Big Data, organisation and analysis

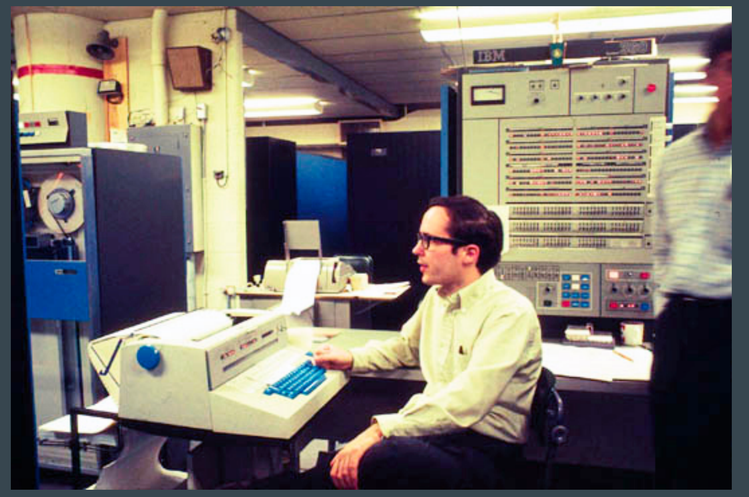
Virtualisation

Steffen M. Noe, Spring 2023

Why virtualisation?

- Server hardware may be “wasted” by dedicating it to one task or process only.
- Reducing server downtime.
- Enhance resilience in disaster recovery situations.
- Increase efficiency in hardware usage.
- Increase productivity (lowering maintenance needs).

A little history of virtualisation?

