Cloud Computing

What is Cloud Computing?

Cloud service models

Cloud deployment models

Goals/benefits

Risks/Challenges

Commercial Products

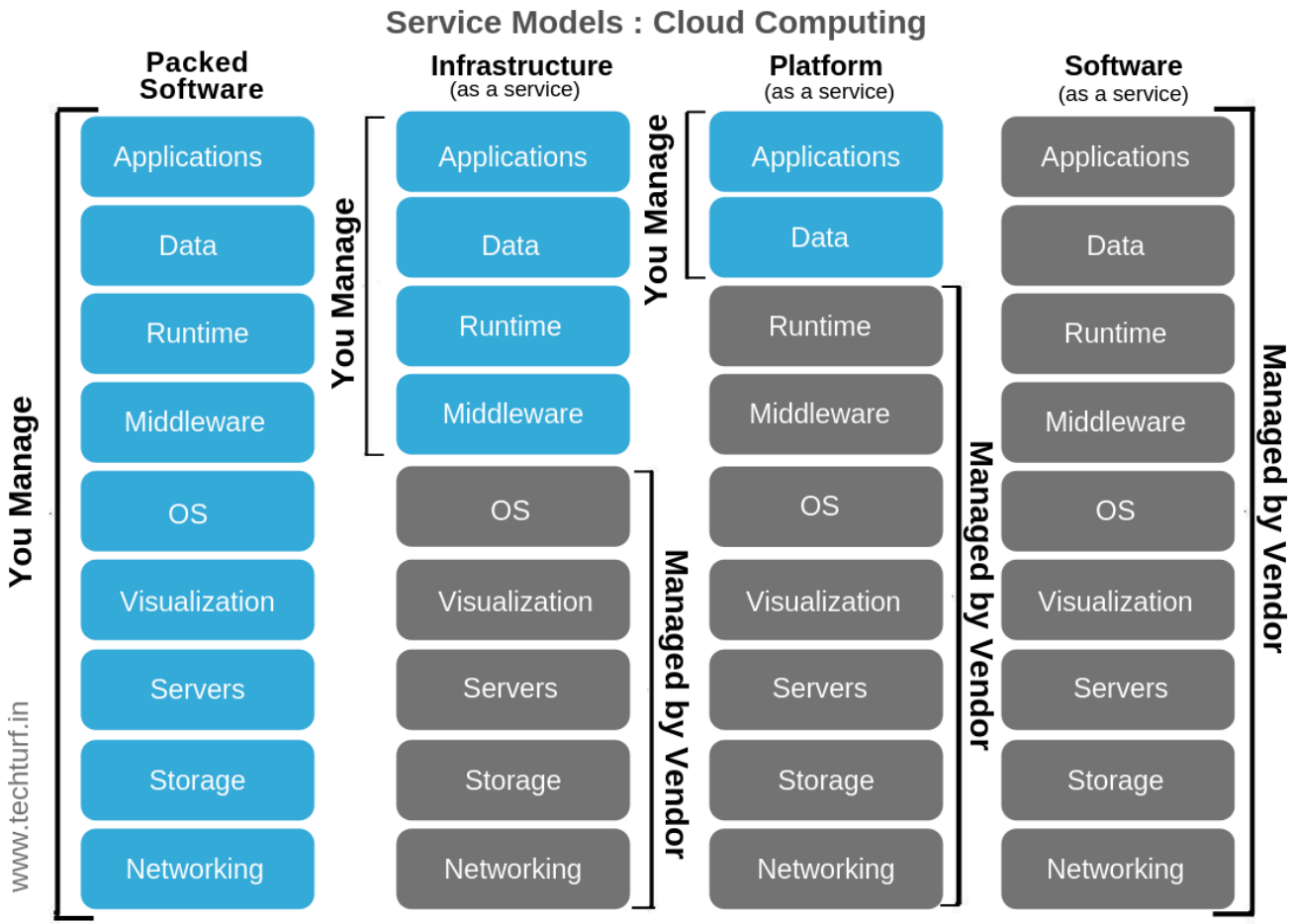
Enabling Technologies

Parallelisation: Amdahl's Law

Virtualisation

Cloud architecture

Yet another picture…



Commercial Products

Amazon AWS
Microsoft Windows Azure
Google Cloud/Google App Engine

Azure Overview (IaaS/PaaS)

- VMs, Linux, Windows
- Storage. e.g. NoSQL Tables
- SQL Database
- Analytics
- Security
- etc.

