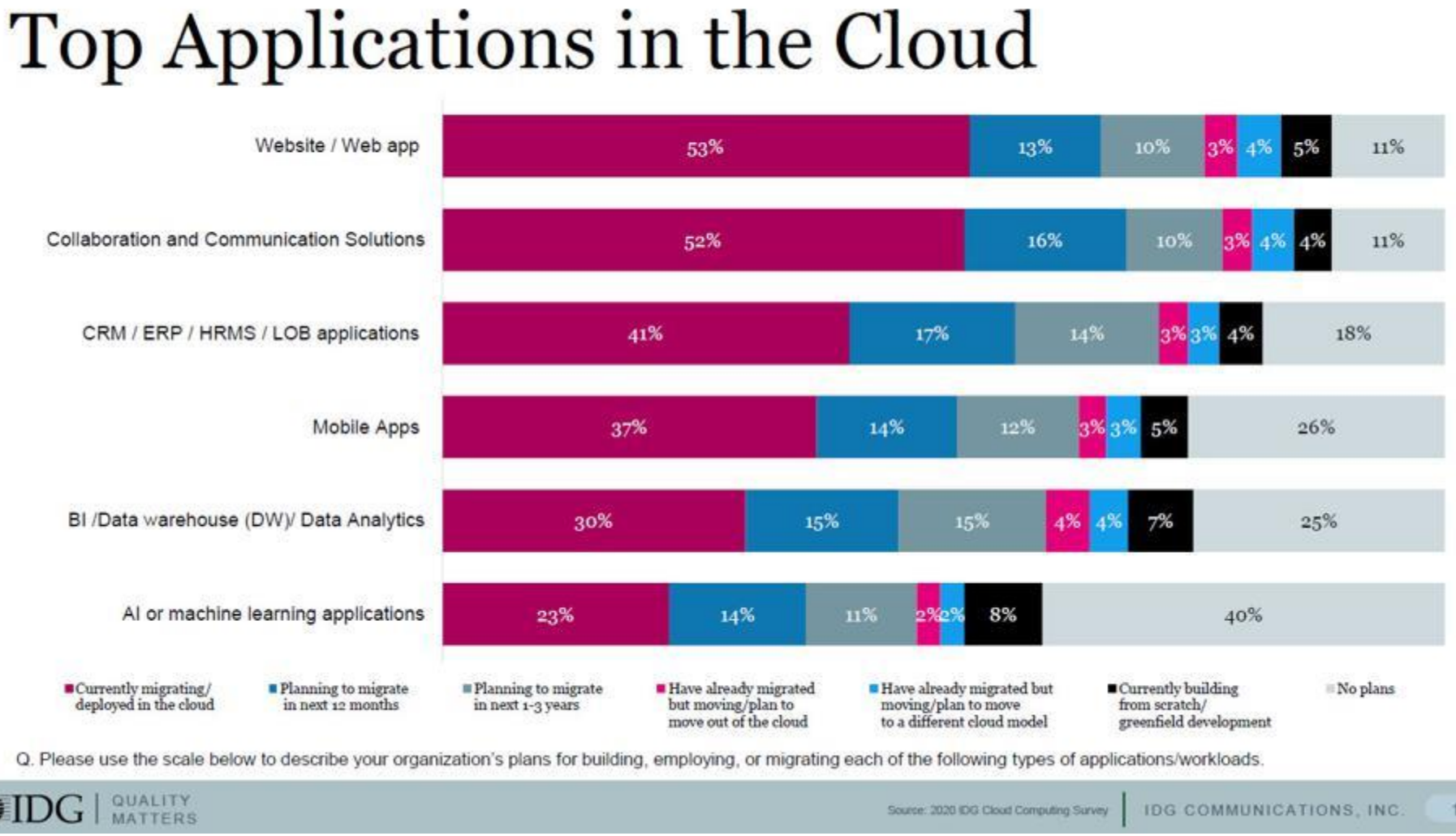
Dynamic usage
- Elastic
provisioning &
service management

