Containers and MicroServices

Mikroteenused ja konteinerarhitektuur

Virtualization

Heino Talvik

Today's Laptop

https://i.dell.com/is/image/DellContent/content/dam/ss2/product-images/dell-client-products/notebooks/latitude-notebooks/14-5430/pdp/laptop-latitude-14-5430-pdp-hero-504x350.psd?qlt=95&fit=constrain,1&hei=500&fmt=jpg&wid=570&hei=400

- Apple 2023 MacBook Pro laptop M2 Pro chip with 12-core CPU and 19-core GPU: 16.2-inch Liquid Retina XDR display, 16GB Unified Memory, 512GB SSD storage
- Dell Intel® Core™ Ultra 7 155H (24 MB cache, 16 cores, up to 4.8 GHz Turbo) RAM 32GB 1TB

Moore's "Law"

- "Transistor density on integrated circuits doubles about every two years"
  - Gordon Moore, co founder of Intel
- This means at the same cost you get:
  - twice as many logical components
  - and higher speed (or less power consumption)

Without Moore's Law: 1 Transistor = 1mm²

1959

Planar Transistor 1mm²

1

60s

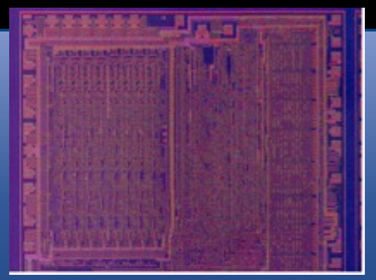
TTL Quad Gate

16

70s

8bit Processor

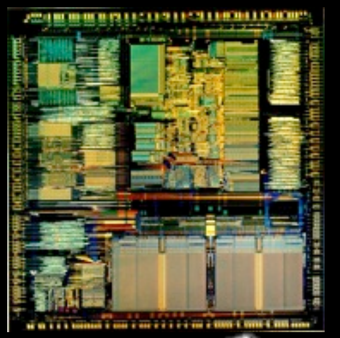
4500



85

32bit Processor

275.000

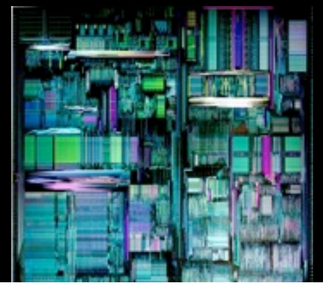


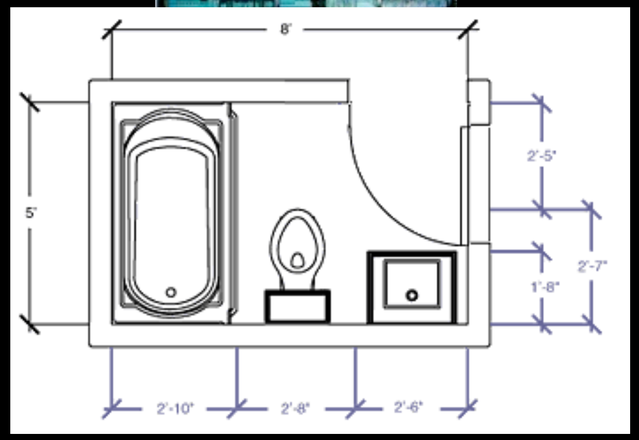
Without Moore's Law: 1 Transistor = 1mm²

95

32bit Processor

5.500.000

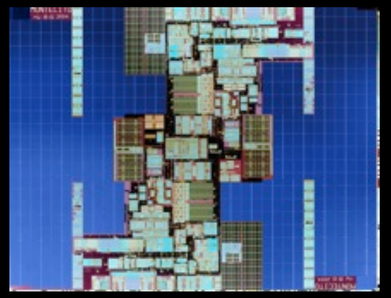




2000s

64bit Processor

221.000.000

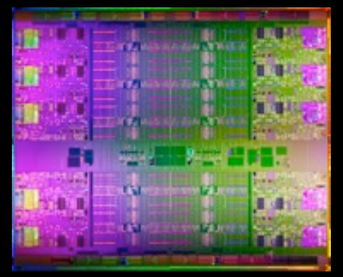


This Room
(- a bit)

2010s

Multi Core Processor

3.000.000.000



Two of those,
plus a bit

Frequency

- Smaller structures, less electrons need to be pushed around → faster transistors
  - 1970-1990 1→50 MHz (as expected)
  - and then things got interesting….

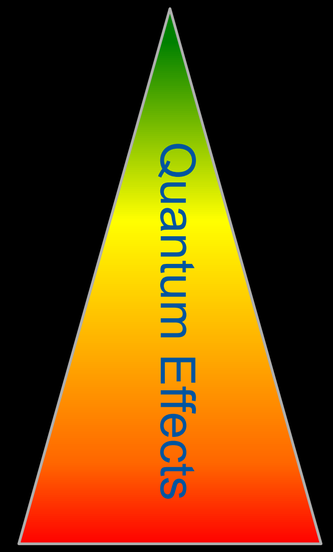
Observations:

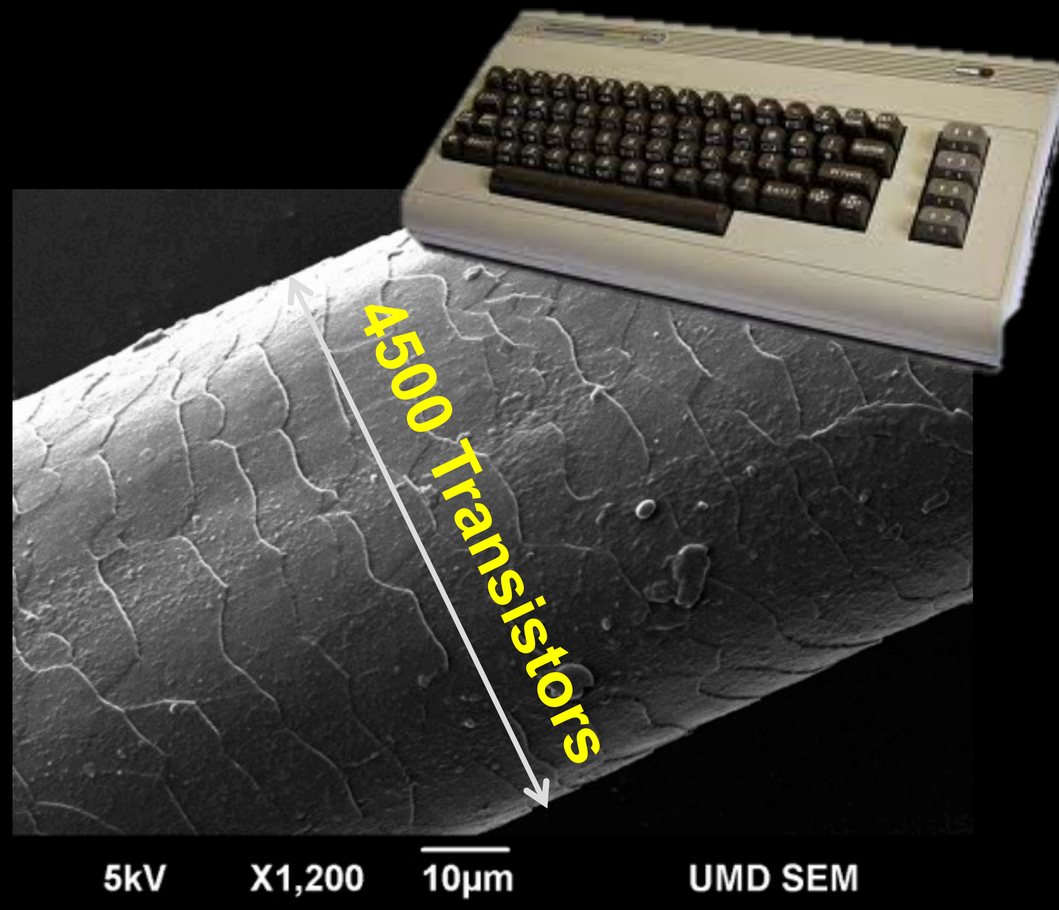
- A 5 Euro music birthday card has more computing power than the Allied Forces during WW-II
- A 2013 smart phone is faster than CERN's computing centre from the mid 90s

- Intuition breaks down → structured approach is needed

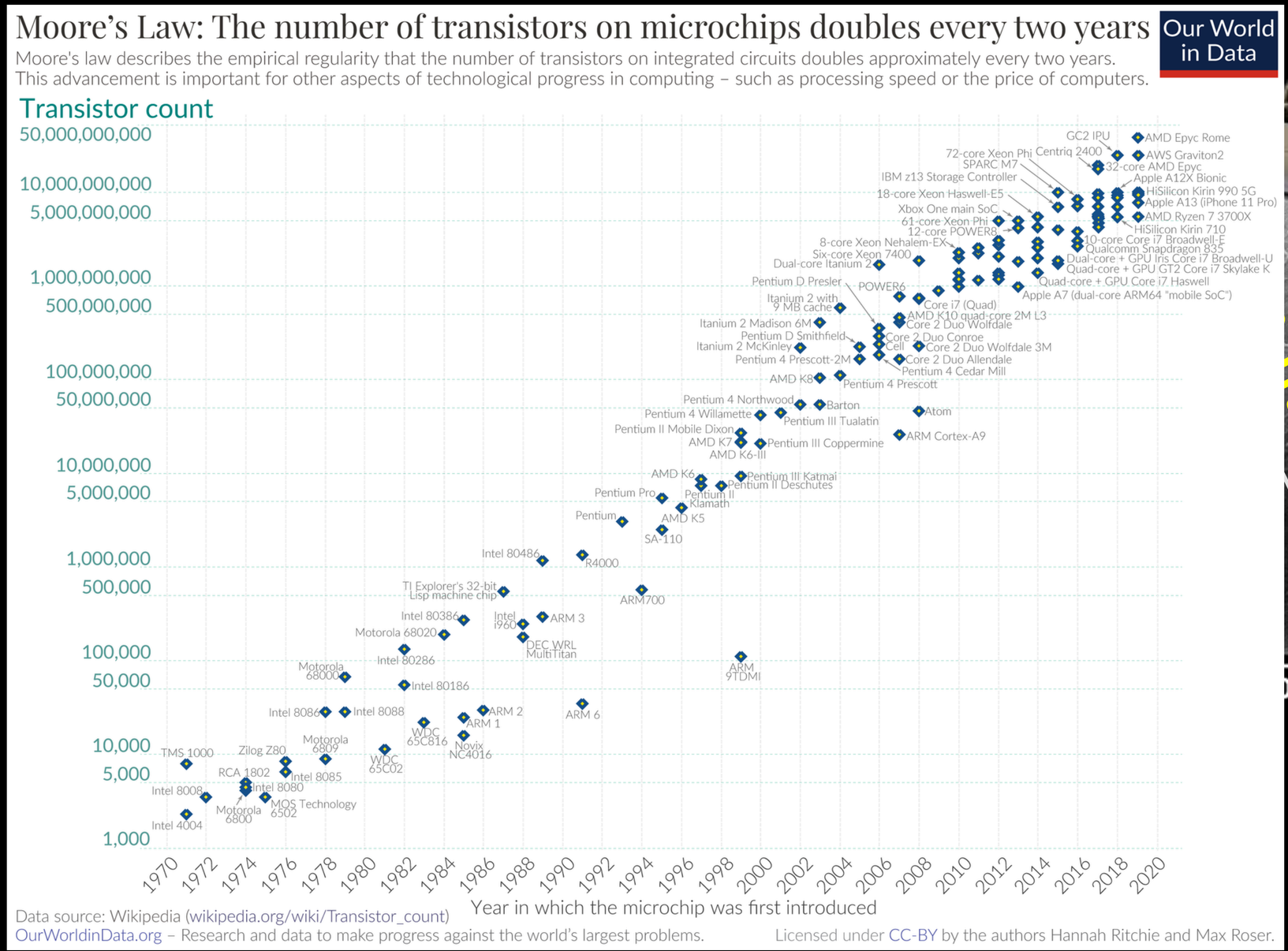
Moore's Law's Future

- Intel's roadmap:
  - 14nm 2014
  - 10nm 2016
  - 7nm 2018
  - 5nm 2020
  - 3nm 2022
- Smallest transistor 2024
  - 1.5 nm using 1 atom