The New York Times Amazon AWS Example

In 2007 NY Times wanted to put all its archives online – 11 million articles, dating back to 1851

4TB of TIFF files (multiple TIFFs per article) needed to be converted to pdfs

100 VMs on EC2, stored data on S3

Time: 24 hours, cost: ~$500

“By leveraging the power of AWS and Hadoop, we were able to utilize hundreds of machines concurrently and process all the data in less than 36 hours.”

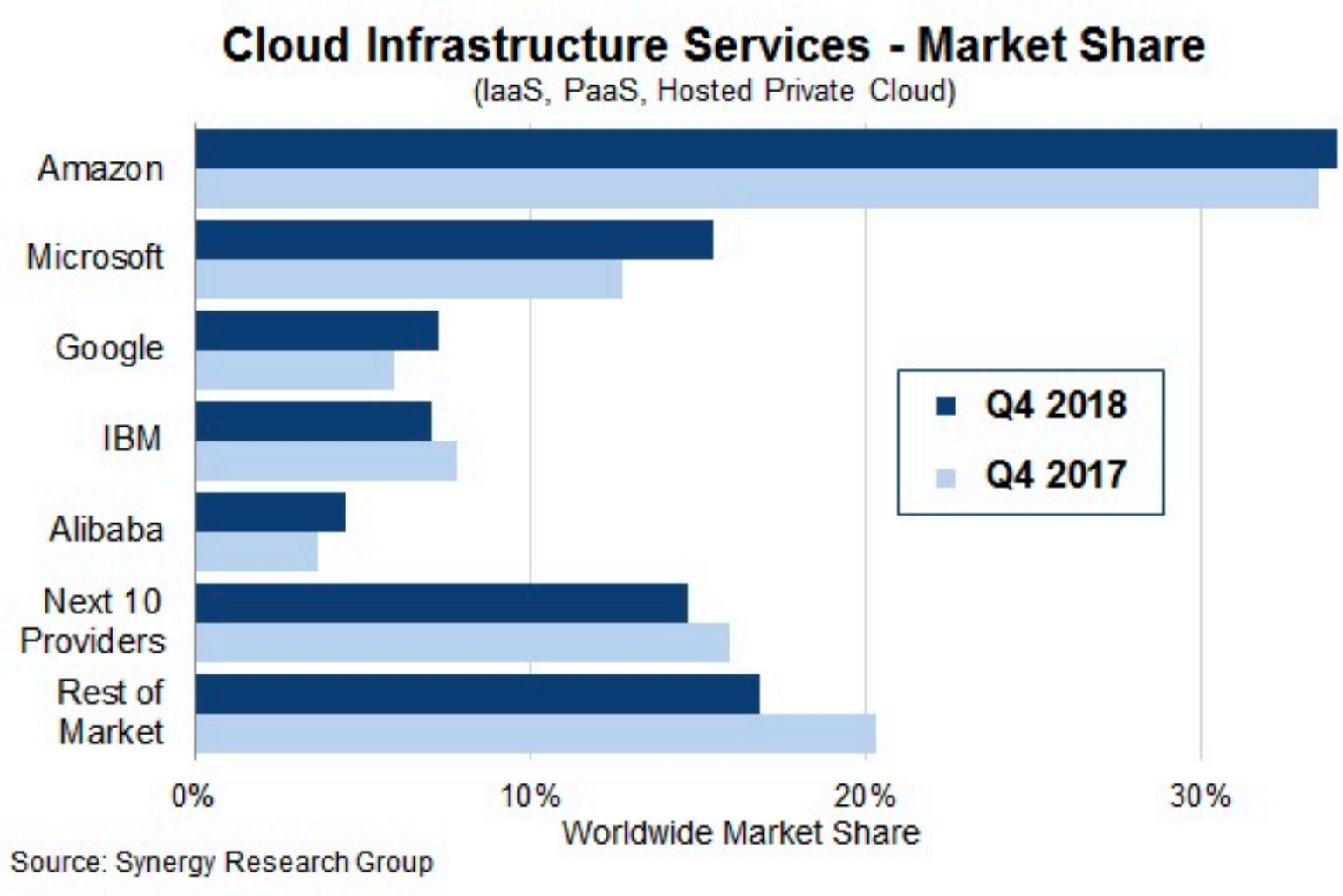

The New York Times.

TITANIC SINKS FOUR HOURS AFTER HITTING ICEBERG; 866 RESCUED BY CARPATHIA, PROBABLY 1250 PERISH; ISMAY SAFE, MRS. ASTOR MAYBE, NOTED NAMES MISSING

The Lost Titanic Setting Toward Out of Belfast Harbor.

PARTIAL LIST OF THE SAVED.

https://open.blogs.nytimes.com/2007/11/01/self-service-prorated-super-computing-fun



https://www.sdxcentral.com/articles/news/aws-remains-dominant-player-in-growing-cloud-market-srg-reports/2019/02/

Rough Comparisons

"If you have a lot of cloud knowledge in-house, no public cloud can compare to the **offerings of AWS**.

If you're mostly in need of **PaaS**, you want seamless hybrid cloud, and you're already using a **lot of Microsoft services**, go with **Azure**.

If you're a major data analyzer, Google's data storage and analytics tools are simply fantastic.

http://dazeinfo.com/2015/05/22/amazon-aws-google-cloud-microsoft-azure/

A diverse set of real-world Java benchmarks shows Google is fastest, Azure is slowest, and Amazon is priciest

