Міністерство освіти і науки України

Національний університет “Львівська політехніка”



Звіт

про виконання лабораторної роботи № 3

iз курсу “Аналітичні та нереляційні бази даних”

**Виконав**:

ст. гр. ІР-41, ІКТА

Білоус Тарас

**Прийняв**:

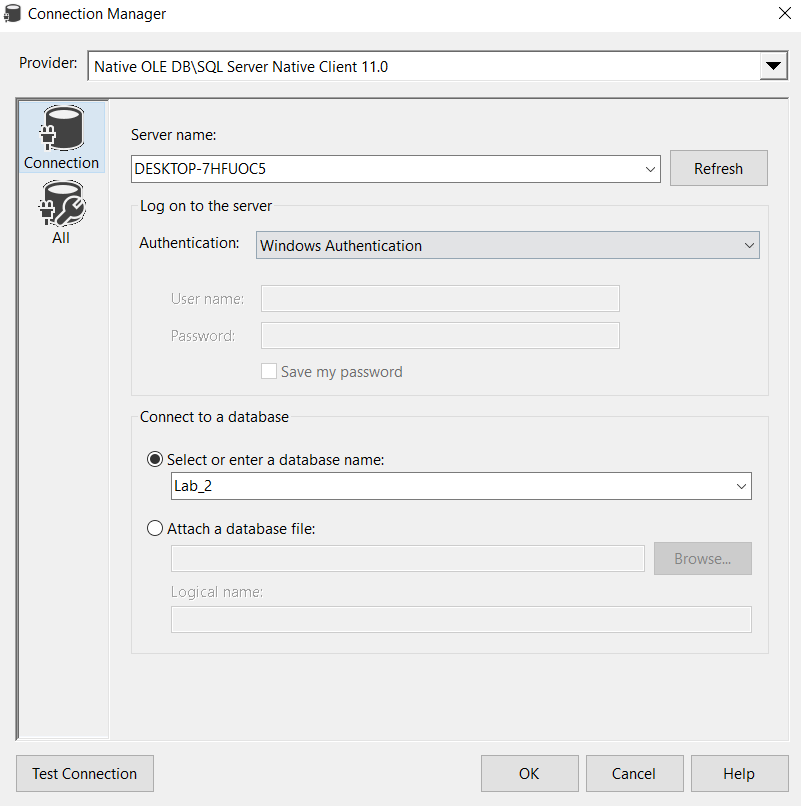
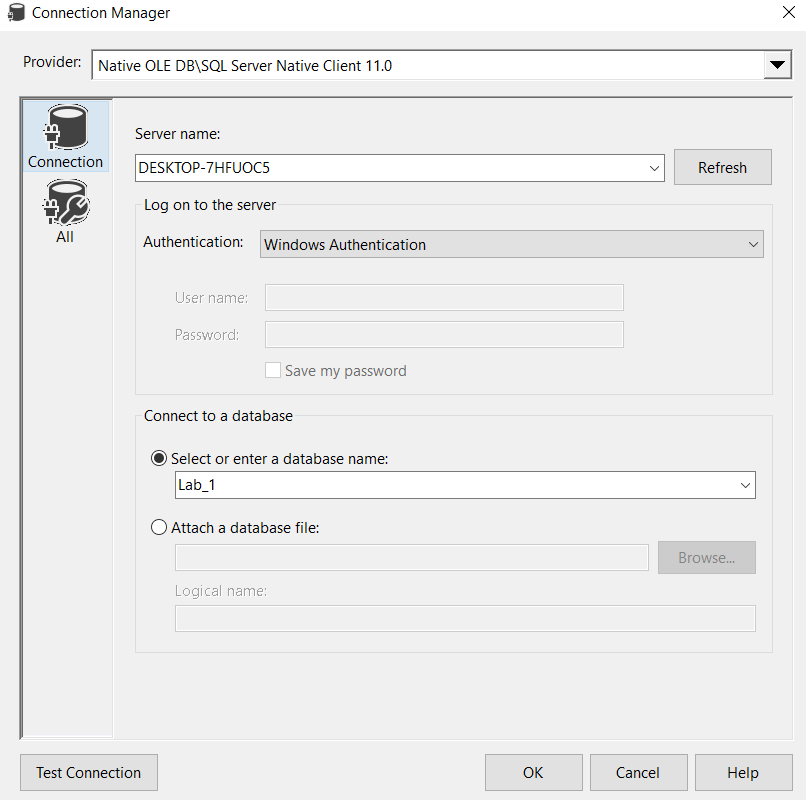
Верес З.Є.

