МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Національний університет “Львівська політехніка”

Інститут комп’ютерних наук та інформаційних технологій

Кафедра інформаційних систем та мереж



**Звіт**

до лабораторної роботи № 5

**«Побудова системи консолідації даних з використанням технології ETL»**

з дисципліни

**«Технології інтеграції інформаційних ресурсів»**

Виконав: ст. гр. ІТІС-11

Стельмах В. Д.

Прийняв:

Щербак С.С.

Львів – 2025

**Лабораторна робота № 5**

**Тема:** Побудова системи консолідації даних з використанням технології ETL

**Мета роботи:** Ознайомитися з основами проєктування та реалізації системи консолідації даних із використанням технології ETL (Extract, Transformation, Loading).

**Порядок виконання роботи:**

1. Налаштувати середовище розробки Python та встановити необхідні бібліотеки (FastAPI, Uvicorn, Pydantic).
2. Розробити вебсервіс на Python з використанням FastAPI, який демонструватиме роботу ETL-процесу для теми "Каталог товарів магазину електроніки".
3. Реалізувати етап Extract: зчитування даних з двох файлів-джерел (source\_products\_branch\_A.json та source\_additional\_info\_branch\_B.json).
4. Реалізувати етап Transform: очищення, стандартизація, об'єднання/оновлення даних та можливі обчислення нових полів.
5. Реалізувати етап Load: збереження трансформованих та консолідованих даних у цільовий файл (consolidated\_electronics\_store.json).
6. Створити ендпоінти API: один для запуску повного ETL-процесу та інший для перегляду консолідованих даних.
7. Продемонструвати роботу вебсервісу, викликаючи ендпоінти та аналізуючи результати на кожному етапі.
8. Задокументувати структуру вебсервісу, приклади його роботи (скріншоти), та надати посилання на репозиторій з кодом програми та прикладами файлів даних у звіті.

**Короткі теоретичні відомості**

ETL (Extract, Transformation, Loading) – це комплекс методів та процесів, що використовуються для перенесення даних з одного або декількох джерел у цільову систему зберігання, зазвичай у сховище даних, для подальшого аналізу та звітності. Процес ETL складається з трьох основних етапів:

* Видобування (Extract): Дані зчитуються з різноманітних джерел. Це можуть бути реляційні бази даних, файли (CSV, JSON, XML), інші вебсервіси тощо. На цьому етапі важливо отримати не тільки самі дані, але й метадані, що описують їх структуру. Існують методи повного видобування та часткового (інкрементного) видобування, яке може відбуватися з або без сповіщення про оновлення.
* Трансформація (Transform): Видобуті дані проходять через низку перетворень для приведення їх до потрібного формату та якості. Це ключовий етап, де дані очищуються, валідуються, стандартизуються, агрегуються, об'єднуються та можуть бути збагачені новими обчисленими значеннями. Основні завдання трансформації включають вибір необхідних даних, очищення, стандартизацію форматів та кодувань, а також розрахунок похідних значень.
* Завантаження (Load): Трансформовані дані завантажуються у цільову систему, наприклад, у сховище даних. Цей етап вимагає оптимізації продуктивності та механізмів відновлення у разі збоїв. Методи завантаження включають початкове завантаження, додаткове (інкрементне) завантаження та повне оновлення.

Технологія ETL є критично важливою для консолідації даних, дозволяючи організаціям отримувати єдине, узгоджене уявлення про свої дані з різних систем, що необхідно для аналізу та прийняття обґрунтованих рішень.