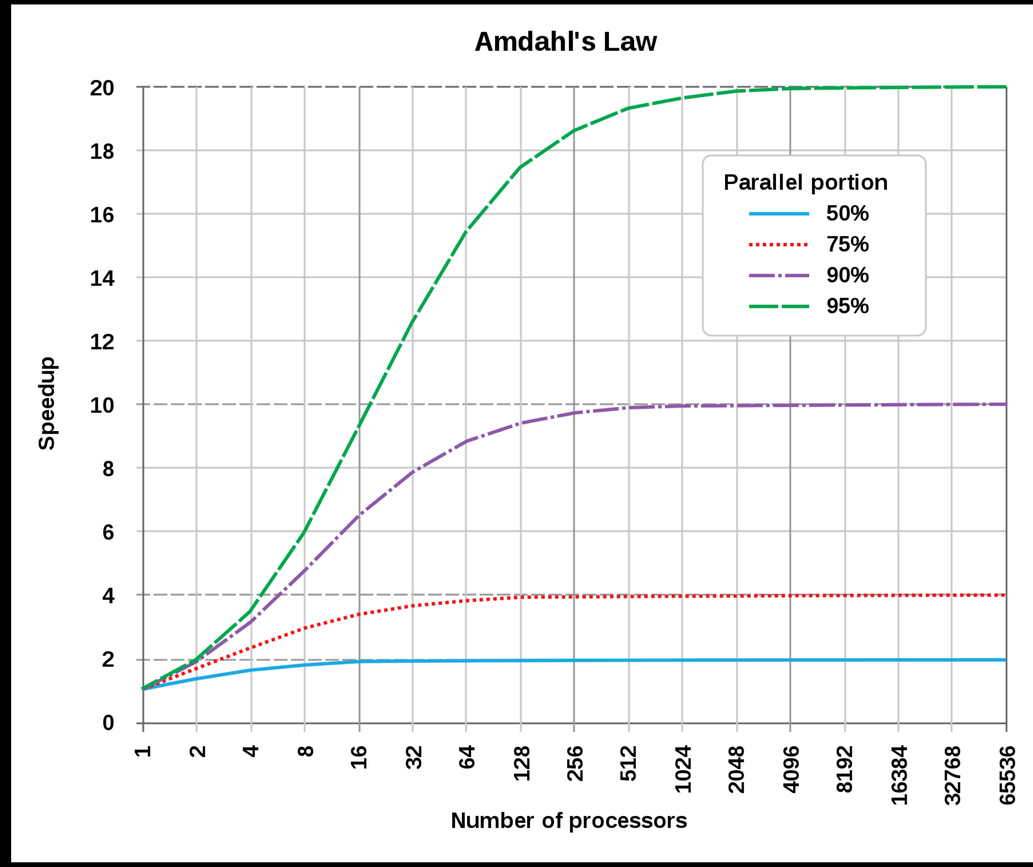


Moore's Law's history

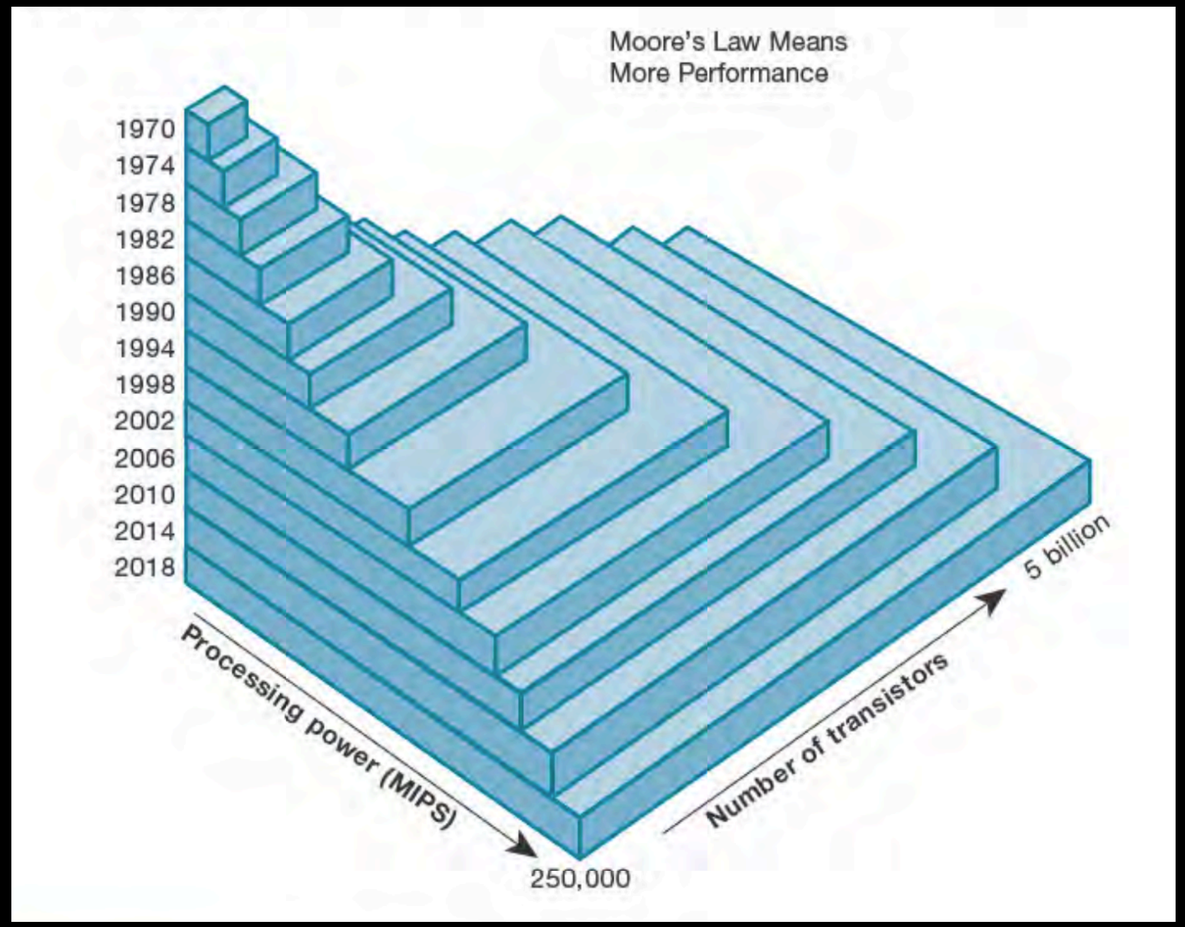


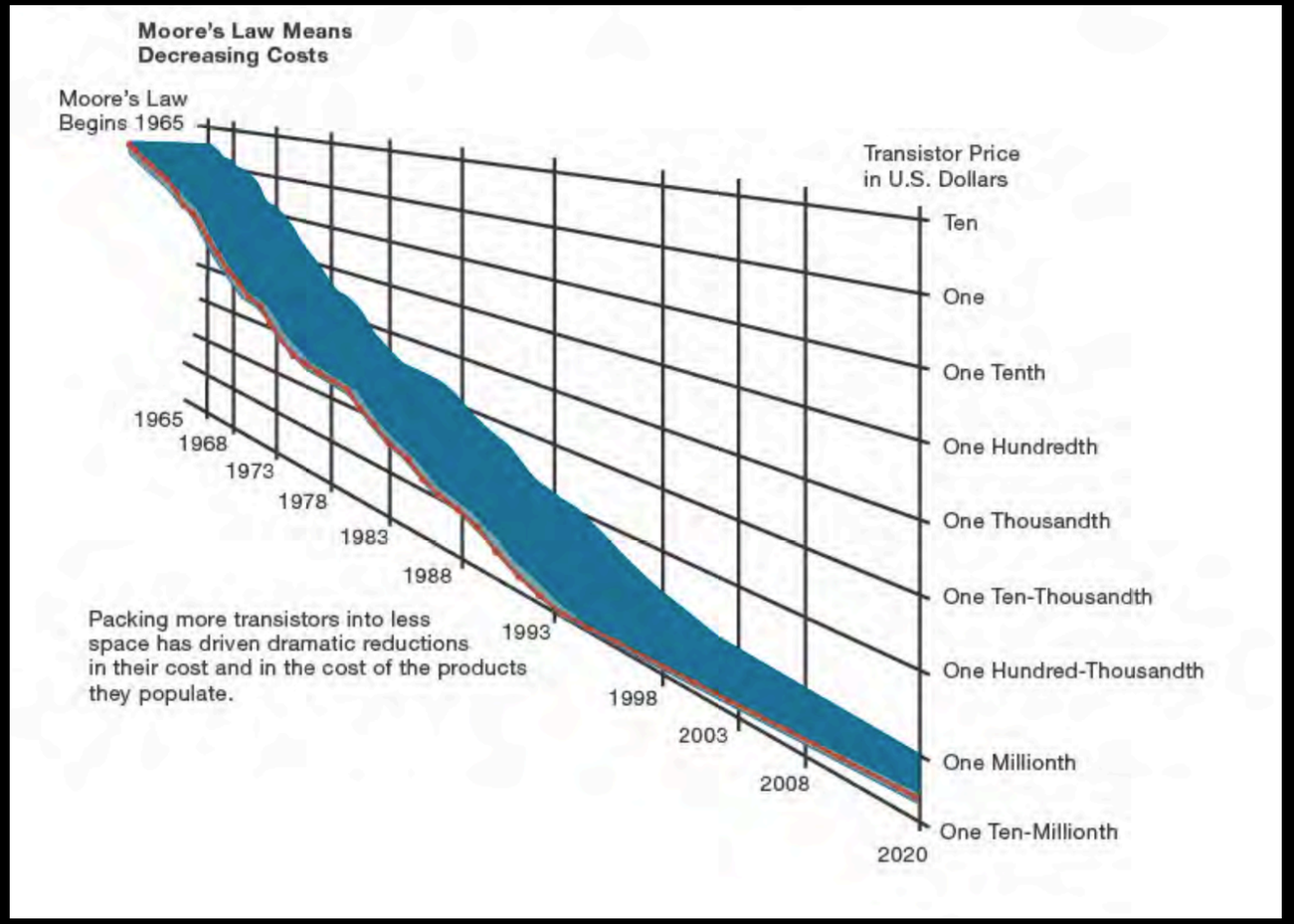
What to do with 5.000.000.000 Transistors?

- Two core concepts have been used to speed up computer architectures
  - Pipelining (working like an assembly line)
  - Parallel processing (working concurrently)
- Many additional smart tricks have been invented
  - Caching, branch prediction, etc.....
- Now: CPUs use all tricks combined
  - Difficult to make full use of the hardware
    - Facing sometimes fundamental limits (Amdahl's Law)



Moore's "Law"


Moore's Law Means
More Performance
1970
1974
1978
1982
1986
1990
1994
1998
2002
2006
2010
2014
2018
Processing power (MIPS)
250,000
Number of transistors
5 billion


Moore's Law Means
Decreasing Costs
Moore's Law
Begins 1965
Transistor Price
in U.S. Dollars
Ten
One
One Tenth
One Hundredth
One Thousandth
One Ten-Thousandth
One Hundred-Thousandth
One Millionth
One Ten-Millionth
1965
1968
1973
1978
1983
1988
1993
1998
2003
2008
2020
Packing more transistors into less space has driven dramatic reductions in their cost and in the cost of the products they populate.

Implication of Moore's Law

If computing speed and capacity double every 24 months, what are the implications in our lives?

Well, the average student is – to one significant figure – about 20 years old.

And the average lifespan in the Estonia – to one significant figure – is about 80 years.

So, the average undergrad student has 60 years to go.

So how much will computing speed and capacity increase during the time you have left?

Double, double, …

60 years / 2 years = 30 doublings

What is 2^{30}?

Consider the computer on your desktop today, compared to the computer on your desktop the day you die.

How much faster will it be?

Can we possibly predict what the future of computing will enable us to do?

HARDWARE

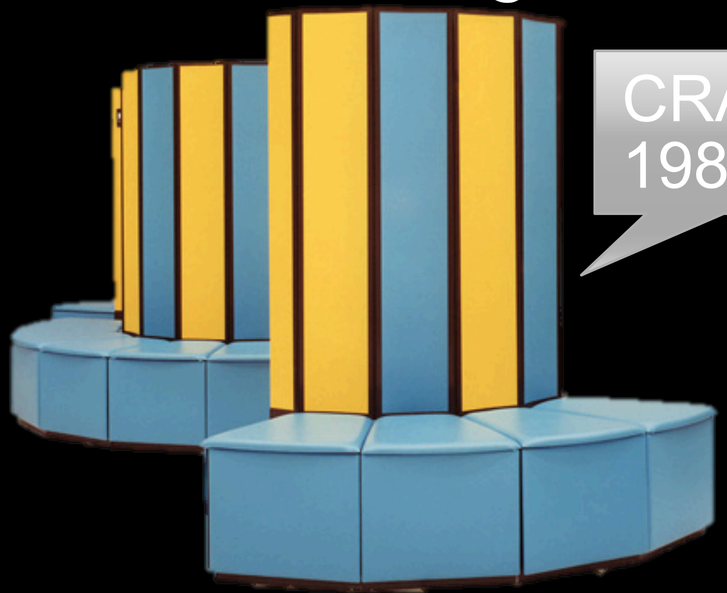
- Supercomputer
- Mainframe
- Server
- Workstation
- Personal computer (PC)
- Mobile device
- Embedded systems

TABLE 3.1 Characteristics of Computers Currently Being Used in Organizations

Type of Computer	Number of Simultaneous Users	Physical Size	Typical Use	Random Access Memory	Typical Cost (in US$)
Supercomputer	One to many	Like an automobile to as large as multiple rooms	Scientific research	5,000+ GB	Up to $200 million
Mainframe	1,000+	Like a refrigerator	Transaction processing, enterprise-wide applications	Up to 3,000+ GB	Up to $10 million
Server	10,000+	Like a DVD player and mounted in a rack to fitting on a desktop	Providing websites or access to databases, applications, or files	Up to 512 GB	Up to $50,000
Workstation	Typically one	Fitting on a desktop to the size of a file cabinet	Engineering, medical, graphic design	Up to 512 GB	Up to $10,000
Personal computer	One	Fitting on a desktop	Personal productivity	512 MB to 32 GB	Up to $5,000
Mobile device	One	Handheld	Personal productivity	512 MB to 16 GB	Up to $1,400

Super Computers?

- Current Super Computers are build from commodity components (3.120.000 cores)
  - adding a high speed, low latency interconnect
  - adding accelerator cards (graphic cards, etc.)



CRAY
1980s

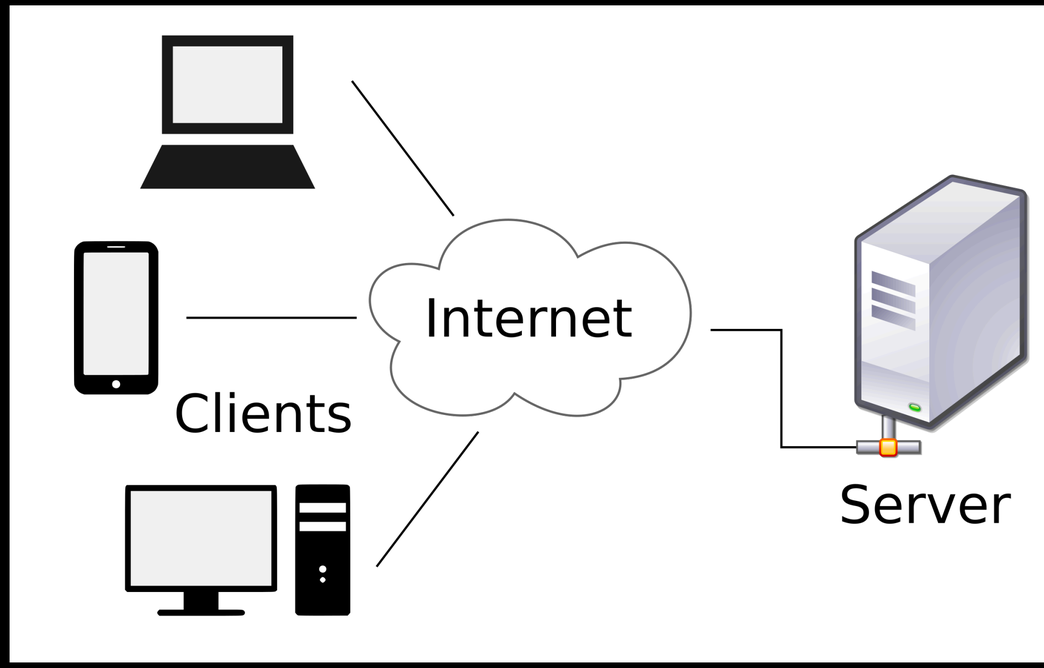
Tianhe-2
2013

Mainframes?

- A mainframe computer, is a computer used primarily by large organizations for critical applications like bulk data processing for tasks such as censuses, industry and consumer statistics, enterprise resource planning, and large-scale transaction processing.
- A mainframe computer is large but not as large as a supercomputer and has more processing power than some other classes of computers, such as minicomputers, servers, workstations, and personal computers.
- Most large-scale computer-system architectures were established in the 1960s, but they continue to evolve. Mainframe computers are used as servers.

Servers?

- … is a piece of computer hardware or software (computer program) that provides functionality for other programs or devices, called “clients".
- This architecture is called the client–server model.
- Servers can provide various functionalities, often called "services", such as sharing data or resources among multiple clients or performing computations for a client.
- A single server can serve multiple clients, and a single client can use multiple servers.
- Typical servers are database servers, file servers, mail servers, print servers, web servers, game servers, and application servers.



Workstation?

