

Proyecto de ETL y Análisis de Empresas del S&P 500: Fase 3 - Almacenamiento en SQL Server

Objetivo:

Implementar la carga de datos limpios de empresas del S&P 500 en una base de datos SQL Server, utilizando Python y SQLAlchemy.

Requisitos:

- Python 3.x
- Librerías: pandas, sqlalchemy, pyodbc

Instrucciones Generales:

1. Configura el entorno e instala las librerías necesarias.
2. Define la estructura de la base de datos en SQL Server para almacenar los datos.
3. Implementa la carga de datos desde archivos CSV a la base de datos utilizando SQLAlchemy y pyodbc.

Fases del Proyecto de Almacenamiento en SQL Server:

1. Configuración del Entorno

Instala las librerías necesarias:

2. Estructura de la Base de Datos

Objetivo: Diseñar la estructura de la base de datos en SQL Server para almacenar los datos de empresas del S&P 500.

Estructura Propuesta:

- **Tabla CompanyProfiles:**

- Symbol (VARCHAR): Símbolo de la empresa (clave primaria).
- Company (VARCHAR): Nombre de la empresa.
- Sector (VARCHAR): Sector al que pertenece la empresa.
- Headquarters (VARCHAR): Ubicación de la sede de la empresa.
- Fecha fundada (VARCHAR): Año de fundación de la compañía.
- **Tabla Companies:**
 - Date (DATE): Fecha (YYY-MM-DD). (clave primaria).
 - Symbol (VARCHAR): Símbolo de la empresa (clave primaria).
 - Close (FLOAT): Precio de cierre del símbolo en una fecha específica.

3. Implementación de la Carga en SQL Server

Objetivo: Cargar los datos limpios de empresas del S&P 500 desde archivos CSV a las tablas de la base de datos en SQL Server.

Tareas:

1. **Conectar a la base de datos SQL Server:**
 - Utilizar sqlalchemy y pyodbc para establecer la conexión.
2. **Cargar datos en las tablas:**
 - Utilizar pandas para leer los archivos CSV y cargar los datos en las tablas Companies y CompanyProfiles.

Resumen:

Este documento guía a los estudiantes a través de la tercera fase del proyecto, que consiste en almacenar los datos limpios de empresas del S&P 500 en una base de datos SQL Server. Los estudiantes deberán:

1. Definir la estructura de la base de datos en SQL Server.
2. Configurar la conexión y utilizar sqlalchemy para cargar los datos desde archivos CSV a las tablas de la base de datos.

Al finalizar, los estudiantes habrán adquirido habilidades en integración de datos y almacenamiento utilizando tecnologías de bases de datos relacionales, como SQL Server, combinadas con herramientas de análisis de datos en Python.