

Laboratorio 1: Preparación del entorno de trabajo

Tópicos selectos en TIC - Introducción a la Ingeniería de Datos

8 de noviembre de 2023

Resumen

El objetivo de este laboratorio es preparar el ambiente de trabajo. Se instalarán todas las herramientas de trabajo que usaremos posteriormente en otros laboratorios.

1. Introducción

Docker es un software de contenedorización que nos permite aislar el software de una manera similar a las máquinas virtuales pero de una manera mucho más ágil. Una imagen de Docker es un snapshot de un contenedor que podemos definir para ejecutar nuestro software. Al exportar nuestras imágenes de Docker a proveedores de nube como Amazon Web Services o Google Cloud Platform, podemos ejecutar nuestros contenedores allí.

Docker define un contenedor como: *"A container is a standard unit of software that packages up code and all its dependencies so the application runs quickly and reliably from one computing environment to another."* Docker.

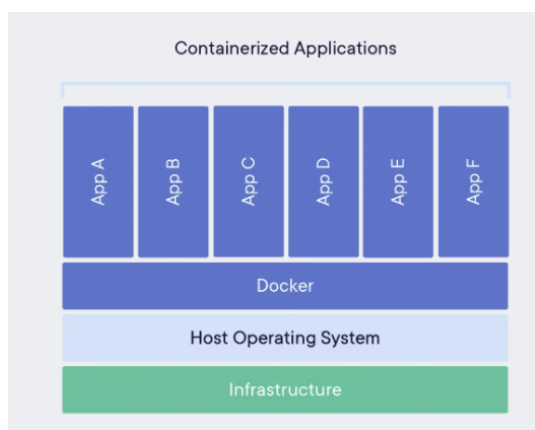


Figura 1: Container

Enlaces útiles:

1. What is a Container?

2. Cheatsheet Docker - 1

Los contenedores Docker no tienen estado: los cambios realizados dentro de un contenedor NO se guardarán cuando el contenedor se elimine y se vuelva a iniciar. Esto es una ventaja porque nos permite restaurar cualquier contenedor a su estado inicial de manera reproducible, pero tendrá que almacenar datos en otro lugar si necesita hacerlo; Una forma común de hacerlo es con volúmenes.

Clonar el repositorio base del laboratorio:
repositorio-base

2. Instalación de herramientas

2.1. Docker

Instalación de Docker.

1. **Windows:** Guía para instalación en Windows
2. **Ubuntu:** Guía para instalación en Ubuntu
3. **Mac:** Guía para instalación en MAC

Verificación de Docker. Abrir la terminal, y correr el siguiente comando:

```
$ docker run hello-world
```

Deberíamos observar una respuesta que contenga el texto: *"Hello from Docker"*

Familiarizarse con el entorno Docker.

2.2. PgAdmin

Instalación de PgAdmin.

1. **Windows:** Guía para instalación en Windows
2. **Ubuntu:** Guía para instalación en Debian - Ubuntu
3. **Mac:** Guía para instalación en MAC

2.3. IDE

Instalación de PyCharm. Guía de instalación para todas las plataformas

2.4. Desarrollo

Explorando Docker. Comandos básicos. Docker run: Contenedoriza una imagen. Crea el entorno aislado.

```
$ docker run [OPTIONS] IMAGE[:TAG|@DIGEST]
```

Ejemplo, container Python en manera interactiva (TTY activado, crea un shell en el container)

```
$ Docker run -it python:3.9
```

Donde encontramos mas imágenes? Docker Hub

Creando Dockerfile. Un Dockerfile es un document de texto que contiene todos los comandos que un usuario puede usar en la linea

de comandos para ensamblar una imagen. Nos permite automatizar el proceso de image building.

Creando un Docker Compose file Compose es una herramienta para definir y ejecutar aplicaciones Docker multicontenedor. Con Compose, se usa un archivo YAML para configurar los servicios de la aplicación. Luego, con un solo comando, crea e inicia todos los servicios desde su configuración.

Descargando el dataset. Usaremos el TLC Trip Record dataset. Podemos descargarlo de acá. Utilizaremos los datos de **Yellow Taxi Trip Records, del mes de enero de 2022**. Enlace de descarga

3. Resultados y discusiones

3.1. Preguntas

1. Explicar las diferencias entre **contenedorización y virtualización**.
2. Cual es la alternativa a Docker compose? En el caso de que queramos lanzar múltiples contenedores que estén en la misma red.
3. Que es el formato PARQUET? Cuales son las ventajas/desventajas frente a otros formatos tradicionales? (CSV, TSV)

Referencias