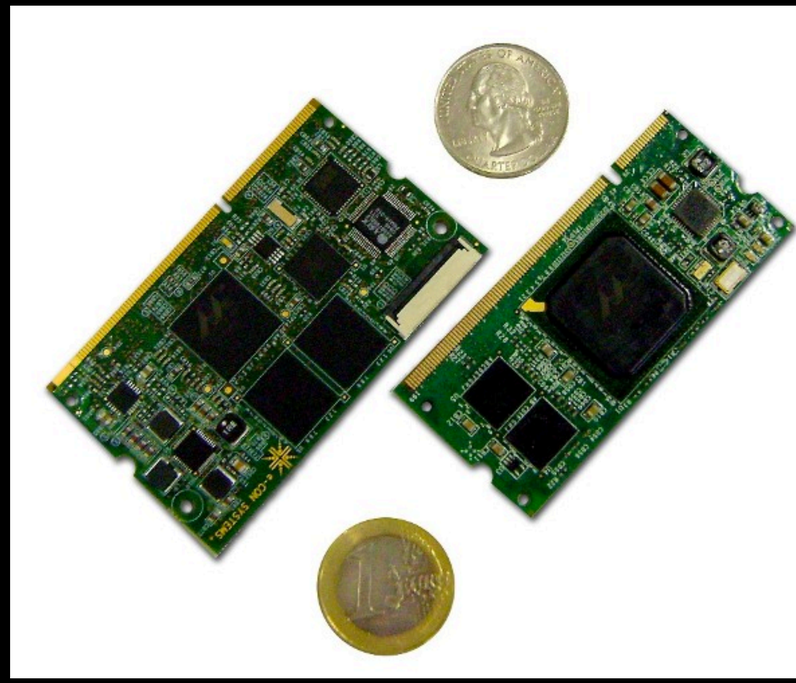
- … is a special computer designed for technical or scientific applications.
- Intended primarily to be used by a single user, they are commonly connected to a local area network and run multi-user operating systems.
- The term workstation has been used loosely to refer to everything from a mainframe computer terminal to a PC connected to a network, but the most common form refers to the class of hardware offered by several current and defunct companies such as Sun Microsystems, Silicon Graphics, Apollo Computer, DEC, HP, NeXT, and IBM which powered the 3D computer graphics revolution of the late 1990s.

Personal Computer?

- … is a computer designed for individual use.
- It is typically used for tasks such as word processing, internet browsing, email, multimedia playback, and gaming.
- Personal computers are intended to be operated directly by an end user, rather than by a computer expert or technician.
- Unlike large, costly minicomputers and mainframes, time-sharing by many people at the same time is not used with personal computers.

Embedded systems?

- … is a computer system—a combination of a computer processor, computer memory, and input/output peripheral devices—that has a dedicated function within a larger mechanical or electronic system.
- It is embedded as part of a complete device often including electrical or electronic hardware and mechanical parts.
- embedded system typically controls physical operations of the machine that it is embedded within. Embedded systems control many devices in common use.
- In 2009, it was estimated that ninety-eight percent of all microprocessors manufactured were used in embedded systems.



Why Virtualization?

- Share resources among many uses
- Allow heterogeneity in environments
- Allow differences in host and guest
- Provide a mechanism for migration, checkpointing, etc
  - Recovery, maintenance
- Provide for elasticity

Challenges to Virtualization

- Isolation
  - Performance w.r.t. resource sharing (CPU, memory, network, disk)
  - Protection w.r.t. resource use
- Support multiple operating systems
- Minimal performance penalty

Granularity 1/3

- Multiplex processes
  - Finely grained, but forces symmetric execution environment
    - Finely grained: Virtualizing at the process level means each process can be treated as an individual unit. This gives a high level of control and precision over how resources are allocated to each process, which is referred to as "fine granularity."
    - Symmetric execution environment: Since you're virtualizing individual processes, all processes must run in the same operating system and environment. The execution environment for each process is the same (symmetric), meaning they share the same OS kernel, libraries, etc. There is no flexibility for running different OSes or environments.

Granularity 2/3

- Multiplex OSes
  - Allows heterogeneous environments
    - Virtualizing entire operating systems (like in full or hardware virtualization, e.g., Virtual Machines) allows each virtualized instance (or VM) to run its own OS. These OSes can be different from each other (heterogeneous environments). For example, you can run Linux on one VM and Windows on another, which provides flexibility that is not possible when only virtualizing processes.
  - High overhead of OS, runtime wastage and start-up latency
    - Virtualizing entire OSes has its drawbacks:
    - High overhead: Each VM requires its own instance of an OS, which consumes a significant amount of resources (CPU, memory, storage). The hypervisor (the layer managing the VMs) also adds overhead.
    - Runtime wastage: Some of the resources allocated to a virtualized OS may not always be fully utilized, leading to inefficiency.
    - Start-up latency: Starting a virtual machine can be slower compared to starting a single process because you're booting an entire OS rather than just loading a program in an already-running OS.

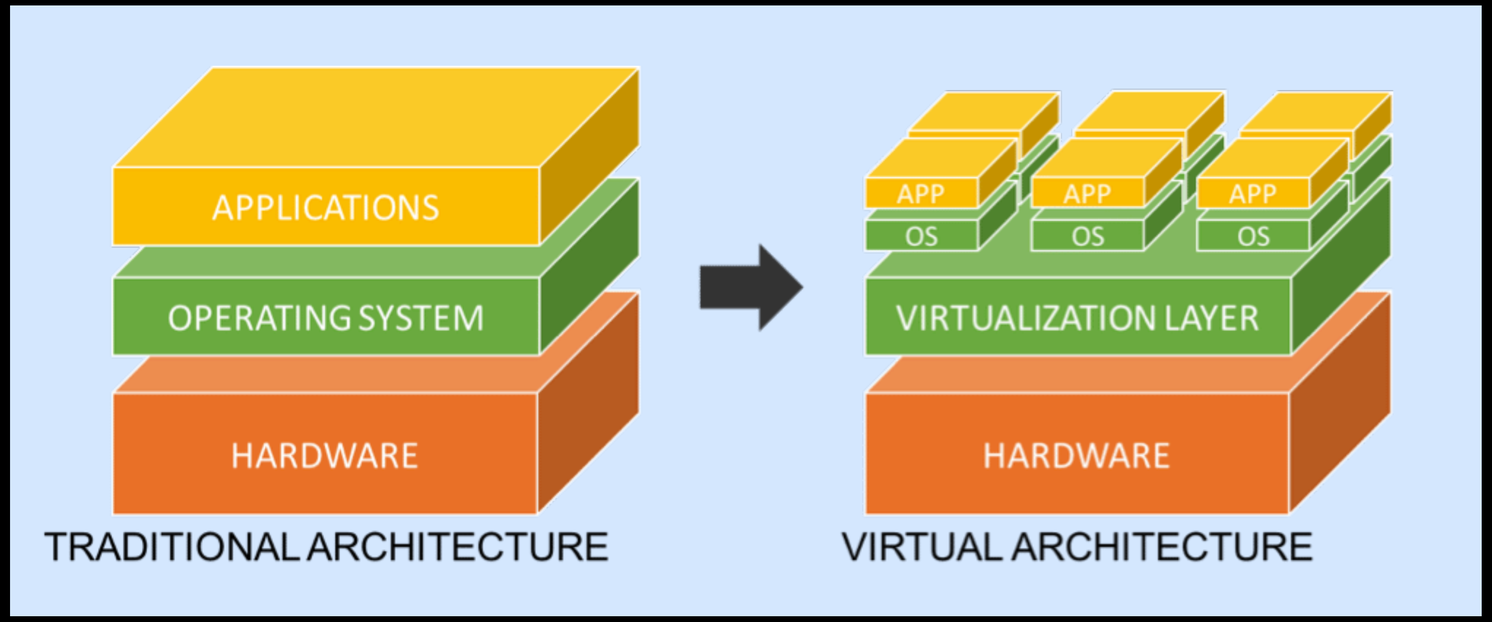
Granularity 3/3

In Summary:

- Virtualizing at the process level is more fine-grained but forces all processes to share the same environment, making it more efficient in terms of resource usage but less flexible.
- Virtualizing at the OS level offers flexibility with heterogeneous environments but comes with significant overhead and inefficiencies.

These trade-offs depend on the goals and requirements of the virtualization scenario.

Virtualisation



- Virtualisation, in computer technology, is the process of grouping and dividing the resources of a computer system into multiple execution environments by applying one or more methodologies or technologies such as hardware or software partitioning, partial or complete machine simulation, and/or emulation.
  - Virtualisation should have no operational impact on applications running under virtualisation
  - Normally operating application should experience no change to its operation when running under virtualisation

Main goals of virtualization

- Increased scalability: Virtualization allows organizations to increase the scalability of their IT infrastructure by running multiple virtual machines on a single physical server. This can help organizations to save money on hardware costs and improve the utilization of their existing resources.
- Improved agility: Virtualization can help organizations to improve the agility of their IT infrastructure by making it easier to deploy new applications and services. This can help organizations to respond more quickly to changing business needs.
- Enhanced security: Virtualization can help organizations to enhance the security of their IT infrastructure by isolating different virtual machines from each other. This can help to protect sensitive data from unauthorized access.
- Reduced costs: Virtualization can help organizations to reduce IT costs by consolidating multiple physical servers into a single virtual server. This can help to reduce hardware costs, power costs, and cooling costs.

The specific goals of virtualization will vary depending on the organization's needs. However, the goals listed above are some of the most common reasons why organizations choose to implement virtualization.

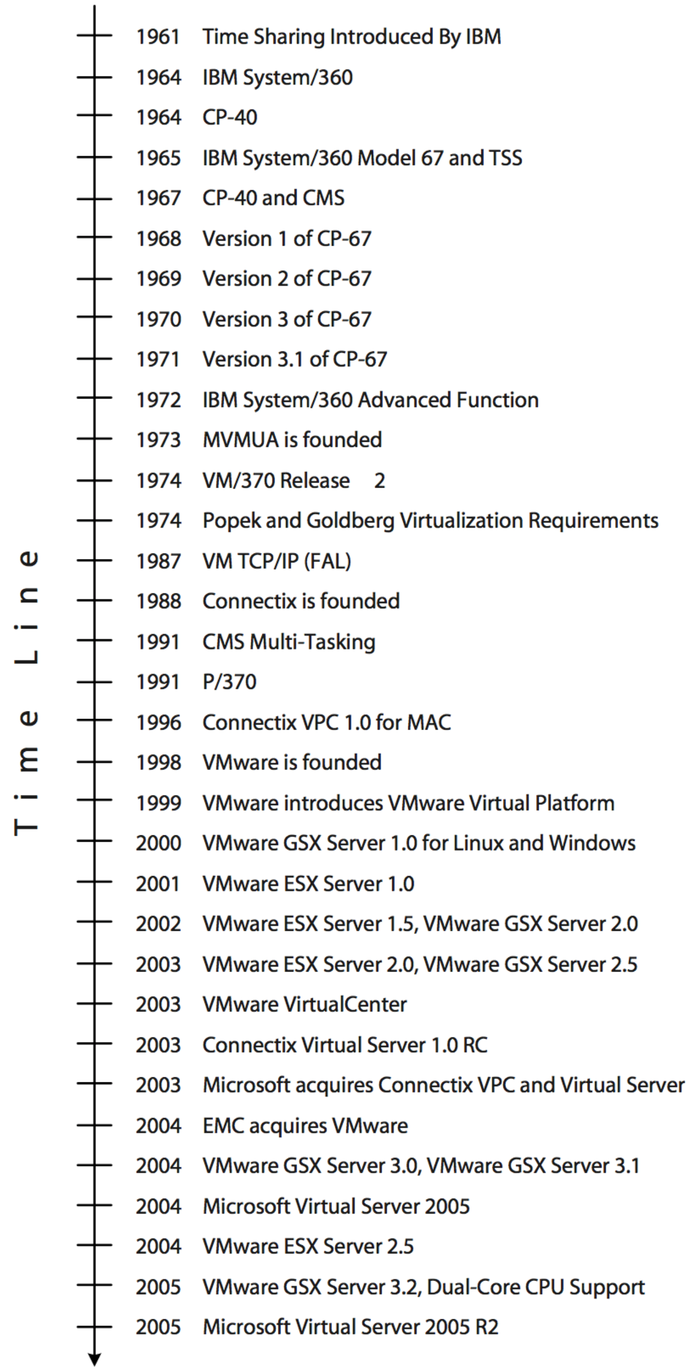
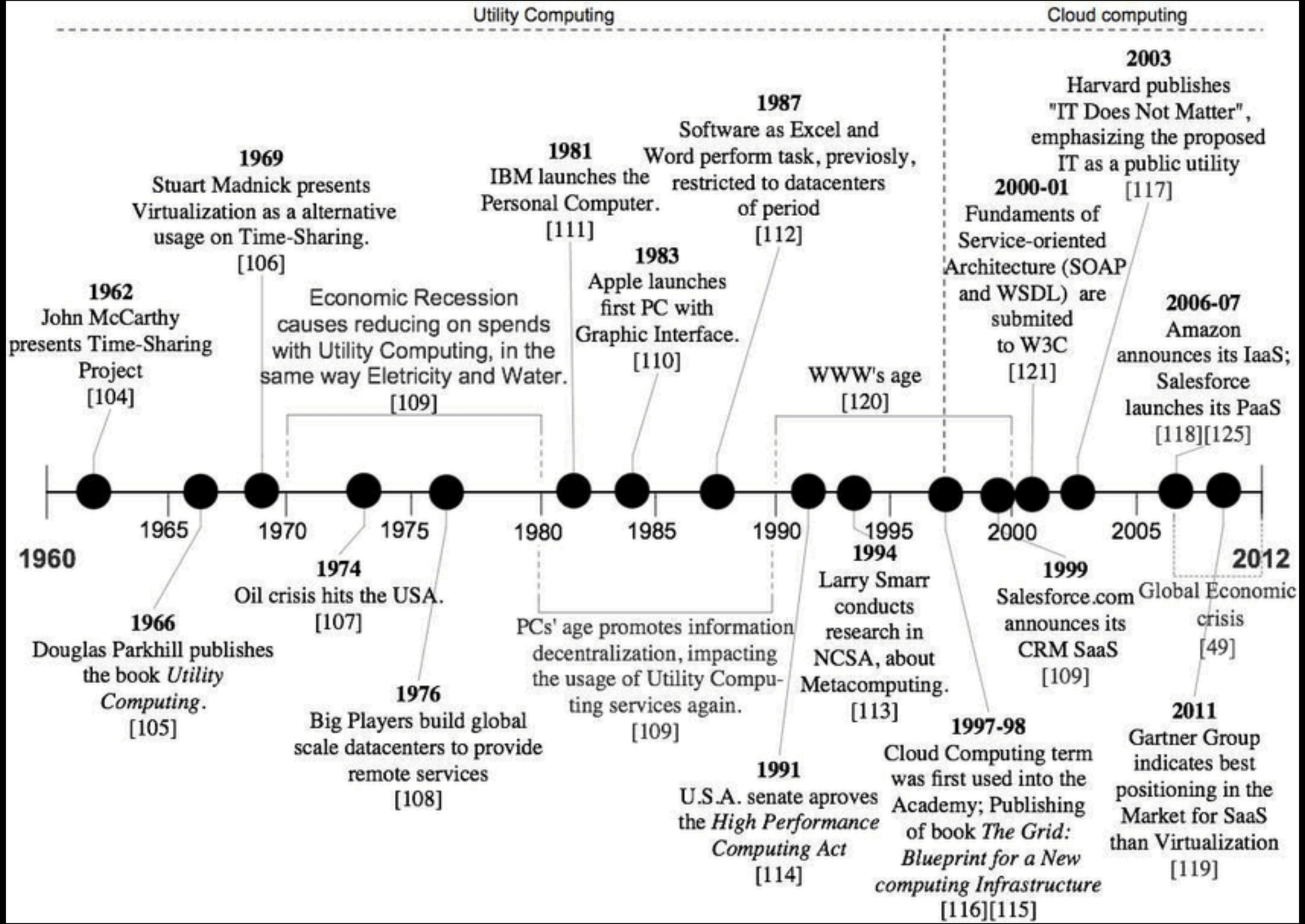
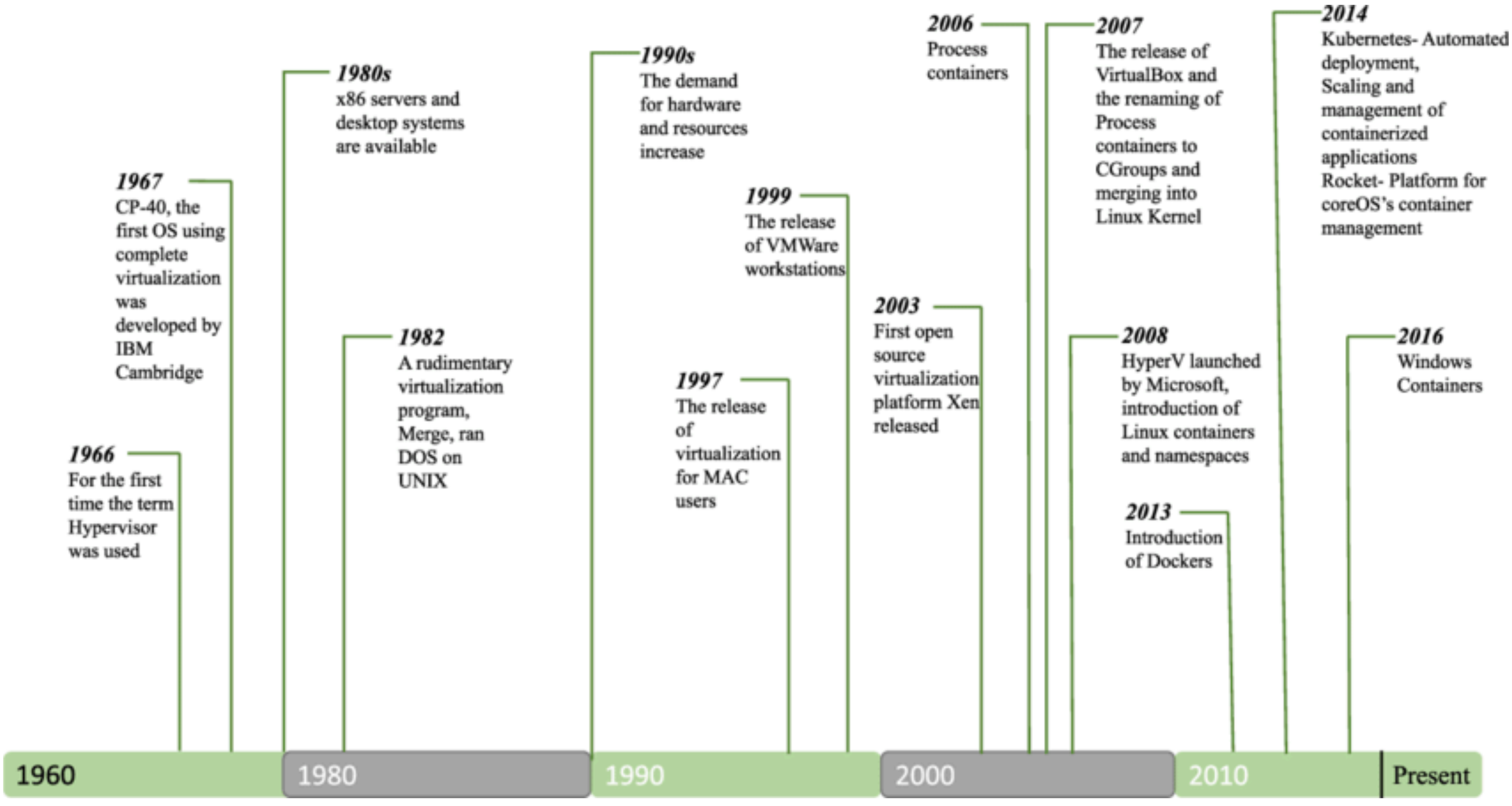
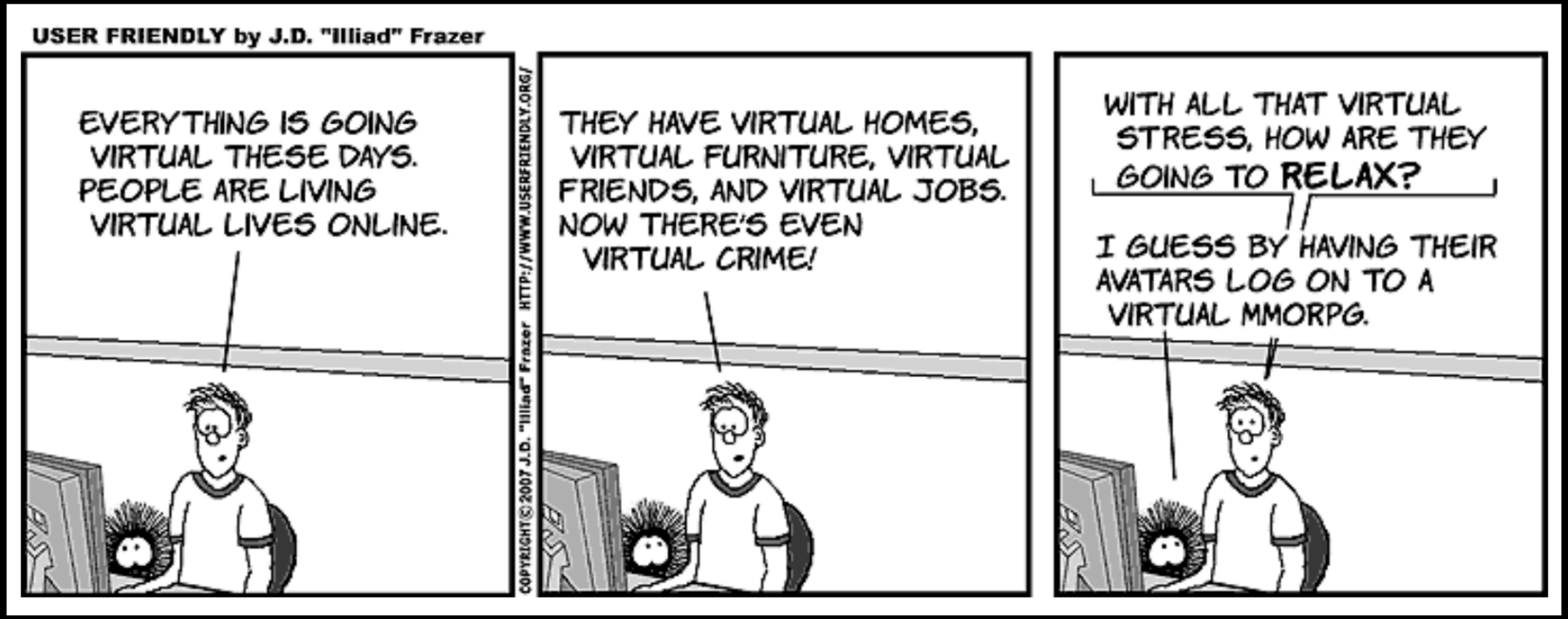


