Containers and Virtual Machines

Everything old is new again (not really)

Michele Finelli m@biodec.com BioDec

Index

Virtual machines

Containers

Enter Docker

Conclusions

Index

Virtual machines

Containers

Enter Docker

Conclusions

- ► IBM! IBM invented everything in the 60's.
- Not really, but in a sense ... it comes from people who wanted to run away from IBM.
- Honeywell 200/2000 systems Liberator allowed people to run a virtual machine that emulated the IBM 14xx systems.

- ► IBM! IBM invented everything in the 60's.
- Not really, but in a sense . . . it comes from people who wanted to run away from IBM.
- Honeywell 200/2000 systems Liberator allowed people to run a virtual machine that emulated the IBM 14xx systems.

- ► IBM! IBM invented everything in the 60's.
- Not really, but in a sense ...it comes from people who wanted to run away from IBM.
- Honeywell 200/2000 systems Liberator allowed people to run a virtual machine that emulated the IBM 14xx systems.

- ► IBM! IBM invented everything in the 60's.
- Not really, but in a sense ...it comes from people who wanted to run away from IBM.
- ► Honeywell 200/2000 systems Liberator allowed people to run a virtual machine that emulated the IBM 14xx systems.

- There were dozens (hundreds?) of different hardware, all of them incompatible which each other.
- ► IBM wrote OS/360 to rule them all but ...
- ...people did not want to rewrite all their (working, legacy) software for the new platform.
- So IBM decided to support virtual machines to run its own OSs inside the OS/360.

- ► There were dozens (hundreds ?) of different hardware, all of them incompatible which each other.
- ► IBM wrote OS/360 to rule them all but . . .
- ...people did not want to rewrite all their (working, legacy) software for the new platform.
- So IBM decided to support virtual machines to run its own OSs inside the OS/360.

- ► There were dozens (hundreds?) of different hardware, all of them incompatible which each other.
- ▶ IBM wrote OS/360 to rule them all but ...
- ...people did not want to rewrite all their (working, legacy) software for the new platform.
- So IBM decided to support virtual machines to run its own OSs inside the OS/360

- ► There were dozens (hundreds?) of different hardware, all of them incompatible which each other.
- ▶ IBM wrote OS/360 to rule them all but ...
- ... people did not want to rewrite all their (working, legacy) software for the new platform.
- So IBM decided to support virtual machines to run its own OSs inside the OS/360

- ► There were dozens (hundreds?) of different hardware, all of them incompatible which each other.
- ▶ IBM wrote OS/360 to rule them all but ...
- ▶ ... people did not want to rewrite all their (working, legacy) software for the new platform.
- So IBM decided to support virtual machines to run its own OSs inside the OS/360.

- To run different computer architectures, with different instruction sets, memory models, etcetera, you need to emulate the whole hardware.
- So basically you have two operating systems one on top of the other.
- The upper OSs are called "operating system", sometimes guest OSs.
- ► The lower one is called **hypervisor** but it is an operating system nonetheless.

- ➤ To run different computer architectures, with different instruction sets, memory models, etcetera, you need to emulate **the whole hardware**.
- So basically you have two operating systems one on top of the other.
- The upper OSs are called "operating system", sometimes guest OSs.
- ► The lower one is called **hypervisor** but it is an operating system nonetheless.

- ➤ To run different computer architectures, with different instruction sets, memory models, etcetera, you need to emulate **the whole hardware**.
- So basically you have two operating systems one on top of the other.
- The upper OSs are called "operating system", sometimes guest OSs.
- ► The lower one is called **hypervisor** but it is an operating system nonetheless.

- ➤ To run different computer architectures, with different instruction sets, memory models, etcetera, you need to emulate **the whole hardware**.
- So basically you have two operating systems one on top of the other.
- The upper OSs are called "operating system", sometimes guest OSs.
- The lower one is called hypervisor but it is an operating system nonetheless.

- To run different computer architectures, with different instruction sets, memory models, etcetera, you need to emulate the whole hardware.
- So basically you have two operating systems one on top of the other.
- ► The upper OSs are called "operating system", sometimes guest OSs.
- ► The lower one is called hypervisor but it is an operating system nonetheless.

- Emulating the hardware can be expensive: both for the hypervisor and the guest OS.
- Moreover, if the emulation is entirely done by software there is always an overhead.
- To speed things up, people started to tinker to add hardware support to virtualization.
- ► And they spectacularly succeeded

- ► Emulating the hardware can be expensive: both for the hypervisor and the guest OS.
- Moreover, if the emulation is entirely done by software there is always an overhead.
- To speed things up, people started to tinker to add hardware support to virtualization.
- ► And they spectacularly succeeded.

