Instituto Superior de Engenharia de Lisboa

Departamento de Engenharia de Eletrónica e Telecomunicações e de Computadores

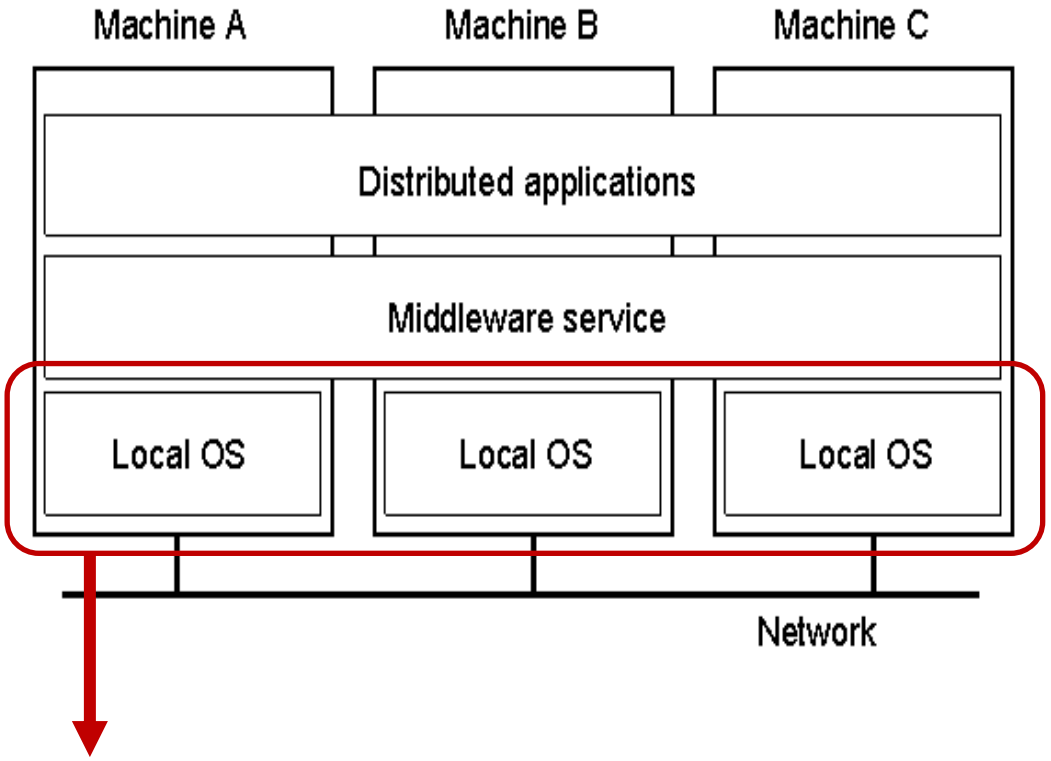
Mestrado em Engenharia Informática e de Computadores (MEIC)
Mestrado em Engenharia Informática e Multimédia (MEIM)

Infraestruturas de suporte à execução

Virtualização e Contentores

Luís Assunção (lass@isel.ipl.pt ; luis.assuncao@isel.pt)
José Simão (jsimao@cc.isel.ipl.pt ; jose.simao@isel.pt)

Middleware



Conceito de *Middleware* numa infraestrutura local

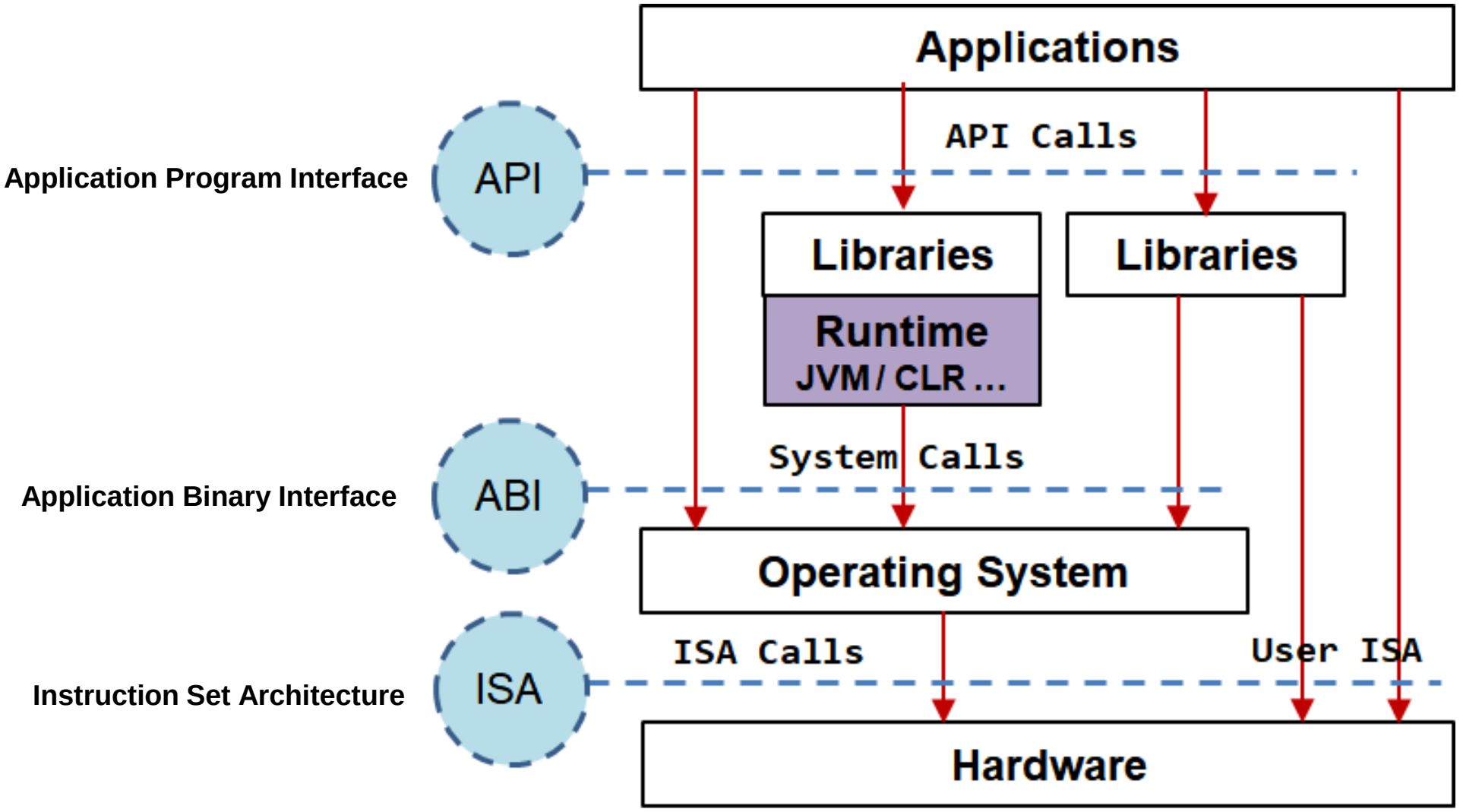
© From: Distributed Systems, Principles and Paradigms - Andrew Tanenbaum