Figure 1.4 Virtualization History Timeline.

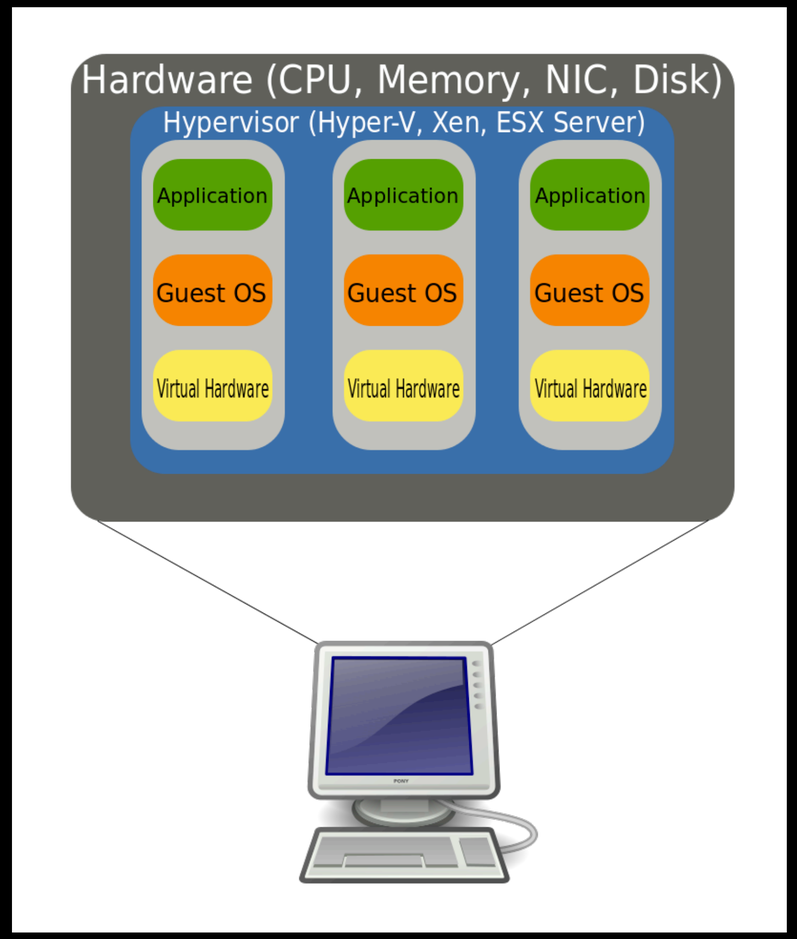

Utility Computing
Cloud computing
1960
1962
John McCarthy presents Time-Sharing Project
[104]
1965
1966
Douglas Parkhill publishes the book *Utility Computing*.
[105]
1969
Stuart Madnick presents Virtualization as a alternative usage on Time-Sharing.
[106]
1970
1974
Oil crisis hits the USA.
[107]
1975
1976
Big Players build global scale datacenters to provide remote services
[108]
Economic Recession causes reducing on spends with Utility Computing, in the same way Eletricity and Water.
[109]
1980
1981
IBM launches the Personal Computer.
[111]
1983
Apple launches first PC with Graphic Interface.
[110]
1985
1987
Software as Excel and Word perform task, previosly, restricted to datacenters of period
[112]
1990
PCs' age promotes information decentralization, impacting the usage of Utility Computing services again.
[109]
1991
U.S.A. senate aproves the *High Performance Computing Act*
[114]
WWW's age
[120]
1994
Larry Smarr conducts research in NCSA, about Metacomputing.
[113]
1995
1997-98
Cloud Computing term was first used into the Academy; Publishing of book *The Grid: Blueprint for a New computing Infrastructure*
[116][115]
2000
2000-01
Fundaments of Service-oriented Architecture (SOAP and WSDL) are submited to W3C
[121]
1999
Salesforce.com announces its CRM SaaS
[109]
2003
Harvard publishes "IT Does Not Matter", emphasizing the proposed IT as a public utility
[117]
2005
2006-07
Amazon announces its IaaS; Salesforce launches its PaaS
[118][125]
2011
Gartner Group indicates best positioning in the Market for SaaS than Virtualization
[119]
2012
Global Economic crisis
[49]


1966
For the first time the term Hypervisor was used
1967
CP-40, the first OS using complete virtualization was developed by IBM Cambridge
1980s
x86 servers and desktop systems are available
1982
A rudimentary virtualization program, Merge, ran DOS on UNIX
1990s
The demand for hardware and resources increase
1997
The release of virtualization for MAC users
1999
The release of VMWare workstations
2003
First open source virtualization platform Xen released
2006
Process containers
2007
The release of VirtualBox and the renaming of Process containers to CGroups and merging into Linux Kernel
2008
HyperV launched by Microsoft, introduction of Linux containers and namespaces
2013
Introduction of Dockers
2014
Kubernetes- Automated deployment, Scaling and management of containerized applications
Rocket- Platform for coreOS's container management
2016
Windows Containers
1960
1980
1990
2000
2010
Present

Virtualization



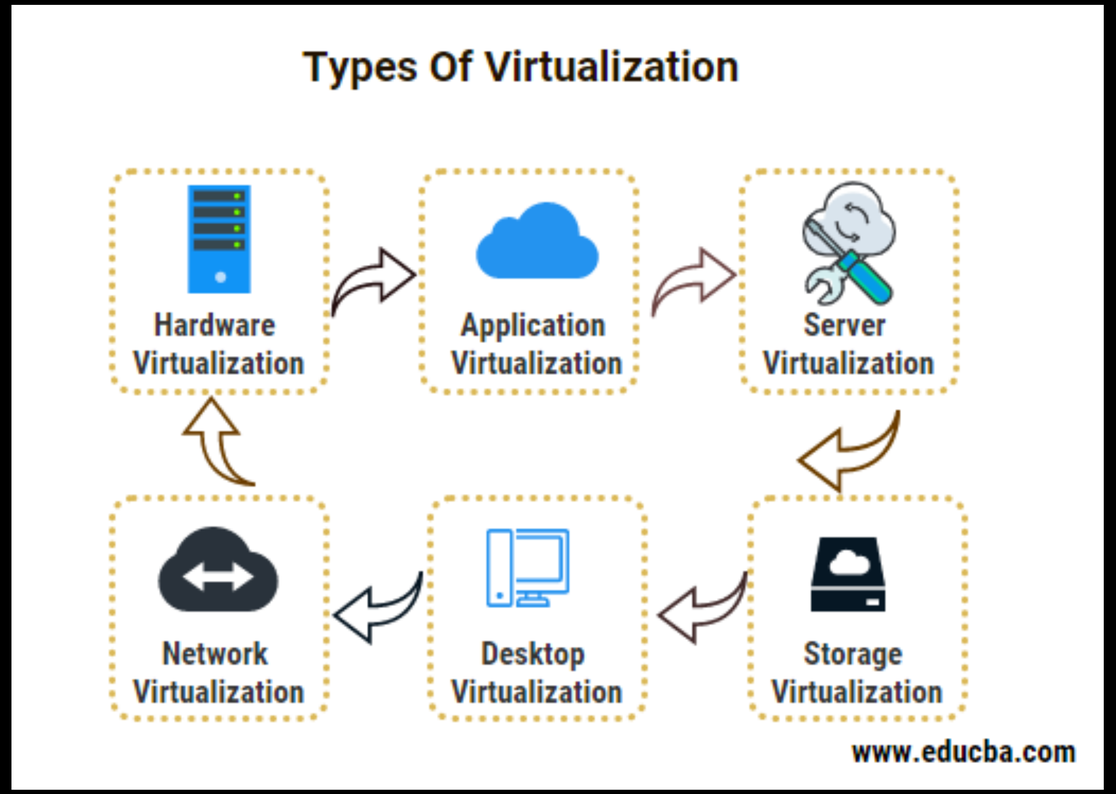
what virtualisation* is?