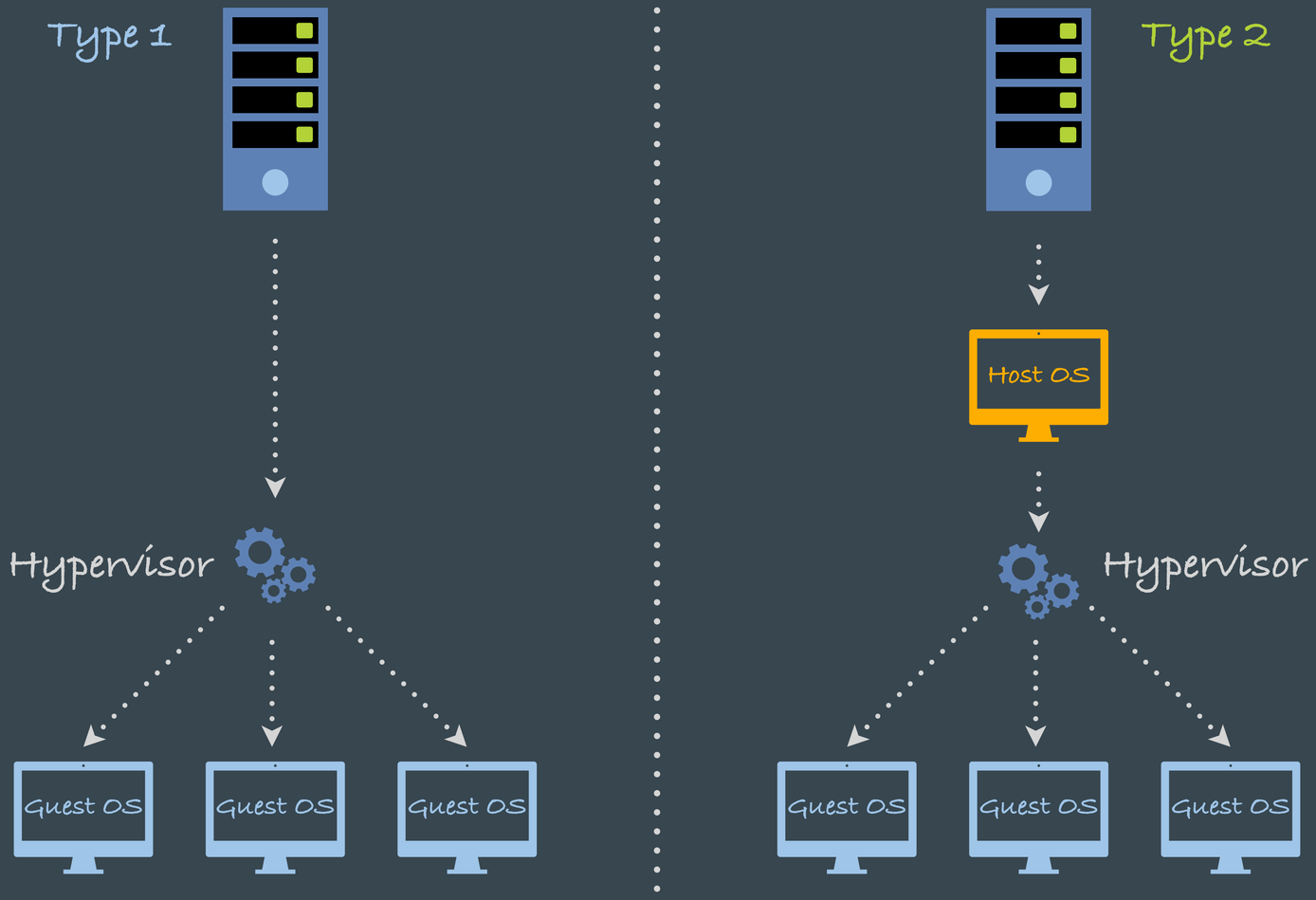


- Virtualisation is not new!
- IBM had first time-sharing and virtual memory system S/360 in 1966.
- First virtual machines on bare-iron were mainframes.
- Internally, IBM favoured “batch-processing” over “time-sharing”
  - batch-processing was known before! A batch of instructions in form of punch cards was provided to the machine and used all the resources!
- Today, we use the “hypervisor” concept — a software layer that either uses “bare-metal” (type 1) or “hosted” (type 2) paradigms.

What is a hypervisor?

It's also called a virtual machine monitor (VMM)

Its a software that allows the creation of virtual machines (VM's) and runs them.



More info - https://linuxhandbook.com/what-is-hypervisor/

Types of computing infrastructure

Big Data and high performance computing (HPC)

Mainframe

IBM z14

Server farm

CRAY XC40

Data center (cloud)

CERN data center

Big Data - Big Iron

Mainframe computers

- Mainframe computers often get called “Big Iron”
- The mainframe units weight about one ton!
- Still a storage solution is required!

What makes it special?

- Mainframes are optimised for high data throughput and complex data intensive workloads!
- Bank transfers are an example
- High reliability, multi layer and redundancy
- Processing terabytes in short time!
- Specialised hardware, bare-iron virtualisation

Server farms

- Use of X86/ARM blade servers common
- Serve different purposes, from supercomputers to cloud systems.
- Very flexible
- Systems can be very complex in terms of the used components
- Providing CPU / GPU / and also Tensor units for data rich applications (AI / ML)
- can provide high computing performance (HPC) but then not necessarily a high data volume handling
- need a storage solution

Data Center

- Data center’s may combine mainframes and server farms
- Adding up the storage infrastructure
- The data center often acts like an abstraction and users do not “see” if they are using a real or virtual resource
- Virtualisation is an integral part of data centres.

Types of virtualisation

Virtual machine

- Software abstraction of a server's hardware
- Installation of the Guest OS into the VM
- Allows different systems on the same physical hardware.
- Need a hypervisor

Can "only" share the real hardware, i.e., if the system is X86 then only OS's compatible with that can be used.
Access to the physical hardware possible, i.e., CPU/GPU
Can also emulate a hardware platform

Container

- Software abstraction that bundles an application into an image
- Fixed Guest OS, mostly Linux
- Isolates multiple user environments on the same host system.
- Uses the host's facilities to access the hardware (no hypervisor)

Can "only" share the real hardware, i.e., if the system is X86 then only Linux systems compatible with that can be used.
Lightweight, hardware abstraction needed

Virtual machines (VM)

Some properties

- VM's can be set up to have different guest OS's like Windows, Linux and Unix (BSD, Solaris,...)
- VM's act like a "real" computer
- VM's are the first generation of cloud computing

Benefits

Resources: once set up it can be reused
Scale: deploy multiple copies (cloud)
Portability: can be relocated between physical servers
Flexibility: faster creation via cloning
Security: cracked systems can be disposed

Types

MS Hypervisor: Split the system in a main partition [hypervisor + host OS (windows)] and child partitions (guest OS's)

Linux KVM: Linux Kernel based VM, runs unmodified Linux or Windows images as hosts

Byte-code VM's: Java (JVM) and Python (PVM) don't use hypervisor technique but transfer compiled byte-code to the host hardware