Львів – 2020

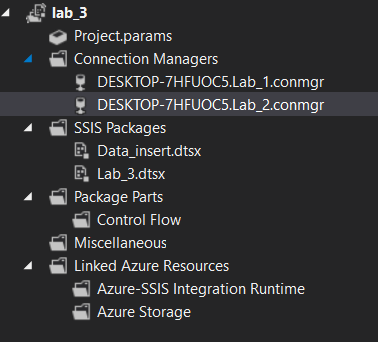
**Лабораторна робота №3**

**Завдання:** реалізувати ETL процес засобами Microsoft SSIS для заповнення таблиці фактів з лабораторної роботи 2 даними з таблиць з бази даних лабораторної роботи 1.

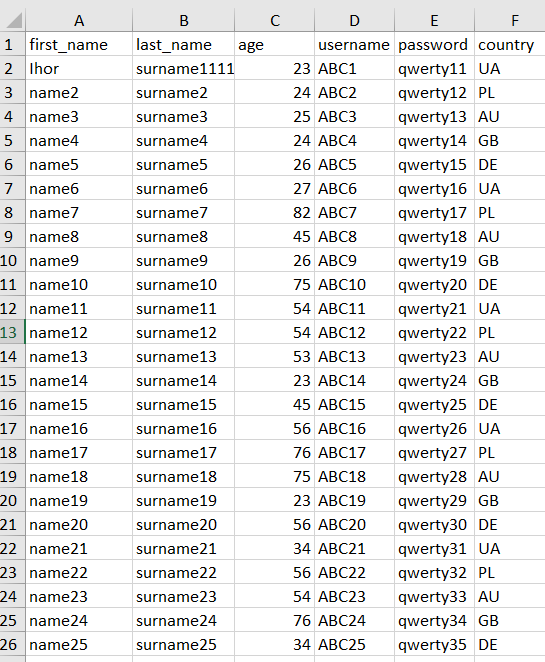
Для початку ми створюємо конекшн менеджери для прямого зв’язку з обома базами даних.



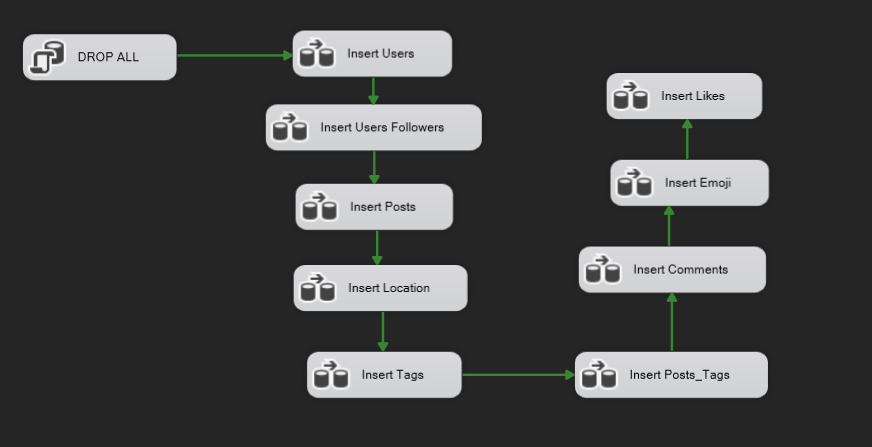
Загальна структура проекту:



У першому SSIS пакеті – ми заповнюємо першу БД на основі даних, що були попередньо занесені у відповідні Excel таблиці.



