ETL proces

By Jakub Poplawski

DEMO 2, GROUP: Wr-43.DB

softserve

Introduction

In previous sprint I was responsible for:

- Stored procedure to generate Commercial Invoices
- Stored procedure to populate Vendors with their products
- Trigger validating VAT tax code upon insert of Vendors into table storing their addresses
- View summarizing Commercial Invoices for a given Warehouse
- Customers' Loyalty Discount program
- Vendors related tables
- Function selecting an appropriate type of Discount on Receipt line
- Table valued function summarizing the information about a selected Member of the Loyalty Discount program

Jakub Popławski

softserve

Assumptions

The goal of my ETL process is to provide data from OLTP database from the previous sprint to the OLAP database with an aim to analyse the flow of warehouse stock.

softserve

Assumptions

Organization – the whole business entity of the Buisness Client.

The Buisness Client buys products from Vendors, and stores them in his Warehouses.

The Buisness Client ships them to his Shops where they are sold to Customers.

Assumptions

Product goes to the Warehouse
from Vendor – **Commercial Invoice** is issued.

Product leaves the Warehouse
to Shop – **Freight Invoice** is issued.

softserve

Assumptions

Quick example

Tab. 1. Summary of the invoice flow as on 4th January

	Vendor	**Warehouse**		Shop
Product X		**Commercial Invoices**	**Freight Invoices**	
02.01.2023		10		
03.01.2023			-3	
04.01.2023		5		
		15	-3	
		BALANCE AS ON **04.01.23:**	**12**	

Tab. 2. Summary of the invoice flow as on 6th January

	Vendor	**Warehouse**		Shop
Product X		**Commercial Invoices**	**Freight Invoices**	
02.01.2023		10		
03.01.2023			-3	
04.01.2023		5		
05.01.2023			-6	
06.01.2023		5		
		20	-9	
		BALANCE AS ON **06.01.23:**	**11**	

softserve

Assumptions

This logic enables us to analyse additive facts.

softserve

Assumptions

The OLAP database consisted of DocumentLines fact tables that merged Commercial Invoices and Freight Invocies.

Warehouse, Vendors, Products, Shops, and Dates Dimension tables were designed.

Fig. 1. Scheme of dimensions and fact tables for the Data Warehouse

Assumptions

Staging area consisted of series of tables that were supported Lookups, and Error tables.

