

Jampp <> ECI

Docker para Big Data



Dante Pawlow

Data Engineer

dante.pawlow@jampp.com

Agenda

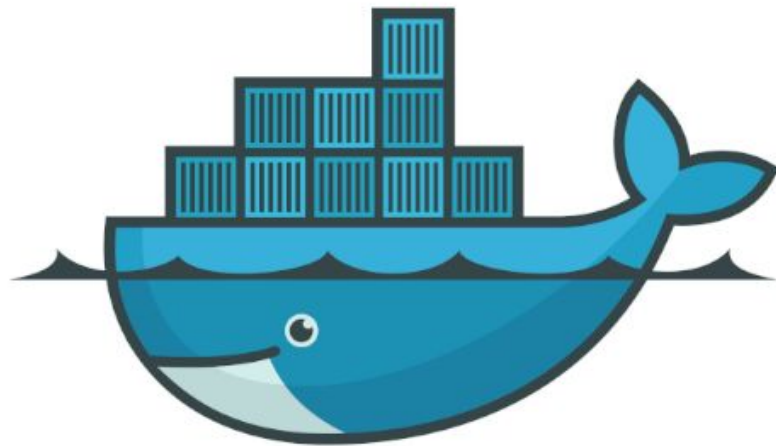


- ¿Qué es docker?
- ¿Cómo levanto mi aplicación?
- ¿Cómo conecto varias aplicaciones entre sí?
- ¿Puedo tener un entorno de Big Data en mi máquina?

¿Qué es Docker?

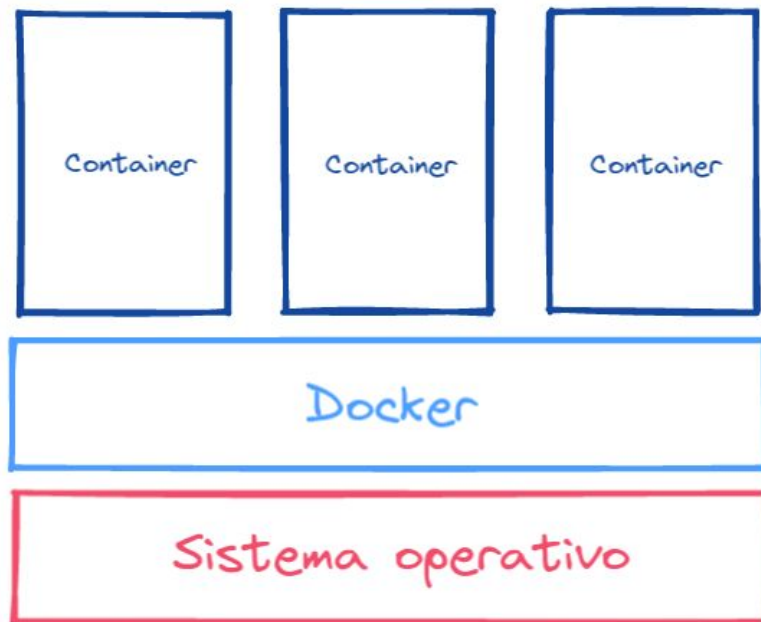
Docker

- Es un proyecto de **código abierto**.
- Sirve para empaquetar aplicaciones junto con sus dependencias de manera estandarizada.
- Es mucho más liviano que usar una VM.
- Su unidad de trabajo son los **containers**.



Container

- Es un paquete ejecutable muy liviano.
- Incluye todo lo que se necesita para ejecutar la aplicación: código, runtime, herramientas y librerías del sistema y configuraciones.
- Aisla al software de su entorno.
- Se crea en tiempo de ejecución a partir de la especificación de una Imagen.



Imagen

- Es la especificación de un container.
- Se define a través de un **Dockerfile**.
- Se puede usar como base de otra Imagen.
- Existen millones de imágenes y Dockerfiles públicas con aplicaciones ya instaladas.

```
ARG HADOOP_VERSION=2.8.5

FROM public.ecr.aws/jampp/hadoop:${HADOOP_VERSION}

ARG HADOOP_VERSION
ARG BUILD_DATE
ARG BUILD_VERSION

RUN mkdir -p /hadoop/dfs/data
VOLUME /hadoop/dfs/data

ADD run.sh /run.sh
RUN chmod a+x /run.sh

EXPOSE 50075

CMD ["/run.sh"]
```

¿Cómo levanto mi aplicación?