- Virtualisation < ~~cloud computing~~
- virtualisation separates operating system from underlaying hardware
- cloud computing separates application from the underlaying hardware
- hypervisor layer on top of operating system

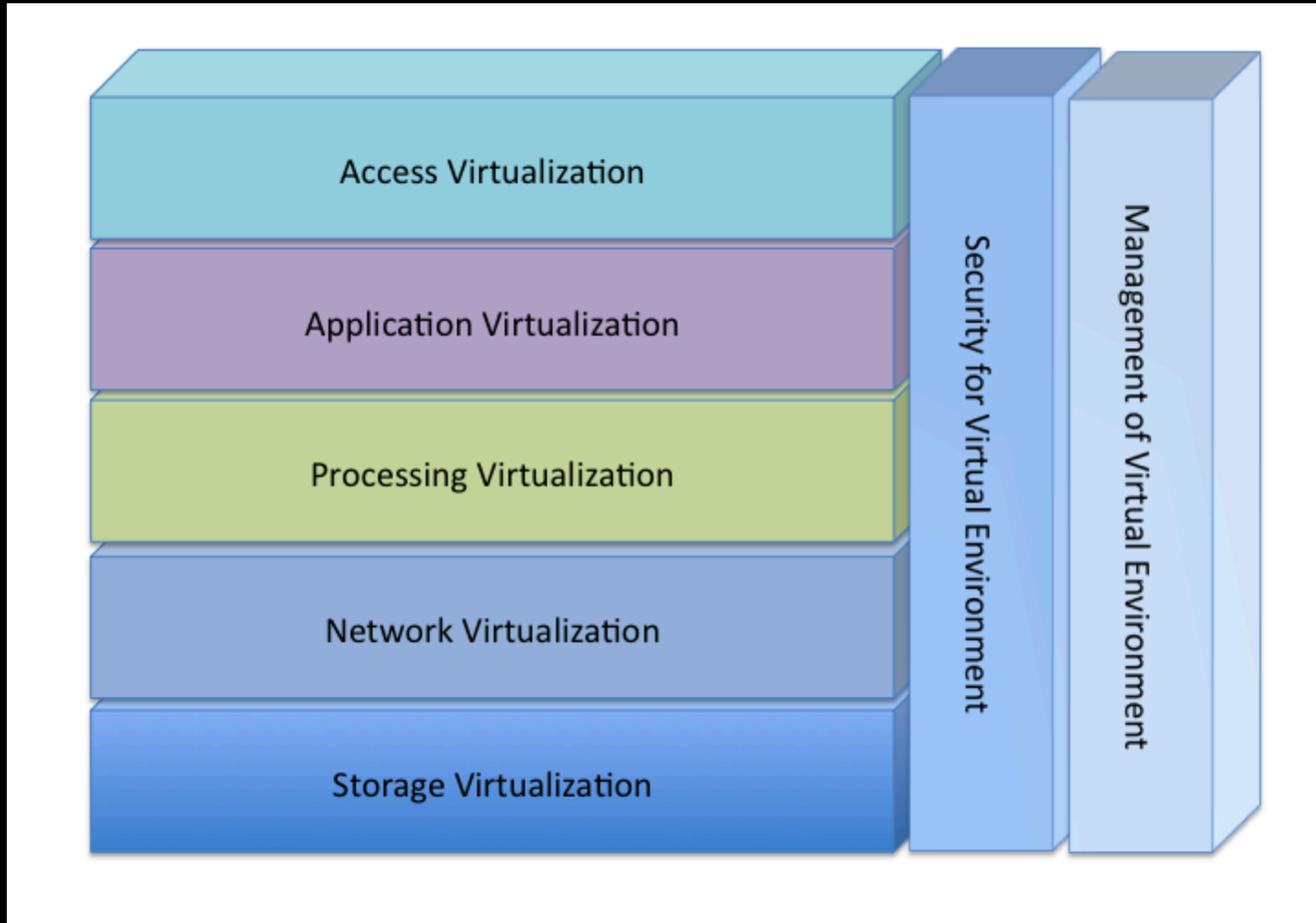
Types of virtualization

- Server Virtualization
- Client & Desktop Virtualization
- Services and Applications Virtualization
- Network Virtualization
- Storage Virtualization



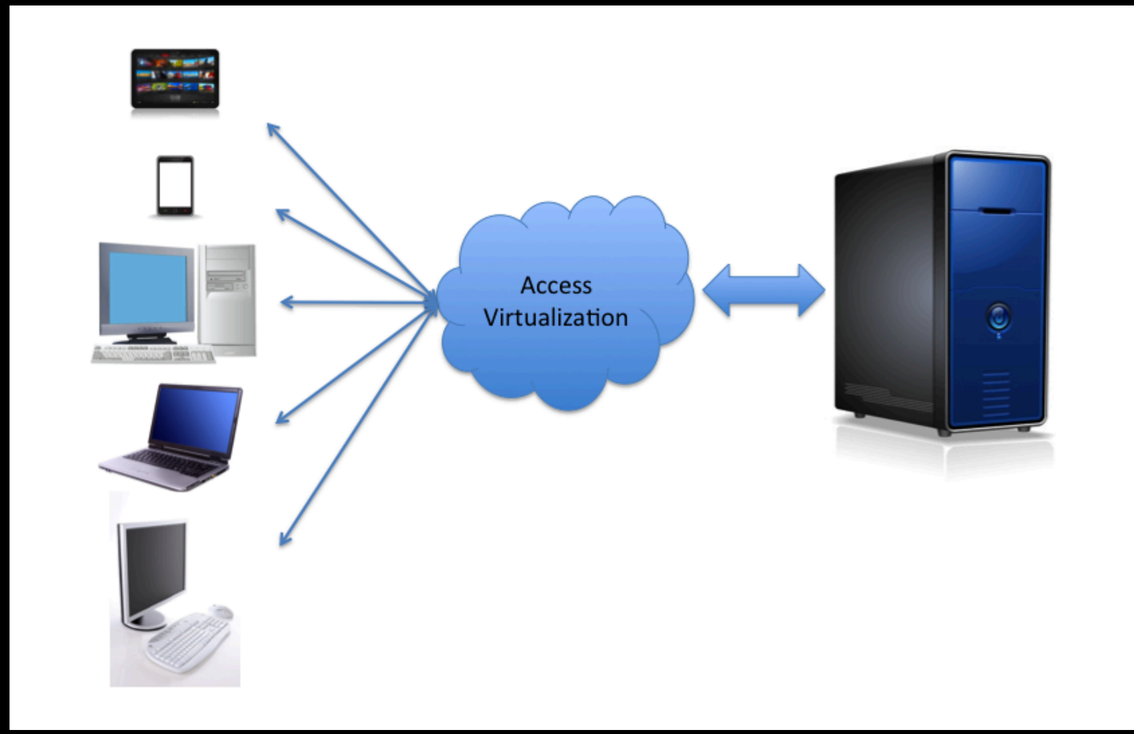
Types of Virtualization

- Access Virtualization—aka Virtual Machines
- Application-Server Virtualization
- Application Virtualization
- Administrative Virtualization
- Network Virtualization
- Hardware Virtualization
- Storage Virtualization



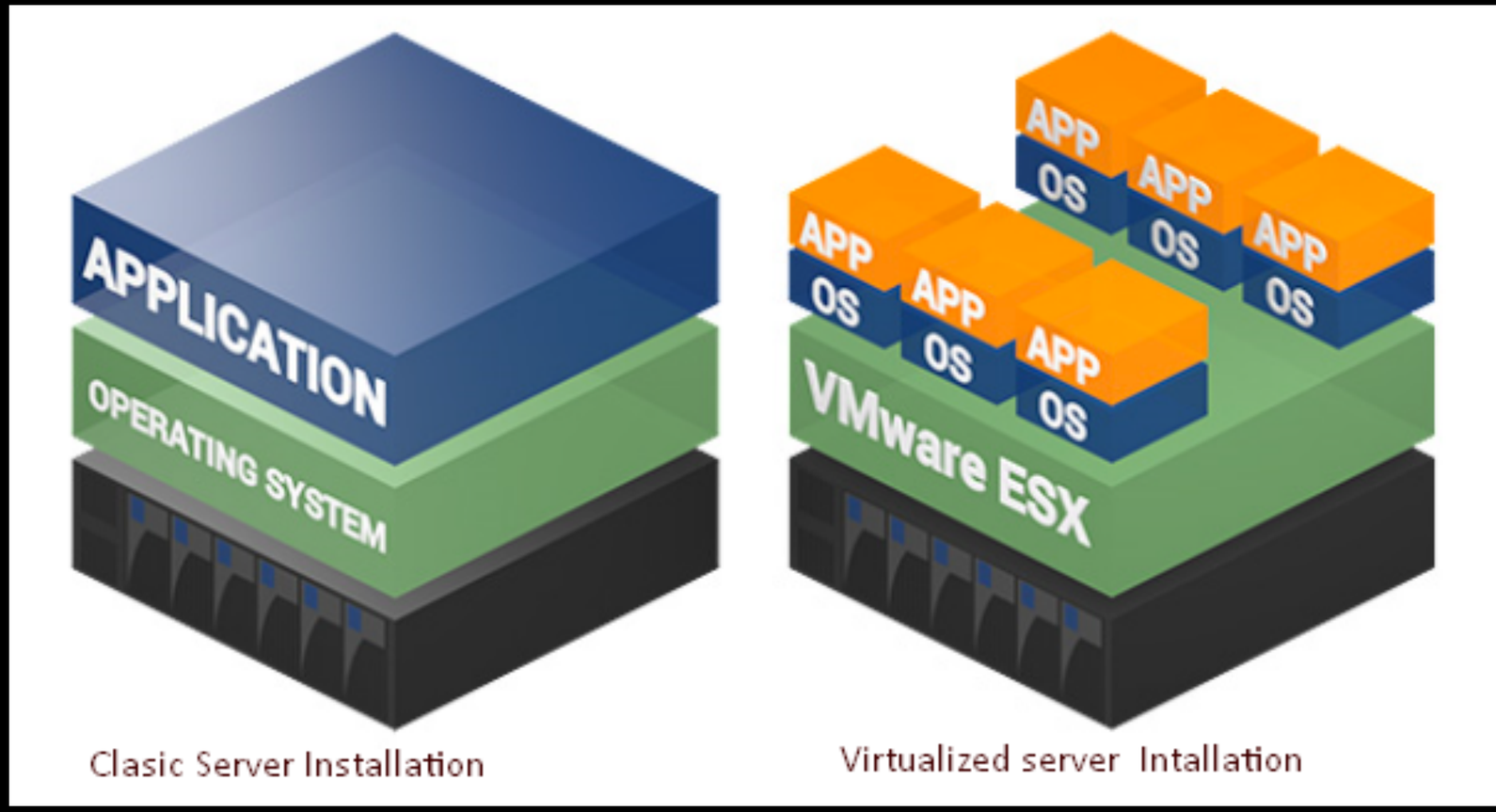
Access virtualization

Hardware and software technology that allows nearly any device to access any application without either having to know too much about the other. The application sees a device it's used to working with. The device sees an application it knows how to display.



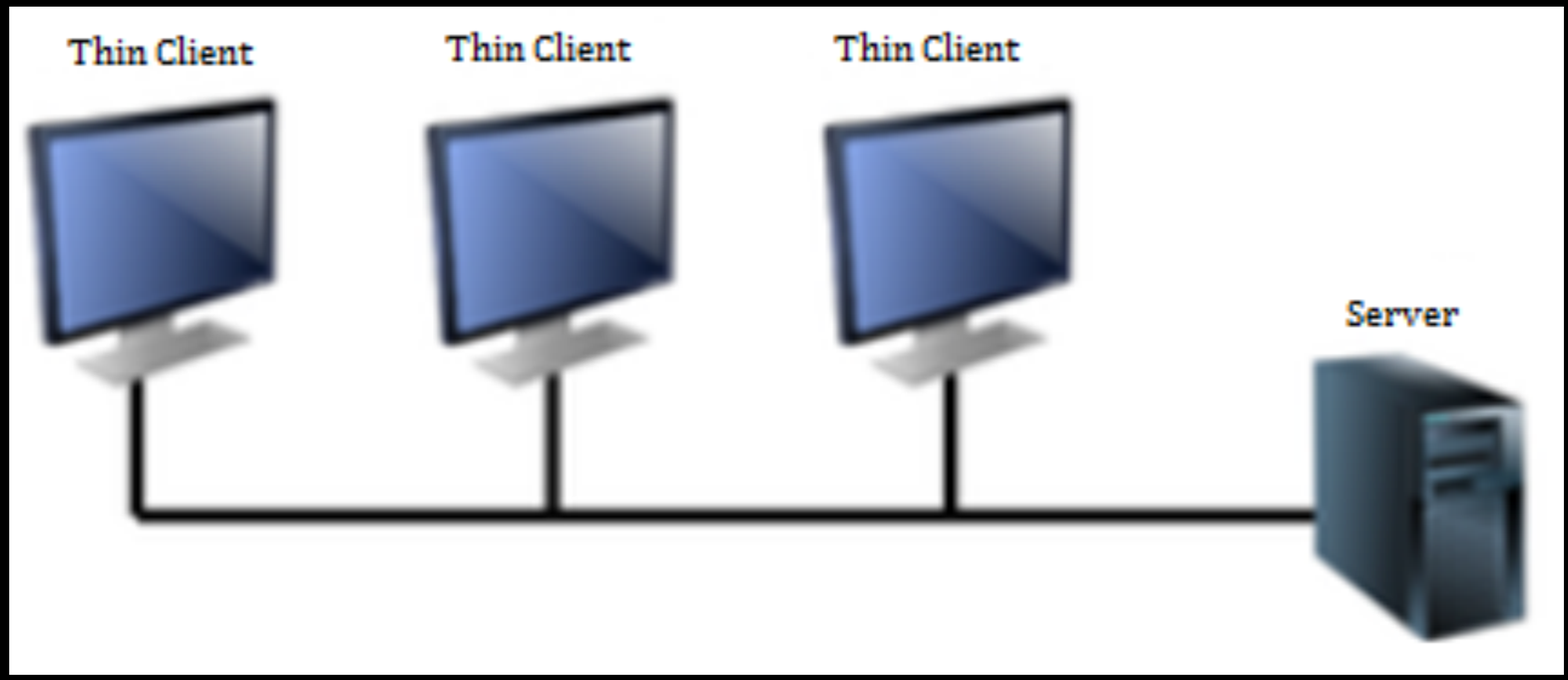
Server virtualization

It is virtualizing your server infrastructure where you do not have to use any more physical servers for different purposes.