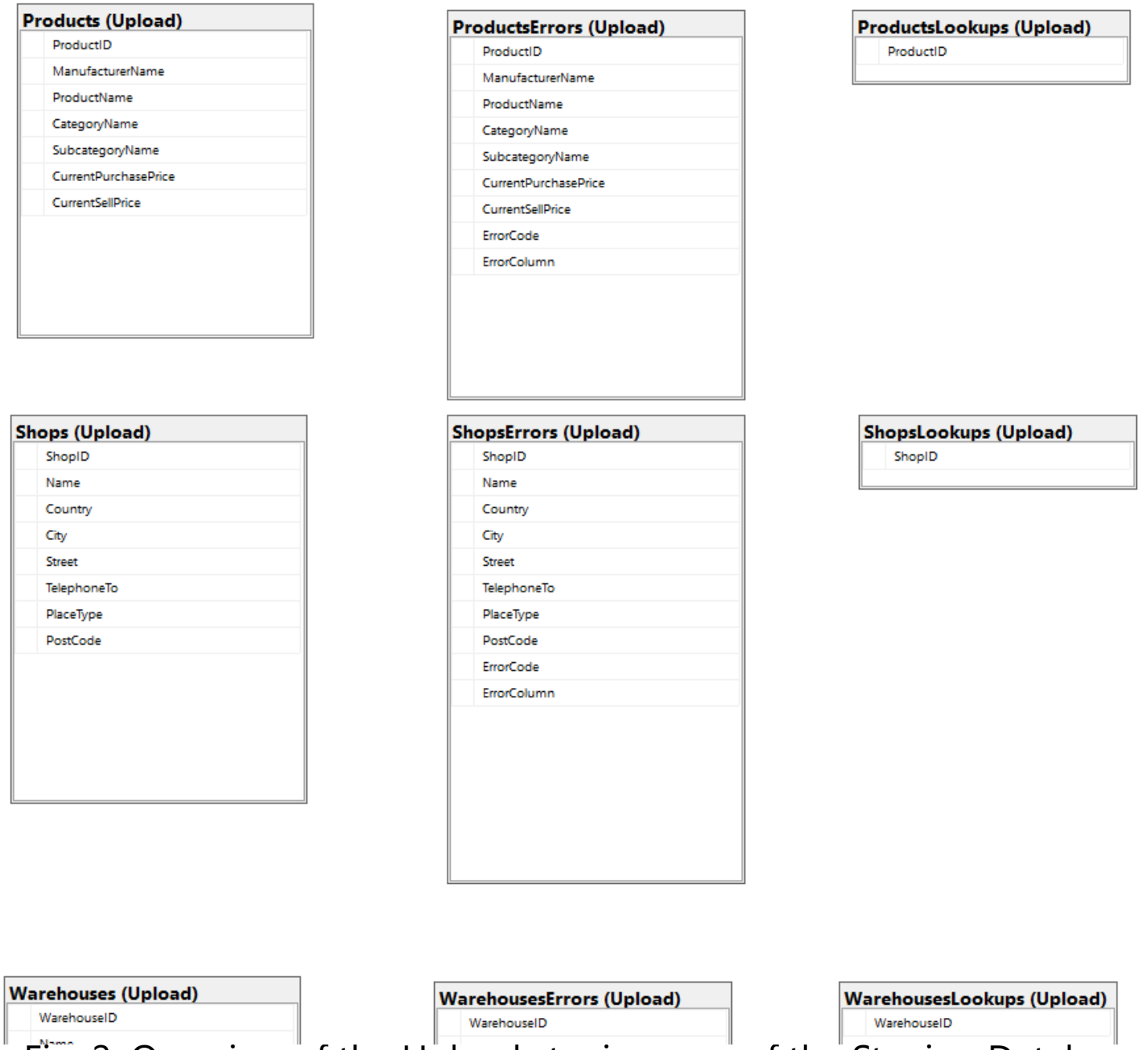

Fig. 2. Overview of the Upload staging area of the Staging Database

ETL Process

The ETL Job was organised in three sections.

First section covered the extraction of data from OLTP database.

Second transformed DocumentDate issue date from DATE data type into INT.

Third uploaded the data into OLAP database.

