INAF

ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

Containerizzazione

Francesco Bedosti – INAF / Istituto di Radioastronomia

INAF

ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

l’alternativa: virtualizzazione

- Un processo produce un computer virtuale, con hardware astratto collegato ad hardware reale.
- Nel computer virtuale possiamo eseguire sistemi operativi ed applicazioni a piacere.
- VMware, QEMU, Parallels, Xen, KVM

INAF

ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

Virtualizzazione :: Pro

- Astrazione hardware.
- Sicurezza: una barriera robusta tra VM e verso l’hardware.
- Portabilità: diversi OS sullo stesso hardware

INAF

ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

Virtualizzazione :: Contro

- Un intero OS per ogni istanza.
- Tempo di avvio elevato.
- Overhead di virtualizzazione.
- Memoria preallocata.

Containerizzazione

- Un container e’ un ambiente di esecuzione autosufficiente e portabile, incapsulato in un unico file o cartella.
- All’interno della macchina reale vengono eseguiti processi che “vedono” solo il container ed un sottoinsieme delle risorse (filesystem, rete, CPU, memoria) della macchina reale.

Containerizzazione :: Pro

Compattezza: il kernel dell'host viene ri-usato da tutti i container

Avvio immediato: il kernel e' già attivo, filesystem già montati, rete già connessa...

Efficiente: accesso a memoria, CPU e filesystem arbitrato dal kernel dell'host

Templating: comodo per fare overlay e montare parti del filesystem in modo incrementale

Containerizzazione :: Contro

- Scarso isolamento: superficie di attacco comunque elevata.
- Portabilità limitata allo stesso sistema operativo ed architettura hardware.

INAF

ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

A cosa serve un container?

- Per aggirare problemi di compatibilità tra versioni e distribuzioni di linux
- Per trasportare ed eseguire suite di software complesse senza dover modificare il setup della macchina
- Per distribuire ed archiviare software
- Per garantire riproducibilità
- Per costruire ambienti di test o sviluppo usa e getta
- Per limitare le risorse a disposizione di un programma

INAF

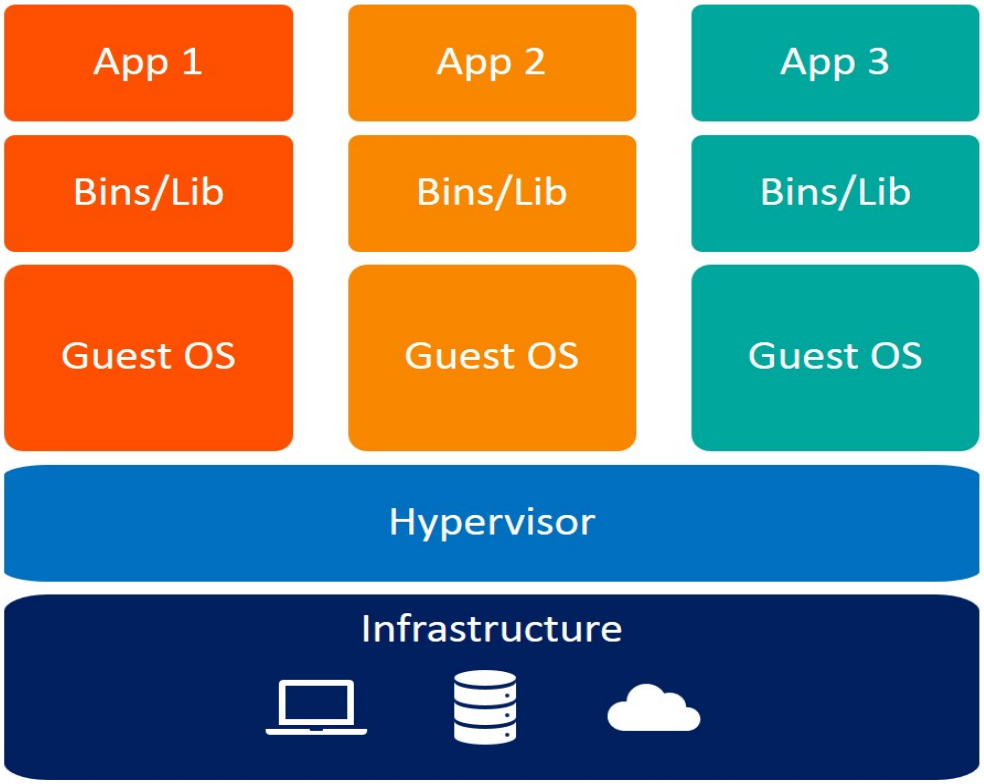
ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