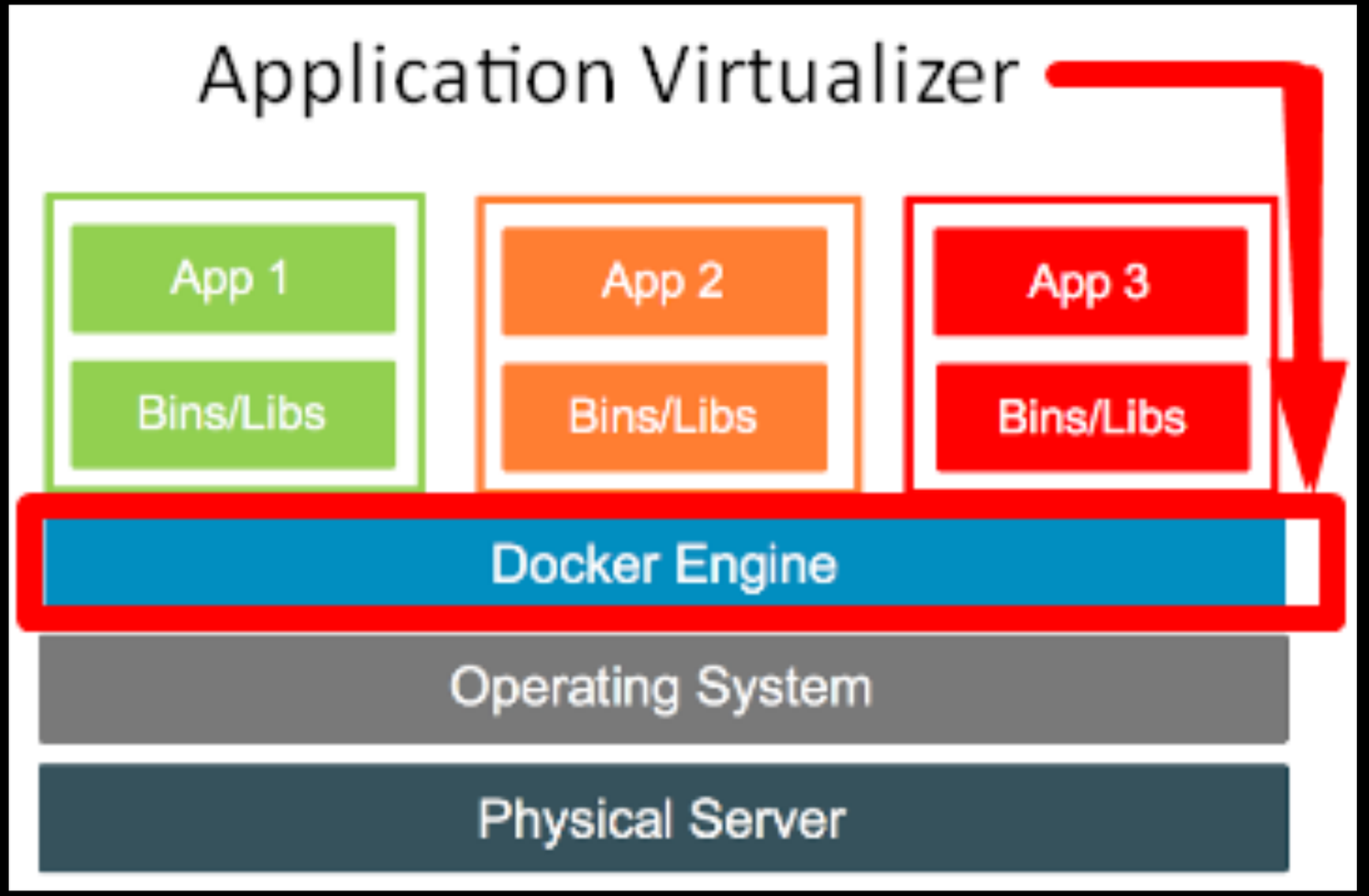
Client & Desktop Virtualization

This is on the user's site where you virtualize their desktops.
Desktops are changed to thin clients and are utilizing the datacenter resources.



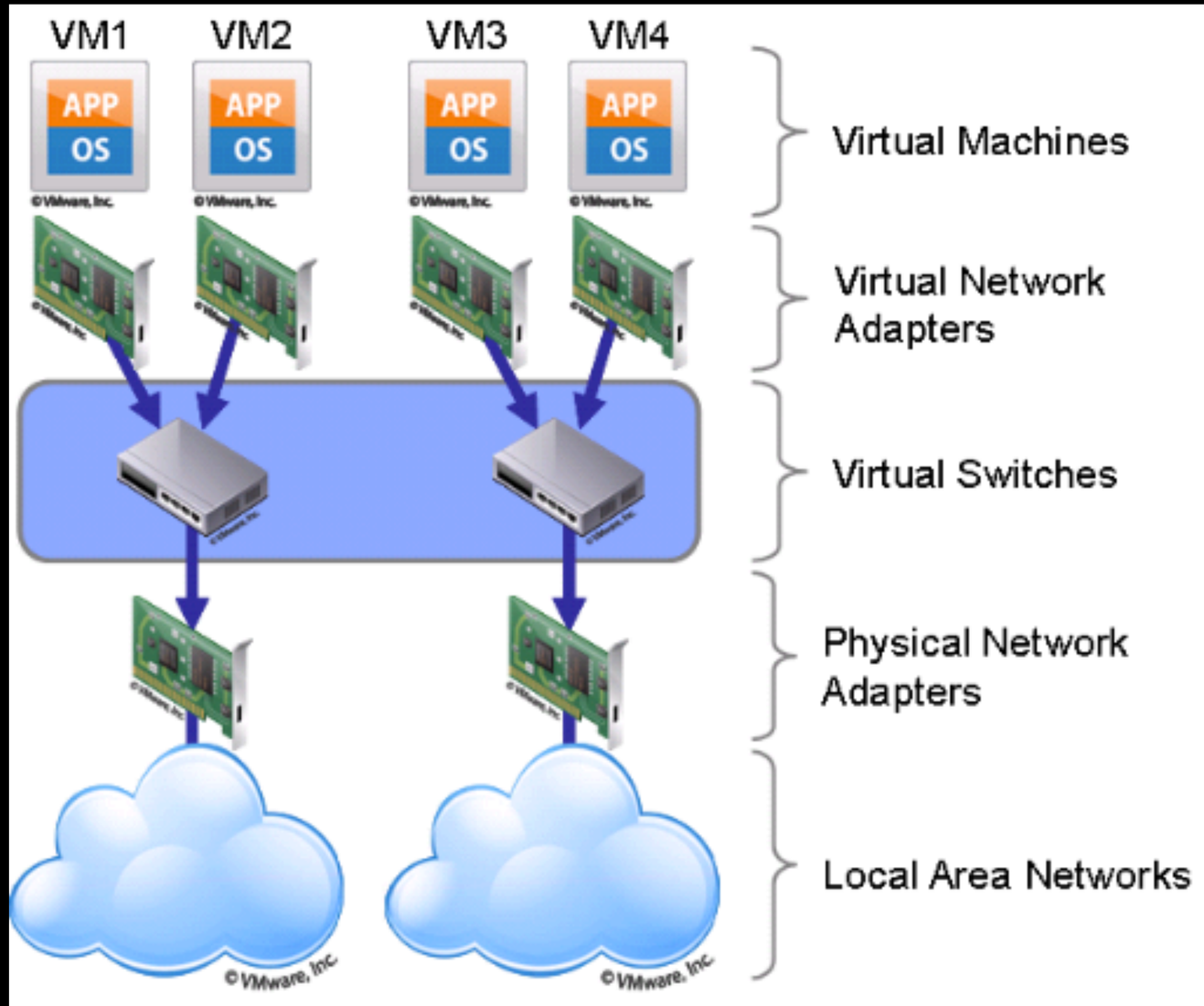
Services and Application Virtualization

The virtualization technology isolates applications from the underlying operating system and from other applications, in order to increase compatibility and manageability.



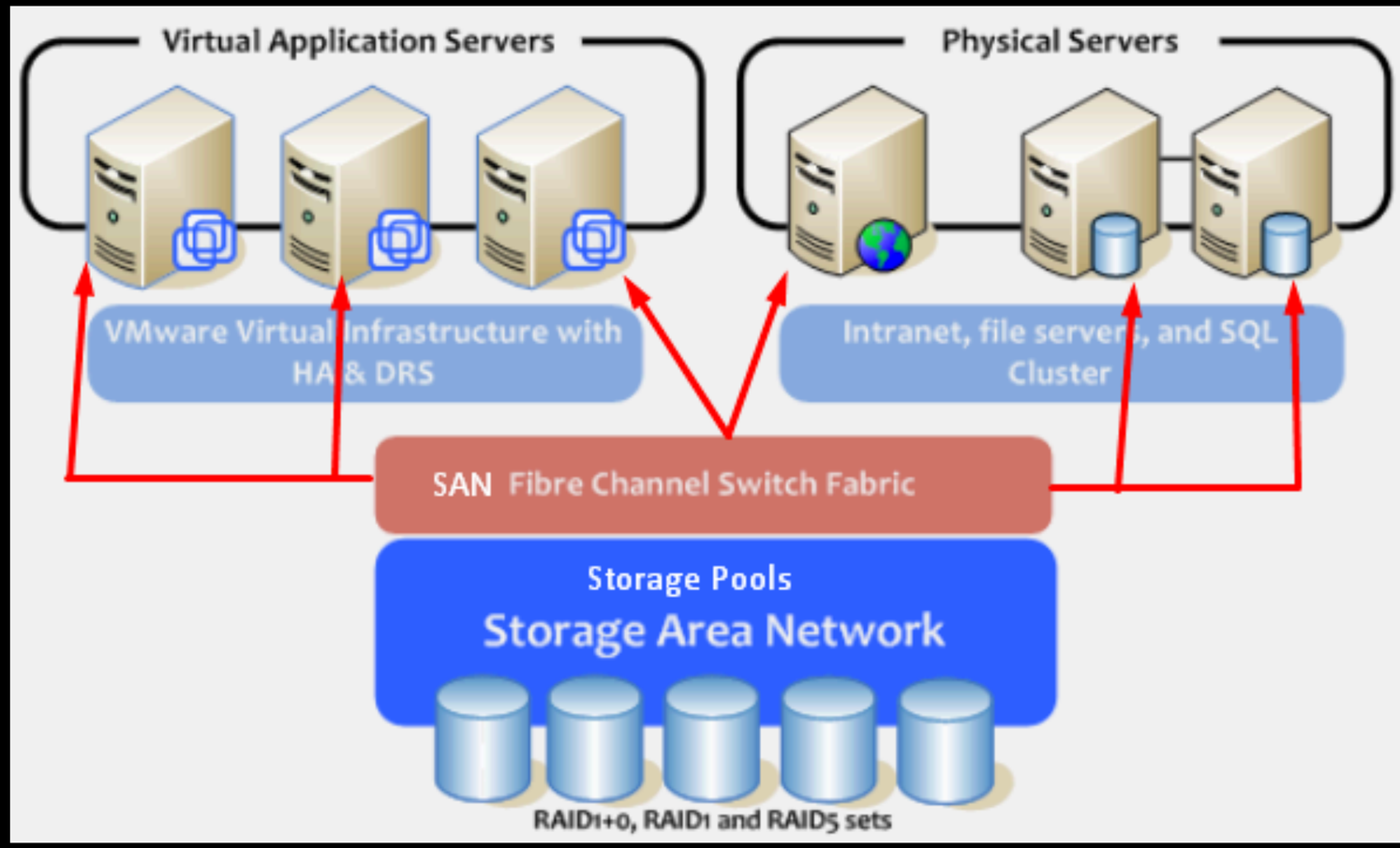
Network Virtualization

Part of virtualization infrastructure, which is used especially to visualize network servers.
It helps to create multiple switches, Vlans, NAT-ing, etc

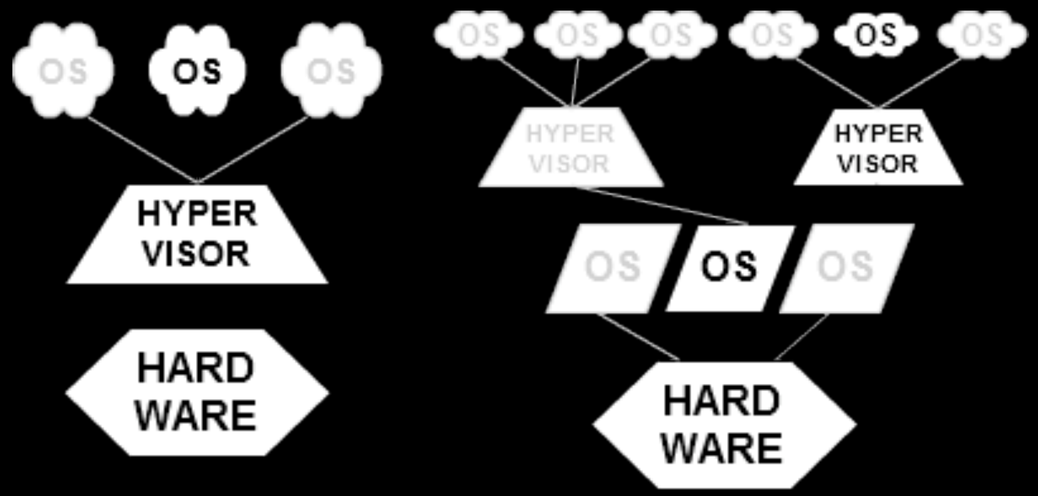


Storage Virtualization

Widely used in datacenters with big storages
Helps to create, delete, allocated storage to different hardware.
This allocation is done through network connection. The leader on storage is SAN.

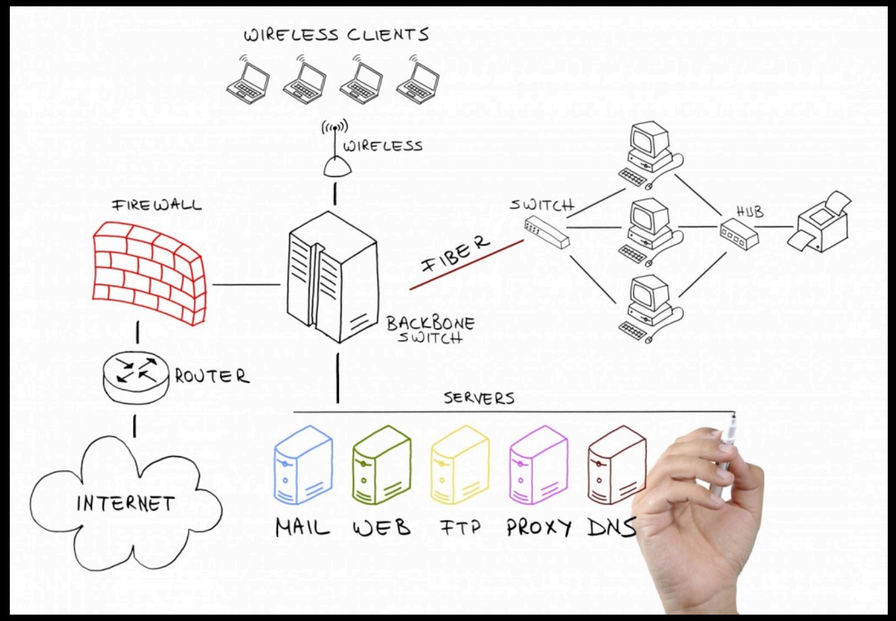


Hypervisor



- Type-1, native or bare-metal hypervisors
  - Run directly on the **host's** hardware to control the hardware and to manage guest operating systems.
  - For this reason, they are sometimes called bare metalhypervisors.
- Type-2 or hosted hypervisors
  - Run on a conventional operating system (OS) just as other computer programs do.
  - A guest operating system runs as a process on the host.
  - Abstract guest operating systems from the host operating system.

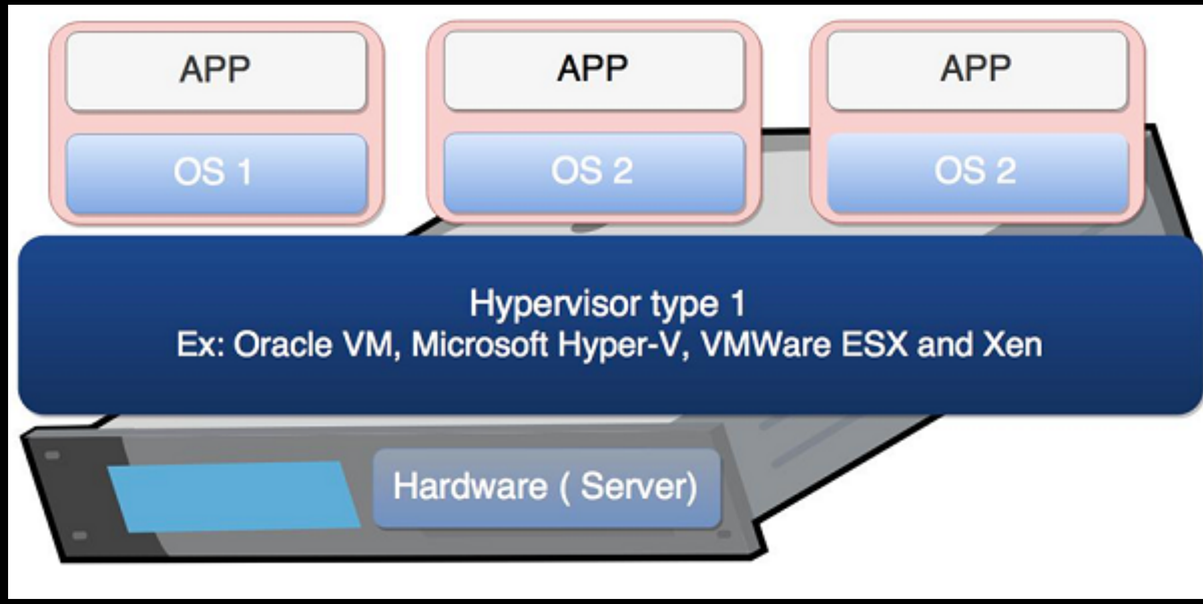
What is IT infrastructure?