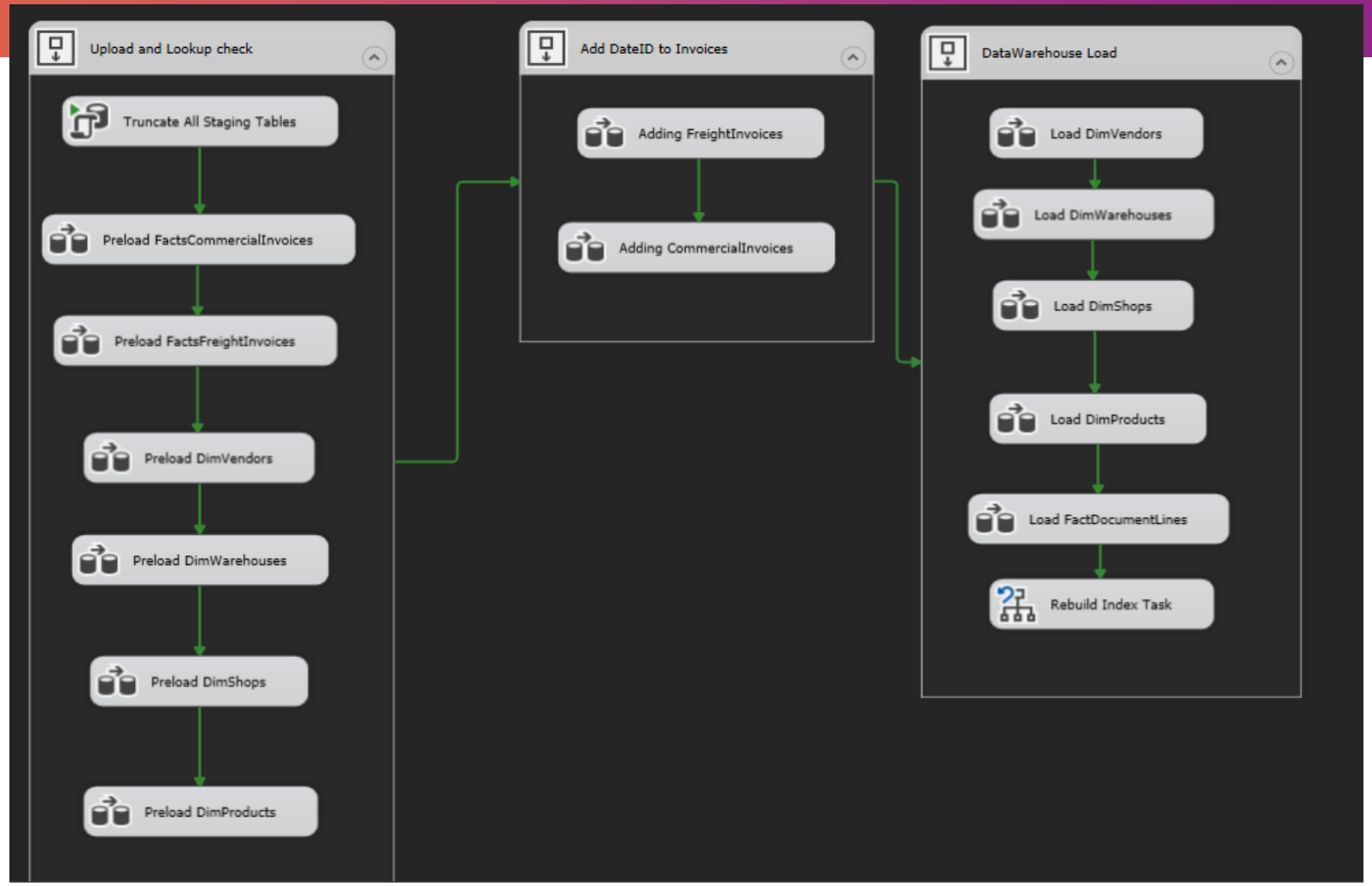

Fig. 3. Overview of the SSIS Job structure

ETL Process

Each upload Data Flow section in the extraction section was equpied with a Lookup transformation that compared the current upload with the Lookup table.

Only new inserts, not present in the Lookups, were accepted into exctration Preload tables.

Each DataFlow had Error Outputs transfered into Error Tables.