- **Alojamento das componentes dos sistemas distribuídos**
  - **Processo de sistema operativo numa máquina física e/ou virtual, *Containers***

Porquê a virtualização

- **Consolidação de recursos**
  - **Menos espaço para mais serviços (menos energia, recursos humanos, ...)**
- **Uso de sistemas legados (*legacy applications*)**
- **Isolamento**
  - **VM e/ou *Containers* isolam falhas de segurança ou erros em componentes de software**
- **Ambientes de desenvolvimento e de investigação**
  - **Facilidade de criação de ambientes com *stack* de software bem definido**
- **Rapidez de aprovisionamento e escalabilidade**
  - **Para melhor acomodar um aumento de carga (ex: número de pedidos)**
- **Migração e balanceamento de carga**
  - **Migração de VMs para consolidar e otimizar a utilização de hardware**
- **Backups e recuperação de desastres**

Interfaces de um sistema computational



Máquinas virtuais de processo ou sistema

- O que é a "máquina virtual" depende da perspetiva do processo ou do sistema operativo
  - Na perspetiva do processo, a máquina é representada pela ABI, e na perspetiva da aplicação é representada pela API
  - Na perspetiva do SO, a máquina é representado pela ISA

- O software que suporta uma "máquina virtual" de processo é designado de *runtime (ex: JVM, CLR)*

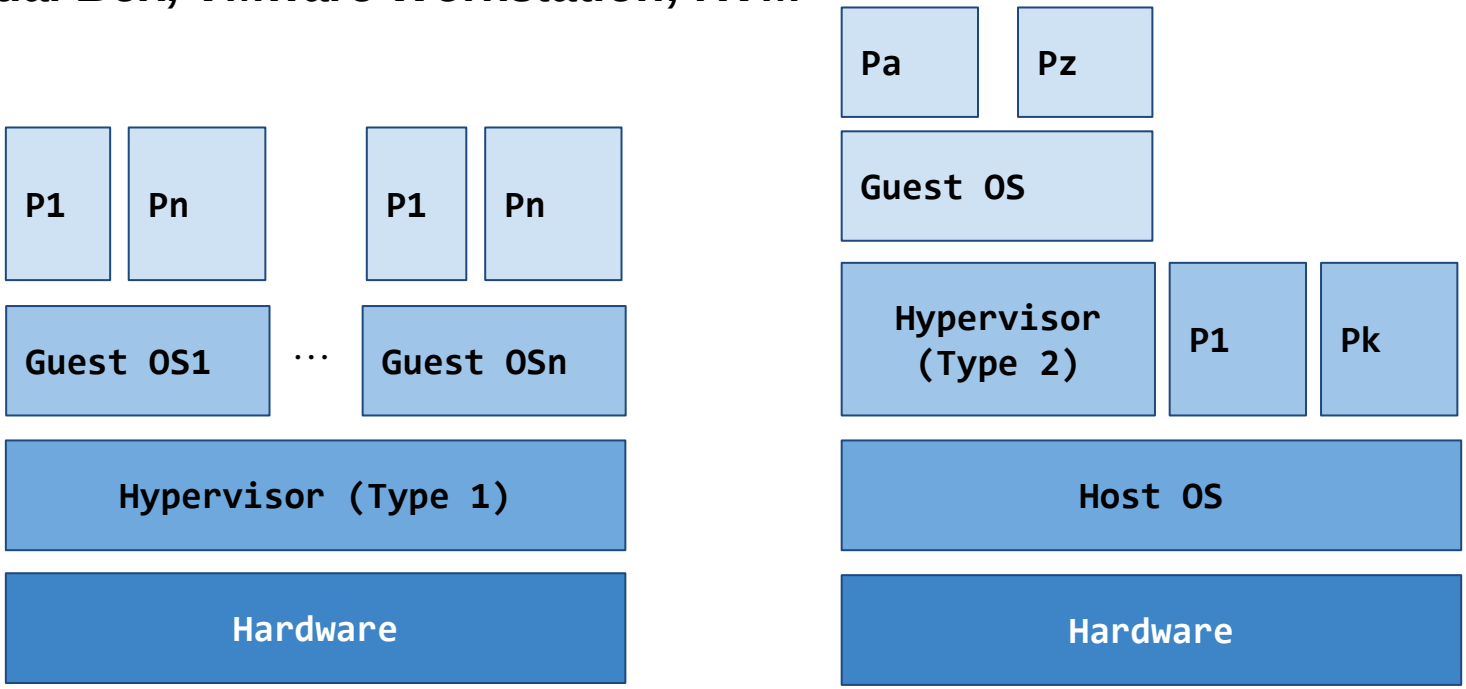
- O software que suporta uma "máquina virtual" de sistema (VM) é referido como *virtual machine monitor* (VMM) ou *hypervisor*.