Optimize
Ownership
use - faster
access to new
assets,
capabilities

Risk and Compliance
Improvement

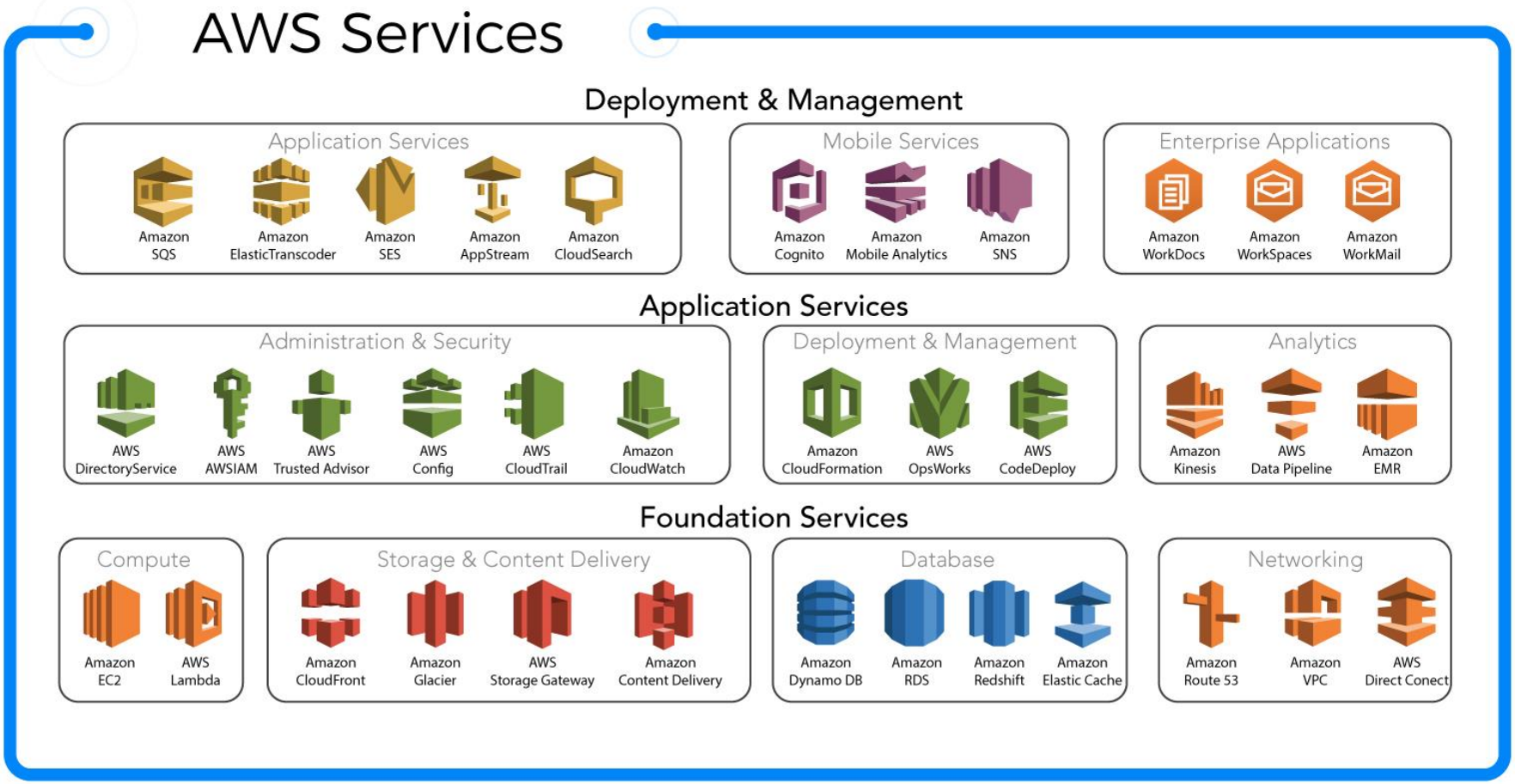
Rapid Provisioning

Top Cloud Applications

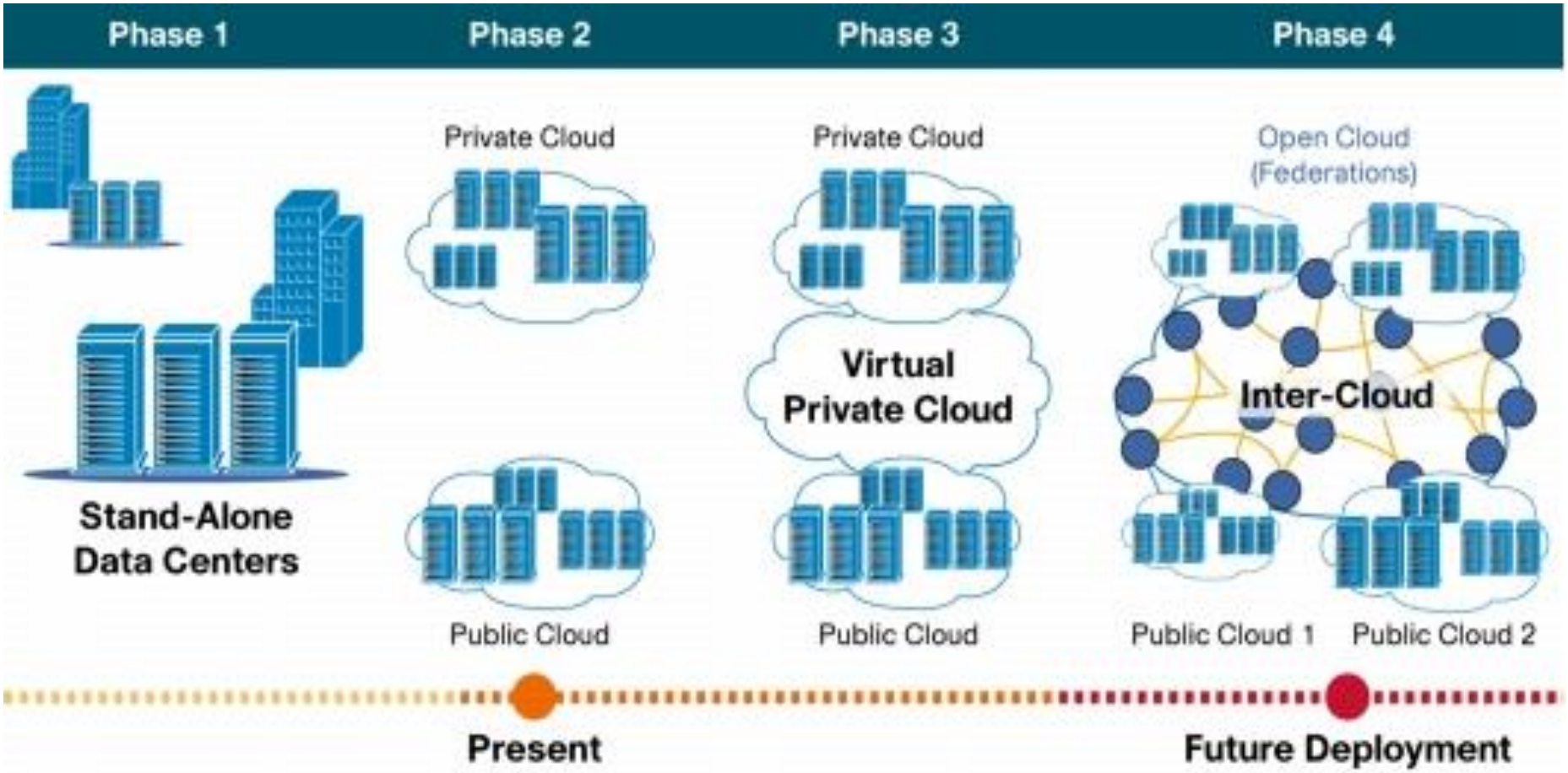


https://lvivity.com/digital-transformation-statistics

Example of AWS Cloud Services

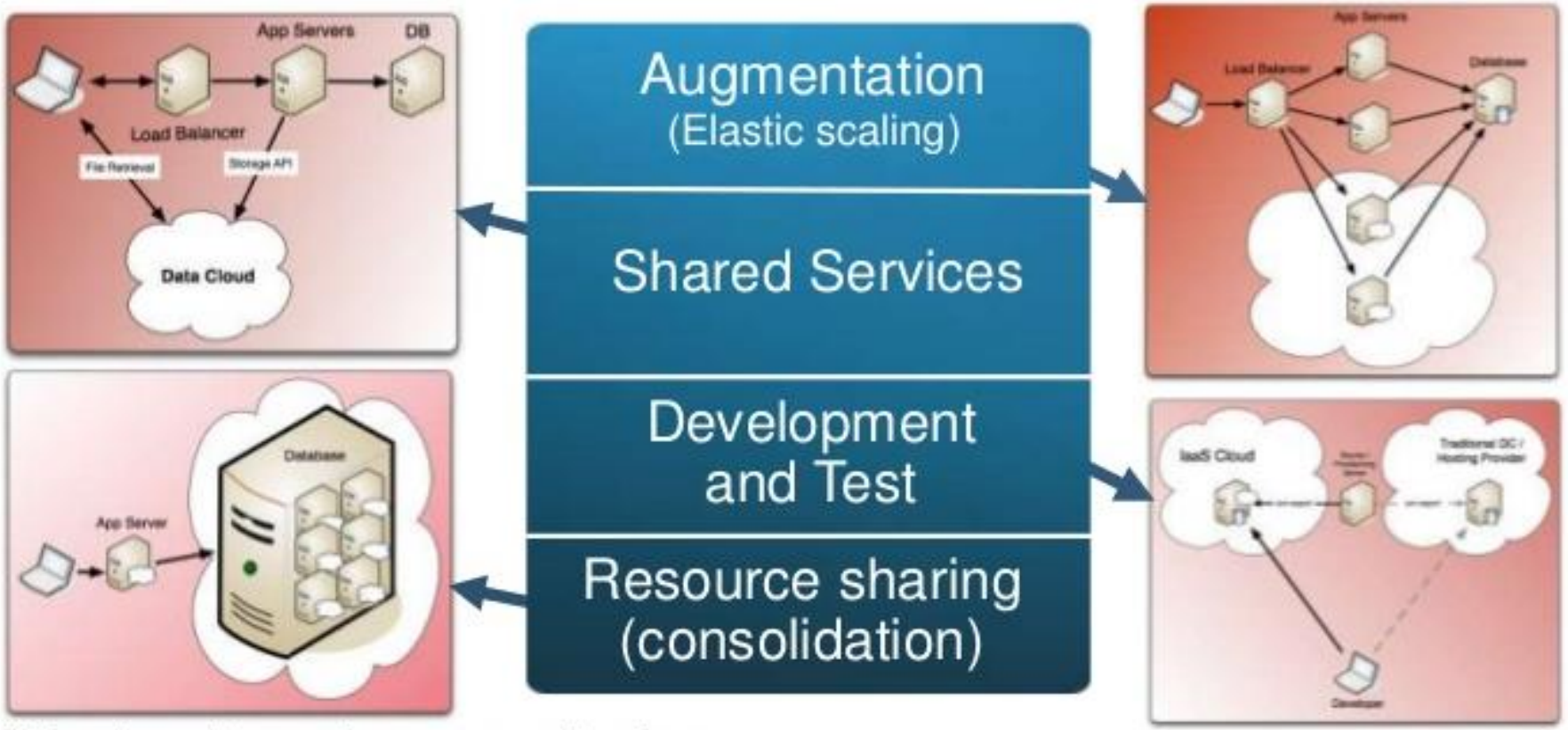


Cloud Adoption Model



Cloud Adoption Model

Starting with Common Use Cases

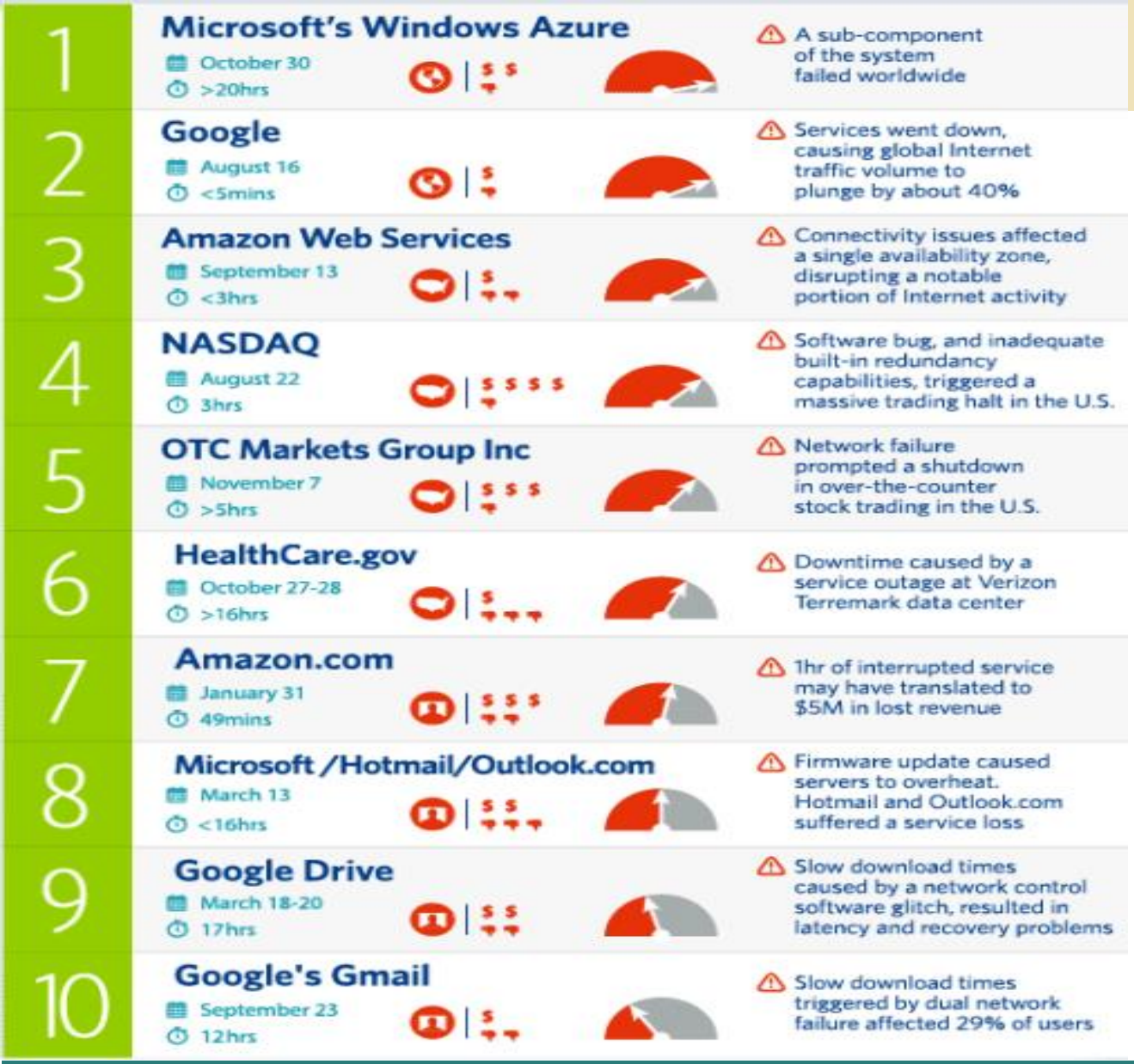


Most enterprises are trying

- **Shared development and test environments**
- **Hardware & Services consolidation**