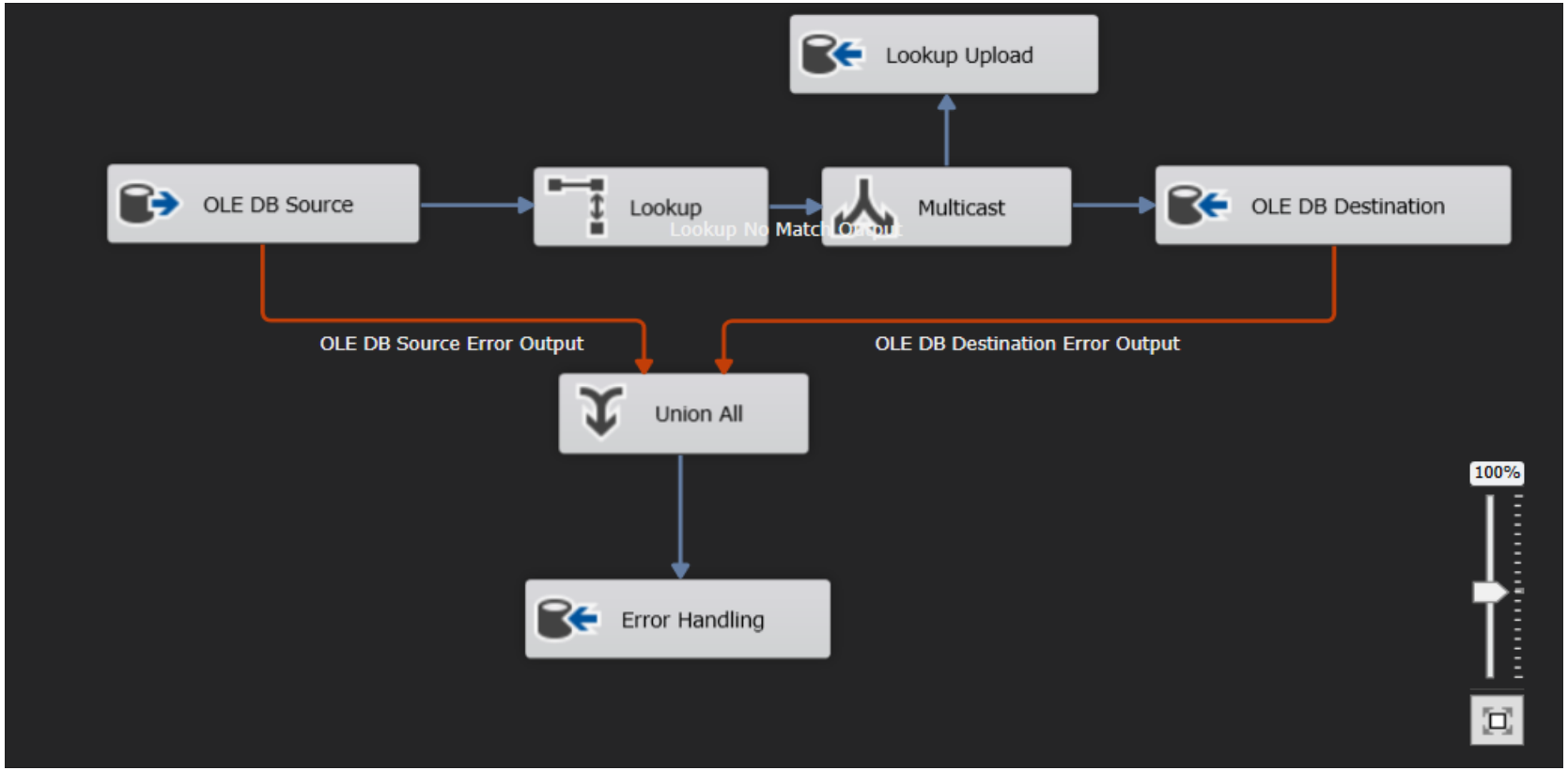

Fig. 4. Scheme of transfer from OLTP database to Staging database

ETL Process

Queries from the the OLTP Source database were designed to perform all initial data aggregation – the queries ressembled as much as possible the inputs into OLAP database.