- Emulating the hardware can be expensive: both for the hypervisor and the guest OS.
- Moreover, if the emulation is entirely done by software there is always an overhead.
- To speed things up, people started to tinker to add hardware support to virtualization.
- ► And they spectacularly succeeded.

- Emulating the hardware can be expensive: both for the hypervisor and the guest OS.
- Moreover, if the emulation is entirely done by software there is always an overhead.
- ➤ To speed things up, people started to tinker to add hardware support to virtualization.
- ► And they spectacularly succeeded

- Emulating the hardware can be expensive: both for the hypervisor and the guest OS.
- Moreover, if the emulation is entirely done by software there is always an overhead.
- To speed things up, people started to tinker to add hardware support to virtualization.
- ► And they spectacularly succeeded.

Hardware-based virtualization

To overcome the overhead of running an OS inside another OS, the hardware was changed to accommodate for special instruction sets:

- The Intel VT-x and AMD AMD-V are processor extensions to the x86 architecture — they were released in years 2005-2006.
- Hardware-based (or hardware-assisted) x86 virtualization nowadays include:
 - 1. Linux KVM.
 - 2. Xen.
 - 3. VMware
 - 4. Microsoft Hyper-V



Hardware-based virtualization

To overcome the overhead of running an OS inside another OS, the hardware was changed to accommodate for special instruction sets:

- The Intel VT-x and AMD AMD-V are processor extensions to the x86 architecture — they were released in years 2005-2006.
- Hardware-based (or hardware-assisted) x86 virtualization nowadays include:
 - 1. Linux KVM.
 - 2. Xen.
 - 3. VMware.
 - 4. Microsoft Hyper-V.



Hardware-based virtualization

To overcome the overhead of running an OS inside another OS, the hardware was changed to accommodate for special instruction sets:

- The Intel VT-x and AMD AMD-V are processor extensions to the x86 architecture — they were released in years 2005-2006.
- Hardware-based (or hardware-assisted) x86 virtualization nowadays include:
 - 1. Linux KVM,
 - 2. Xen,
 - 3. VMware,
 - 4. Microsoft Hyper-V.



- It is not an overstatement to say that the success of hardware-based virtualization and Linux support for it is demonstrated by the mere existence of cloud technologies at all.
- ► Cloud computing would have been **impossible** without hardware-based virtualization and **free software**.
- So, obviously, hardware-based virtualization must be the best thing that can happen to software development, since Fortran 77, right ?

- It is not an overstatement to say that the success of hardware-based virtualization and Linux support for it is demonstrated by the mere existence of cloud technologies at all.

- It is not an overstatement to say that the success of hardware-based virtualization and Linux support for it is demonstrated by the mere existence of cloud technologies at all.
- Cloud computing would have been impossible without hardware-based virtualization and free software.
- So, obviously, hardware-based virtualization must be the best thing that can happen to software development, since Fortran 77, right ?

- It is not an overstatement to say that the success of hardware-based virtualization and Linux support for it is demonstrated by the mere existence of cloud technologies at all.
- Cloud computing would have been impossible without hardware-based virtualization and free software.
- ► So, obviously, hardware-based virtualization must be the best thing that can happen to software development, since Fortran 77, right ?

Index

Virtual machines

Containers

Enter Docker

Conclusions

- ► History of 4.2BSD shows that Bill Joy on 18 March 1982 added a system call called **chroot**.
- From the man page: chroot change root directory

- ► History of 4.2BSD shows that Bill Joy on 18 March 1982 added a system call called **chroot**.
- From the man page: chroot change root directory

- ► History of 4.2BSD shows that Bill Joy on 18 March 1982 added a system call called **chroot**.
- ► From the man page: *chroot* change root directory

Why?

chroot (and its user space command) was created for the following reasons:

- to run untrusted code.
- to cleanly build code from scratch,
- to isolate environments.

Why?

chroot (and its user space command) was created for the following reasons:

- ► to run untrusted code,
- to cleanly build code from scratch,
- to isolate environments.

Why?

chroot (and its user space command) was created for the following reasons:

- ► to run untrusted code,
- ▶ to *cleanly* build code from scratch,
- to isolate environments.

Why?

chroot (and its user space command) was created for the following reasons:

- ► to run untrusted code,
- ▶ to cleanly build code from scratch,
- to isolate environments.

Jails

In 1999 we added a partitioning facility to FreeBSD called jail(2). It reuses the chroot(2) implementation, but prevents well-documented means to escape chroot confinement. Jail offers semi-permeable partitioning of the file system, process, and networking namespaces, and removes all super-user privileges that would affect objects not entirely inside the jail.