A cosa NON serve un container?

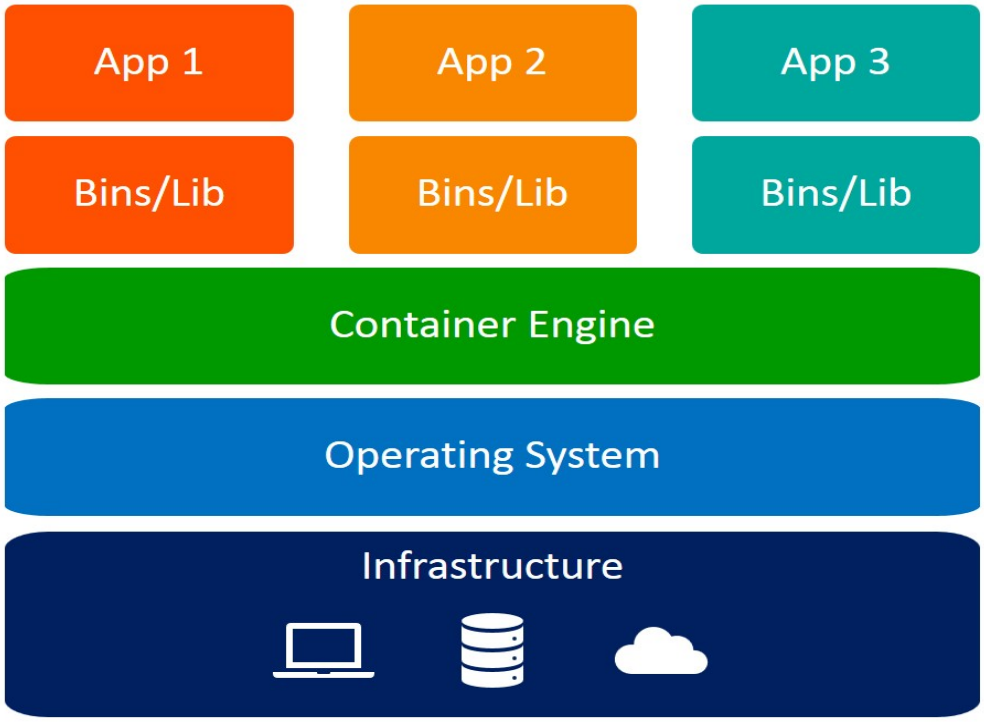
- Ad eseguire software per architetture o sistemi operativi diversi
- Ad isolare l'esecuzione di software per ridurre i rischi di sicurezza

INAF

ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS



Virtual Machines



Containers

Implementazioni

- Chroot (1982) – *confinamento del filesystem*
- BSD Jails (2000), Linux Cgroups (2008) - *confinamento delle risorse*
- *LXC, LXD (2008)*
- Docker (2013) - *virtualizzazione di micro-servizi*
- Singularity (2015) – *HPC*
- Shifter (2015) - *basato su Docker, per HPC*
- .... e molti altri

INAF

ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

Singularity

- Nato specificamente per calcolo scientifico @LBNL
- Semplice, per l’admin e per l’utente
- Rootless, normale processo utente senza demoni o librerie di sistema
- Nessuna modifica dell’infrastruttura
- Immagini in un singolo file o folder
- BYOE (Bring Your Own Environment)
- Ben schedulabile con torque/PBS/slurm
- Supporta X11, MPI, GPU, InfiniBand

INAF

ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

Come costruire un container

singularity build debian8.sif library://debian:8

da:
- library
- Singularity Image File (SIF)
- sandbox directory
- ricetta

a:
- Singularity Image File (SIF)
- Sandbox directory

INAF

ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

Come eseguire un container?