http://www.infoworld.com/article/2610403/cloud-computing/ultimate-cloud-speed-tests--amazon-vs--google-vs--windows-azure.html

Enabling technologies

2000s - different technologies converged and are combined together to enable the emergence of cloud computing

- Broadband
- **Cluster Computing - Parallelism**
- Service-Oriented Architectures
- Web 2.0
- Multi-tenancy
- **Virtualization**

Exercise

Are the following services IAAS, PAAS or SAAS?

1. Azure Linux Virtual Machine
2. Azure Blob Storage
3. Azure SQL Database
4. Azure Managed Disks
5. Azure Face API
6. Azure Translator Text API
7. Engine Yard
8. Facebook
9. Gmail
10. Heroku
11. Dropbox
12. Google App Engine

Parallel Computing - Divide & Conquer

- Many calculations are carried out simultaneously
- Large problems can often be divided into smaller ones, which are solved in parallel
- Originally mostly in high-performance computing
- Today dominant paradigm – multi-core processors
- Due to slower increase in processor frequency (heat & power)

Wikipedia Featured Article on Parallel Computing

https://www.maketecheasier.com/why-cpu-clock-speed-isnt-increasing/

Different levels of parallelism

Bit-level
  Increasing **processor word size**, 8-bit … 64-bit

Instruction-level
  e.g., instruction pipelining, multiple instructions can partially overlap

Data parallelism
  Same task performed on different batches of data

Task parallelism
  Distributing tasks (threads, processes) across processors

Multiprocessing

The use of two or more CPUs within a single computer system.