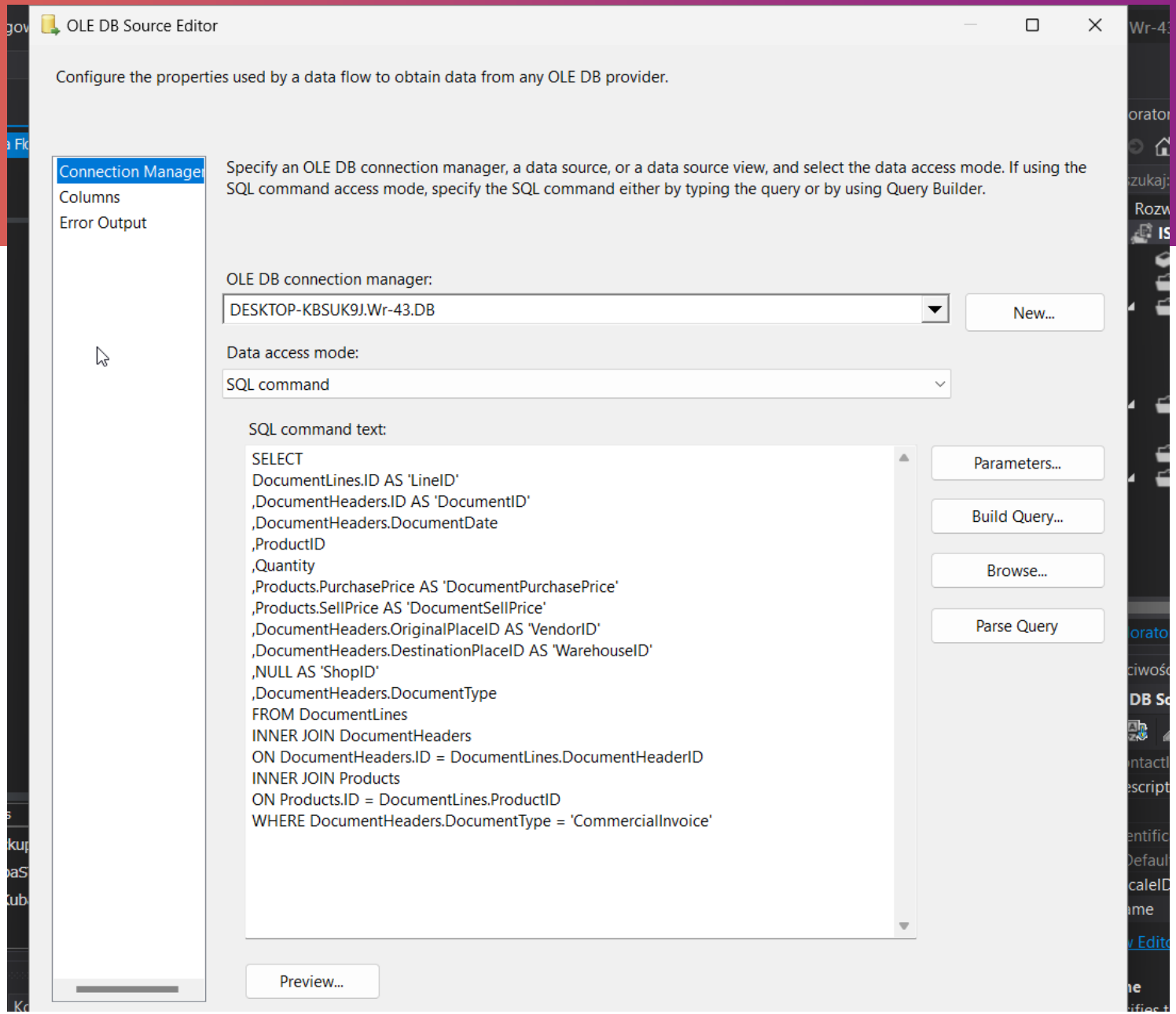

Fig. 5. Example of the Source query from the OLTP database

ETL Process

Lookup assigments compared the IDs of the current upload with the one present in the Lookup tables, that resulted from the previous upload, thus enabling increment upload through the Staging area.

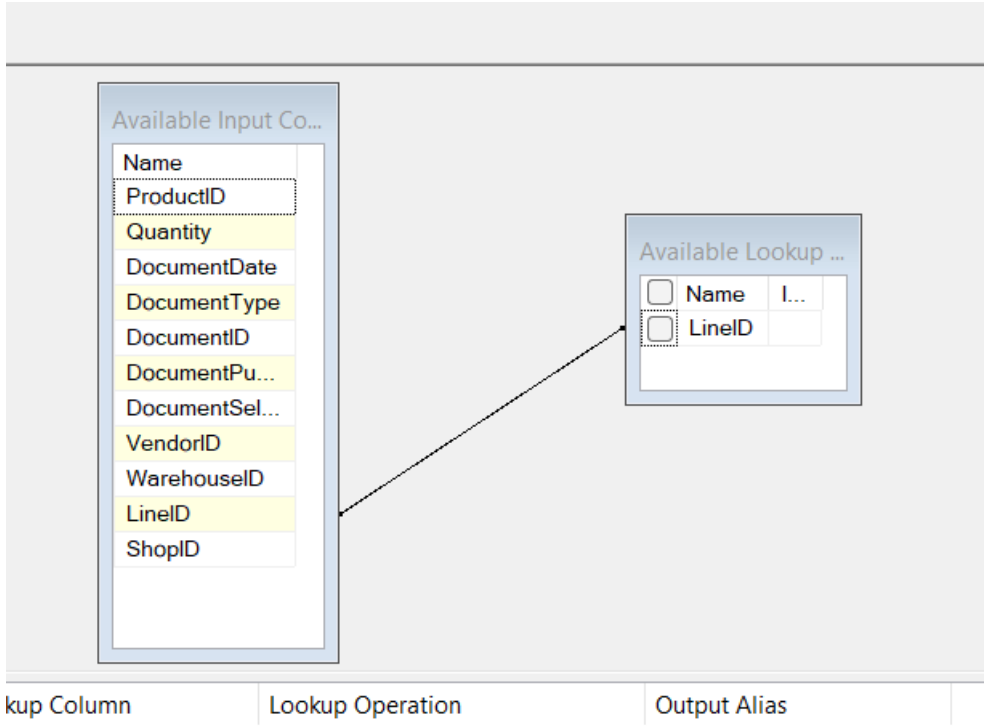

Fig. 6. Example of column comparison of Upload data with the Look-up table

softserve

ETL Process

Error outputs had standard additional columns of ErrorCode and ErrorColumn for the purpose of errors documentation in the Error tables.