Building Systems to be Shared Securely, Poul-Henning Kamp, ACM Queue July/August 2014.



FreeBSD Jails

The FreeBSD "Jail" facility provides the ability to partition the operating system environment, while maintaining the simplicity of the UNIX "root" model. In Jail, users with privilege find that the scope of their requests is limited to the jail, allowing system administrators to delegate management capabilities for each virtual machine environment. Creating virtual machines in this manner has many potential uses: the most popular thus far has been for providing virtual machine services in Internet Service Provider environments

Confining the Omnipotent Root, Poul-Henning Kamp and Robert Watson, Sane 2000.

- "(...) to delegate management capabilities for each virtual machine environment."
- "Creating virtual machines in this manner (...)"
- "(...) has many potential uses; the most popular thus far has been for providing virtual machine services in Internet Service Provider environments."



- "(...) to delegate management capabilities for each virtual machine environment."
- "Creating virtual machines in this manner (...)"
- "(...) has many potential uses; the most popular thus far has been for providing virtual machine services in Internet Service Provider environments."



- "(...) to delegate management capabilities for each virtual machine environment."
- ► "Creating virtual machines in this manner (...)"
- "(...) has many potential uses; the most popular thus far has been for providing virtual machine services in Internet Service Provider environments."



- "(...) to delegate management capabilities for each virtual machine environment."
- ► "Creating virtual machines in this manner (...)"
- "(...) has many potential uses; the most popular thus far has been for providing virtual machine services in Internet Service Provider environments."





- "(...) to delegate management capabilities for each virtual machine environment."
- ► "Creating virtual machines in this manner (...)"
- "(...) has many potential uses; the most popular thus far has been for providing virtual machine services in Internet Service Provider environments."



Zones

In year 2005, Sun Microsystems launched Solaris Zones:

Zones provide a means of virtualizing operating system services, allowing one or more processes to **run in isolation** from other activity on the system.

This isolation **prevents processes** running within a given zone **from monitoring or affecting processes running in other zones**.

Virtualization and Namespace Isolation in Solaris (PSARC/2002/174), September 7, 2006.



Zones (continued)

A zone is a "sandbox" within which one or more applications can run without affecting or interacting with the rest of the system.

It also provides an abstraction layer that **separates applications from physical attributes** of the machine on which they are deployed, such as physical device paths and network interface names.

Virtualization and Namespace Isolation in Solaris (PSARC/2002/174), September 7, 2006.



From the paper:

- So, you are basically running Solaris which is good ...if you are Sun Microsystem.
- ► But people wanted to consolidate **Microsoft Windows**.
- ▶ And to run Linux, too.

► From the paper:

- So, you are basically running Solaris which is good ...if you are Sun Microsystem.
- But people wanted to consolidate Microsoft Windows.
- And to run Linux, too.



► From the paper:

- So, you are basically running Solaris which is good ...if you are Sun Microsystem.
- But people wanted to consolidate Microsoft Windows.
- And to run Linux, too.



► From the paper:

- So, you are basically running Solaris which is good ...if you are Sun Microsystem.
- ► But people wanted to consolidate Microsoft Windows.
- ► And to run Linux, too.



► From the paper:

- So, you are basically running Solaris which is good ...if you are Sun Microsystem.
- ► But people wanted to consolidate **Microsoft Windows**.
- And to run Linux, too.



From the paper:

- So, you are basically running Solaris which is good ...if you are Sun Microsystem.
- ► But people wanted to consolidate Microsoft Windows.
- ► And to run Linux, too.



- Being tied to the same OS of the underlying server meant that by year 2006 the containers adoption hit a wall.
- Some kind of support (i.e. Solaris' branded zone) allowed cross OS support, but without the simplicity of full machine virtualization
- ► The market was dominated by Microsoft Windows machines, and Linux-based hypervisors became the de facto standard for running them.

- ▶ Being tied to the same OS of the underlying server meant that by year 2006 the containers adoption hit a wall.
- Some kind of support (i.e. Solaris' branded zone) allowed cross OS support, but without the simplicity of full machine virtualization
- The market was dominated by Microsoft Windows machines, and Linux-based hypervisors became the de facto standard for running them.



- ▶ Being tied to the same OS of the underlying server meant that by year 2006 the containers adoption hit a wall.
- Some kind of support (i.e. Solaris' branded zone) allowed cross OS support, but without the simplicity of full machine virtualization.
- The market was dominated by Microsoft Windows machines, and Linux-based hypervisors became the de facto standard for running them.