The ability of a system to support **more than one processor** and/or the ability to **allocate tasks between them**

Execution of **multiple concurrent processes** in a system, with each process running on a separate CPU or core, as opposed to a single process at any one instant

Parallel Computing speed up

- Reducing the time to **obtain the solution**
- Portion of the computation which can be parallelised determines the overall speed up
- **Amdahl's law -** find the maximum expected improvement to an overall system when only part of the system is improved

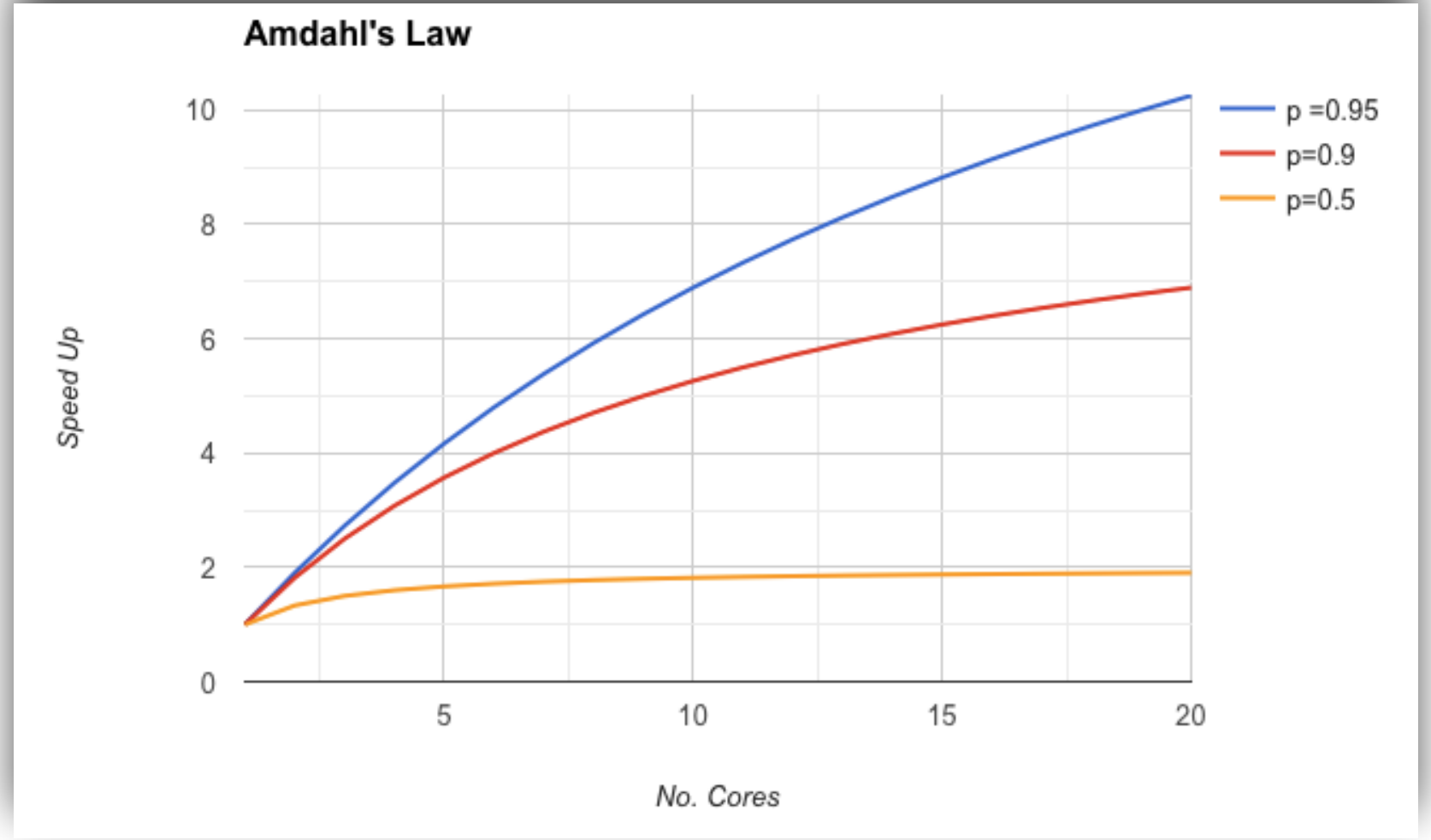
Amdahl's Law

Provides an estimate of the speedup achievable by parallelisation as further resources (CPUs, cores) are added.