Послідовність виконання логіки:



Заради збереження цілісності індексів та унікальних ключів у таблицях – перед запуском виконується команда, що очищає усі таблиці і занулює PKs.

DELETE FROM [Users\_Followers];

DELETE FROM [Posts\_Tags];

DBCC CHECKIDENT('Posts\_Tags', RESEED, 0);

DELETE FROM [Tags];

DBCC CHECKIDENT('Tags', RESEED, 0);

DELETE FROM [Comments];

DBCC CHECKIDENT('Comments', RESEED, 0);

DELETE FROM [Location];

DBCC CHECKIDENT('Location', RESEED, 0);

DELETE FROM [Likes];

DBCC CHECKIDENT('Likes', RESEED, 0);

DELETE FROM [Emoji];

DBCC CHECKIDENT('Emoji', RESEED, 0);

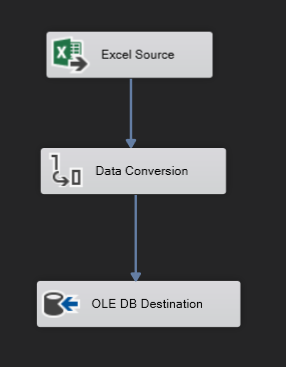
DELETE FROM [Posts];

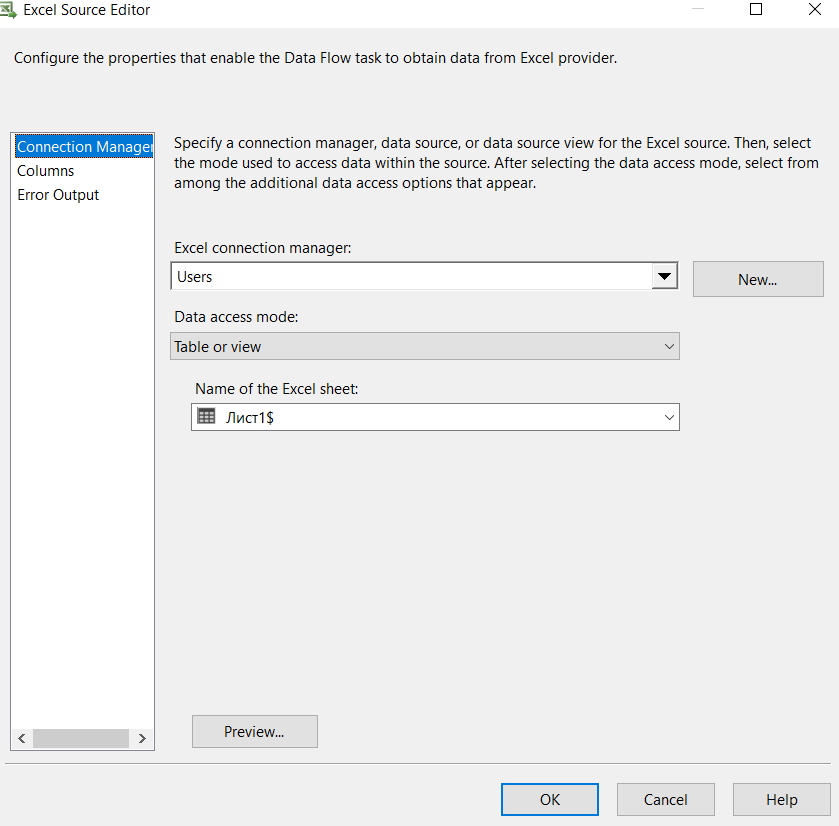
DBCC CHECKIDENT('Posts', RESEED, 0);