Cloud: Pros and Cons

Cloud Downtime



Top Ten Cloud Computing Outages in 2013

Sample SLAs

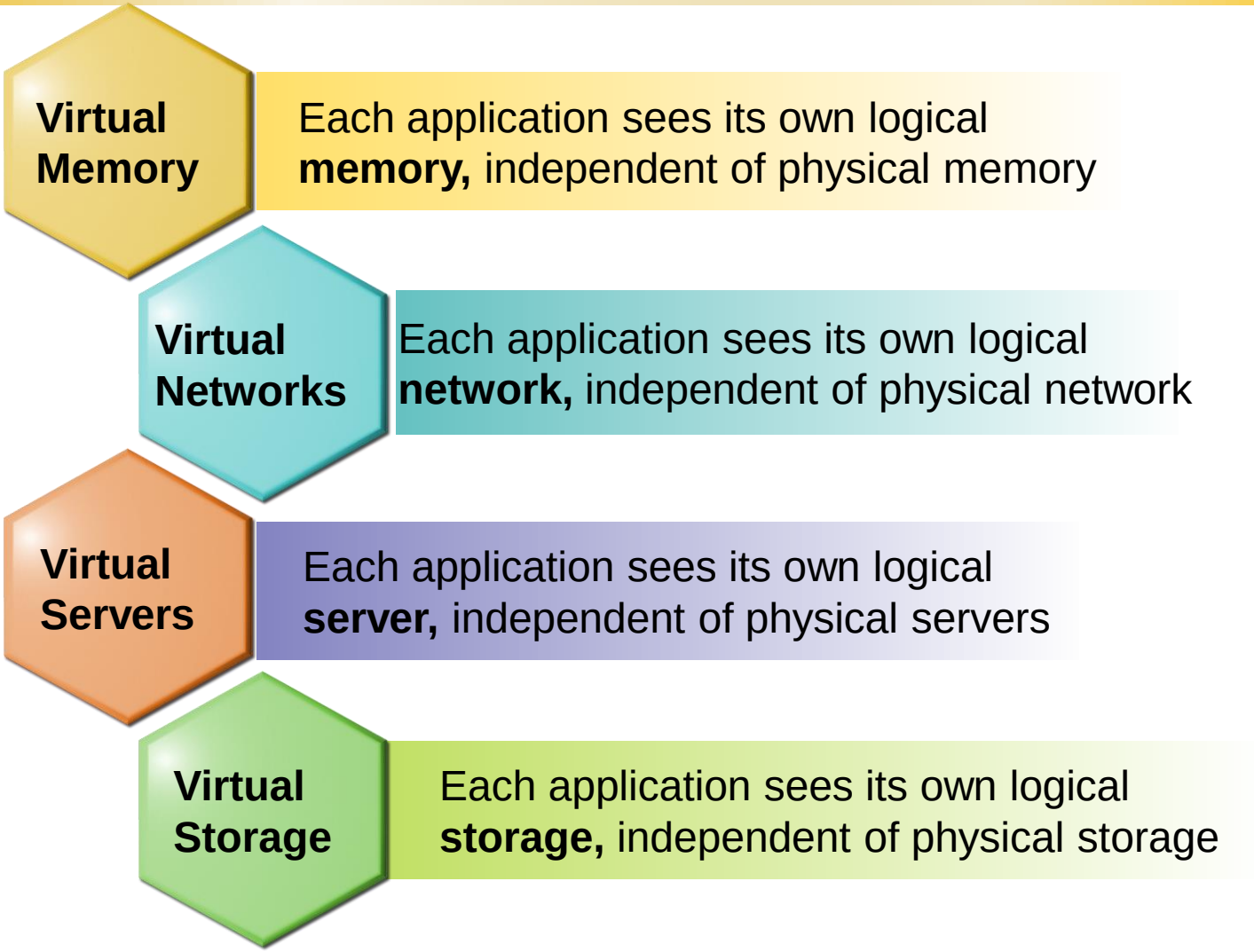
	Rackspace	Amazon
Uptime	100%	99.95%
Time-to-resolve	1 hour	Not specified
Timespan	Current period	"Service Year"
Credit back when failing to meet SLA	"Users receive 5% of the fees for each 30 minutes of network downtime, up to 100% of the fees"	"If the Annual Uptime Percentage for a customer drops below 99.95% for the *Service Year*, that customer is eligible to receive a Service Credit equal to 10% of their bill."

99.999% vs 99.9% uptime

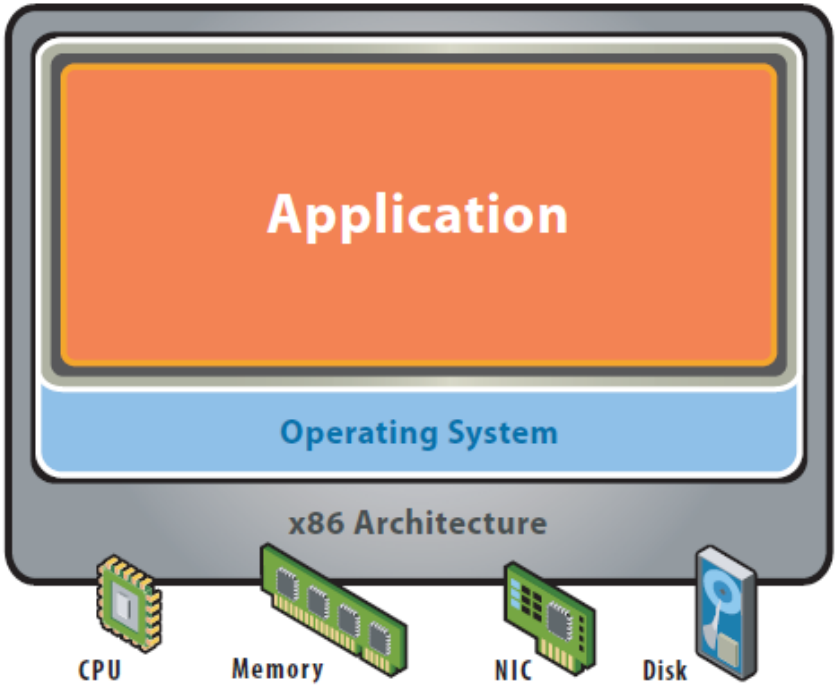
Five nine reliability

Availability	Downtime / Year	Downtime / Month	Downtime / Week	Downtime / Day
99.999%	5.256 Mins	0.438 Mins	0.101 Mins	0.014 Secs
99.995%	26.28 Mins	2.19 Mins	0.505 Mins	0.072 Secs
99.990%	52.56 Mins	4.38 Mins	1.011 Mins	0.144 Secs
99.950%	4.38 Hrs	21.9 Mins	5.054 Mins	0.72 Secs
99.900%	8.76 Hrs	43.8 Mins	10.108 Mins	1.44 Mins
99.500%	43.8 Hrs	3.65 Hrs	50.538 Mins	7.2 Mins
99.250%	65.7 Hrs	5.475 Hrs	75.808 Mins	10.8 Mins
99.000%	87.6 Hrs	7.3 Hrs	101.077 Mins	14.4 Mins

Virtualization Concepts



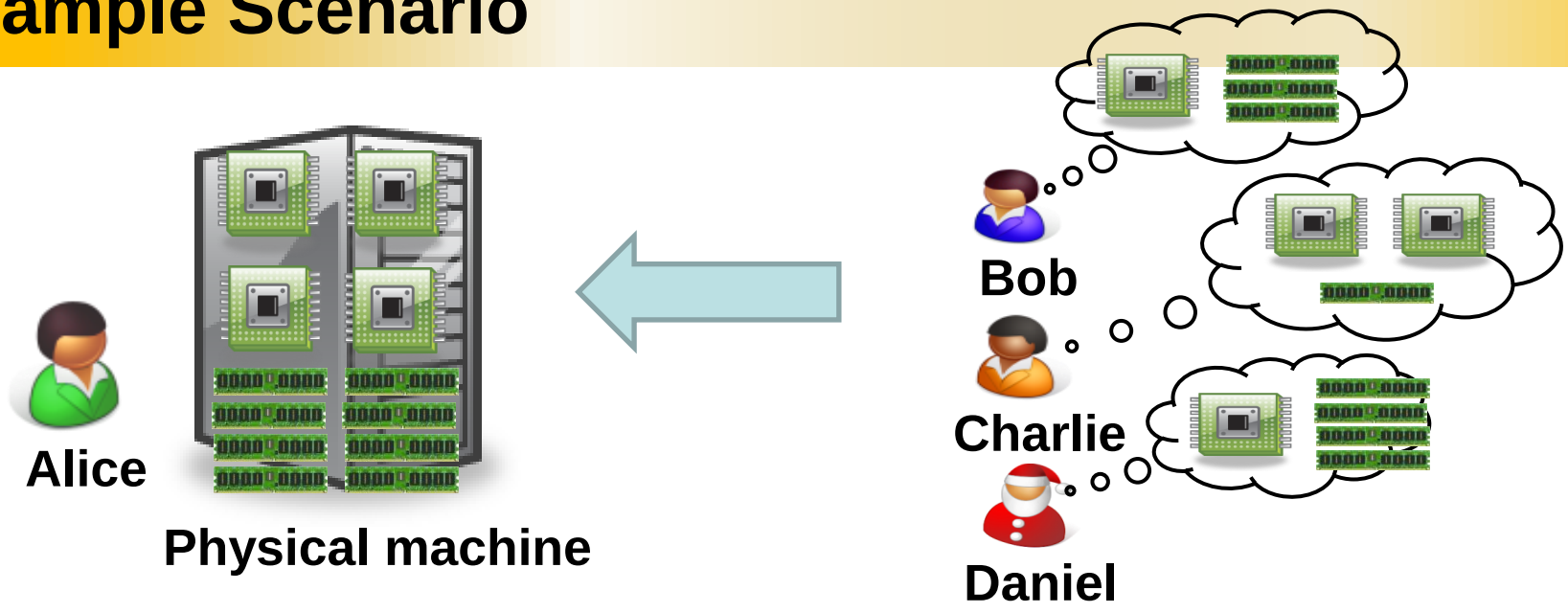
No Virtualized Server



- Single OS per machine
- Software and hardware tightly coupled
- Running multiple applications on same machine often creates conflict
- Inflexible and costly infrastructure

Server Virtualization

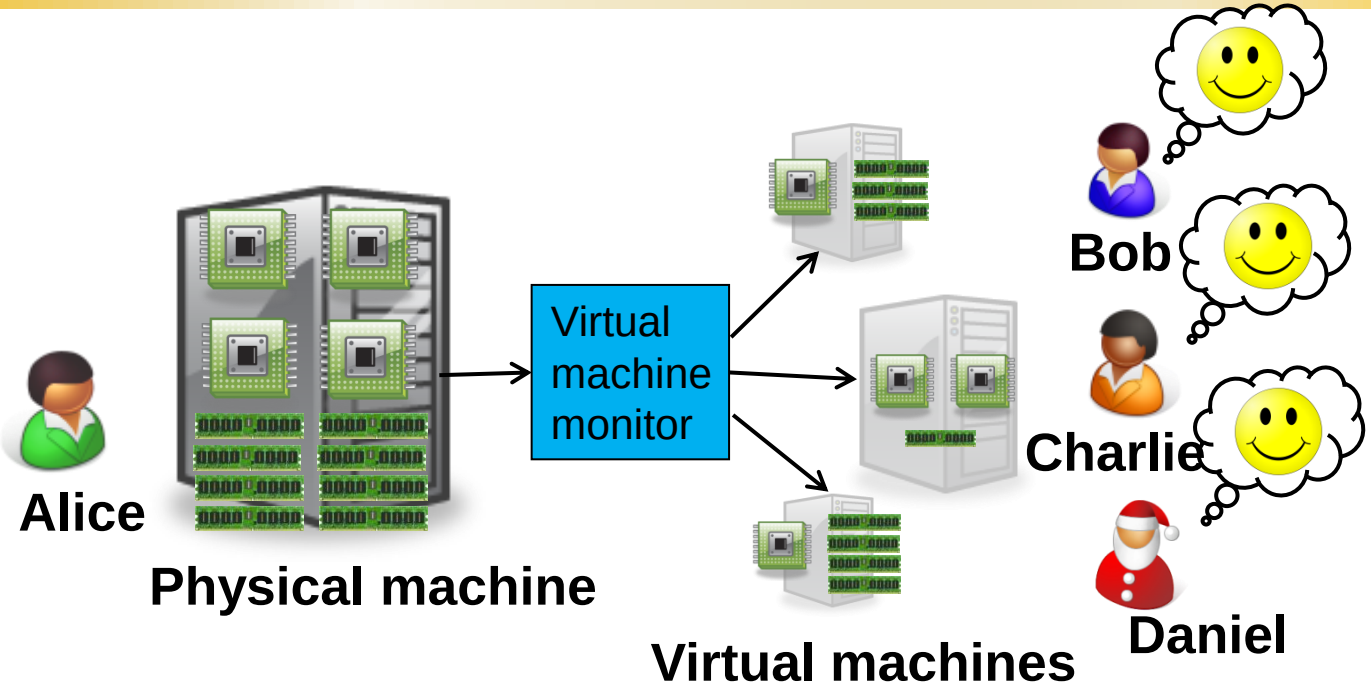
Sample Scenario



- Suppose Alice has a machine with 4 CPUs and 8 GB of memory, and three customers:
  - Bob wants a machine with 1 CPU and 3GB of memory
  - Charlie wants 2 CPUs and 1GB of memory
  - Daniel wants 1 CPU and 4GB of memory
- What should Alice do?