$$S_{latency}(s) = \frac{1}{(1-p) + \frac{p}{s}}$$

***p* is the proportion that is parallelisable**
***s* is the number of cores available**
Check, $p = 0$, $p = 1$, $s = \infty$

e.g.
If 90% of a task is parallelisable
It takes 20 minutes to run on a single CPU
It cannot be speeded up to less than two minutes



From spreadsheet on Moodle

p = 0.5: can't do better than 2x

p = 0.95: more cores continue to add benefits

Try p= 0.99

Issues

Coordination of concurrent computation

Overhead, reduces theoretical speedup

Locking of shared resources

Deadlocks – processes competing for resources forced to wait for additional resources held by other processes and none of the processes can finish

Livelock – processes continually change the state while waiting for resource

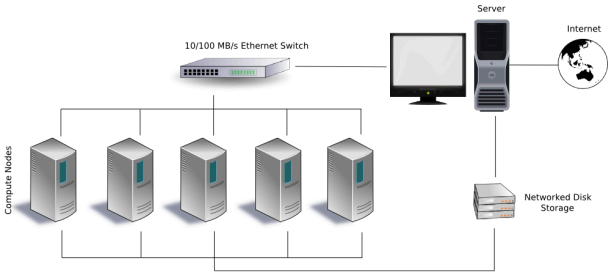
Lock starvation – lower priority threads never getting to the resource

Clustering

Group of independent resources **interconnected** and **work as a single system**

- **Reduced failure rates**
- **Increased availability and reliability** – by providing redundancy and failover

Each node set to perform the same task, controlled and scheduled by software.



- **Server cluster**
- **Database cluster**

Rely on high-speed dedicated network connections to communicate about workload distribution, task scheduling, data sharing and system synchronisation

Virtualisation (in more detail)

What is virtualisation?