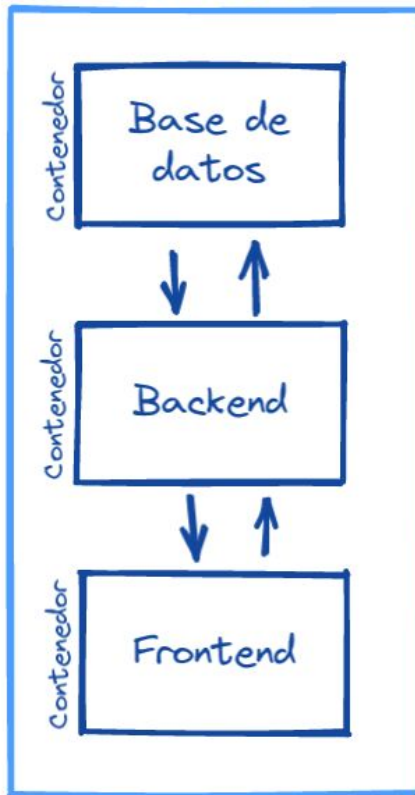
Acá pensaba mostrar un ejemplo
compartiendo mi pantalla, pero
tal vez agregue una slide con
algunos de los comandos que se
usan

¿Cómo conecto varias
aplicaciones entre sí?

Docker Compose

- Herramienta para definir las interacciones entre varios containers
- Pueden ser varias instancias de la misma aplicación o aplicaciones distintas
- Es simplemente un archivo de configuración
- Permite definir y restringir las conexiones entre containers

Red de docker



(acá muestro ejemplo de un
docker-compose.yml y de cómo
interactúan dos containers)

Caso de uso: entorno de Big Data con Docker Compose

Nuevamente acá comparto mi pantalla, pero tal vez agrego una slide con un pequeño esquema de la infraestructura que replica nuestro docker env

Conclusiones



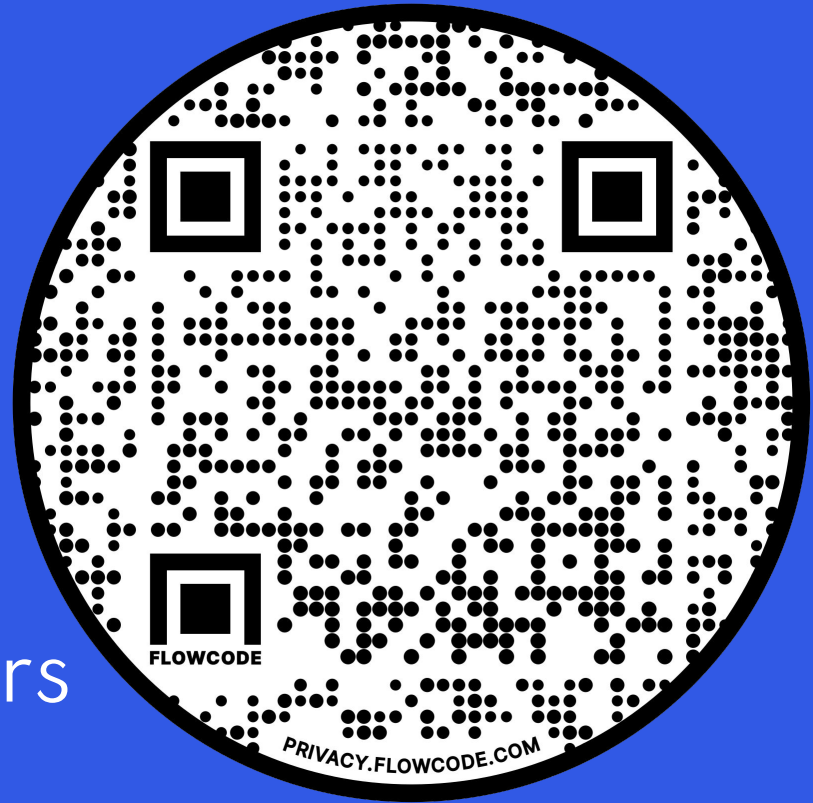
- ✓ Facilita el aislamiento del entorno
- ✓ Facilita la portabilidad de aplicaciones
- ✓ Las imágenes públicas facilitan el set-up
- ✓ Se pueden tener varias aplicaciones corriendo en la misma máquina sin conflictos

We are hiring!

www.jampp.com/jobs



Data Analysts
ML Engineers
Performance
Engineers
Data Engineers
Cloud Engineers
Full Stack Engineers
....& more!



¡Gracias!