

Universidad San Carlos de Guatemala
Facultad de ingeniería.
Ingeniería en ciencias y sistemas



Práctica 1:

ETL con Python: de dataset crudo a
tabla relacional lista para análisis

PONDERACIÓN: 10 pts

 **Tiempo estimado: 16 hrs**

Índice

| | |
|--|----------|
| 1. MARCO FORMATIVO | 3 |
| 1.1. Valor..... | 3 |
| 1.2. Competencia(s) | 3 |
| 1.3. Objetivo SMART..... | 3 |
| 2. Enunciado de la Práctica..... | 4 |
| 2.1 Descripción del problema a resolver..... | 4 |
| 2.2 Alcance de la práctica | 4 |
| 2.3 Requerimientos técnicos | 5 |
| 3. Entregables..... | 5 |
| 4. Material de apoyo | 6 |
| 5. Recursos y herramientas a utilizar | 6 |
| 6. Cronograma..... | 7 |
| 7. Rúbrica de Calificación..... | 7 |
| 7.1 Requisitos para optar a la calificación | 7 |
| 7.2 Resumen de Puntuaciones..... | 8 |
| 7.3 Comentarios Generales..... | 8 |
| Detalle de la Calificación | 9 |

1. MARCO FORMATIVO

1.1. Valor

| | |
|-------------------------------------|--|
| Nombre del valor | ¿Cómo se aplica en tu laboratorio? |
| Responsabilidad académica y técnica | Cumplimiento riguroso de los lineamientos del proceso ETL, documentación del avance y entrega funcional del modelo implementado. |

1.2. Competencia(s)

| | |
|------------------------|--|
| Tipo de Competencia | Generales y específica |
| Competencia General | <ul style="list-style-type: none"> ● Aplica principios de ingeniería y ciencias de la computación integrando fundamentos teóricos y prácticos en la formulación y resolución adecuada de problemas de información ● Comunica resultados de análisis de datos utilizando diversos formatos y lenguajes dirigidos a diferentes audiencias profesionales y sociales |
| Competencia Específica | Analiza datos estructurados y no estructurados utilizando procesos de preparación, transformación y exploración para identificar patrones, relaciones y oportunidades de mejora |

1.3. Objetivo SMART

| SMART | Definición | Objetivo redactado |
|------------------------------|---|---|
| Específico (¿Qué?) | El objetivo es concreto y tangible. | Desarrollar un proceso ETL completo en Python que extraiga, transforme y cargue datos desde una fuente proporcionada hacia un modelo de inteligencia de negocios implementado en SQL Server. |
| Medible (¿Cuánto?) | El objetivo tiene una medida objetiva de éxito. | El objetivo se considerará cumplido al contar con una aplicación funcional que ejecute correctamente las tres fases del proceso ETL y cargue los datos en el modelo definido, verificando su contenido mediante consultas analíticas. |
| Alcanzable (¿Cómo?) | El objetivo debe ser posible con los recursos disponibles. | El desarrollo se realizará con Python y SQL Server, herramientas accesibles para los estudiantes, aplicando los conocimientos adquiridos sobre ETL y modelado de datos. |
| Realista (¿Para qué?) | El objetivo contribuye a metas más amplias. | El objetivo es fomentar la aplicación práctica de conceptos de ingeniería de datos, integrando análisis, modelado y procesamiento para soportar decisiones empresariales. |
| A Tiempo (¿Cuándo?) | El objetivo tiene fecha límite o mejor aún un cronograma de hitos de progreso | La práctica deberá completarse y entregarse según las fechas establecidas en el cronograma del curso del semestre 2026-1. |

2. Enunciado de la Práctica

2.1 Descripción del problema a resolver

Las organizaciones actuales generan grandes volúmenes de información provenientes de diversas fuentes. Para aprovechar estos datos de forma estratégica, se requiere implementar un proceso de integración y análisis estructurado.

En esta práctica, el estudiante deberá **diseñar e implementar un proceso ETL (Extracción, Transformación y Carga)** utilizando Python y SQL Server.

El escenario simula el manejo de información proveniente de registros de vuelos, donde los datos se encuentran en múltiples fuentes heterogéneas y deben ser estandarizados para su análisis.