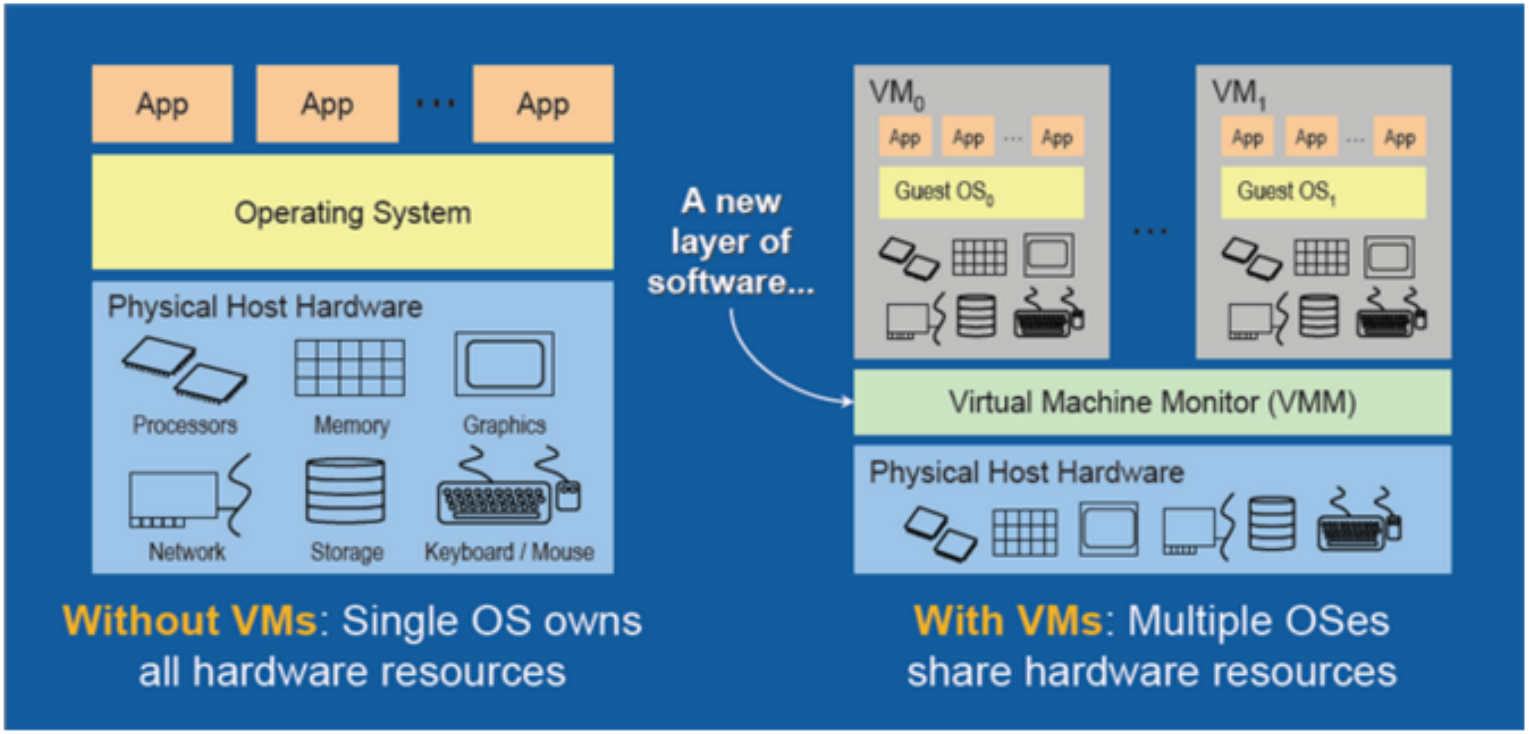
Types

VMs and VMMs

Benefits

Issues

Impact on Cloud



https://software.intel.com/en-us/articles/the-advantages-of-using-virtualization-technology-in-the-enterprise

Virtualisation

Creation of virtual instances of IT resources

Physical IT resources can provide multiple virtual images of themselves so that **underlying processing capabilities** can be shared by multiple users

Key idea: **Decoupling hardware from software** – hardware requirements can be simulated by emulation software running in virtualized environments

Virtualisation types

Virtualisation of servers, storage, networks, power

Simulate interface to a physical object using:

- **Multiplexing** – multiple virtual objects from one instance of a physical object
- **Aggregation** – create one virtual object from multiple physical objects (eg RAID)
- **Emulation** – create virtual object from a different type of physical object (disk emulates RAM)

Virtual machine (VM)

Isolated environment that appears to be a whole machine but actually only has access to a portion of the computer resources

Process VM – for individual process

Application VM – runs on normal OS and provides platform independent host for a single application, eg JVM

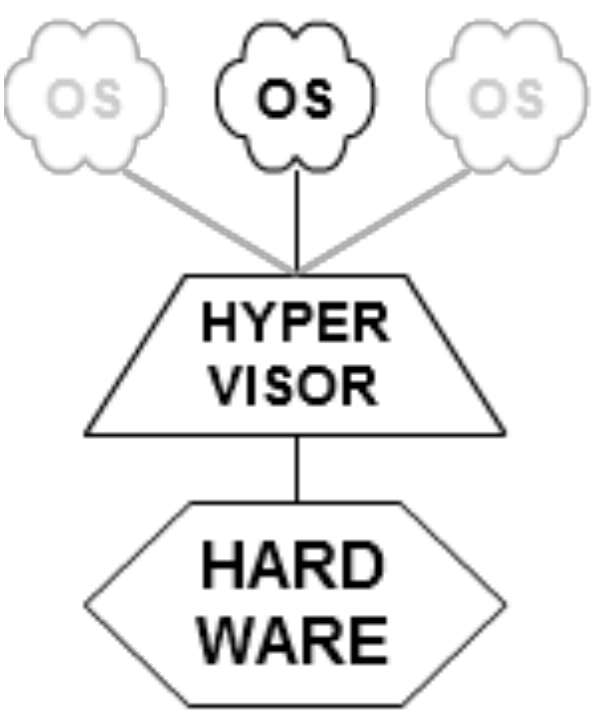
System VM – complete system that can run multiple applications

