

Universidad de San Carlos de Guatemala

Facultad de Ingeniería

Escuela de Ciencias y Sistemas

Seminario de Sistemas 2

Segundo Semestre 2024

Ing. Luis Alberto Vettorazzi Espana

Ing. Fernando Jose Paz Gonzales

Aux. Jose Fernando Alvarez

Morales

Aux. Sergio Enrique Cubur



# Proceso ETL

Primera practica de laboratorio

<b>Objetivos</b>	<b>2</b>
Objetivos generales	2
Objetivos específicos	2
<b>Descripción</b>	<b>2-3</b>
Descripción general	2
Aplicación	2
Consultas	3
<b>Documentación</b>	<b>3</b>
<b>Restricciones</b>	<b>3</b>
<b>Entrega</b>	<b>5</b>

# 1. Objetivos

## 1.1. Objetivos Generales

1. Aprender el proceso de ETL
2. Brindar resultados con la información obtenida

## 1.2. Objetivos Específicos

1. Utilizar el lenguaje Python para el procesamiento de información
2. Limpiar datos
3. Utilizar SQL Server para la creación de un Datawarehouse

# 2. Descripción

## 2.1. Descripción General

Se le solicita que realice un proceso de ETL (Extraer, Transformar, Cargar) de los archivos que se les brindaran. En base a los archivos usted deberá generar el modelo que se adapte mejor, y posteriormente, realizar las consultas que se le soliciten.

## 2.2. Aplicación

Se realizará una aplicación de consola en Python, la cual tendrá las siguientes opciones:

- a) **Borrar modelo:** Se ejecutará un script y se borrará cualquier tabla que utilice para el desarrollo de la práctica.
- b) **Crear modelo:** Se ejecutará el script de creación de las tablas del modelo que el estudiante haya desarrollado.
- c) **Extraer información:** Se solicitará la ruta de los archivos de carga, para proceder a extraer la información de estos.
- d) **Cargar información:** Se ejecutará un script que transforme la información de los archivos de entrada y luego se cargue al modelo desarrollado.
- e) **Realizar consultas:** Ejecutará un script con las consultas solicitadas y guardará los resultados en un archivo de texto.

### 2.3. Consultas

1. Debe mostrar un SELECT COUNT(\*) de todas las tablas para ver que si realizo la carga en las tablas del modelo.
2. Porcentaje de pasajeros por género.
3. Nacionalidades con su mes año de mayor fecha de salida.

Nacionalidad	01-2021	02-2021	03-2021
Guatemala	23	12	32

4. Count de vuelos por país.
5. Top 5 aeropuertos con mayor número de pasajeros.
6. Count dividido por estado de vuelo.
7. Top 5 de los países más visitados.
8. Top 5 de los continentes más visitados.
9. Top 5 de edades dividido por género que más viajan.
10. Count de vuelos por MM-YYYY.

## 3. Documentación

1. Diagrama del modelo realizado
2. Manual técnico (Descripción del modelo creado identificando cual es la tabla de hechos y cuáles son las tablas de dimensiones)

## 4. Restricciones

1. La aplicación debe ser desarrollada en el lenguaje de programación Python.
2. La aplicación será desarrollada en consola.
3. Toda la información y el modelo deben ser creados en SQL Server.
4. Copias obtendrán una nota de 0 y reportarán a la Escuela de Ciencias y Sistemas.

## 5. Entrega

1. **FECHA DE ENTREGA:** Domingo 11/08/2024 antes de las 23:59. No se aceptarán entregas a partir de esa hora.
2. Adjuntar lo solicitado en un archivo .zip con el siguiente formato [SS2]Practica1\_carnet.zip o repositorio correspondiente.
3. Subir el archivo comprimido o repositorio en la tarea asignada en UEDI.