- ▶ Being tied to the same OS of the underlying server meant that by year 2006 the containers adoption hit a wall.
- Some kind of support (i.e. Solaris' branded zone) allowed cross OS support, but without the simplicity of full machine virtualization
- ► The market was dominated by Microsoft Windows machines, and Linux-based hypervisors became the *de facto* standard for running them.



- Control groups (cgroup).
- Namespaces.
- Copy-on-write storage:
 - ► AUFS, overlay (file level),
 - device mapper thin provisioning (block level),
 - ► BTRFS, ZFS (FS level).

- ► Control groups (cgroup).
- Namespaces.
- Copy-on-write storage:
 - ► AUFS, overlay (file level).
 - device mapper thin provisioning (block level),
 - ► BTRFS, ZFS (FS level).

- ► Control groups (cgroup).
- Namespaces.
- Copy-on-write storage:
 - ► AUFS, overlay (file level).
 - device mapper thin provisioning (block level),
 - ► BTRFS, ZFS (FS level).

- ► Control groups (cgroup).
- Namespaces.
- Copy-on-write storage:
 - ► AUFS, overlay (file level),
 - device mapper thin provisioning (block level),
 - ► BTRFS, ZFS (FS level).



- ▶ Even when you don't run containers . . .
- ... you are in a container.
- Your host processes still execute in the root namespaces and cgroups

- ► Even when you don't run containers . . .
- ...you are in a container.
- Your host processes still execute in the root namespaces and cgroups

- ► Even when you don't run containers . . .
- ... you are in a container.
- Your host processes still execute in the root namespaces and cgroups

- ► Even when you don't run containers . . .
- ... you are in a container.
- Your host processes still execute in the root namespaces and cgroups

Index

Virtual machines

Containers

Enter Docker

Conclusions

- Running a KVM/QEMU virtual machine is matter of installing libvirtd and running virsh start
- How do you do the same, in container land?
- ► In many different ways.

- ► Running a KVM/QEMU virtual machine is matter of installing libvirtd and running virsh start
- How do you do the same, in container land?
- ▶ In many different ways.

- Running a KVM/QEMU virtual machine is matter of installing libvirtd and running virsh start
- ▶ How do you do the same, in container land?
- In many different ways.

- Running a KVM/QEMU virtual machine is matter of installing libvirtd and running virsh start
- ▶ How do you do the same, in container land?
- ▶ In many different ways.

Container run-times

Alternatives:

- ▶ LXC/LXD.
- systemd-nspawn
- ► The Docker engine.
- ▶ rkt
- ▶ runC

Container run-times

Alternatives:

- ► LXC/LXD.
- systemd-nspawn
- ► The Docker engine.
- ▶ rkt
- ▶ runC

Container run-times

Alternatives:

- ► LXC/LXD.
- ► systemd-nspawn
- ► The Docker engine.
- ▶ rkt
- ▶ runC

Container run-times

Alternatives:

- ► LXC/LXD.
- ► systemd-nspawn
- ► The Docker engine.
- rkt.
- ▶ runC

Container run-times

Alternatives:

- ► LXC/LXD.
- ► systemd-nspawn
- ► The Docker engine.
- ▶ rkt
- ▶ runC

Container run-times

Alternatives:

- ► LXC/LXD.
- ► systemd-nspawn
- ► The Docker engine.
- ▶ rkt
- ▶ runC

- In March 2013 a company called dotCloud launched a product to run containers in Linux. The product was aimed at developers and was meant to be easy.
- They succeeded quite spectacularly.
- And renamed the company as Docker.

- ► In March 2013 a company called dotCloud launched a product to run containers in Linux. The product was aimed at *developers* and was meant to be easy.
- They succeeded quite spectacularly.
- And renamed the company as Docker.

- In March 2013 a company called dotCloud launched a product to run containers in Linux. The product was aimed at developers and was meant to be easy.
- ► They succeeded quite spectacularly.
- And renamed the company as Docker.



- In March 2013 a company called dotCloud launched a product to run containers in Linux. The product was aimed at *developers* and was meant to be easy.
- They succeeded quite spectacularly.
- ► And renamed the company as **Docker**.

The sweet spot

- Docker offered a solution to a problem that was beginning to be very relevant . . .
- ► ...the problem of *linking*.

The sweet spot

- ► Docker offered a solution to a problem that was beginning to be very relevant . . .
- ► ...the problem of *linking*.

The sweet spot

- ► Docker offered a solution to a problem that was beginning to be very relevant . . .
- ▶ ...the problem of *linking*.

- Linking per se was (is) a solved problem at least since the 70s' but ...
- ...it is the software that has become more complicated: a modern (let's say web) project has dozens of different components, all of them talking each other over the network
- The fault could be in part attributed by hipster driven development . . .
- ...but reality is that there has been a huge driver towards distributed software components.