- O sistema operativo de uma VM é o *guest* (convidado)

- O software que suporta a VM é o *host* (hospedeiro)

Hypervisor e execução privilegiada

- O *hypervisor* de um ambiente virtualizado é classificado do Tipo 1 ou do Tipo 2, se respetivamente, não depende, ou depende da existência de um sistema operativo
- O Tipo 1 (ou *bare metal*) interage diretamente com o *hardware* e não necessita de um sistema operativo, introduzindo menos *overhead. Exemplos: Citrix/Xen, VMware ESXi; Microsoft Hyper-V*
- O Tipo 2 executa-se sobre um sistema operativo, tirando partido da transparência que esse sistema tem a diferentes *hardwares. Exemplos: Microsoft Virtual PC, Oracle Virtual Box, VMware Workstation; KVM*



Cloud Computing

- ***"A model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.", NIST 2011***

Service Class	Access & Tools	Service contents
SaaS Software as a Service	Web Browser	**Cloud Applications:** *Social Networks, Email, Office suites (Google docs), ERP, CRM, IAM (Identity and Access Management), Video processing, …*
PaaS Platform as a Service	Development Environments	**Cloud Platform:** Programming languages, frameworks, *Mashups* editors, Web APIs, Data Storage models (Relational, NoSQL), *Data Analytics, …*
IaaS Infrastructure as a Service	Virtualization Manager	**Cloud Infrastructure:** Computer servers, Data Storage, Firewall, Load Balancer, IP Addressing, VPN, ….

Virtualização: Organizações, Clouds públicas, …

Centros de dados das organizações

. . .