Server Virtualization

Sample Scenario



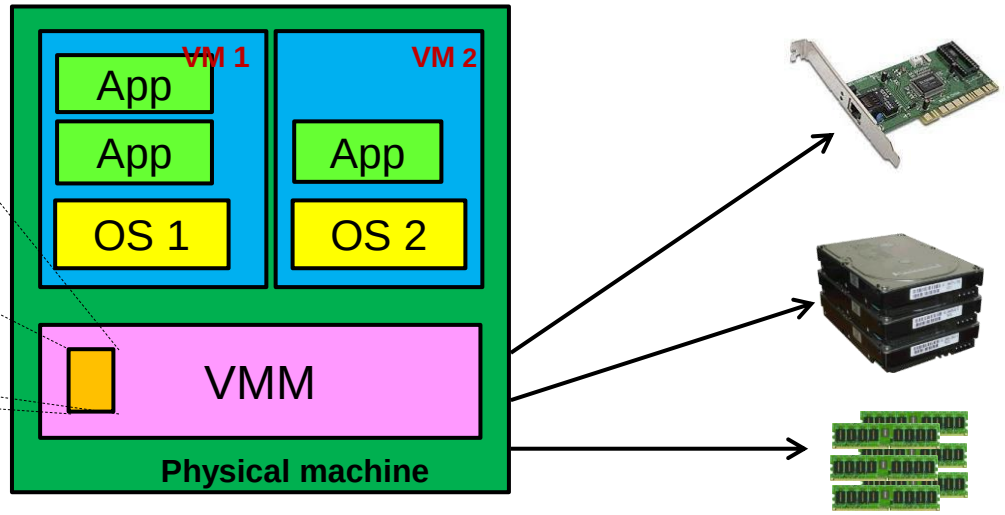
- Alice gives each customer a virtual machine (VM) with the requested resources
- From each customer's perspective, it appears as if they had a physical machine all by themselves (isolation)

Server Virtualization

Hypervisor (VMM)

VM	Virt	Phys
1	0-99	0-99
1	299-399	100-199
2	0-99	300-399
2	200-299	500-599
2	600-699	400-499

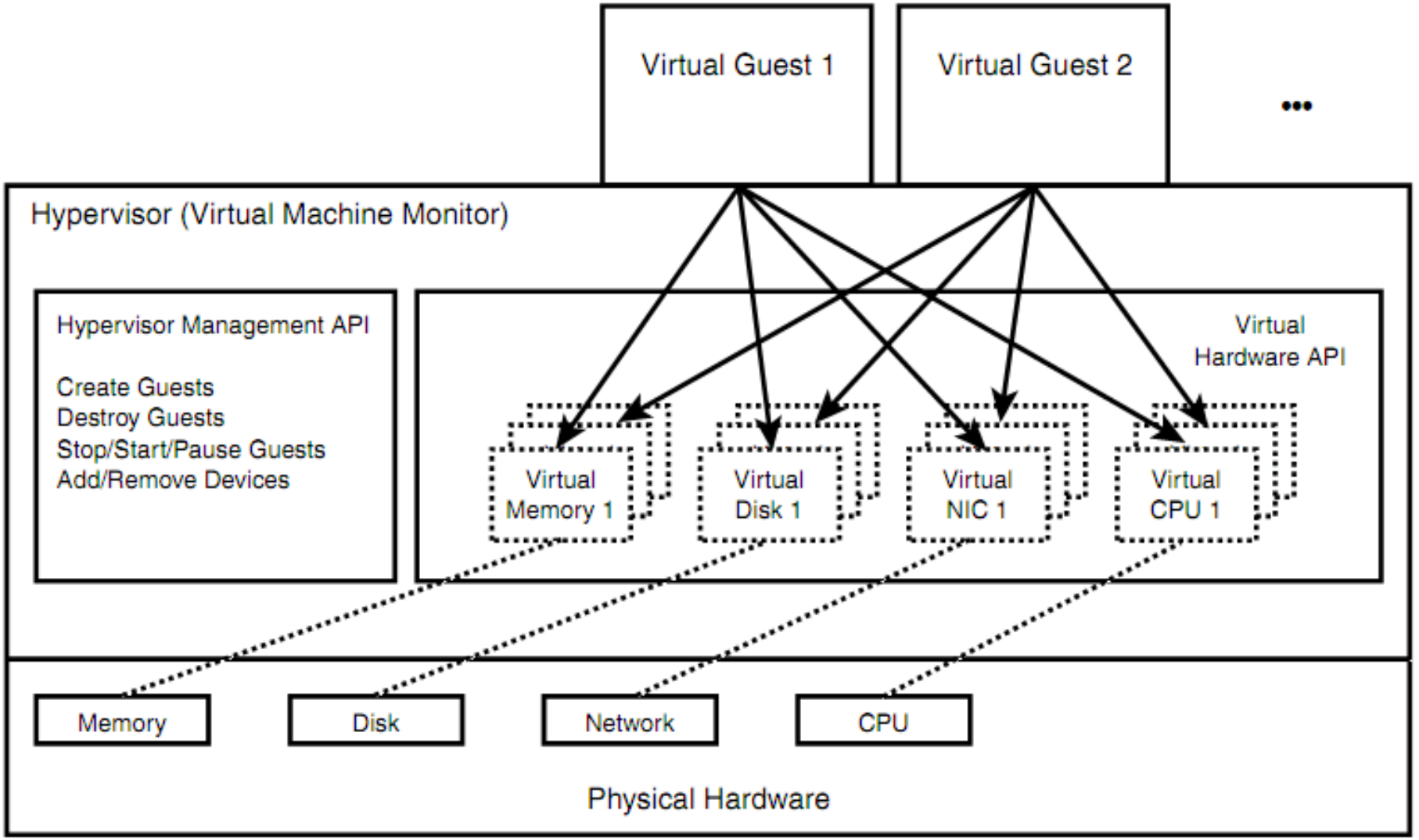
Translation table



- Resources (CPU, memory, ...) are virtualized
- VMM ("Hypervisor") has translation table that maps each VM request for virtual to physical resources

Server Virtualization

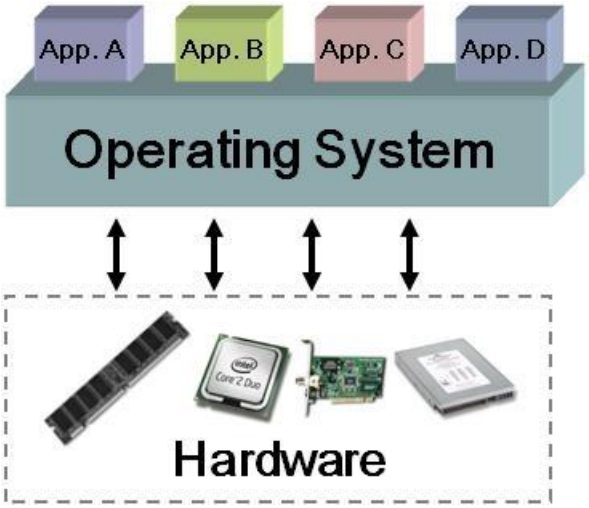
Hardware, Hypervisor and VM



Server Virtualization

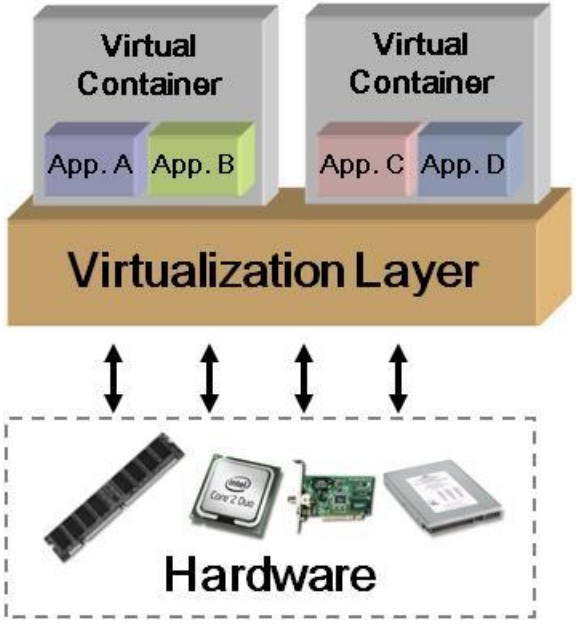
Non-Virtualized system

A single OS controls all hardware platform resources



Virtualized system

Multiple VMs on a single physical machine



Isolated VM

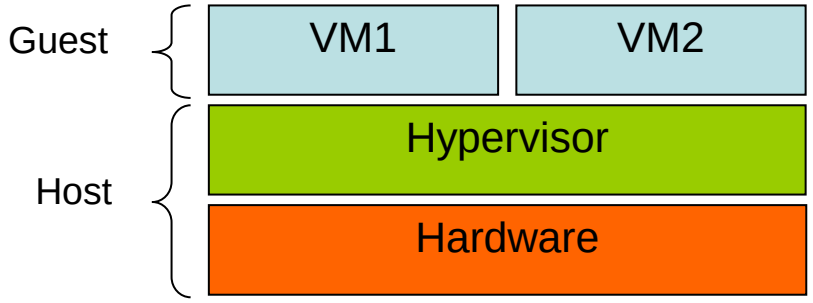
Hypervisor/ virtualization manager

to manage and monitor VMs

Server Virtualization

Two Hypervisor Types

bare-metal hypervisor (type 1)