Per avviare una shell:

singularity shell debian8.sif

Per eseguire il comando predefinito:

singularity run debian8.sif

Per eseguire un comando:

singularity exec debian8.sif cat /etc/os-release

Library

Le **Library** sono collezioni pubbliche di container gia pronti (docker:// http:// shub://)

- singularity build kern.sif docker://kernsuite/base:dev

Al Volo?

```
usr@host:~$ singularity exec library://debian:8 cat /etc/debian_version
8.11
```

INAF

ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

Ricette

Una ricetta e' una definizione completa ed autosufficiente per costruire un container

- **Compatta**: un semplice file di testo puo descrivere container di grandi dimensioni
- **Riproducibile e trasparente**: consente anche a terzi di ricreare il container conoscendone esattamente il contenuto
- **Modificabile**: puo essere usata come base per ottenere container diversi

INAF

ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

Costruire con una ricetta

singularity build debian8.sif Singularity.recipe

INAF

ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

Esempio di ricetta

```
BootStrap: library
From: debian:8

%post
    apt-get -y update
    apt-get -y install saods9

%environment
    export PATH=~/mysoftware:$PATH

%labels
    Author Ada Lovelace
    Version 1.42
```

Bind

- Automaticamente, $HOME , /sys:/sys , /proc:/proc, /tmp:/tmp, /var/tmp:/var/tmp, /etc/resolv.conf:/etc/resolv.conf, /etc/passwd:/etc/passwd, e $PWD sono disponibili all'interno del container.
- Si possono aggiungere altri folder con la direttiva –bind

singularity shell --bind /netfs/mydata debian8.sif

Overlay

Si puo sovrapporre un folder scrivibile ad un container in sola lettura, rendendolo cosi usabile in lettura-scrittura senza modificarlo.

```
singularity shell --overlay le_mie_modifiche/ debian8.sif
```

INAF

ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

cgroups

Usando direttive cgroups e' possibile limitare l'accesso a CPU, memoria e I/O in modo fine.

```
[memory]
      524288000

[cpu]
      # parti su 1024
      shares = 512

      # us su 100ms
      quota = 20000

      # specifici cores
      cpus = "0-1"

[blockIO]
      weight = 1000
      leafWeight = 1000
```

INAF

ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

Servizi e demoni

Una instance permette di eseguire un servizio in un container che continua ad esistere in background

```
singularity pull library://debian:8
singularity instance start   debian_8.sif esemplare1

singularity instance start  library://debian:8 esemplare1

singularity instance list
singularity exec instance://esemplare1 cat /etc/os-release
singularity shell instance://esemplare1
singularity instance stop esemplare1
```

Crittografia

- Il root file system puo essere crittografato, con password o con chiave RSA.
- Il container puo essere firmato con GPG

INAF

ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

Fakeroot

Permette ad un normale utente di creare ed eseguire un container ottenendo internamente i privilegi di root.

- L’utente viene mappato come root
- Gli altri utenti del container vengono mappati come utenti non esistenti nel sistema reale (uid + 131072)
- Il container non puo fare bind di porte <1024
- Permette di creare container senza essere root

INAF

ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

MPI

- Compatibile con OpenMPI e MPICH
- Le versioni di MPI dentro e fuori dal container devono essere allineate
- I container vengono avviati con mpirun

```
mpirun -n 5 singularity exec debian8_mpi.sif /opt/mpi_app
```

INAF

ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

Grazie per l’attenzione

Francesco Bedosti – INAF / Istituto di Radioastronomia