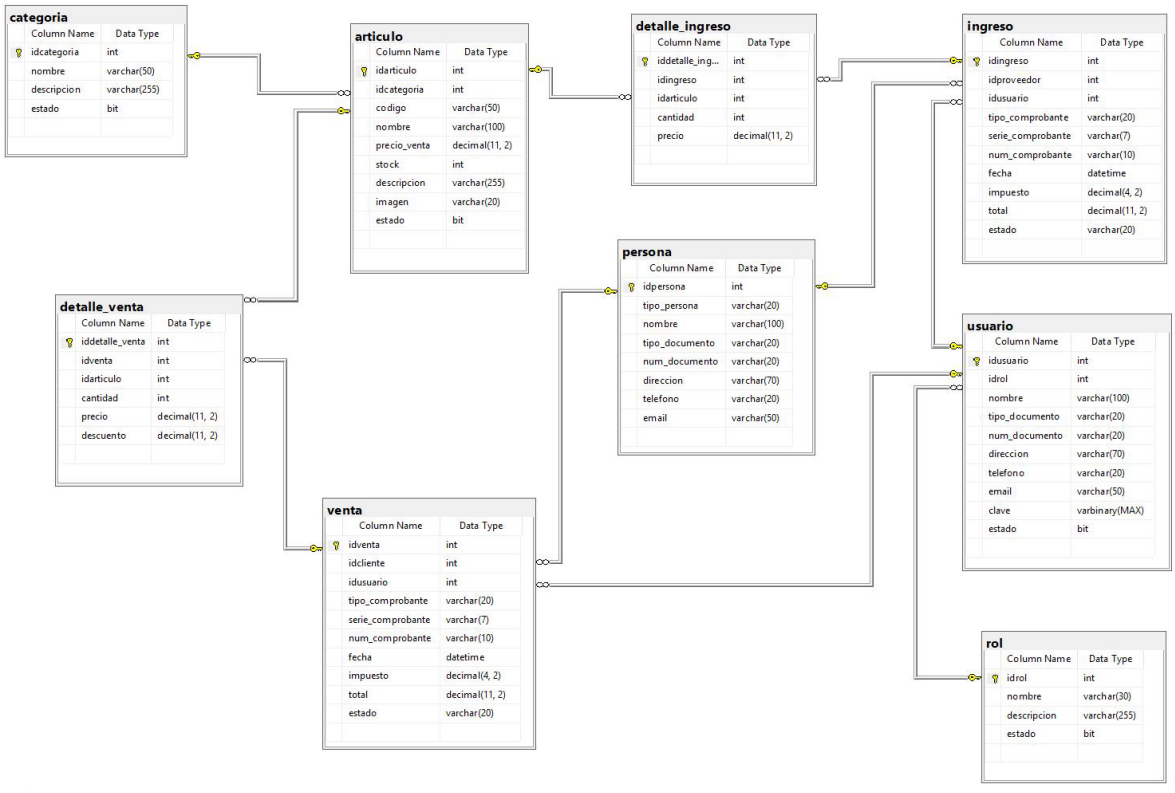
Proyecto Final – Ciencia de Datos en Python

Proyecto: Ingeniería de Datos con Python
Tema: Python, Pandas, SQL, ETL, AWS
Fecha y Hora de Entrega: 12/04/2024 23:55
Formato de Entrega: Archivos de Construcción y Video.
Grupo: Grupos de 2 o 3 personas
Calificación: Presentación por medio de Vídeo

DESCROPCIÓN: Para este proyecto usted deberá desarrollar un pipeline de ingeniería de datos utilizando Python, SQL y AWS como herramientas de desarrollo, su proyecto tener los siguientes componentes:

Sistema transaccional: el cual será consturido de forma automática utilizando AWS, boto3 y SQL. El sistema transaccional que deberá desarollar es el siguiente:



Ingestión de datos transaccionales: Deberá poblar la base de datos utilizando Python.

Pr**eguntas de Negocio:** Deberá plantear 5 preguntas de negocio para determinar que estrucutra de datos (data warehouse) es la más adecuada para dar solucion a las preguntas que se planteen.

Arquitectura de datos: Deberá definir el códugo DDL para la estrucura de datos adecuada que de solución a las preguntas planteadas anteriormente.

ETL: utilizando Python deberá llenar la estructura de datos definida anteirormente a partir del sistema transaccional inicial.

Analytics: Utilizando Markdown, numpy, pandas, matplotlib y seaborn, deberá elaboraro un notebook con las respuetas a la preguntas planteadas de forma consistente y lo más profundo que sea posible.

DETALLES TECNICOS: A continuación se describen los detalles técnicos mínimos que su projecto debe cumplir:

- Para desarrollar el sistema transaccional podrá utilzar cualquier gestor de base de datos SQL que esté disponible en RDS.
- El procesamiento puede realizarlo en un maquina local o en una instancia de EC2 corriendo Python.
- La salida deberá ser sobre en RDS pero en un sistema gestor de base de datos que no sea el que utilzó para el sistema transaccional.
- El Notebook debe contener detalles sobre los procesos necesarios para responder las preguntas de negocio planteadas.
- Notar que no puede usar SQL para hacer la construcción de ningína estructura salvo para leer de tablas almacenadas en las bases de datos es decir SELECT * FROM tabla. Cualquier otra necesidad de procesamiento deberá hacerla en Python con pandas y librerías de procesamiento de información

ENTREGA: Como entrega deberá publicar todos los archivos utilizados por medio de un link de Git, incluyendo la presentación utilzada en el video, los notebooks utiilzados, los y los archivos adicionales que requiera. Adicinalmente deberá hacer un video de 5 a 7 minitos máximo donde explique todos los pasos que realizó para desarrollar su proyecto, es decir describir todos los elementos de su proyecto.