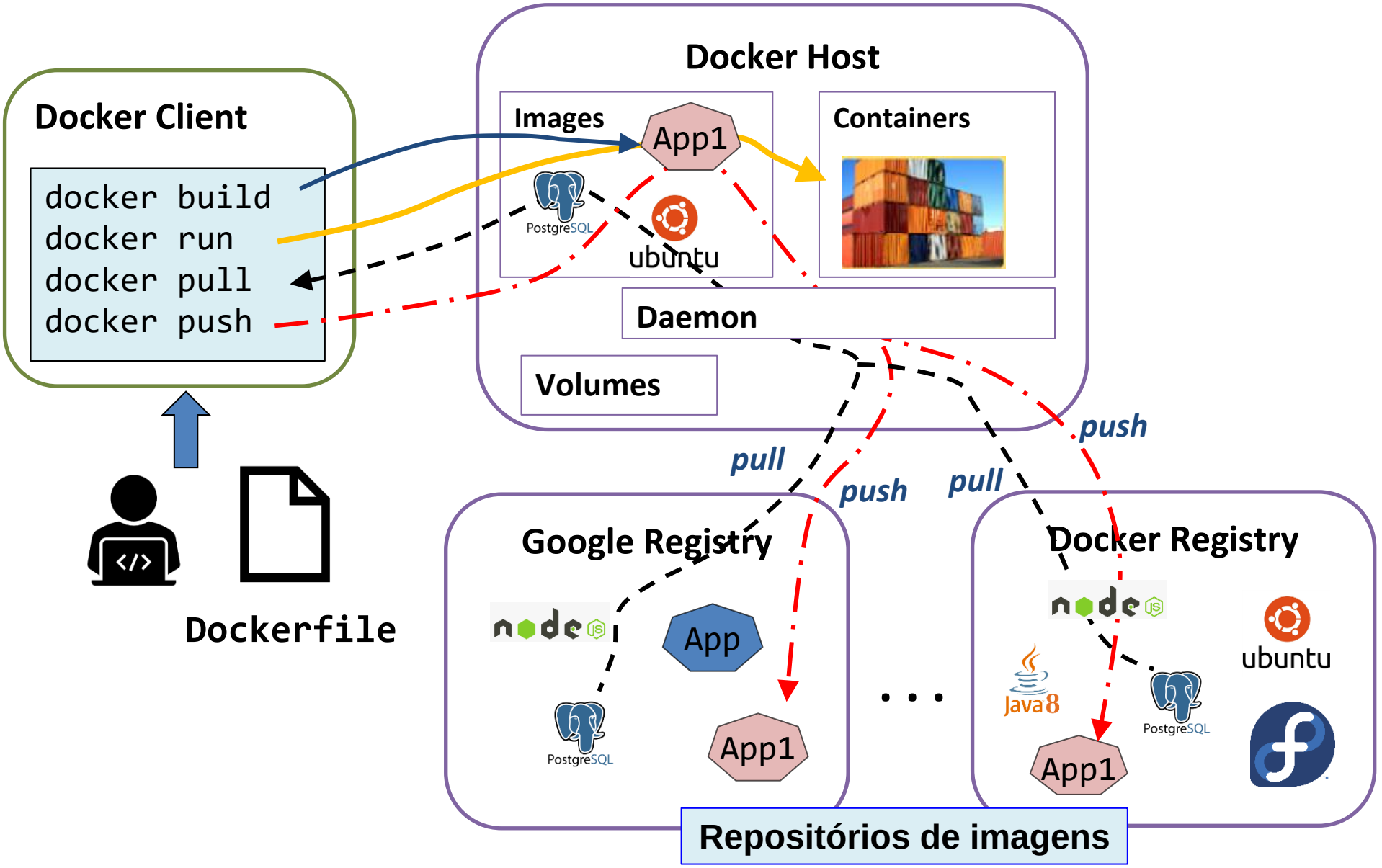
Deployment genérico de software

- Uma aplicação pode ter várias dependências que precisam de estar presentes no sistema alvo
  - *runtime*; bibliotecas; outras aplicações
- O *deployment* pode ser feito com base em imagens de VMs
  - No entanto, as imagens podem não ser portáveis entre diferentes *infraestruturas* com suporte de virtualização, nomeadamente na *Cloud*
- Uma abordagem diferente tem sido usada nos últimos anos com *containers* para executar imagens binárias, partilhadas de forma independente entre infraestruturas heterogéneas
  - Um *container* facilita o processo de desenvolvimento, teste, *deployment* e operação de sistemas:
    - Separação de responsabilidades entre componentes do sistema
    - Divisão de tarefas entre equipas
    - Isolamento e transparência face aos recursos computacionais
    - Automatização das operações de *deployment* e monitorização em produção
  - Mas! continuam a existir os desafios de segurança, coordenação e interação entre as partes bem como o tratamento das falhas parciais

Empacotamento em imagens de *containers*

- Os *containers* são, no essencial, processos que correm no contexto do sistema operativo, geridos por um *runtime*
- Fornecem um isolamento inferior ao das VMs mas superior ao de processos regulares do sistema operativo, virtualizando o acesso ao sistema de ficheiros e utilizando os recursos (CPU, Mem, I/O) com algumas restrições
- Os *containers* executam imagens binárias, previamente construídas, e comprometidas com uma *Application Binary Interface* (ABI) (ex: linux, *windows*)
  - Para determinados ambientes de execução, incluindo *runtimes*, *middlewares* e aplicações, existem normalmente imagens para diferentes plataformas de *hardware*: arm64, amd64, …
- Existem várias concretizações de *runtimes* de *containers*:
  - Linux *Containers*: LXC, LXD, RKT, CRI-O, . . .
  - Docker (que iremos usar como base para desenvolver aplicações distribuídas)

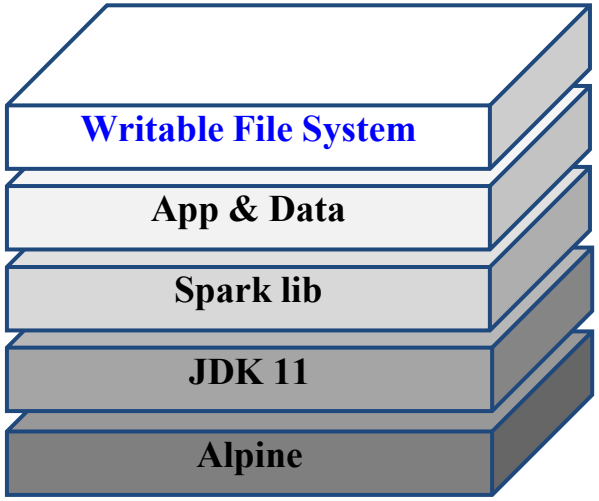
Sistema Docker



Imagens e *Containers*

- Uma imagem de *container* representa um ou mais ficheiros (obtidos de um Docker *Registry*, ex: *hub.docker.com*), e usados localmente para iniciar um *container*
- Existem formatos diferentes de imagens mas a generalidade das ferramentas suporta o formato aberto OCI
  (*Open Container Image*: https://www.opencontainers.org/)
  - Cada imagem depende de outra e adiciona algum *middleware* e/ou aplicação
  - Existem imagens base Linux e Windows
- Quando o *container* se inicia tem disponível um *file system* isolado do *host*
  - As alterações no *file system* não são persistidas

Se a App escrever num ficheiro e o *container* terminar os dados são perdidos



Exemplo: Instalar Docker *engine* em CentOS

***Host* com sistema Linux (CentOS 8), numa VM GCP**

https://docs.docker.com/engine/install/centos/

- `sudo yum install -y yum-utils`
- `sudo yum-config-manager --add-repo \`
  `https://download.docker.com/linux/centos/docker-ce.repo`
- `sudo yum install docker-ce docker-ce-cli containerd.io`
- `sudo systemctl start docker`
- `sudo docker run hello-world`

Para evitar usar `sudo` ao interagir com o *deamon* docker, é possível adicionar o utilizador Linux da VM GCP ao grupo privilegiado `docker` e `ativar` as mudanças no grupo

- `sudo usermod -aG docker $USER`
- `sudo newgrp docker`

Exemplo: *release* do SO dentro do *container*

- **Execução de *shell* "`/bin/bash`" em *container* com imagem de sistema operativo Fedora**

```
docker run -i -t fedora
```