Virtual machine monitor (VMM)

Or hypervisor – software that partitions the resources into one or more virtual machines / creates and runs virtual machines

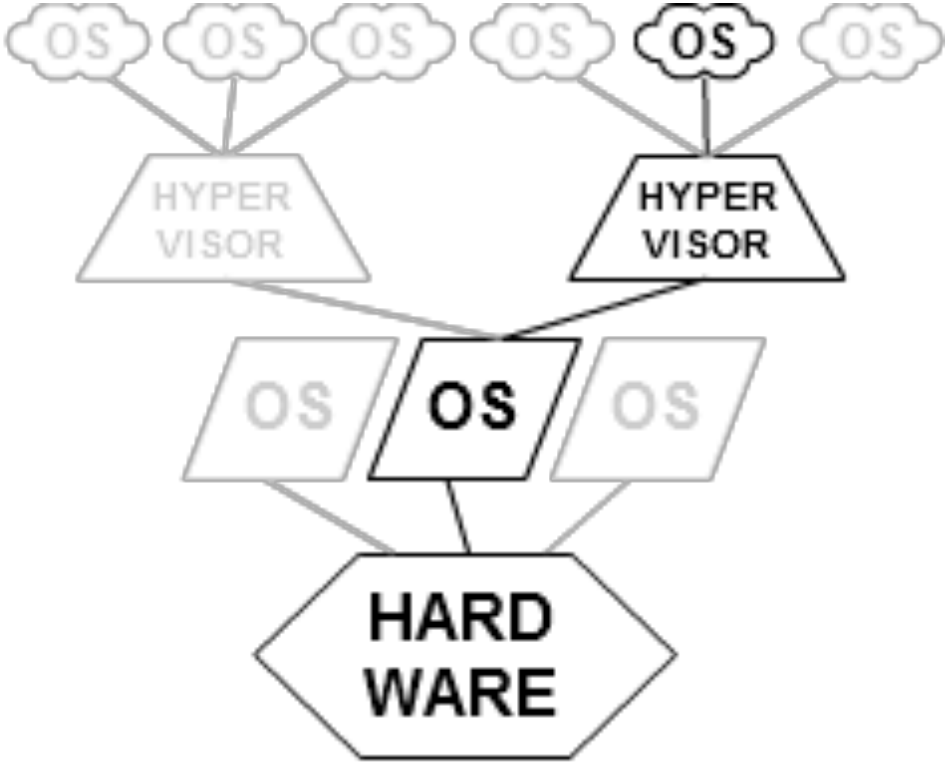
Guest OS

Host OS



TYPE 1

native
(bare metal)



TYPE 2

hosted

Examples of host OS and guest OS combinations

https://en.wikipedia.org/wiki/Comparison_of_platform_virtualization_software

Benefits of Virtualization

- Use fewer physical servers
  - cheaper hardware cost, less energy use, less generated heat
  - Faster server provisioning
- Easier to isolate applications
  - One application/one (virtual) server
  - Extend lifetime of legacy applications
- Decreased dependency on (specific) hardware
  - Change hardware vendors, disaster recovery, increased uptime

Side effects/Risks

- Software might behave differently in virtualized environments – performance degradation
- More powerful hardware needed
- Complex root cause analysis – what caused the problem?
- Security implications - - patching and maintenance, hypervisor infected by malware

Use of virtualisation in cloud

1. Key component in cloud computing
2. Creation of an intelligent abstraction layer which hides the complexity of underlying hardware or software/not locked into any individual vendor
3. More efficient use of hardware
4. Quicker provisioning - elasticity

Cloud Computing

What is Cloud Computing?

Cloud service models

Cloud deployment models

Goals/benefits

Risks/Challenges

Commercial Products

Enabling Technologies

Parallelisation: Amdahl's Law

Virtualisation

Cloud Computing: Methodology, Systems, and Applications
edited by Lizhe Wang, Rajiv Ranjan, Jinjun Chen, Boualem Benatallah