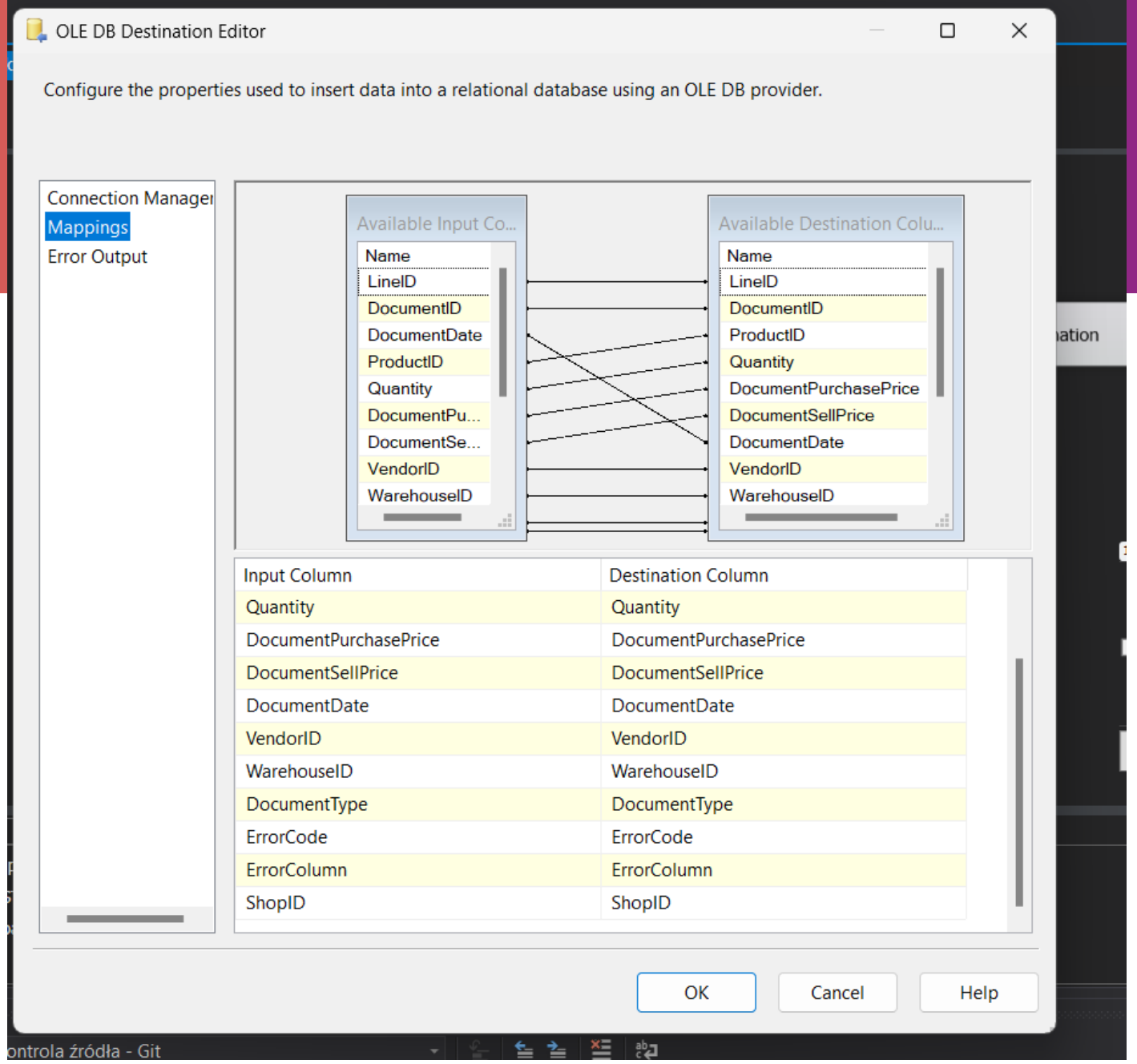

Fig. 7. Input of the Error output into Error table

ETL Process

Freight Invoices and Commercial Invoices data were transformed with a query that converted the DATE data type of the DocumentDate into INT data type.