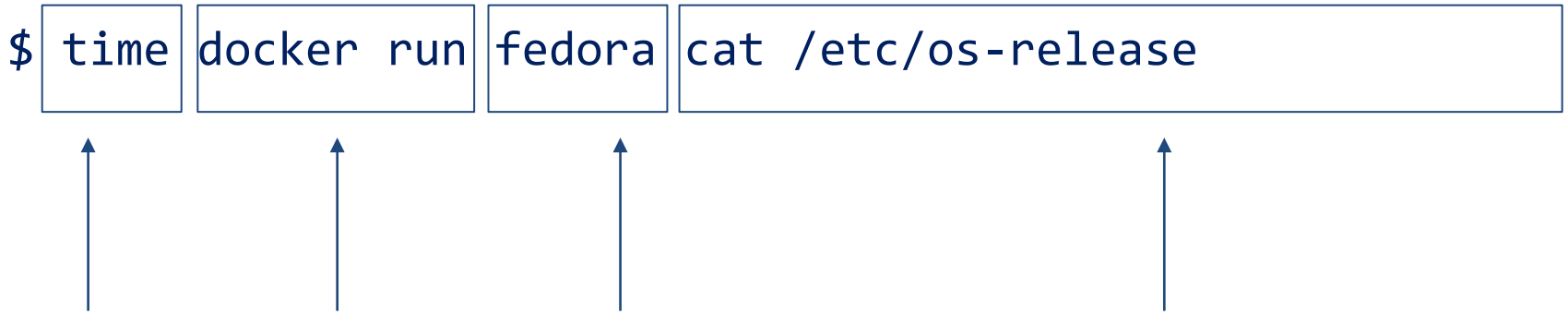
```
-i iterativo
-t pseudo-terminal (/bin/bash)
```

Primeira vez que a imagem é usada localmente é feito download do repositório de imagens

```
Unable to find image 'fedora:latest' locally
latest: Pulling from library/fedora
4c69497db035: Already exists
Digest: sha256:ee55117b3058f2f12961184fae4b9c392586e400487626c6bd0d15b4eae94ecc
Status: Downloaded newer image for fedora:latest
[root@57cad972cfe4 /]# cat /etc/os-release
NAME=Fedora
VERSION="31 (Container Image)"
ID=fedora
VERSION_ID=31
VERSION_CODENAME=""
PLATFORM_ID="platform:f31"
PRETTY_NAME="Fedora 31 (Container Image)"
```

Shell dentro do container

Iniciar *container* e execução de comando




```
real    0m0.599s
user    0m0.029s
sys     0m0.022s
```

- **A execução de um *container* `fedora` e, dentro deste, o comando `cat /etc/os-release`, demorou no total ~0.6 seg.**
- **As VMs podem demorar algumas dezenas de segundos a iniciar**

Os 4 passos do ciclo de desenvolvimento e produção

1. **Desenvolver a aplicação usando a linguagem de programação, bibliotecas e ambiente de execução que sejam apropriados**

2. **Criar a imagem binária com a aplicação desenvolvida, referindo uma imagem base que tenha apenas o sistema operativo ou já com alguns dos componentes (bibliotecas, *middleware*, etc.)**

3. **Publicar a imagem num repositório (*Docker Registry*) de imagens**

4. **Em qualquer sistema, com suporte para *containers*, é possível lançar em execução um *container* a partir da imagem binária publicada no ponto 3**

Exemplo de construção de imagem a partir de JAR

Artefacto: ChatServiceImpl-1.0-jar-with-dependencies.jar

Dockerfile:

Dockerfile

```
FROM openjdk:11

RUN mkdir /usr/chatserver

WORKDIR /usr/charserver

COPY ChatServiceImpl-1.0-jar-with-dependencies.jar .

CMD ["java", "-jar", "ChatServiceImpl-1.0-jar-with-dependencies.jar"]
```