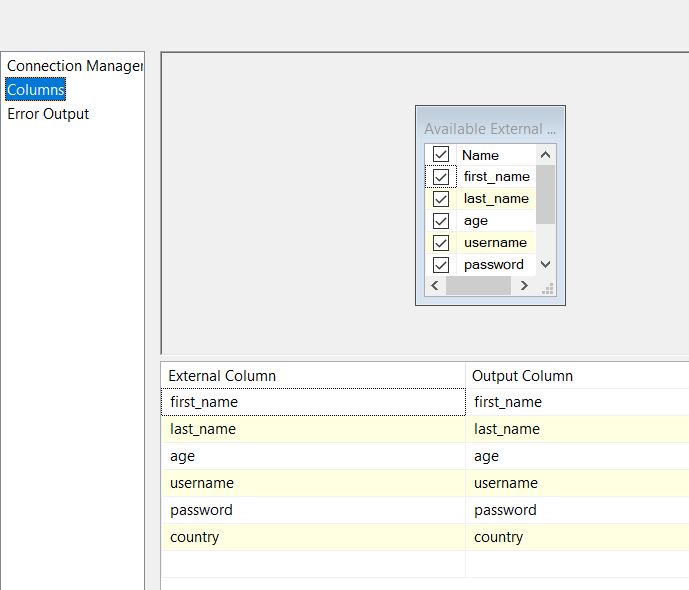
DELETE FROM [Users];

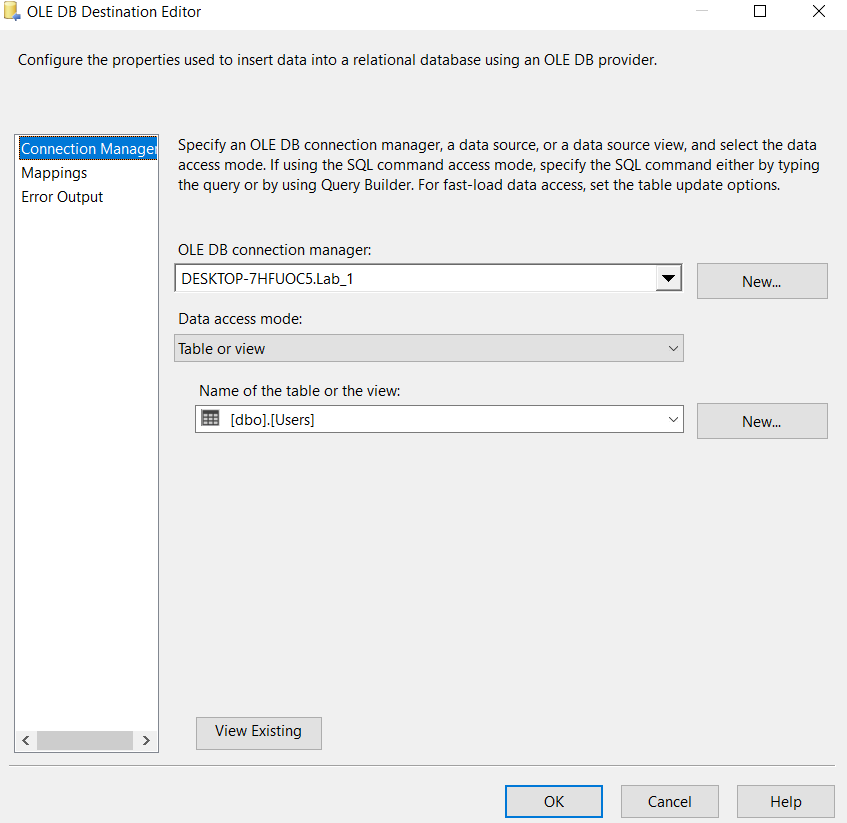
DBCC CHECKIDENT('Users', RESEED, 0);

Загальна структура кожного інсерту – витягування даних із ексель файлу, конвертація даних(якщо необхідно) і послідовне занесення даних у відповідну таблицю.