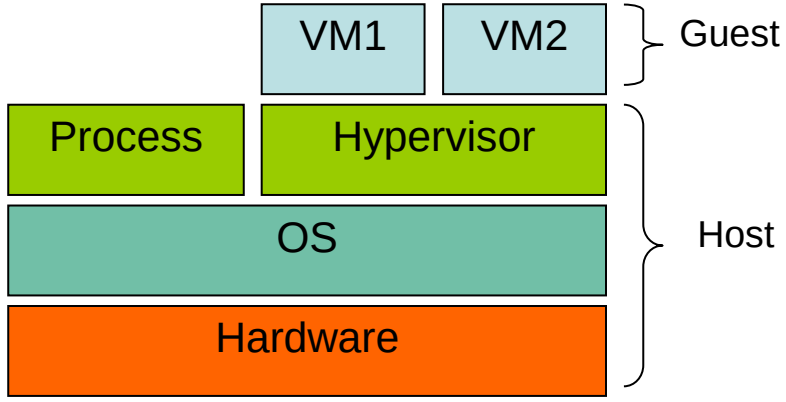


VMware ESXi, Microsoft Hyper-V

Hypervisors run directly on the system hardware

- **More efficient and more secure**
- **Less hardware support**

hosted hypervisor (type 2)



VMware Workstation Player, Oracle VirtualBox, Linux KVM

Hypervisors run on a host OS that provides I/O and memory management

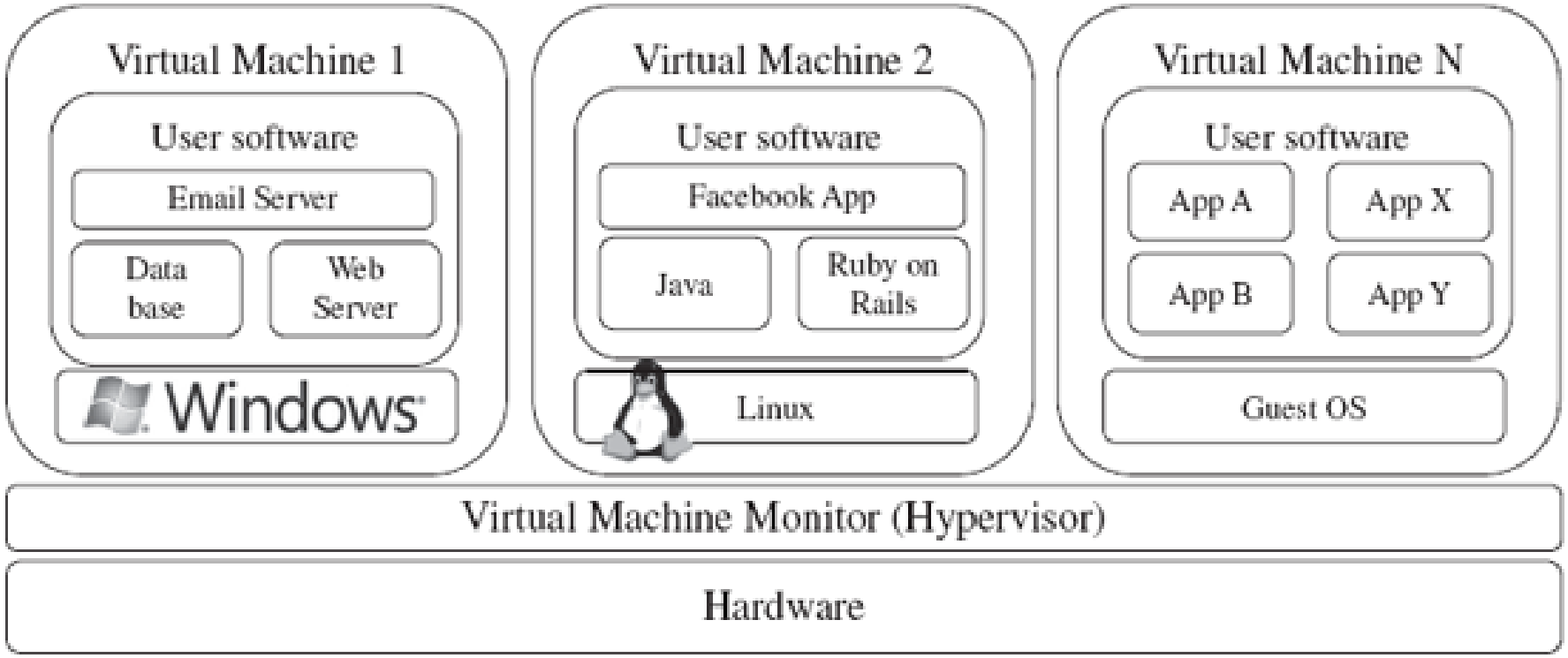
- **Easy to install and manage**
- **more overhead, less secure**

Server Virtualization Techniques

- *Guest machine* is the virtual machine (VM)
- *Full virtualization*
  - Unmodified guest OS
  - Privileged instructions are translated into safe codes on the fly
  - Execution may slow
- *Para-virtualization* like ***Xen***
  - Modified guest OS and co-work with hypervisor
  - Any privileged instructions are replaced by hypervisor calls
- *Hardware-assisted virtualization (VT)*
  - Unmodified guest OS
  - Require Intel VT-x or AMD-v support
  - Simplify VMM and give better performance
  - **Used by most hypervisors nowadays**

Server Virtualization

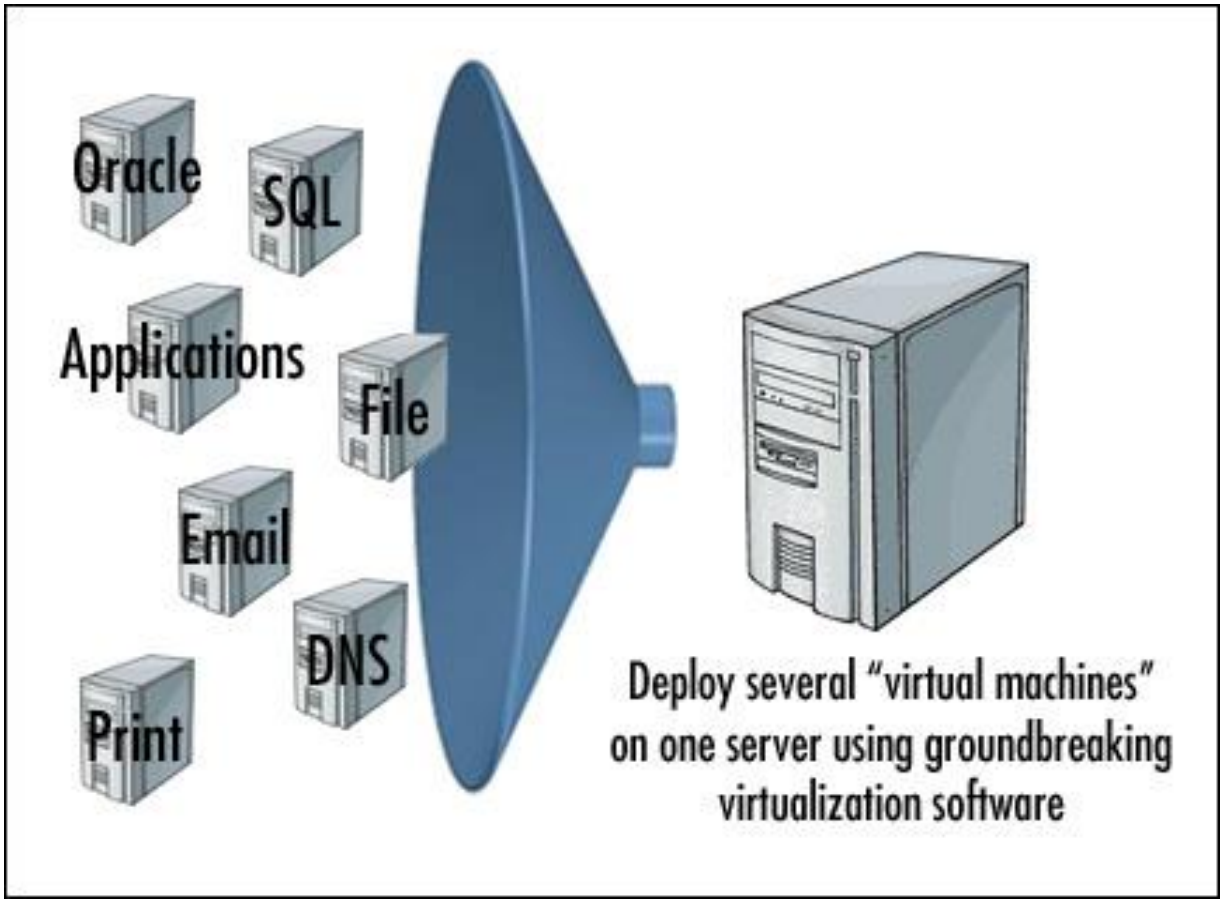
Guest Virtual Machines



Server Virtualization

Resource Consolidation

Run many logical (virtual) machines on the same physical machine and divide system resources between virtual machines