Imagem base

Diretoria de trabalho dentro do *container*

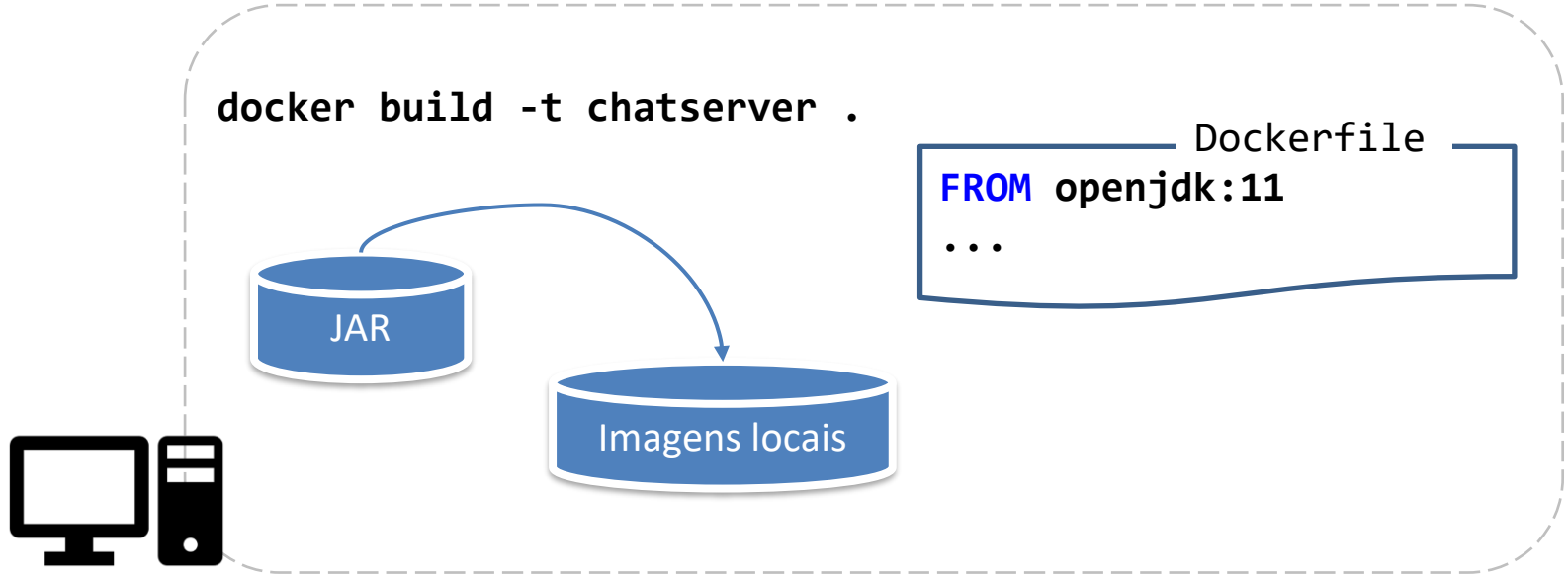
Copia JAR do *host* para imagem

Comando a executar quando se iniciar o *container*

https://docs.docker.com/develop/develop-images/dockerfile_best-practices/

“Dockerizar” uma aplicação

- O ficheiro *`Dockerfile`* refere as imagens base e as aplicações que têm de ser copiadas do host para a imagem
- O comando *`docker build`* constrói novas imagens a partir de um ficheiro *`Dockerfile`*



- Imagem fica guardada localmente com um nome e uma *tag* (no exemplo, `chatserver:latest`)

Exemplo de execução do comando build

```
$ docker build -t chatserver .
Sending build context to Docker daemon  15.29MB
Step 1/5 : FROM openjdk:11
11: Pulling from library/openjdk
bb7d5a84853b: Pull complete
da1c1e7baf6d: Pull complete
1d2ade66c57e: Pull complete
Digest: sha256:d2118a2aed78004c41277c179441ed86eea5e2b07fd0ac5a029f145208daf1b3
Status: Downloaded newer image for openjdk:11 ---> 189a7454500c
Step 2/5 : RUN mkdir /usr/chatserver
 ---> Running in ce3042270d5b
Removing intermediate container ce3042270d5b
 ---> 1869b3c02554
Step 3/5 : WORKDIR /usr/chatserver
 ---> Running in c7d1fe6bb3c5
Removing intermediate container c7d1fe6bb3c5
 ---> 9e33d4cf1725
Step 4/5 : COPY ChatServiceImpl-1.0-jar-with-dependencies.jar .
 ---> c21a903afd87
Step 5/5 : CMD ["java", "-jar", "ChatServiceImpl-1.0-jar-with-dependencies.jar", "chatserver.chatApp"]
 ---> Running in b5dbb6e78bb9
Removing intermediate container b5dbb6e78bb9
 ---> 09cb66386e39
Successfully built 09cb66386e39
Successfully tagged chatserver:latest
```

Execução dos passos descritos no Dockerfile

Ciclo de vida de imagem num *container*

- **Inicío de execução**
  - **exposição do porto 9000 do *container* como porto 8000 do *host***

```
$ docker run -i -t -p 8000:9000 chatserver
Oct 18, 2021 10:29:54 AM chatserver.ChatServer main
INFO: Server started, listening on 9000
*** server await termination
```

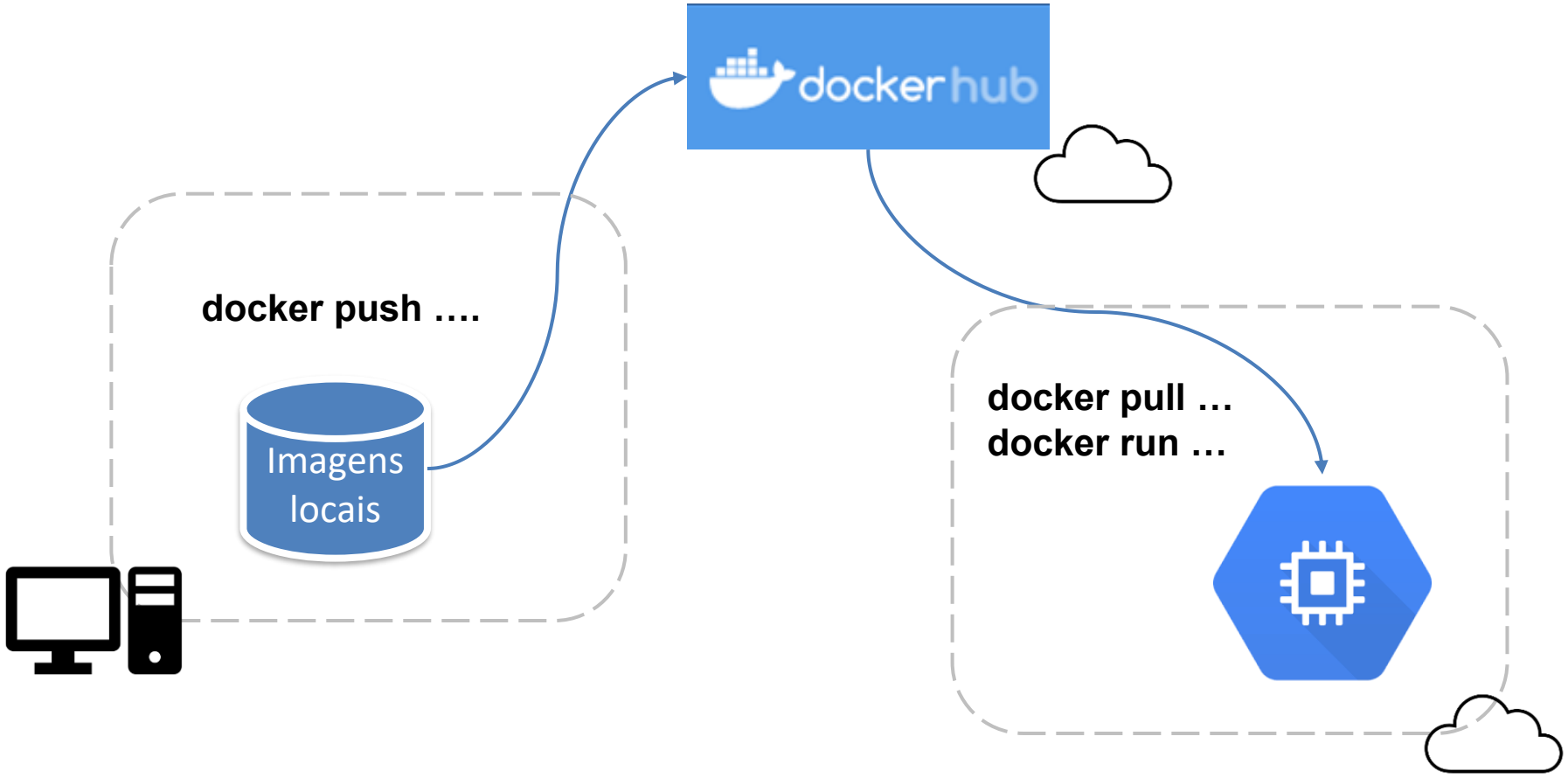
- **Observação de execução**

```
$ docker ps -all
CONTAINER ID   IMAGE         COMMAND                  CREATED          STATUS
PORTS                     NAMES
5a2efe12833e   chatserver    "java -jar ChatServi…"   9 seconds ago    Up 8 seconds
0.0.0.0:8000->9000/tcp    exciting_agnesi
```