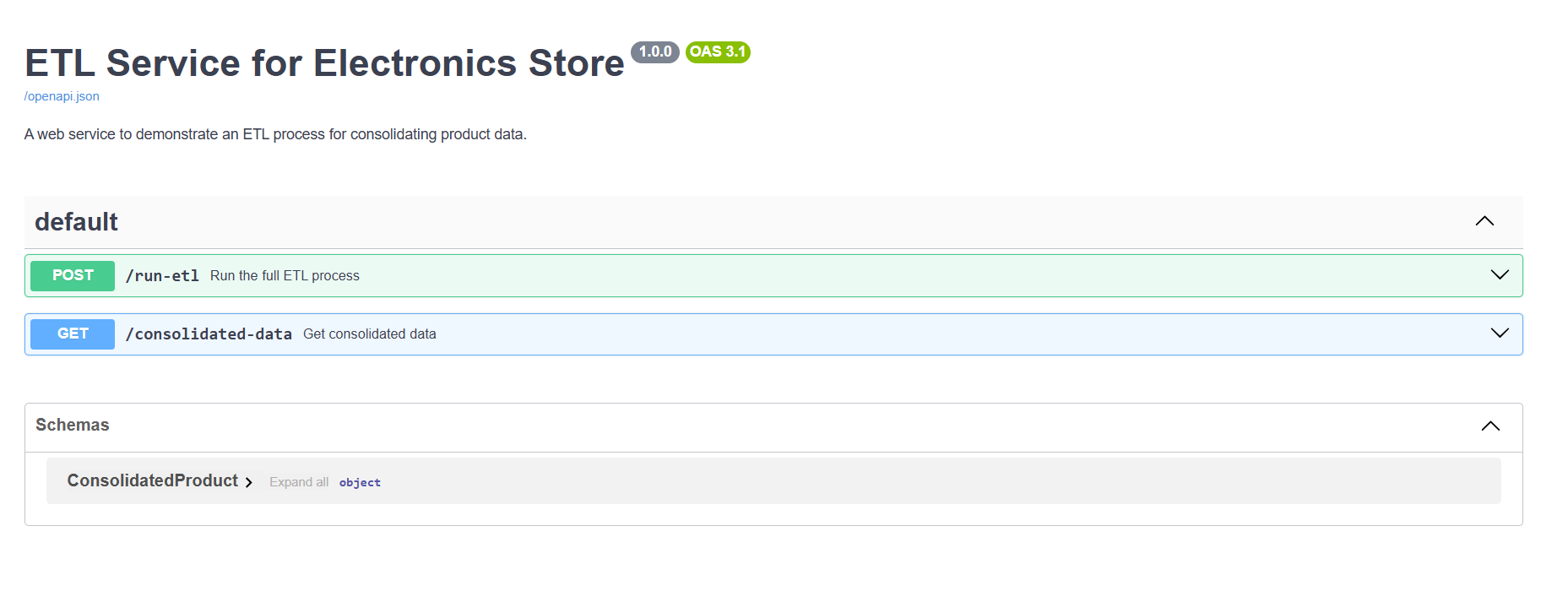
**Результати виконання роботи**

**Демонстрація роботи ETL-сервісу**

Робота ETL-сервісу демонструється шляхом виклику його ендпоінтів.

Загальний вигляд документації API (Swagger UI) та запуск ETL-процесу (POST /run-etl):

Представлено головну сторінку документації API в Swagger UI, де видно ендпоінт /run-etl. Показано приклад виклику цього ендпоінту для запуску повного ETL-процесу та відповідь сервера про успішне завершення.



Зображення, що містить текст, знімок екрана, програмне забезпечення, число

Вміст, створений ШІ, може бути неправильним.

Рис. 1. Запуск ETL-процесу через Swagger UI.

Консольний вивід (логи) сервера під час виконання ETL-процесу:

Демонструється фрагмент логів сервера (консольний вивід FastAPI/Uvicorn), де видно повідомлення про етапи Extract, Transform та Load, кількість оброблених записів та можливі попередження чи помилки під час обробки.

Зображення, що містить текст, знімок екрана

Вміст, створений ШІ, може бути неправильним.

Рис. 2. Консольні логи сервера під час виконання ETL-процесу.

Перегляд консолідованих даних (GET /consolidated-data):

Після успішного виконання ETL-процесу, показано виклик ендпоінту /consolidated-data через Swagger UI (або інший HTTP-клієнт) та JSON-відповідь сервера, що містить список консолідованих товарів.

Зображення, що містить текст, знімок екрана, програмне забезпечення, Комп’ютерна піктограма

Вміст, створений ШІ, може бути неправильним.

Рис. 3. Перегляд консолідованих даних через API.

Вміст цільового файлу (consolidated\_electronics\_store.json) після ETL:

Представлено фрагмент вмісту файлу consolidated\_electronics\_store.json, який було створено або оновлено в результаті роботи ETL-процесу. Це показує фінальний результат завантаження даних.

Зображення, що містить текст, знімок екрана, меню

Вміст, створений ШІ, може бути неправильним.

Рис. 4. Фрагмент консолідованих даних у цільовому файлі.

**Аналіз роботи ETL-сервісу**

Розроблений вебсервіс успішно демонструє основні етапи ETL-процесу:

* Видобування: Дані коректно зчитуються з визначених JSON файлів-джерел.
* Трансформація: Реалізовано логіку очищення даних (наприклад, видалення товарів з нульовою ціною/кількістю), стандартизації категорій, об'єднання інформації з двох джерел (оновлення існуючих товарів та додавання нових), а також розрахунок похідного поля price\_change\_percentage.
* Завантаження: Консолідовані та трансформовані дані успішно зберігаються у цільовий JSON-файл.
* Ендпоінти /run-etl та /consolidated-data функціонують згідно з очікуваннями, дозволяючи ініціювати процес та переглядати його результат. Логування надає інформацію про хід виконання процесу.

**Посилання на вихідний код програми та файли даних**

Повний вихідний код розробленого ETL-сервісу на Python з використанням FastAPI, а також приклади файлів-джерел (source\_products\_branch\_A.json, source\_additional\_info\_branch\_B.json) та згенерований цільовий файл (consolidated\_electronics\_store.json) розміщено у репозиторії GitHub за наступним посиланням:

<https://github.com/VitaliiStelmakh/TIIR/blob/main/lab_5>

**Висновки**

Під час виконання цієї лабораторної роботи я ознайомився з концепцією та основними етапами технології ETL (Extract, Transformation, Loading). Було розроблено вебсервіс на FastAPI, який моделює ETL-процес для консолідації даних про товари з різних джерел. Ця робота дозволила отримати практичний досвід у видобуванні, трансформації (включаючи очищення, стандартизацію та об'єднання) та завантаженні даних, а також зрозуміти важливість ETL для підготовки даних до аналізу.