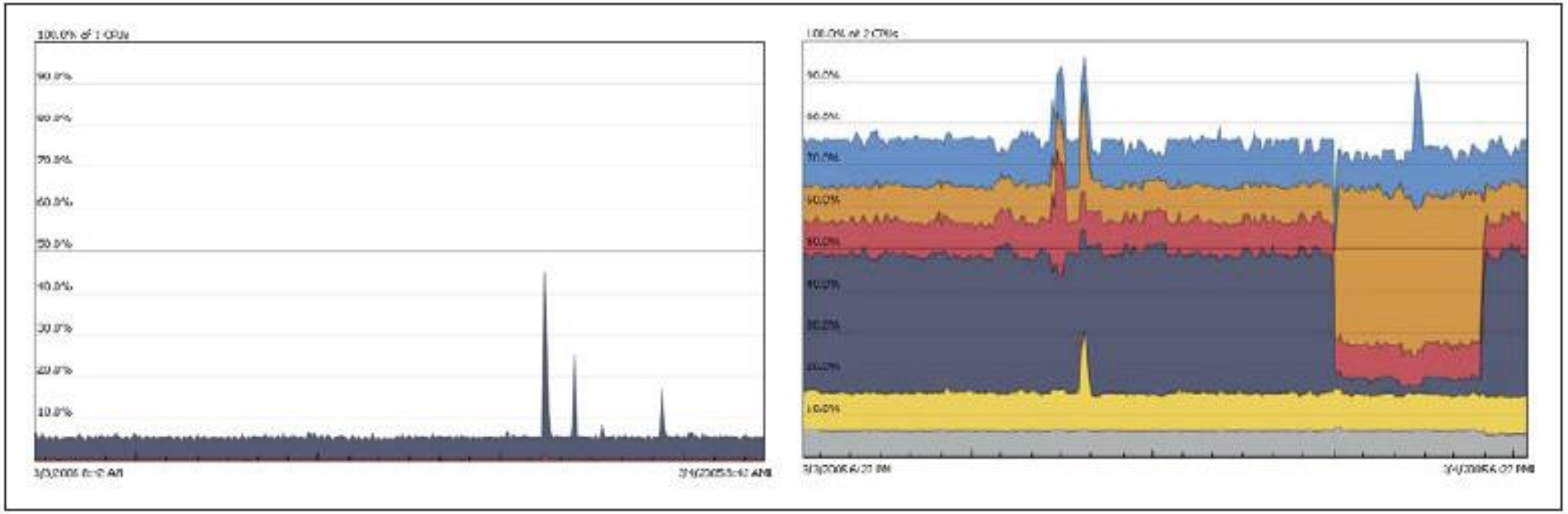


Server Virtualization

Higher Utilization

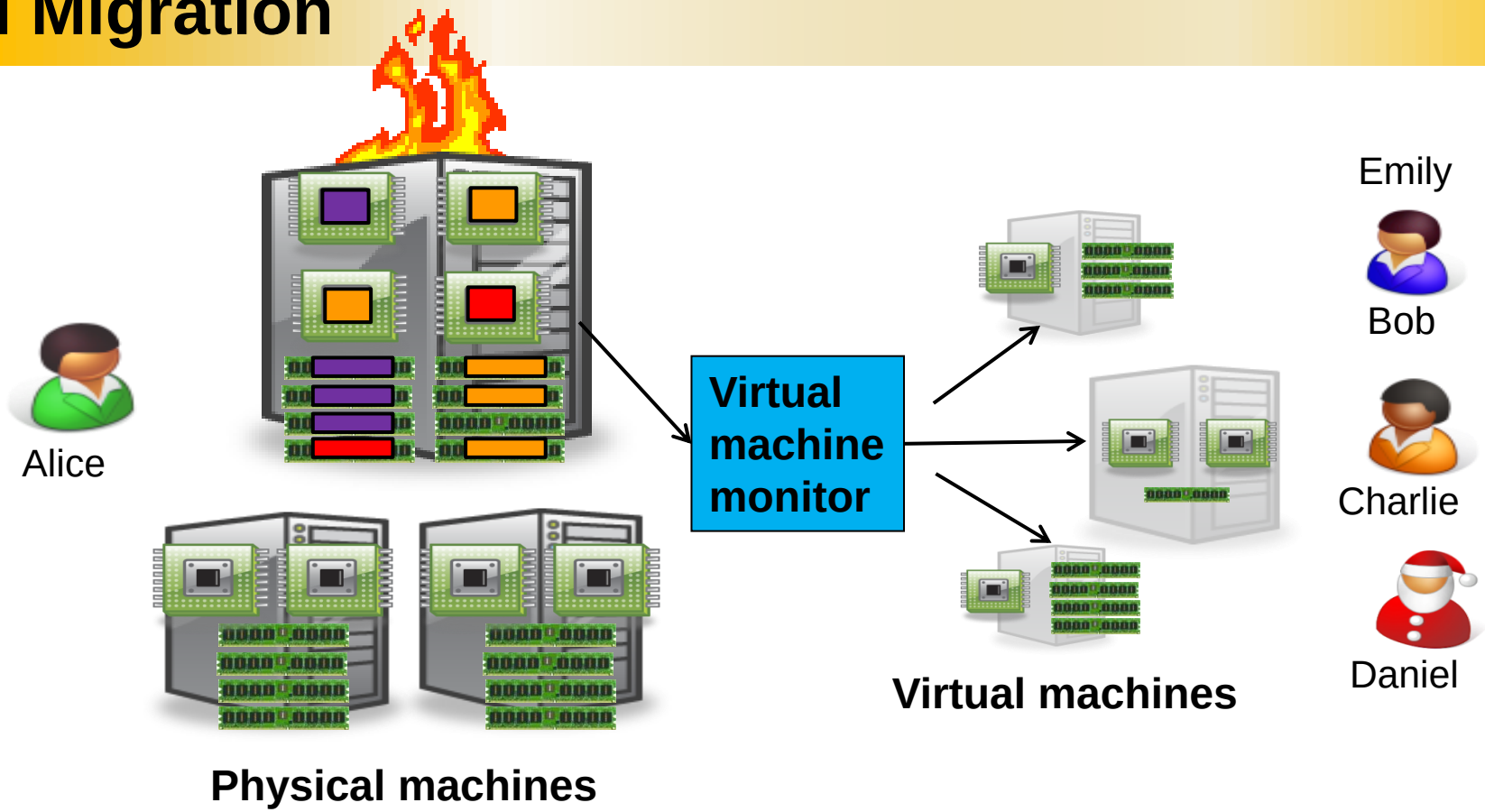
Consolidation of workloads from underutilized servers onto a single server to achieve higher utilization

Dedicated Server **Virtualized Server**



Server Virtualization

VM Migration

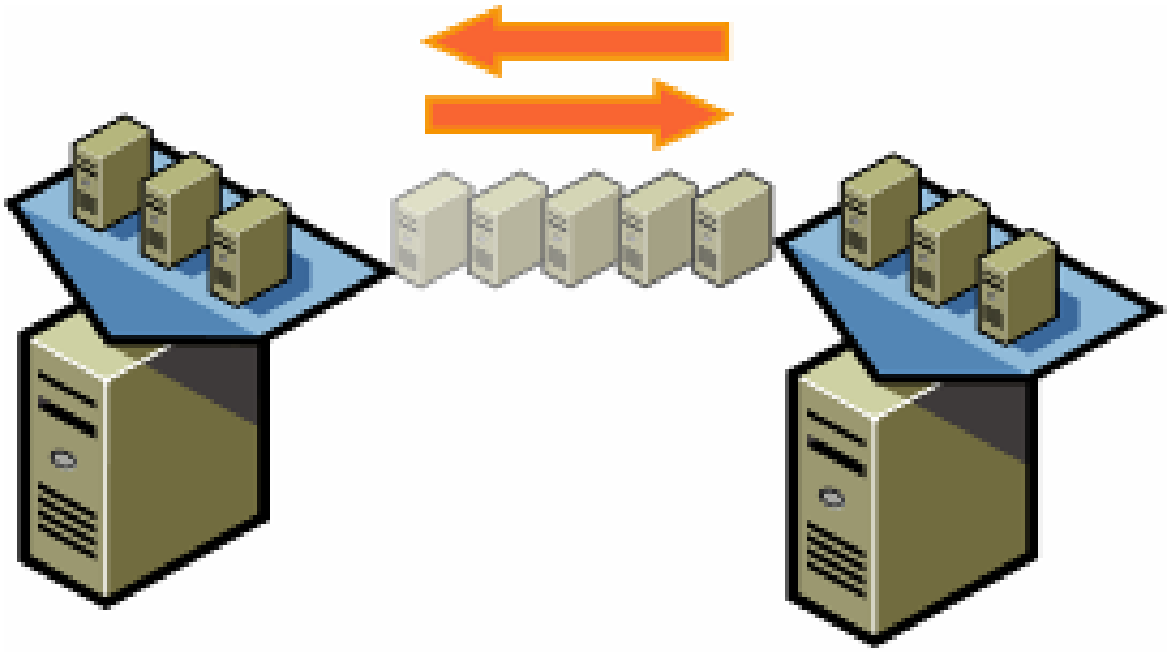


- What if the machine was shut down
- Alice can **migrate** a VM to another physical machine without any customer noticing

Server Virtualization

VM Migration

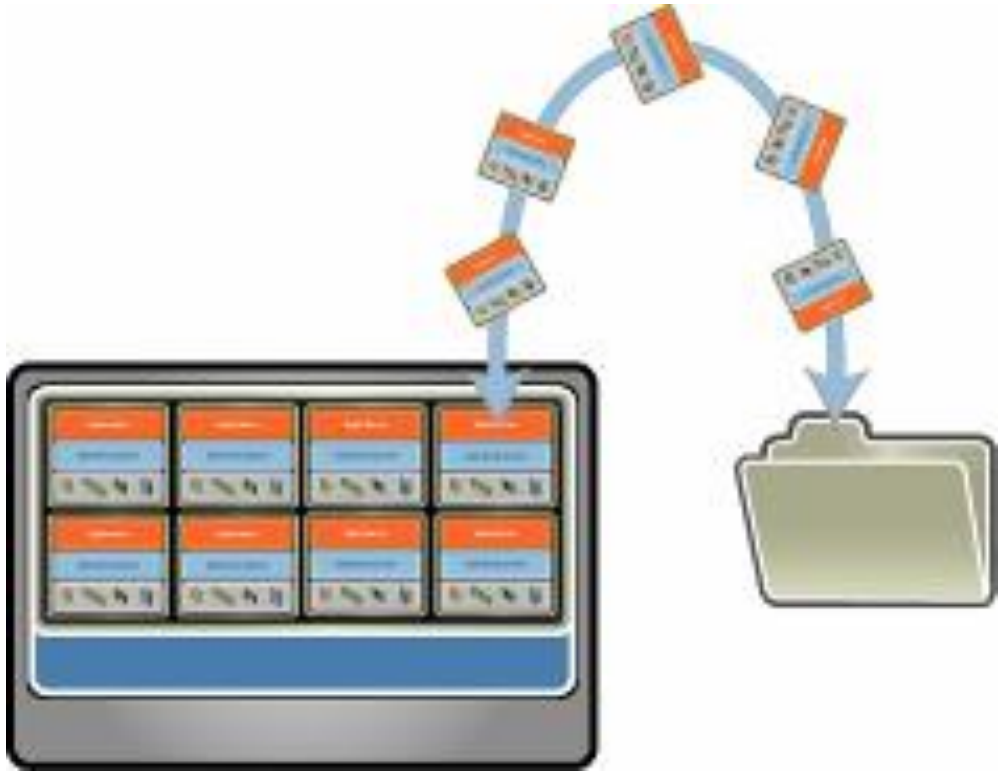
- Migrate a virtual machine to similar or different physical server
- Live migration of virtual machines with (almost) zero downtime



Server Virtualization

Snapshot as Backup

- Entire state of the virtual machine can be saved to files
- Move and copy virtual machines as easily as moving and copying files