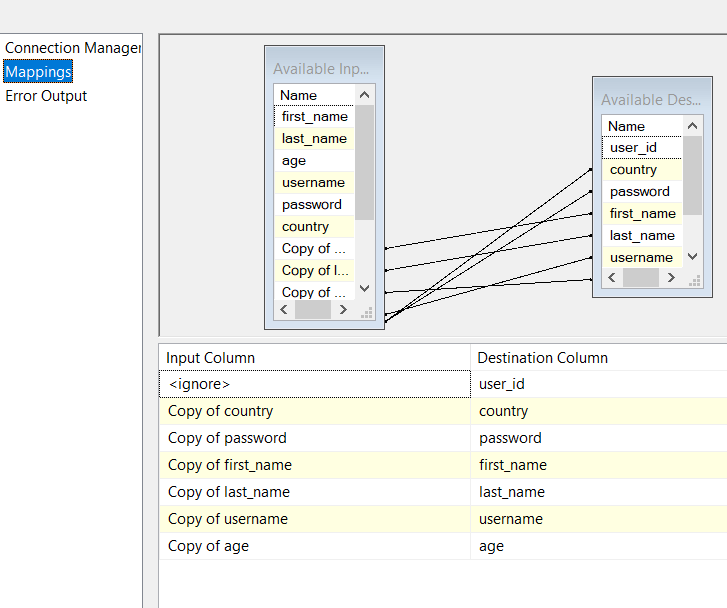




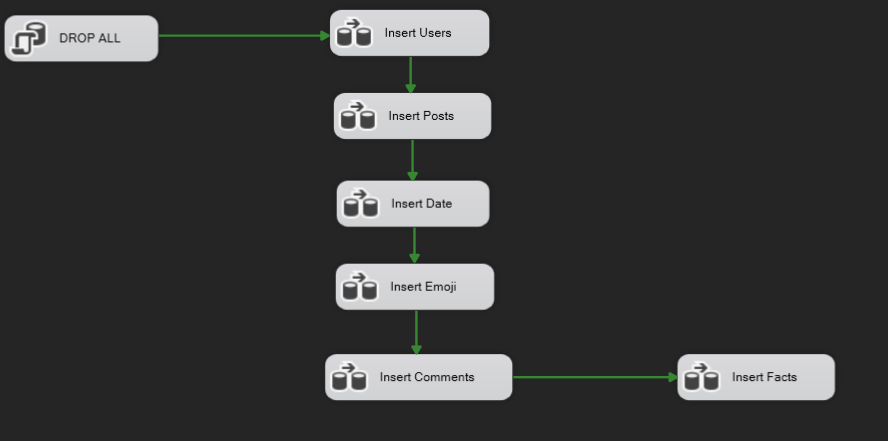




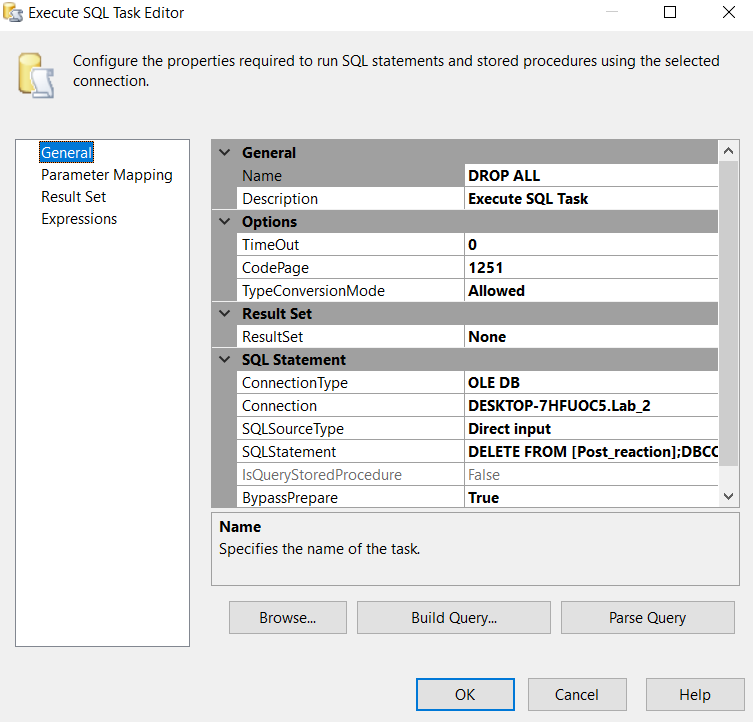
Оскільки айді у нас автоінкрементується, то вона не міститься у ексель таблицях та ігнорується як поле.