- … is defined broadly as a set of information technology (IT) components that are the foundation of an IT service; typically physical components (computer and networking hardware and facilities), but also various software and network components

virtualized infrastructure

- type-1 hypervisor
- decrease vulnerability
- enhanced availability with clusters

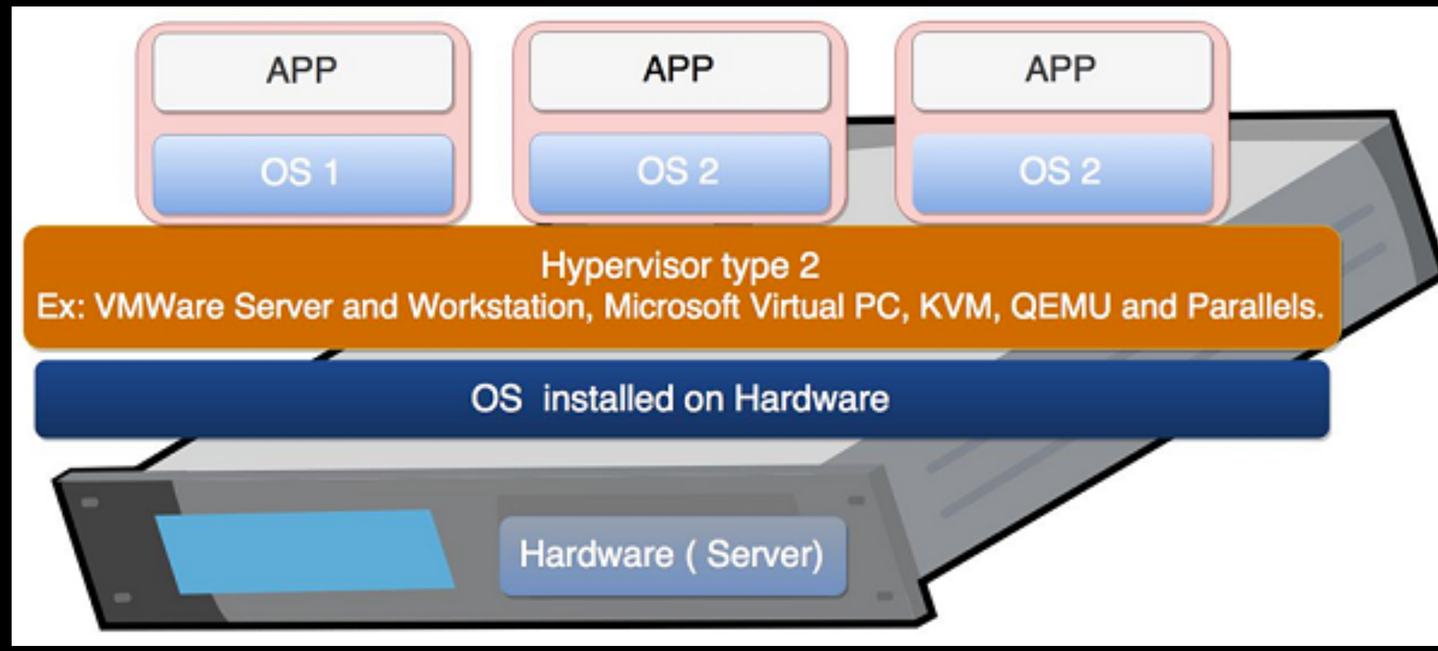


Virtualmachine management software

- for management of type 1 hyperwizer
- can handle several hyperwizers
- can turn on/off virtual machine instances based on load
- move instances from one VM to another
- can launch new VM if some errors come out (fault tolerance)
- can over allocate resources

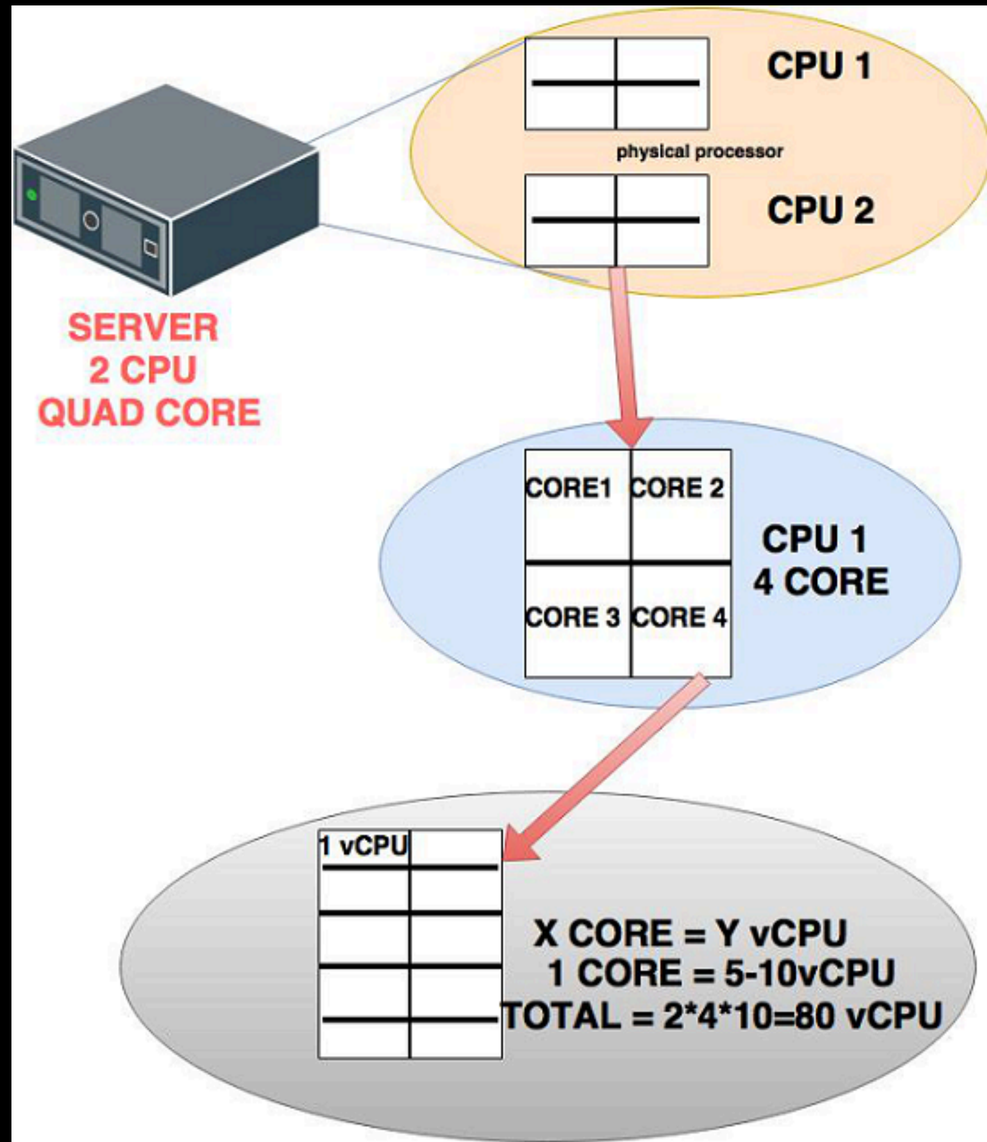
Hosted hypervisors

- Hardware has operating system
- on top of that there is host operating system
- on top of that there is hosted hypervisor
- on top of that there are guest operating systems

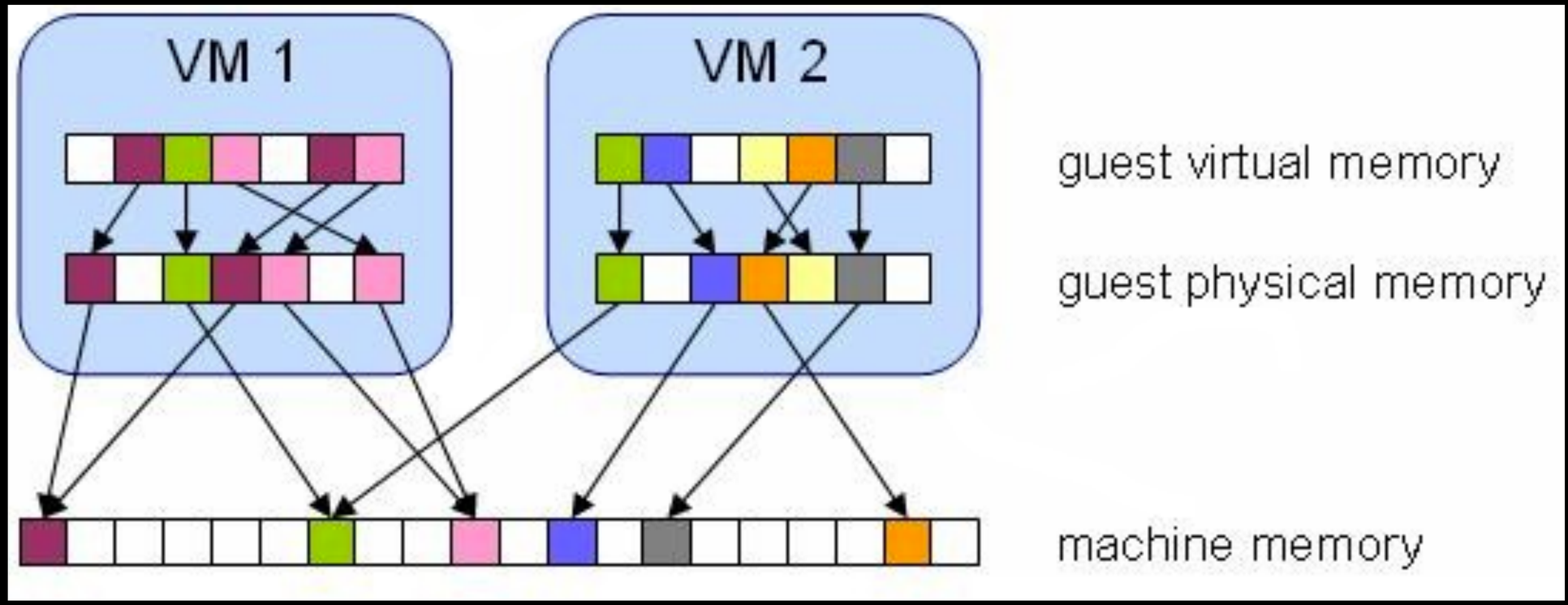


CPU virtualization

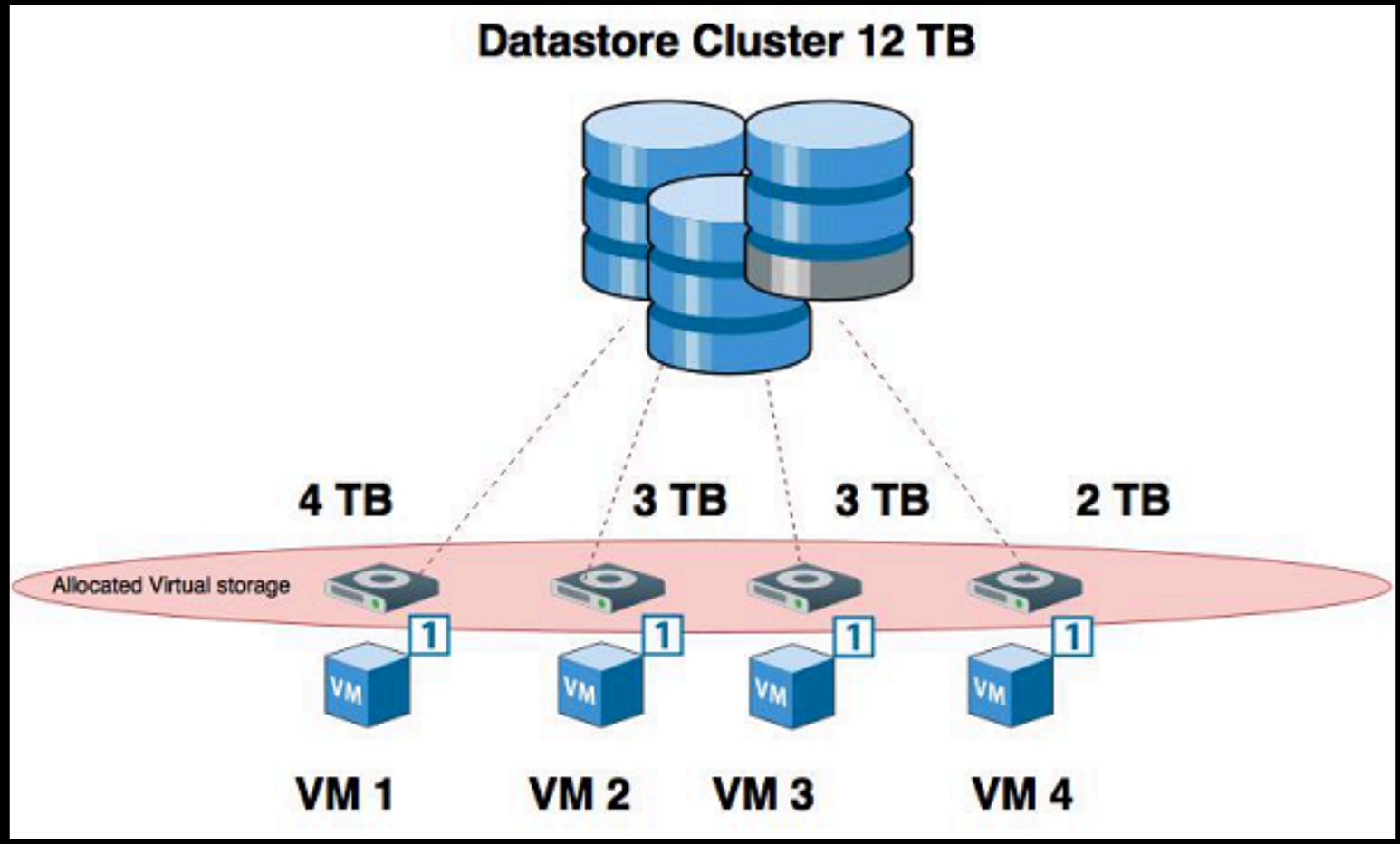
- Hypervisor is abstacting physical CPU into virtual CPUs.
- Multiple VMs allows to "time share" a given physical processor core.