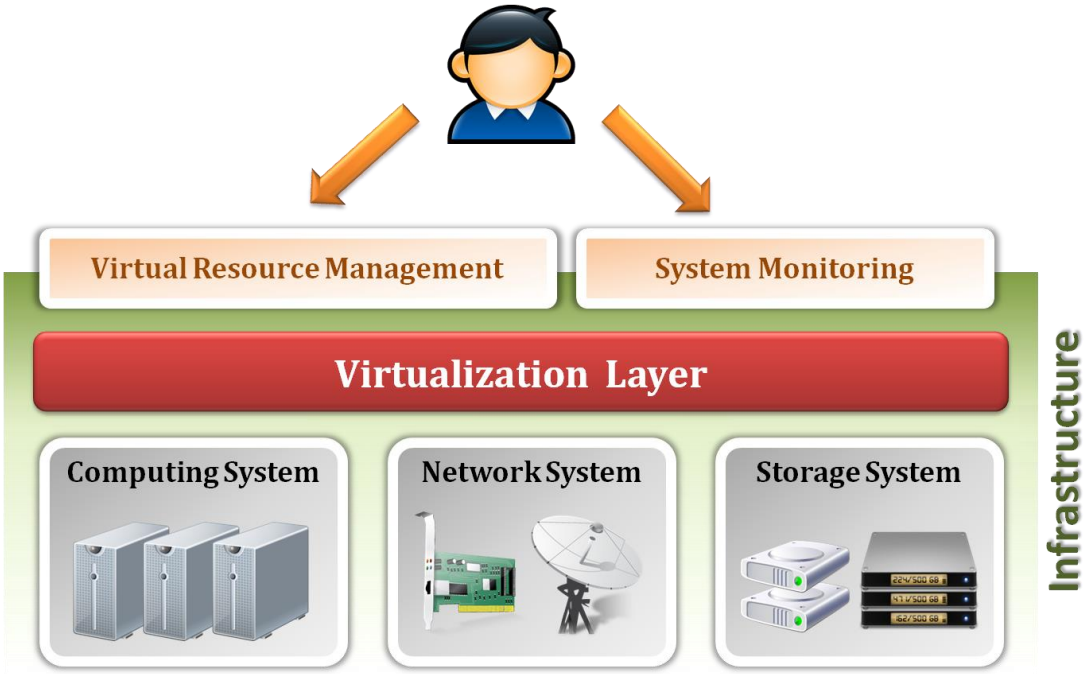


Server Virtualization

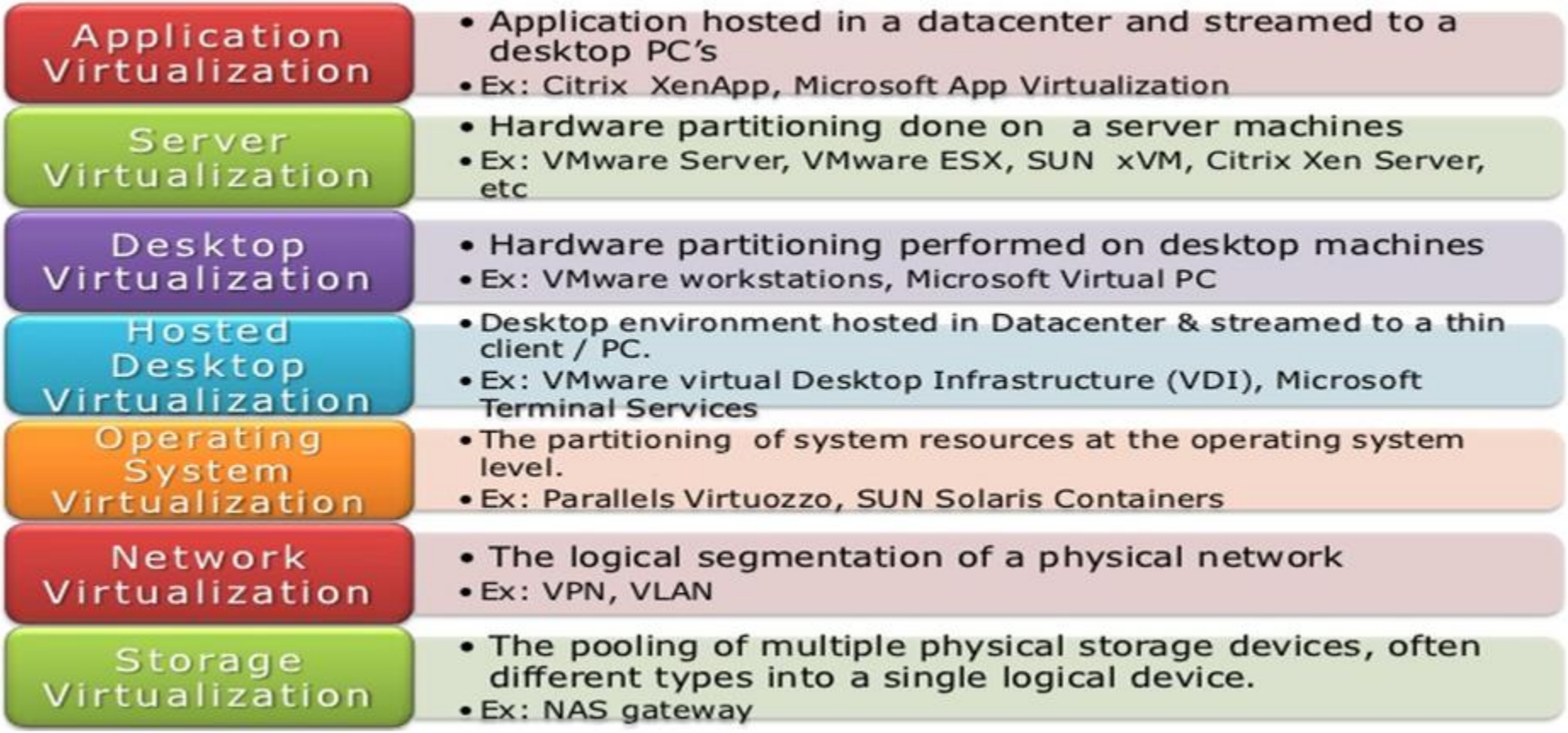
Summary

- **Virtualization** is an enabling technique to provide mapping of physical resources to multiple logical resources
- Virtual instances of IT resources: CPU, Memory, Network, Disk

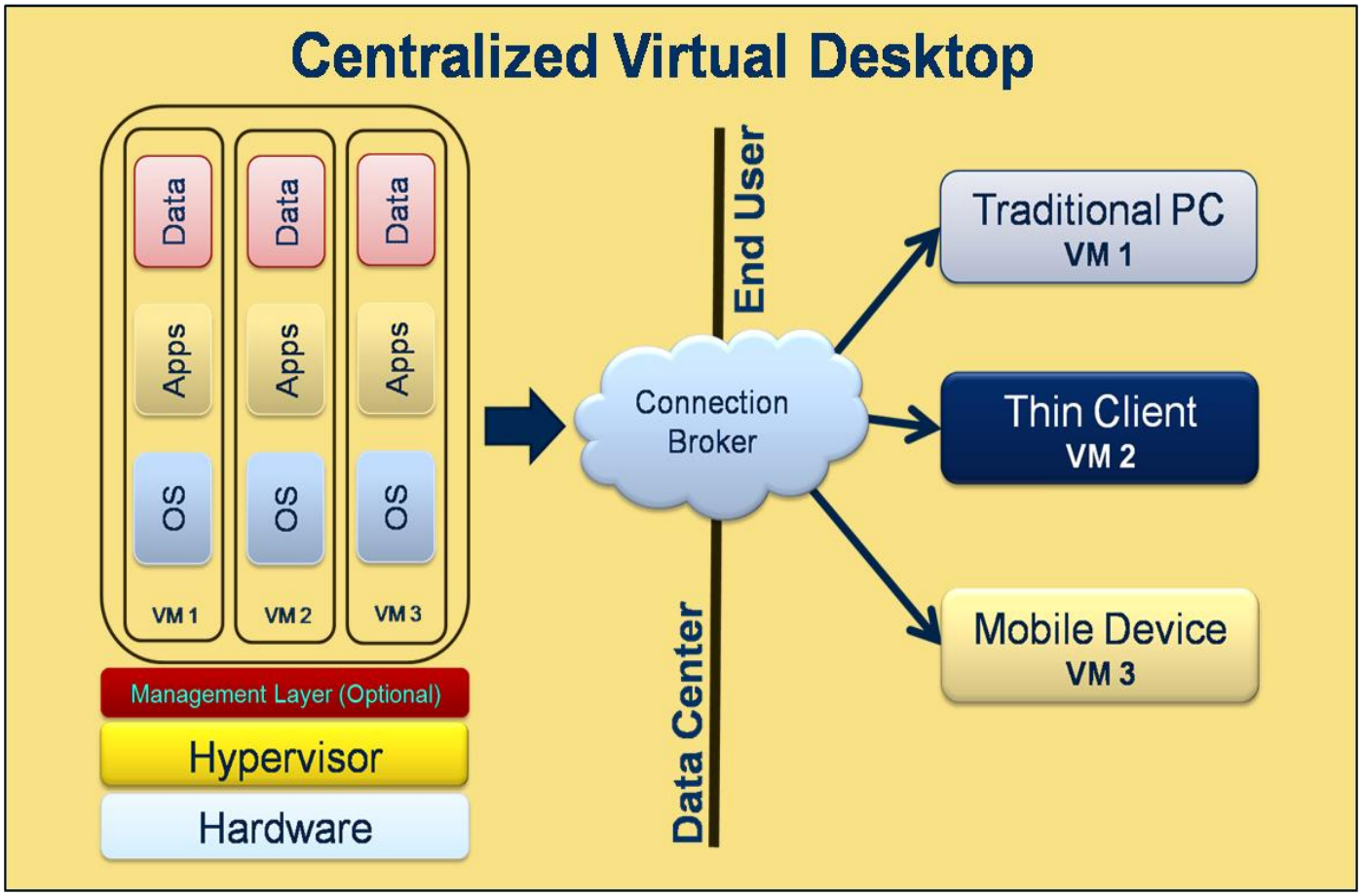
- **Isolate users from the underlying resources**
- **Isolate users from each other**
- **Isolate from fault and security**