- **Destruição do container em execução**

```
$ docker kill 5a2
5a2
```

Push, Pull e execução *anywhere*



Publicar imagem em repositório Dockerhub (1)

- **Para publicar do Dockerhub é preciso estar autenticado**

```
$ docker login
Login with your Docker ID to push and pull images from Docker Hub. If you don't have
a Docker ID, head over to https://hub.docker.com to create one.
Username: user
Password: password
Login Succeeded
```

- **A imagem local tem de ser marcada (*tag*) com o formato adequado ao repositório remoto (<user>/<image name>[:<tag>])**

```
$ docker tag chatserver jslaisel/chatserver
```

alternativa

```
$ docker tag chatserver jslaisel/chatserver:v1
```

```
REPOSITORY          TAG       IMAGE ID       CREATED          SIZE
jslaisel/chatserver v1        a19f8ffe21da   34 minutes ago   484MB
jslaisel/chatserver latest    a19f8ffe21da   34 minutes ago   484MB
chatserver          latest    a19f8ffe21da   34 minutes ago   484MB
```

- **O comando *`build`* pode usar este formato para publicar em repositório externo**

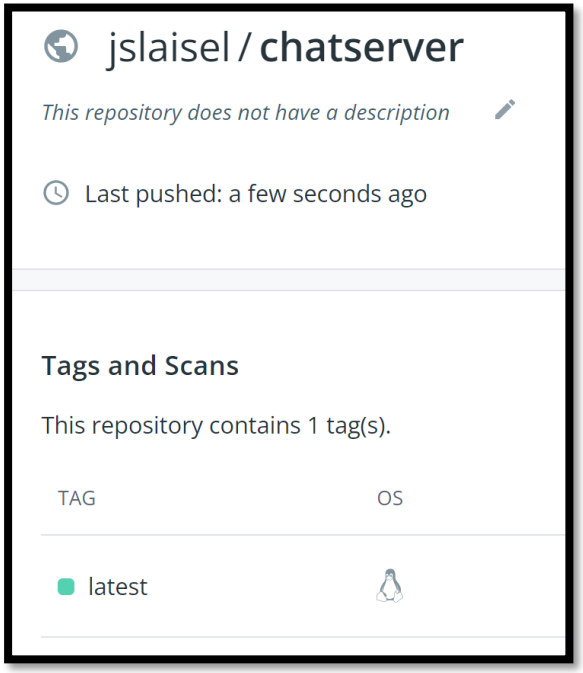
```
$ docker build -t jslaisel/chatserver:v1 .
```

Publicar imagem em repositório Dockerhub (2)

- **A imagem pode agora ser publicada no repositório**

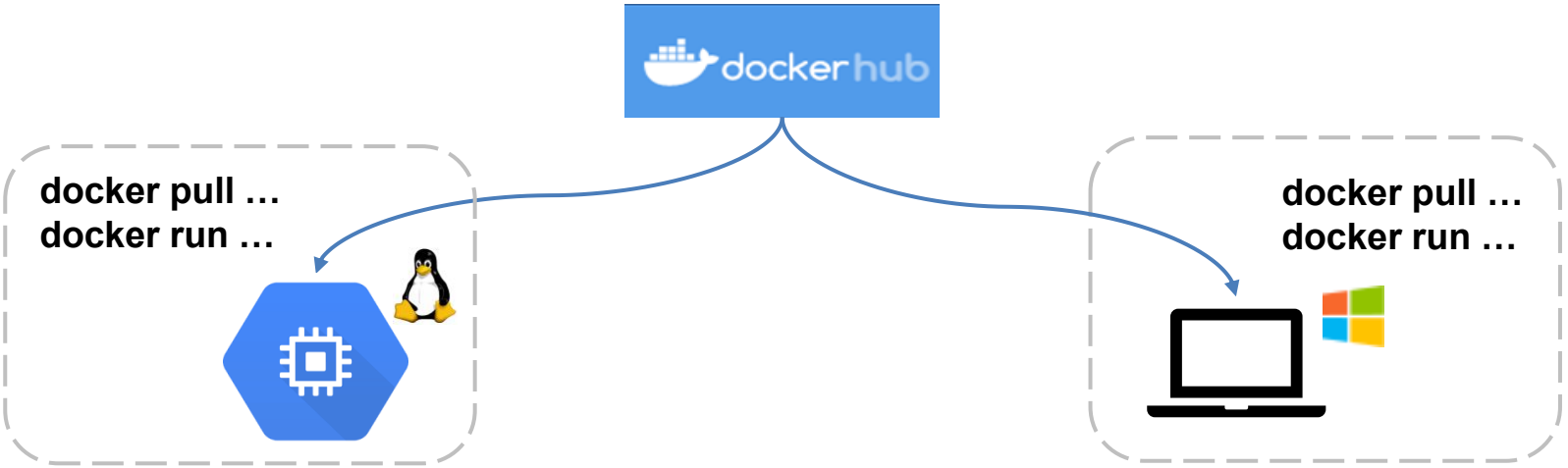
```
$ docker push jslaisel/chatserver
Using default tag: latest
The push refers to repository [docker.io/jslaisel/chatserver]
...
```

