UNIVERSIDAD GALILEO DE GUATEMALA

CUARTO TRIMESTRE

PRODUCT DEVELOPMENT

ING. CARLOS ZELADA

AUX. OBED ESPINOZA

# MANUAL TECNICO

JHERSON SAZO

21000341

INDICE

Tabla de contenido

[MANUAL TECNICO 1](#_Toc90739045)

[1. Aspectos generales 3](#_Toc90739046)

[2. Tecnologías 3](#_Toc90739047)

[a. Docker: 3](#_Toc90739048)

[b. Streamlit: 3](#_Toc90739049)

[c. Mysql; 3](#_Toc90739050)

[d. Airflow: 3](#_Toc90739051)

[3. Diagramas 4](#_Toc90739052)

[a. Diagrama ETL 4](#_Toc90739053)

[b. Proceso ETL 5](#_Toc90739054)

[c. Diagrama Relacional 5](#_Toc90739055)

[4. Deployment de la aplicación 6](#_Toc90739056)

[a. Estructura del proyecto: 6](#_Toc90739057)

[b. Ejecución 7](#_Toc90739058)

[5. Requerimientos 8](#_Toc90739059)

## Aspectos generales

En este manual se describen las tecnologías y flujos utilizados para llevar a cabo una aplicación en la que se pueda consultar de forma dinámica los casos de COVID, proporcionados a través de archivos csvs.

## Tecnologías

En esta sección se describen las tecnologías utilizadas en la aplicación con una pequeña descripción de cada una.

### Docker:

Docker es una plataforma de software de código abierto para crear, implementar y administrar contenedores de aplicaciones virtualizados en un sistema operativo (SO) común, con un ecosistema de herramientas aliadas

### Streamlit:

Streamlit es una biblioteca de Python de código abierto que facilita la creación y el intercambio de hermosas aplicaciones web personalizadas para el aprendizaje automático y la ciencia de datos.

### Mysql;

Es un sistema de gestión de bases de datos relacional desarrollado bajo licencia dual: Licencia pública general/Licencia comercial por Oracle Corporation y está considerada como la base de datos de código abierto más popular del mundo,1​2​ y una de las más populares en general junto a Oracle y Microsoft SQL Server, todo para entornos de desarrollo web.

### Airflow:

Apache Airflow es una herramienta de tipo workflow mánager que utiliza el lenguaje de programación Python, sus funciones principales son: gestionar, monitorizar y planificar flujos de trabajo, usada como orquestador de servicios.

## Diagramas

En esta sección se brinda un detalle del flujo que realiza airflow para realizar el proceso de ETL.

### Diagrama ETL

Diagrama

Descripción generada automáticamente

### Proceso ETL

El proceso consiste en tres sensores que se mantienen escuchando una carpeta asignada a una conexión de tipo File, donde cuando llega el archivo especificado, puede avanzar en el flujo.

El primer flujo es el mas importante, debido a que en este se realiza la carga de los países y sus provincias, esta carga a base de datos solo se realiza en este punto, los dos flujos siguientes solo realizan carga a su tabla correspondiente.

Cada procesamiento de archivos realiza las siguientes tareas:

* Elimina registros que tengan latitud y longitud vacíos.
* Que no tengan ningún caso registrado.
* Elimina caracteres especiales que pueden interferir en la relación de las tablas.

Cada ingesta de datos a base de datos para los casos confirmados, fallecidos y recuperados realiza lo siguiente:

* Se dispone en memoria la llave correspondiente a su país/provincia.
* Consulta a un diccionario por la llave correspondiente, y realiza una resta de los casos detectados el día anterior con el actual para tener el registro correcto de nuevos casos por día.

## Diagrama Relacional

Se muestra el modelo relacional utilizado en la aplicación.

Diagrama

Descripción generada automáticamente

## Deployment de la aplicación

### Estructura del proyecto:

Escala de tiempo

Descripción generada automáticamente con confianza baja

### Ejecución

Para ejecutar la aplicación, se requiere abrir una terminal en el directorio del proyecto, y ejecutar docker-compose up.

Si se cumplen correctamente la instalación de las dependencias de cada contenedor, deberá aparecer en consola lo siguiente:

Texto

Descripción generada automáticamente

Validar en los puertos 8080 del host:

Interfaz de usuario gráfica, Texto, Sitio web

Descripción generada automáticamente

Validar en el puerto 8501 para el dashboard en streamlit:

Imagen que contiene Texto

Descripción generada automáticamente

## Requerimientos

Para que la aplicación pueda ser replicada de forma local o en un servidor remoto, este debe cumplir con los siguientes programas instalados:

* Python 3.7 o superior
* Docker cli, Docker server
* Docker compose