Container systems

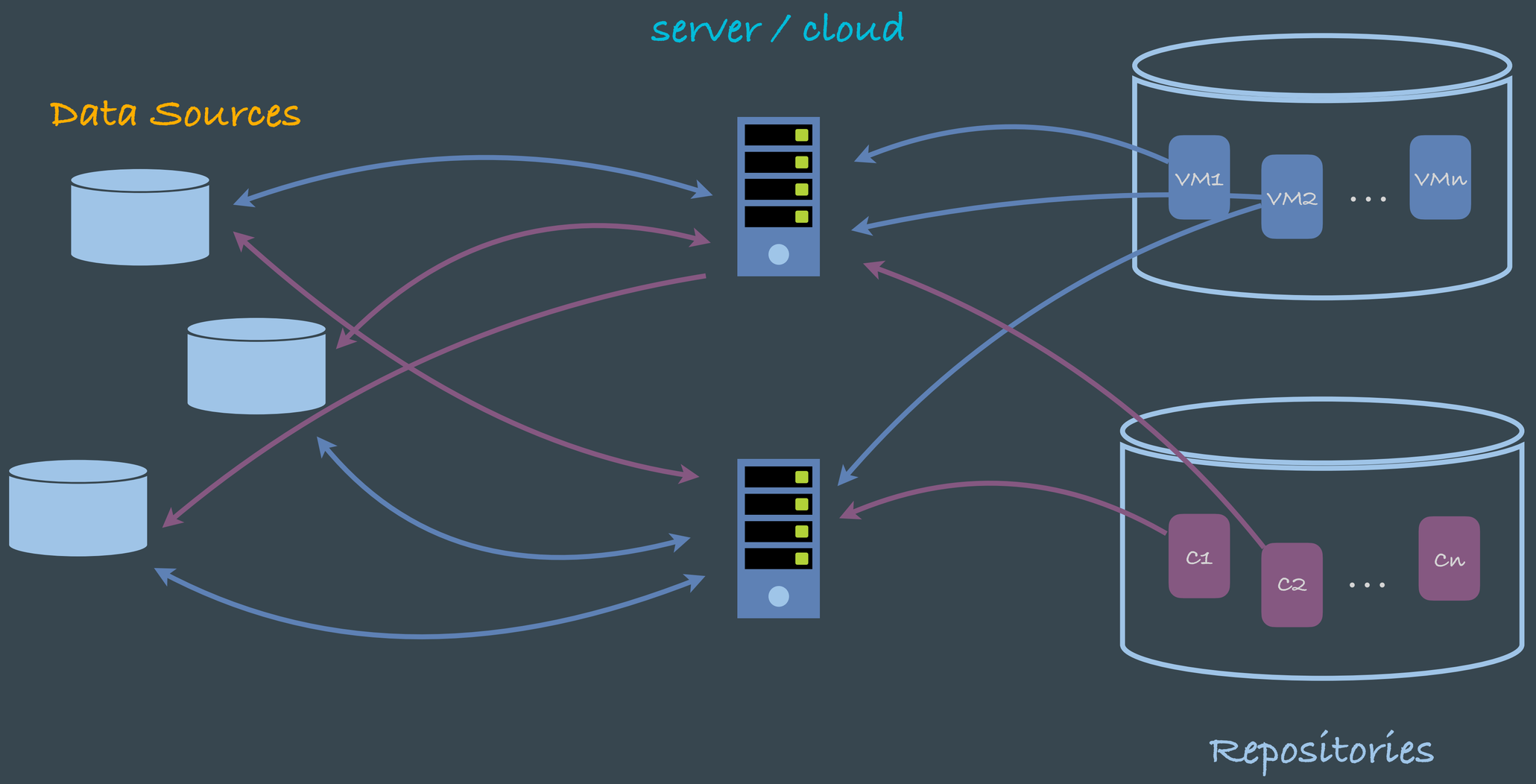
Some properties

- Containerisation is the latest evolution in cloud computing
- Containers bundle applications and their dependencies into one image
- Containers use internally one OS (Linux or Windows are possible)

Benefits

Efficiency: far less resources that VM's needed
Speed: provides fast micro-services, start's up much faster than VM's
Scale: deploy multiple copies (cloud)
Portability: between different servers and clouds, truly write once - run everywhere
Flexibility: faster creation via cloning, micro-service's can be chained and combined
Security: isolating applications from each other and the host
Easy Management: can use automatised installation, upgrade and roll-back techniques (Kubernetes)

Virtualisation in data processing



Useful links

https://www.suse.com/c/mainframe-versus-server-farm-comparison/

Challenge

- Try to understand what a container is
- How is it different from the virtual machine?
- Watch the video of Liz Rice on “Containers from scratch” - https://www.youtube.com/watch?v=8fi7uSYlOdc