El objetivo es construir un modelo multidimensional de inteligencia de negocios que permita ejecutar consultas analíticas y generar información útil para la toma de decisiones.

2.2 Alcance de la práctica

Obligatorio:

- Implementar una aplicación en **Python** que ejecute las tres fases del proceso ETL: extracción, transformación y carga.
- Crear un **modelo multidimensional de datos en SQL Server** que soporte consultas analíticas.
- Desarrollar consultas SQL que permitan validar la correcta carga de datos y generar indicadores relevantes (por ejemplo, número de vuelos, destinos más frecuentes, distribución por género, etc.).
- Documentar el proceso con diagramas y explicación técnica del modelo multidimensional implementado.
- Entregar un repositorio con la estructura establecida para el curso y un **README.md** con detalles del desarrollo y resultados.

Opcional / Recomendado:

- Incorporar manejo de excepciones y registro de errores en la aplicación.
- Implementar visualizaciones básicas con herramientas externas (Power BI, Tableau o Python con Matplotlib).
- Optimizar las consultas SQL para mejorar el rendimiento del modelo multidimensional.

2.3 Requerimientos técnicos

- Lenguaje de programación: **Python (versión 3.10 o superior)**
- Motor de base de datos: **Microsoft SQL Server**
- Librerías sugeridas: **pandas, pyodbc, sqlalchemy, datetime**
- Conocimiento y aplicación del **proceso ETL**: extracción de archivos fuente, transformación (limpieza, homologación y estandarización), y carga en tablas destino.
- Diseño e implementación de un **modelo multidimensional o relacional** para inteligencia de negocios.
- Uso de **consultas SQL analíticas** (agrupaciones, conteos, filtrado, etc.) para validación y obtención de resultados.
- Organización del proyecto en un repositorio con nombre **SS22S2026_#carnet**, carpeta **Practica1**, y documentación técnica en formato Markdown.

3. Entregables

| Tipo | Descripción |
|---|--|
| Aplicación ETL en Python | Script funcional que realice la extracción, transformación y carga de los datos desde las fuentes proporcionadas hacia el modelo multidimensional de base de datos. |
| Script de Base de Datos (SQL Server) | Archivo con las sentencias SQL necesarias para la creación de las tablas, relaciones y estructuras del modelo multidimensional de inteligencia de negocios. |
| Modelo de Datos | Diagrama que muestre las entidades, atributos y relaciones del modelo multidimensional implementado, con enfoque relacional o dimensional según corresponda. |
| Consultas Analíticas | Archivo con las consultas SQL utilizadas para validar la carga de datos y obtener información relevante (conteos, top 5, distribuciones, etc.). |
| Documentación Técnica | Documento en formato PDF o README.md que incluya descripción del proceso ETL, diseño del modelo multidimensional, pasos de ejecución y resultados obtenidos. |
| Repositorio del Curso | Carpeta Practica1 dentro del repositorio SS22S2026_#carnet , con todos los archivos organizados y nombrados conforme a los lineamientos del laboratorio. |

4. Material de apoyo

- Documentación oficial de Python <https://docs.python.org/3/>
- Tutorial de procesos ETL con Python y Pandas <https://realpython.com/python-etl/>
- Documentación de Pandas <https://pandas.pydata.org/docs/>
- Guía de conexión Python–SQL Server (pyodbc / SQLAlchemy) <https://learn.microsoft.com/sql/connect/python/>
- Fundamentos de SQL Server y consultas analíticas <https://learn.microsoft.com/sql/t-sql/>
- Introducción al modelado dimensional <https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/>

5. Recursos y herramientas a utilizar

Software:

- Python 3.10 o superior
- Microsoft SQL Server
- Visual Studio Code o PyCharm (IDE recomendado)

Plataformas:

- GitHub (para repositorio y control de versiones)
- UEDI (para entrega oficial de la práctica)

Lecturas recomendadas:

- Documentación oficial de Python y Pandas
- Guías de modelado dimensional y relacional
- Manual de SQL Server: creación de bases de datos y consultas analíticas
- Artículos técnicos sobre procesos ETL y gobierno de datos

6. Cronograma

| Tipo | Fecha Inicio | Fecha Fin |
|------------------------|--------------|-----------|
| Asignación de Práctica | Semana 3 | Semana 3 |
| Elaboración | Semana 3 | Semana 5 |
| Calificación | Semana 5 | Semana 5 |