Forms of Virtualization



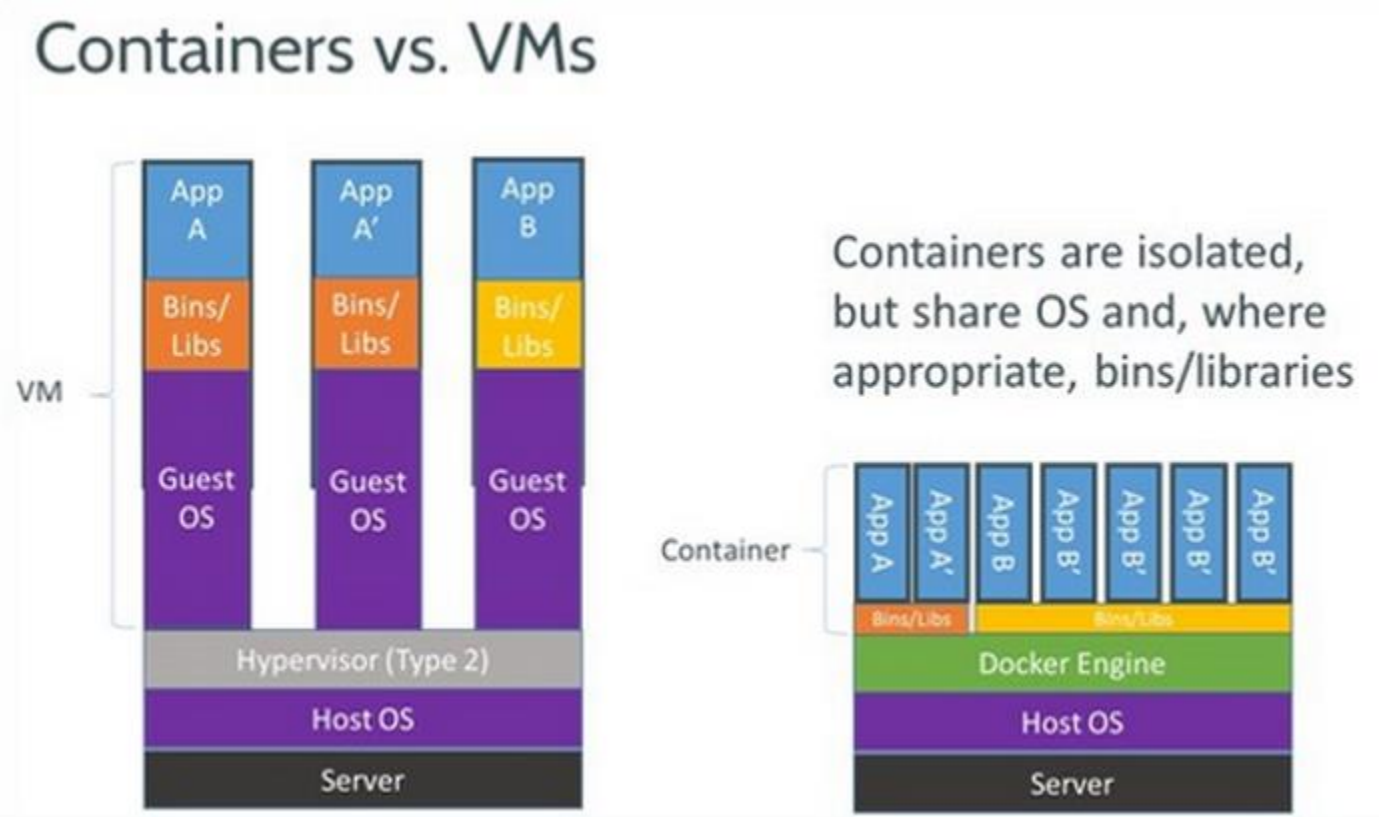
Desktop Virtualization



OS Virtualization

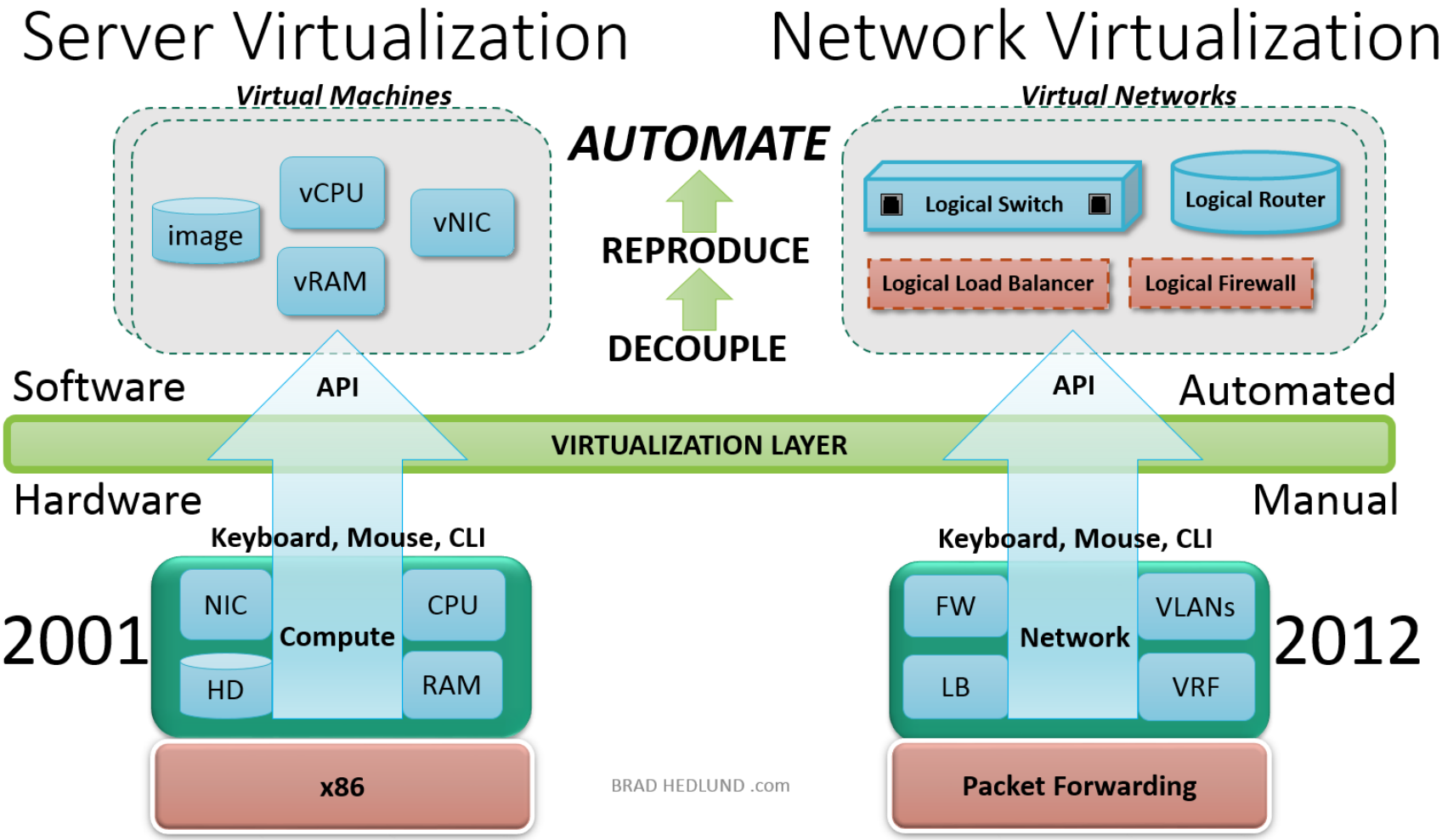
Container

- More light-weight than virtual machine
- Docker is popular



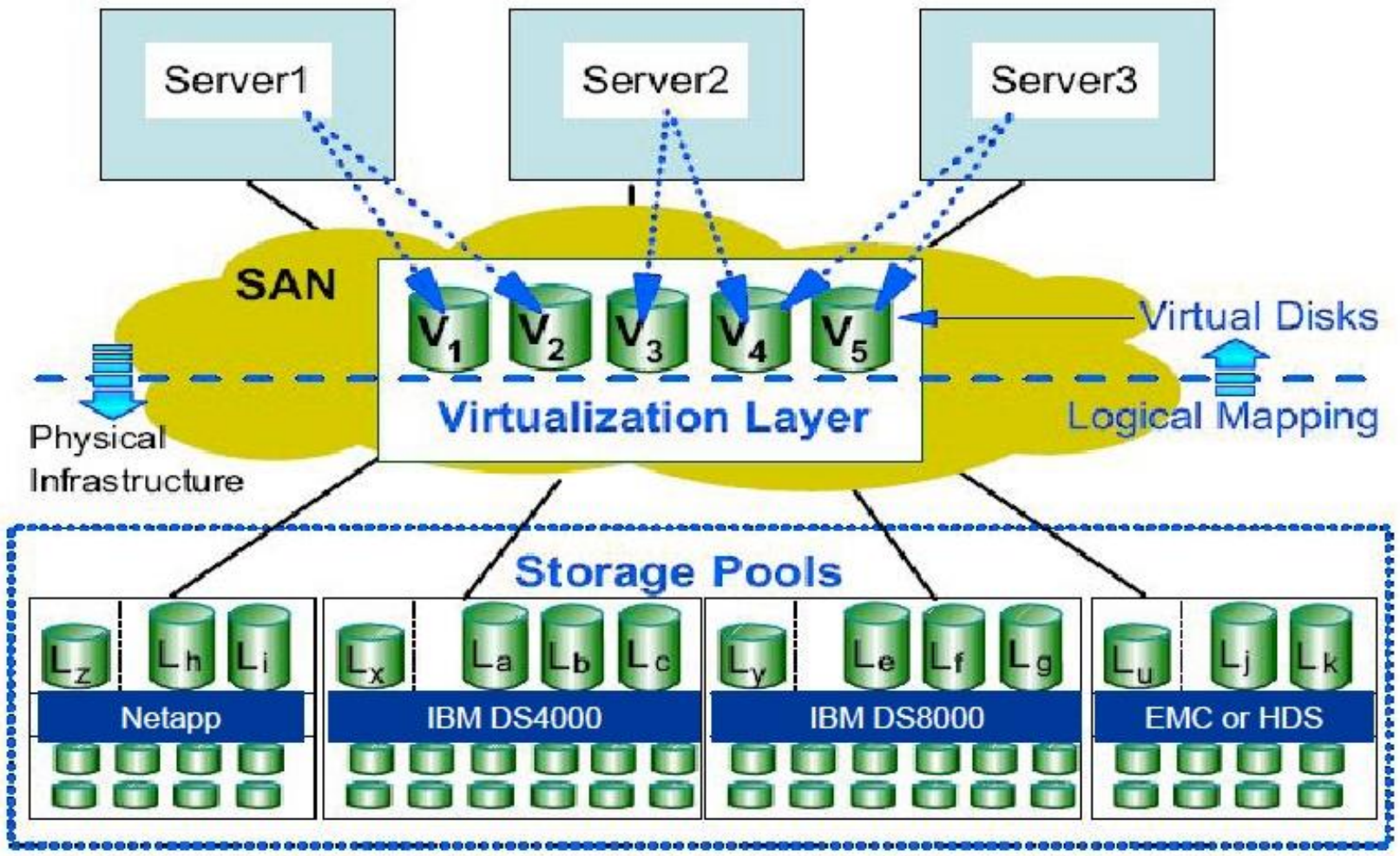
Network Virtualization

Virtual networks are placed on top of physical network elements



Storage Virtualization

Logical storage appears and behaves as physical storage directly connected to host, by combining multiple storage devices into a single storage pool



Virtualization in Cloud