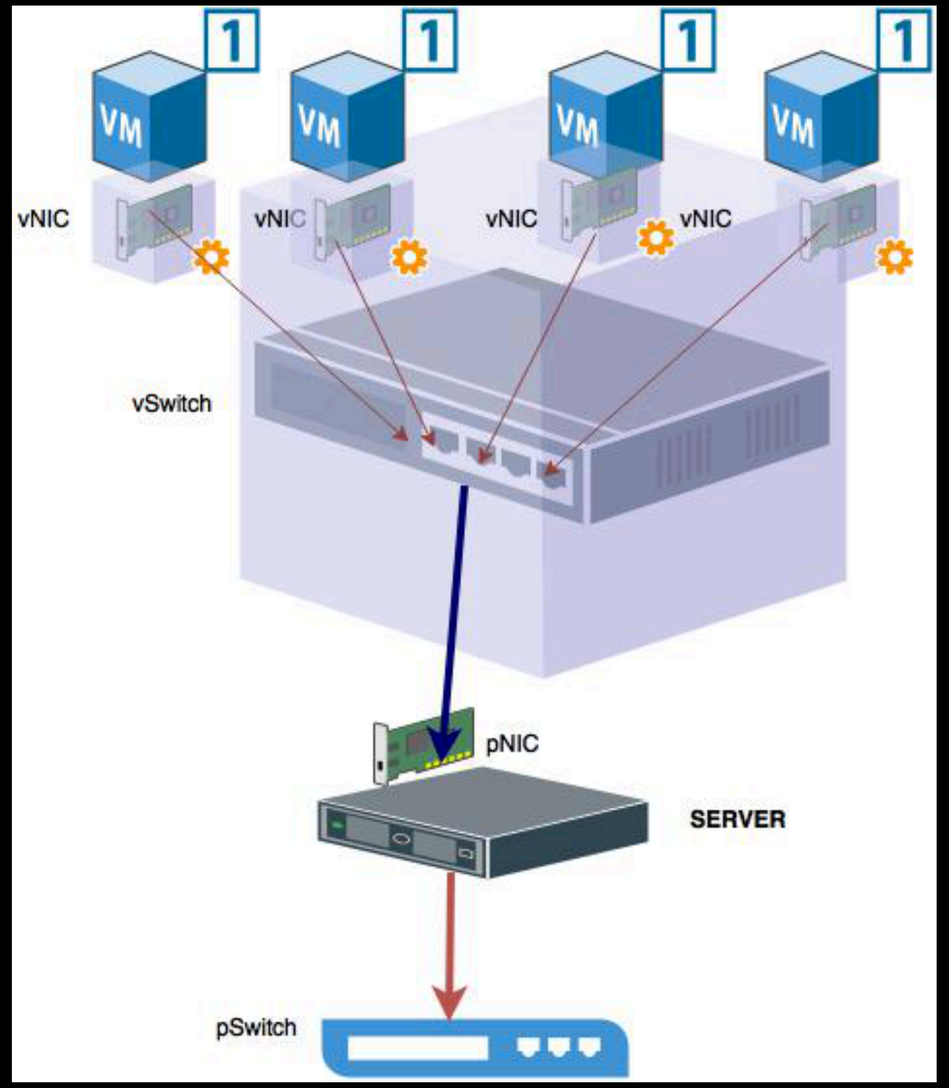
Memory virtualization



Storage virtualization



Network virtualization



Forms of Virtualization

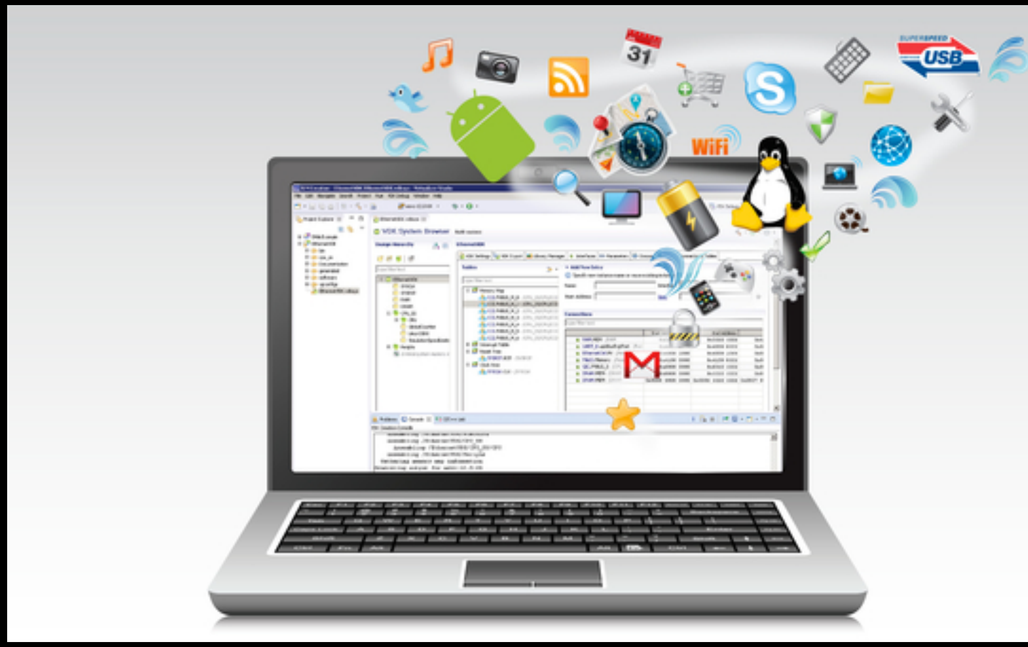
- Application
  - Configuration of resources so that one application may appear to be more than one instance.
  - Multiple personalities and configurations
- Emulation
  - Performed through emulation of a machine's instructions.
  - Tends to be quite slow and inefficient.
  - Can be extremely versatile.
  - Can virtualize foreign hardware and architectures.
- Virtual Machine Interpreters
  - These interpret their own specific language.
  - Generally have a virtual environment "sandbox".
- API Virtualization
  - Replaces the API layer with a virtual subsystem library.
  - Applications execute native machine instructions. - Replaces foreign API libraries with libraries and interfaces to the host facilities and libraries.

Choosing the Right Virtualization

- Operating System Virtualization is highly efficient

Virtualization software

- meant for moving existing servers to virtual environment
- virtual appliances to create your own virtual machines (preinstalled OS or routers or firewalls or swiches)



benefits of virtualization from the **business** point of view

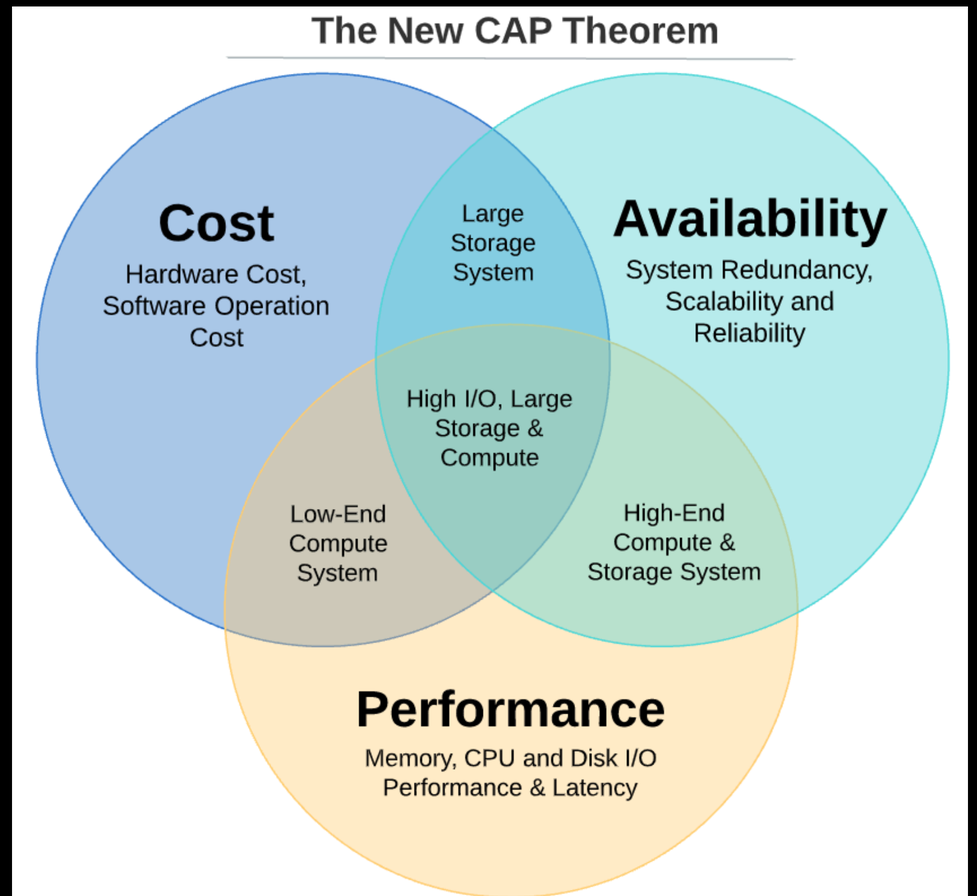
- Reduction in hardware costs and maintenance
- Easier to manage, maintain and upgrade
- Faster recovery (back in business)
- Multiple platforms on one server
- Vendor independency
- Environment friendly
- Faster server provisioning, increased flexibility
- Easier access from anywhere
- More secure applications
- Longer technology lifespans

Benefits of virtualization from the **datacenter** point of view

- Less heat buildup
- Reduced cost
- Efficient HW utilization
- Faster redeploy
- Easier backups
- Greener pastures
- Better testing
- No vendor lock-in
- Better disaster recovery
- Single-minded servers
- Easier migration to cloud
- Adaptive computing

Advantages of virtualization

- Using Virtualization for Efficient Hardware Utilization



Advantages of virtualization

- Using Virtualization for Efficient Hardware Utilization
- Using Virtualization to Increase Availability

Advantages of virtualization

- Using Virtualization for Efficient Hardware Utilization
- Using Virtualization to Increase Availability
- Disaster Recovery

Advantages of virtualization

- Using Virtualization for Efficient Hardware Utilization
- Using Virtualization to Increase Availability
- Disaster Recovery
- Save Energy

Advantages of virtualization

- Using Virtualization for Efficient Hardware Utilization
- Using Virtualization to Increase Availability
- Disaster Recovery
- Save Energy
- Deploying Servers too fast

Advantages of virtualization

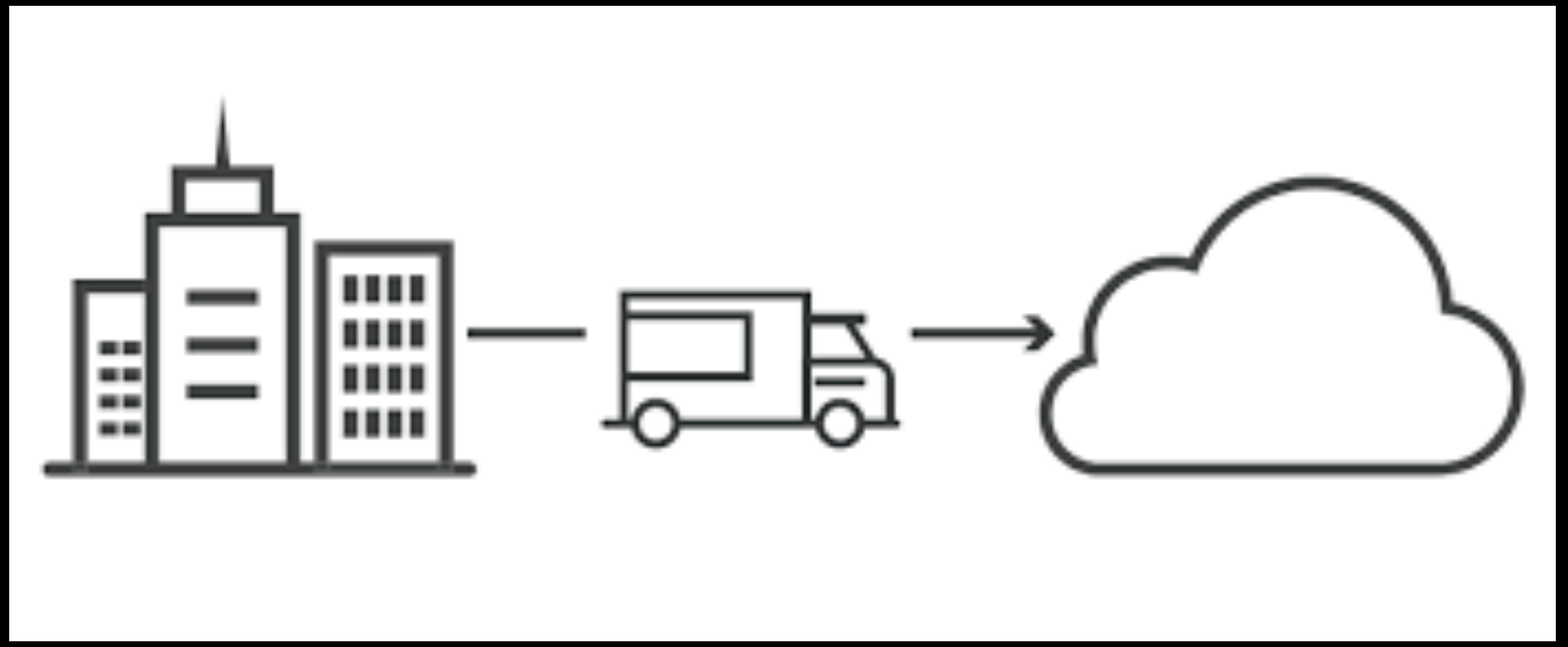
- Using Virtualization for Efficient Hardware Utilization
- Using Virtualization to Increase Availability
- Disaster Recovery
- Save Energy
- Deploying Servers too fast
- Save Space in your Server Room or Datacenter

Advantages of virtualization

- Using Virtualization for Efficient Hardware Utilization
- Using Virtualization to Increase Availability
- Disaster Recovery
- Save Energy
- Deploying Servers too fast
- Save Space in your Server Room or Datacenter
- Testing and setting up Lab Environment

Advantages of virtualization

- Shifting all your Local Infrastructure to Cloud in a day



Advantages of virtualization

- Shifting all your Local Infrastructure to Cloud in a day
- Possibility to Divide Services

Advantages of virtualization

- Shifting all your Local Infrastructure to Cloud in a day
- Possibility to Divide Services
- Disadvantages of Virtualization
  - Extra Costs

Advantages of virtualization

- Shifting all your Local Infrastructure to Cloud in a day
- Possibility to Divide Services
- Disadvantages of Virtualization
  - Extra Costs
  - Software Licensing

Advantages / disadvantages of virtualization

- Shifting all your Local Infrastructure to Cloud in a day
- Possibility to Divide Services
- Disadvantages of Virtualization
  - Extra Costs
  - Software Licensing
  - Learn the new Infrastructure

Virtualization

- Increased Security
- Managed Execution
- Sharing
- Aggregation
- Emulation
- Isolation
- Portability

Virtualization Improves Security

- Virtualization helps compartmentalize functionality
- Functions which use to share a common machine can be isolated
- Web server compromise doesn't lead to DNS compromise, etc.
- Security services can be isolated from public services.
- Security monitoring can be “out of band” to what's monitored.
- Independent security subsystems can be “bolted on”

Virtualization Degrades Security

- Merely consolidating systems does nothing to improve security
- Adding a hypervisor layer and addition OS complicates things
- Complexity is the enemy of security
- Hypervisors may have their own vulnerabilities
- Attacks between virtual machines are possible
- Networking becomes much more complicated