

Universidad de San Carlos de Guatemala

Facultad de Ingeniería

Escuela de Ciencias y Sistemas

Seminario de Sistemas 2

Sección A

Segundo Semestre 2023

Ing. Luis Alberto Vettorazzi Espana

Aux. Freddy Alejandro Monterroso



USAC-Delivery

Proceso ETL

Primera práctica de laboratorio

Objetivos	2
Objetivos generales	2
Objetivos específicos	2
Descripción	2-3
Descripción general	2
Aplicación	2
Consultas	3
Documentación	3
Restricciones	3
Entrega	5

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 brindarán con la información de pedidos de la empresa USACDelivery en Guatemala. Las entregas de pedidos en este país tienen sus particularidades y variaciones, por lo que es esencial que el proceso de ETL se adapte a estas condiciones específicas.

En base a los archivos proporcionados, deberá generar un modelo que se ajuste de manera óptima a las características del sistema de entregas de pedidos en Guatemala. Esto implica considerar factores como las zonas geográficas de entrega, los tiempos promedio de entrega en diferentes regiones, los posibles obstáculos logísticos y las preferencias de los clientes guatemaltecos en cuanto a horarios y modalidades de entrega. Una vez que haya completado la etapa de transformación y haya generado un modelo sólido, podrá realizar las consultas solicitadas.

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. Top 5 Clientes con más entregas realizadas
3. Top 5 de productos con el costo de envío más alto.
4. Top 5 de ciudades con más entregas
5. Top 5 de envíos con estado Pendiente
6. Top 5 de envíos con estado Entregado
7. Total de entregas realizadas por ciudad.
8. Total de Entregas por días de la semana
9. Total de Entregas por mes
10. Total de Entregas por año

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) en PDF.

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. La entrega es individual.

6. Entrega

1. Crear un repositorio en GitHub con el siguiente formato SS2-Practica1_carnet, de manera privada y compartirlo con siguiente usuario: **freddy-201212854**
2. **FECHA DE ENTREGA:** domingo 20/08/2023 antes de las 23:59 PM. No se aceptarán entregas a partir de esa hora y tomar en cuenta que se tomará el último commit.
3. Subir el enlace de su repositorio en la tarea asignada en UEDI.