- ▶ Linking per se was (is) a solved problem at least since the 70s' but . . .
- ...it is the software that has become more complicated: a modern (let's say web) project has dozens of different components, all of them talking each other over the network
- ► The fault could be in part attributed by hipster driven development . . .
- ...but reality is that there has been a huge driver towards distributed software components.

- Linking per se was (is) a solved problem at least since the 70s' but . . .
- ...it is the software that has become more complicated: a modern (let's say web) project has dozens of different components, all of them talking each other over the network.
- ► The fault could be in part attributed by hipster driven development . . .
- ...but reality is that there has been a huge driver towards distributed software components.



- ▶ Linking per se was (is) a solved problem at least since the 70s' but . . .
- ...it is the software that has become more complicated: a modern (let's say web) project has dozens of different components, all of them talking each other over the network.
- ► The fault could be in part attributed by *hipster driven development* . . .
- ...but reality is that there has been a huge driver towards distributed software components.



- ▶ Linking per se was (is) a solved problem at least since the 70s' but . . .
- ...it is the software that has become more complicated: a modern (let's say web) project has dozens of different components, all of them talking each other over the network.
- ► The fault could be in part attributed by *hipster driven development* . . .
- ▶ ...but reality is that there has been a huge driver towards distributed software components.



Index

Virtual machines

Containers

Enter Docker

Conclusions

- ► Right now (February 2016) the ecosystem is complex, fragmented and always changing.
- Death by a thousand frameworks.
- But software systems will always be more and more distributed
- Containers are here to stay.

- ► Right now (February 2016) the ecosystem is complex, fragmented and always changing.
- Death by a thousand frameworks.
- But software systems will always be more and more distributed.
- Containers are here to stay.

- ► Right now (February 2016) the ecosystem is complex, fragmented and always changing.
- Death by a thousand frameworks.
- But software systems will always be more and more distributed.
- Containers are here to stay.

- Right now (February 2016) the ecosystem is complex, fragmented and always changing.
- Death by a thousand frameworks.
- But software systems will always be more and more distributed.
- Containers are here to stay.

- ► Right now (February 2016) the ecosystem is complex, fragmented and always changing.
- Death by a thousand frameworks.
- But software systems will always be more and more distributed.
- ► Containers are here to stay.

Questions



```
DI2016 Incontro DevOps Italia 2016, the 1st of April, in Bologna: there will be talks on Docker and containers, too.

WWW http://www.incontrodevops.it/
```

Me m@biodec.com@gaunilone

*license of the slides:

http://creativecommons.org/licenses/by-sa/4.0/



Thanks for your attention!

IDI2016 Incontro DevOps Italia 2016, the 1st of April, in Bologna: there will be talks on Docker and containers, too.

WWW http://www.incontrodevops.it/

Twitter @incontrodevops

Me m@biodec.com @gaunilone

*license of the slides:

http://creativecommons.org/licenses/by-sa/4.0/



```
IDI2016 Incontro DevOps Italia 2016, the 1st of April, in Bologna: there will be talks on Docker and containers, too.
```

```
WWW http://www.incontrodevops.it/
```

Twitter @incontrodevops

Thanks for your attention!

Me m@biodec.com @gaunilone

*license of the slides:

http://creativecommons.org/licenses/bv-sa/4.0/



```
Thanks for your attention!

IDI2016 Incontro DevOps Italia 2016, the 1st of April, in Bologna: there will be talks on Docker and containers, too.

WWW http://www.incontrodevops.it/
```

Me m@biodec.com@gaunilone

*license of the slides:



```
Thanks for your attention!

IDI2016 Incontro DevOps Italia 2016, the 1st of April, in Bologna: there will be talks on Docker and containers, too.
```

WWW http://www.incontrodevops.it/

Twitter @incontrodevops

Me m@biodec.com @gaunilone

*license of the slides:

http://creativecommons.org/licenses/bv-sa/4.0/



```
Thanks for your attention!

IDI2016 Incontro DevOps Italia 2016, the 1st of April, in Bologna: there will be talks on Docker and containers, too.

WWW http://www.incontrodevops.it/
Twitter @incontrodevops

Me m@biodec.com@gaunilone
```



```
Thanks for your attention!

IDI2016 Incontro DevOps Italia 2016, the 1st of April, in Bologna: there will be talks on Docker and containers, too.

WWW http://www.incontrodevops.it/
Twitter @incontrodevops
Me m@biodec.com@gaunilone

*license of the slides:
http://creativecommons.org/licenses/by-sa/4.0/
```