https://hub.docker.com/repository/docker/jslaisel/chatserver



- **Mais comandos**
  - **https://www.docker.com/sites/default/files/d8/2019-09/docker-cheat-sheet.pdf**

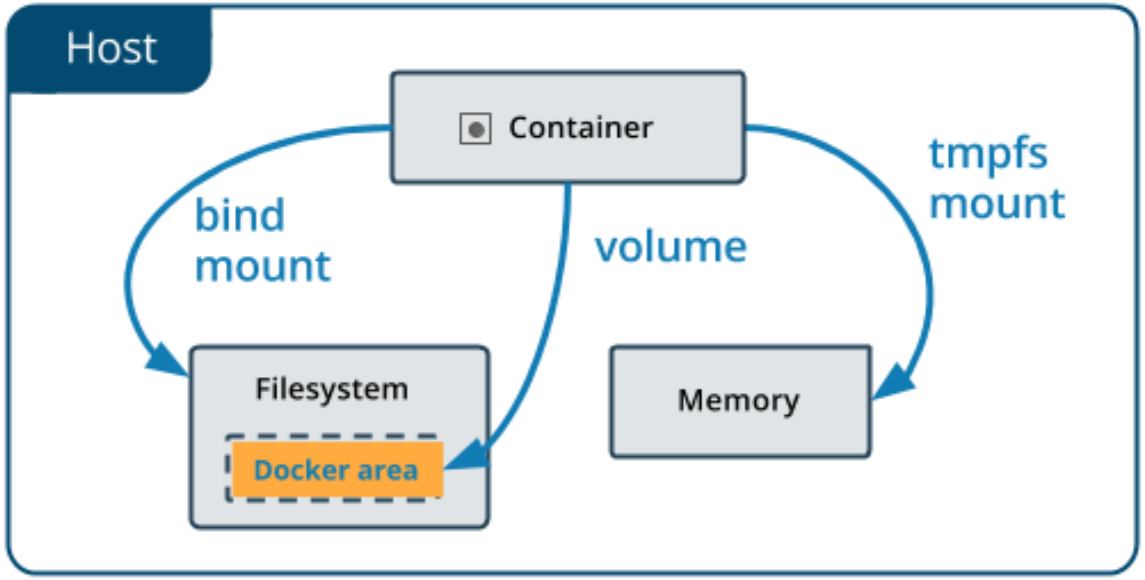
Exemplo de *pull* e *run*




```
$ docker pull jslaisel/chatserver
Using default tag: latest
latest: Pulling from jslaisel/chatserver
1a0005db7778: Already exists
...
Digest: sha256:fcd930dd63f881b7310bf929a24867b89ba808dfcd14284d1b6510cda561b1b1
Status: Downloaded newer image for jslaisel/chatserver:latest
docker.io/jslaisel/chatserver:latest
$ docker run jslaisel/chatserver
Oct 28, 2021 12:15:26 PM chatserver.ChatServer main
INFO: Server started, listening on 9000
*** server await termination
```

Persistência de dados

- **A camada de escrita do *container* não persiste as alterações feitas no sistema de ficheiros após o container ser destruído**
- **Para persistir dados ao nível do sistema de ficheiros do *host*, o *container* tem de partilhar uma pasta com o sistema de ficheiros do *host***



https://docs.docker.com/storage/volumes/

Exemplo com *volumes*

containers a partilhar um volume de nome *volshare*

```
$ docker volume create volshare
volshare
$ docker run --name fedora-1 -it -v volshare:/myshare fedora
[root@80c38b8c7d31 /]# ls /myshare
[root@80c38b8c7d31 /]# date > /myshare/fedora-1.txt
[root@80c38b8c7d31 /]# ls /myshare
fedora-1.txt  fedora-2.txt
[root@80c38b8c7d31 /]# cat /myshare/fedora-2.txt
Fri May 29 08:29:22 UTC 2020
[root@80c38b8c7d31 /]#
```

```
$ docker run --name fedora-2 -it -v volshare:/extshare fedora
[root@fafa04f367f1 /]# ls /extshare
fedora-1.txt
[root@fafa04f367f1 /]# date > /extshare/fedora-2.txt
[root@fafa04f367f1 /]# ls /extshare
fedora-1.txt  fedora-2.txt
[root@fafa04f367f1 /]# cat /extshare/fedora-1.txt
Fri May 29 08:26:51 UTC 2020
[root@fafa04f367f1 /]#
```

```
$ docker volume ls
DRIVER      VOLUME NAME
local       volshare
```