Другий SSIS пакет створений для заповнення БД у вигляді зірки на основі даних із 1 БД.



По аналогії із попереднім пакетом – першим кроком ми видаляємо усі дані із наших таблиць.



DELETE FROM [Post\_reaction];

DBCC CHECKIDENT('Post\_reaction', RESEED, 0);

DELETE FROM [Users];

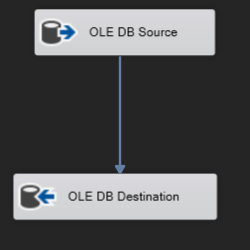
DELETE FROM [Posts];

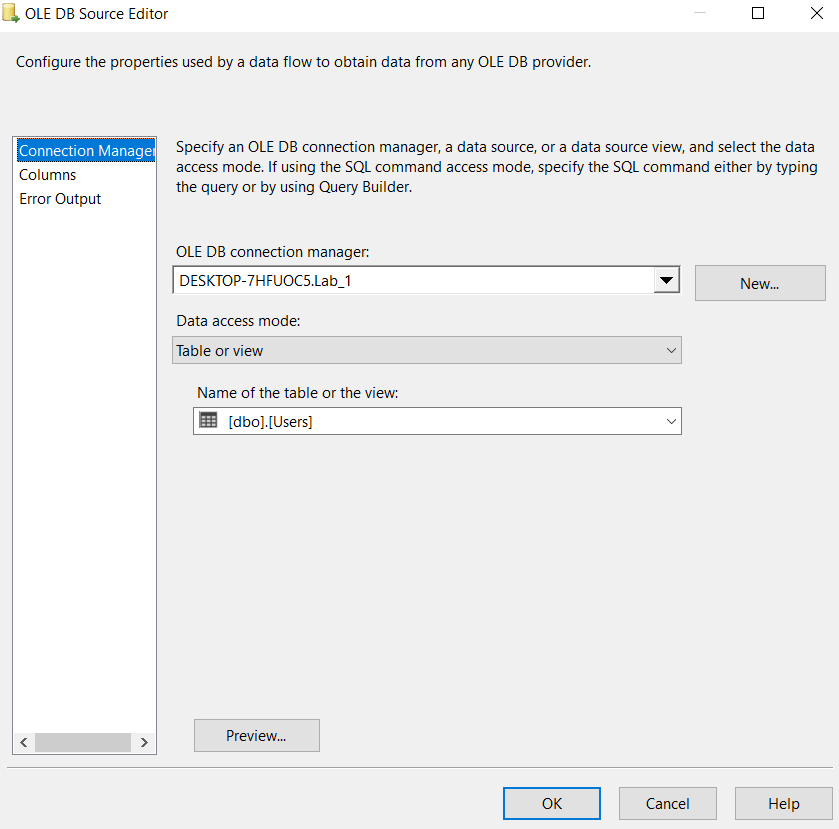
DELETE FROM [Date];