7. Rúbrica de Calificación

7.1 Requisitos para optar a la calificación

| Tema | Descripción | Cumple (Sí/No) |
|--|--|----------------|
| Cumplimiento de la tecnología establecida | El proceso ETL debe desarrollarse en Python y el modelo multidimensional de inteligencia de negocios implementarse en SQL Server . | |
| Uso de herramientas requeridas | Se debe evidenciar el uso de librerías de Python (pandas, pyodbc o SQLAlchemy), conexión a base de datos y ejecución de consultas SQL. | |
| Gestión y entregas de práctica | El repositorio SS22S2026_#carnet debe contener la carpeta Practica1 con código fuente, scripts SQL, documentación y README.md según lineamientos. | |
| Documentación obligatoria | Incluir documentación técnica en formato PDF o Markdown que describa el proceso ETL, el modelo multidimensional de datos y los resultados obtenidos. | |
| Pruebas y funcionalidad mínima | La aplicación debe ejecutar correctamente la extracción, transformación y carga de datos, generar el modelo multidimensional y validar resultados mediante consultas analíticas. | |

7.2 Resumen de Puntuaciones

| Área | Puntos Totales | Puntos Obtenidos |
|--|----------------|------------------|
| 1. Habilidades | | |
| Desarrollo del proceso ETL (estructura y ejecución correcta) | 15 | |
| Documentación técnica (claridad, orden y coherencia) | 15 | |
| Cumplimiento de entregables (formato, nomenclatura y organización del repositorio) | 10 | |
| Sub-Total Habilidades | 40 | |
| 2. Conocimientos | | |
| Modelo de datos en SQL Server (estructura, integridad y relaciones correctas) | 25 | |
| Consultas analíticas (validación de carga, indicadores y resultados esperados) | 35 | |
| Sub-Total Conocimientos | 60 | |
| TOTAL | 100 | |

7.3 Comentarios Generales

—

—

—

Detalle de la Calificación

| No. | Criterio de evaluación | Punteo máximo | Satisfactorio (100%–61%) | Necesita mejorar (60%–0%) | Punteo obtenido |
|----------|------------------------------------|---------------|--|--|-----------------|
| 1 | Habilidades | 40 | | | |
| 1.1 | Desarrollo del proceso ETL | 15 | Implementa correctamente las fases de extracción, transformación y carga en Python. El código es funcional, estructurado y sin errores críticos. | No completa las tres fases del proceso ETL o presenta fallos que impiden su ejecución. | |
| | Rango del punteo: | | 15/15–9 | 15/8–0 | |
| 1.2 | Documentación técnica | 15 | La documentación describe el proceso ETL, el modelo multidimensional de datos y resultados con claridad, formato ordenado y lenguaje técnico adecuado. | La documentación es incompleta, desorganizada o carente de explicación técnica. | |
| | Rango del punteo: | | 15/15–9 | 15/8–0 | |
| 1.3 | Cumplimiento de entregables | 10 | Entrega completa con repositorio estructurado, archivos correctamente nombrados y README detallado. | Faltan archivos, el repositorio no sigue el formato indicado o hay errores en la presentación. | |
| | Rango del punteo: | | 10/10–6 | 10/5–0 | |

| | | | | | |
|--------------------------------|--------------------------------------|------------|---|---|--|
| Sub-total Habilidades | | 40 | | | |
| 2 | Conocimientos | 60 | | | |
| 2.1 | Modelo de datos en SQL Server | 25 | El modelo multidimensional relacional o dimensional está correctamente diseñado e implementado, con integridad referencial y relaciones coherentes. | El modelo multidimensional presenta errores de diseño, relaciones incorrectas o no ejecuta correctamente en SQL Server. | |
| | Rango del punteo: | | 25–15 | 14–0 | |
| 2.2 | Consultas analíticas | 35 | Las consultas SQL validan la carga de datos y generan resultados analíticos correctos (conteos, indicadores, top 5, etc.). | Las consultas son incompletas, incorrectas o no evidencian comprensión del modelo multidimensional . | |
| | Rango del punteo: | | 35–21 | 20–0 | |
| Sub-total Conocimientos | | 60 | | | |
| Total General | | 100 | | | |