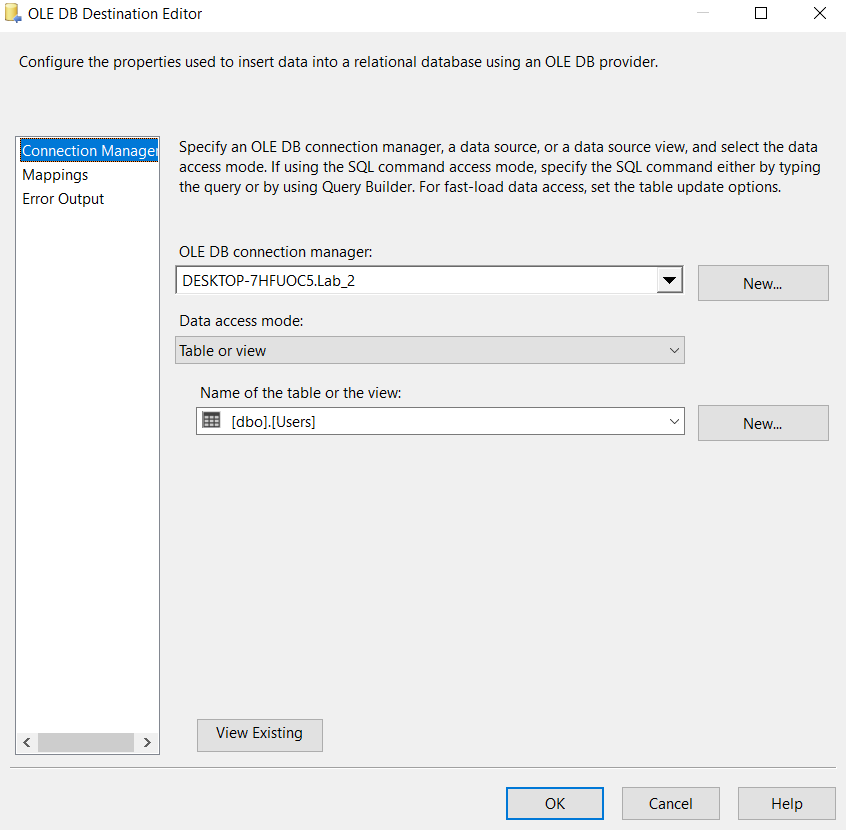
DELETE FROM [Comments];

DELETE FROM [Emoji];

Типовий вигляд кожного дата флоу таску – із джерела даних (таблиця із 1 БД) ми отримуємо наші дані, групуємо чи перетворюємо їх і заносимо у відповідний діменшн у 2 БД.







Для агрегації даних і заповнення статистичної таблиці фактів – ми використовуємо даний запит , обраховуючи основні, а також допоміжні поля.

USE Lab\_1;

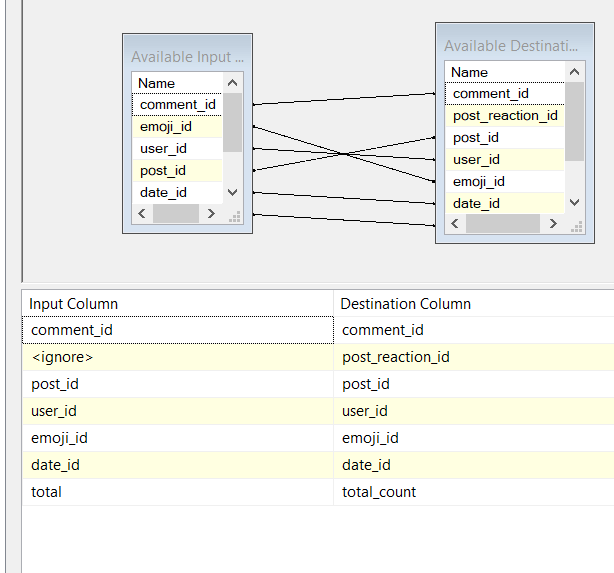
SELECT Likes.like\_id date\_id, Likes.user\_id, Likes.post\_id, Likes.emoji\_id, Comments.comment\_id, COUNT(\*) total

FROM Likes

JOIN Comments ON Likes.post\_id = Comments.post\_id AND Likes.user\_id = Comments.user\_id

GROUP BY Likes.like\_id, Likes.user\_id, Likes.post\_id, Likes.emoji\_id, Comments.comment\_id

ORDER BY post\_id;



**Висновок:** в результаті виконання даної лабораторної роботи я навчився використовувати засоби Microsoft SSIS для реалізації ETL процесу.