The purpose of the transformation was to build a connection with the Dim.Date table in the OLAP database through DateID key.

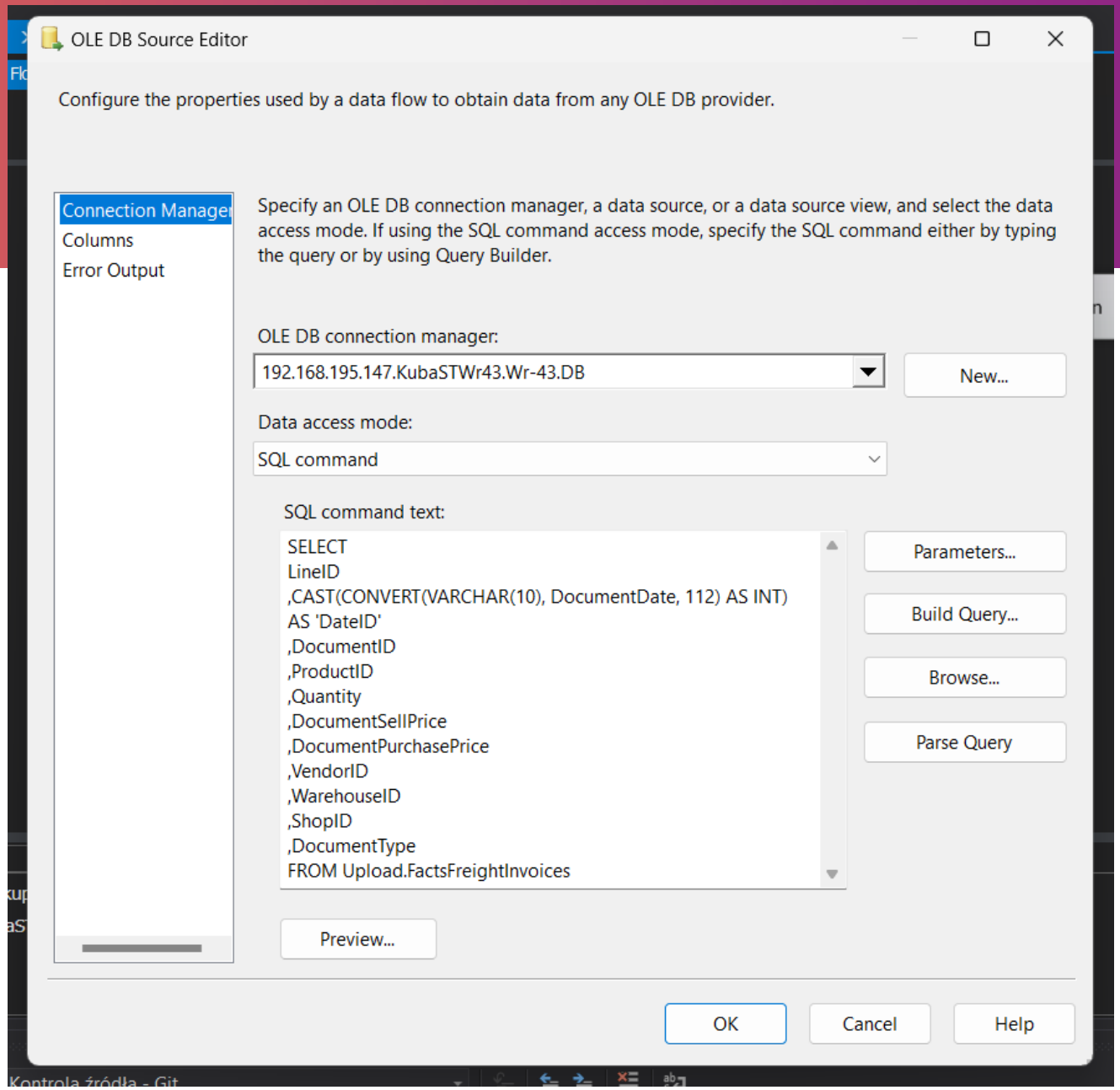

Fig. 8. Transformation is finalised by upload to the AddDate section of Staging

ETL Process

Aggregated nad transformed data were transfered from the Staging database into OLAP database in the third section of the ETL Job.

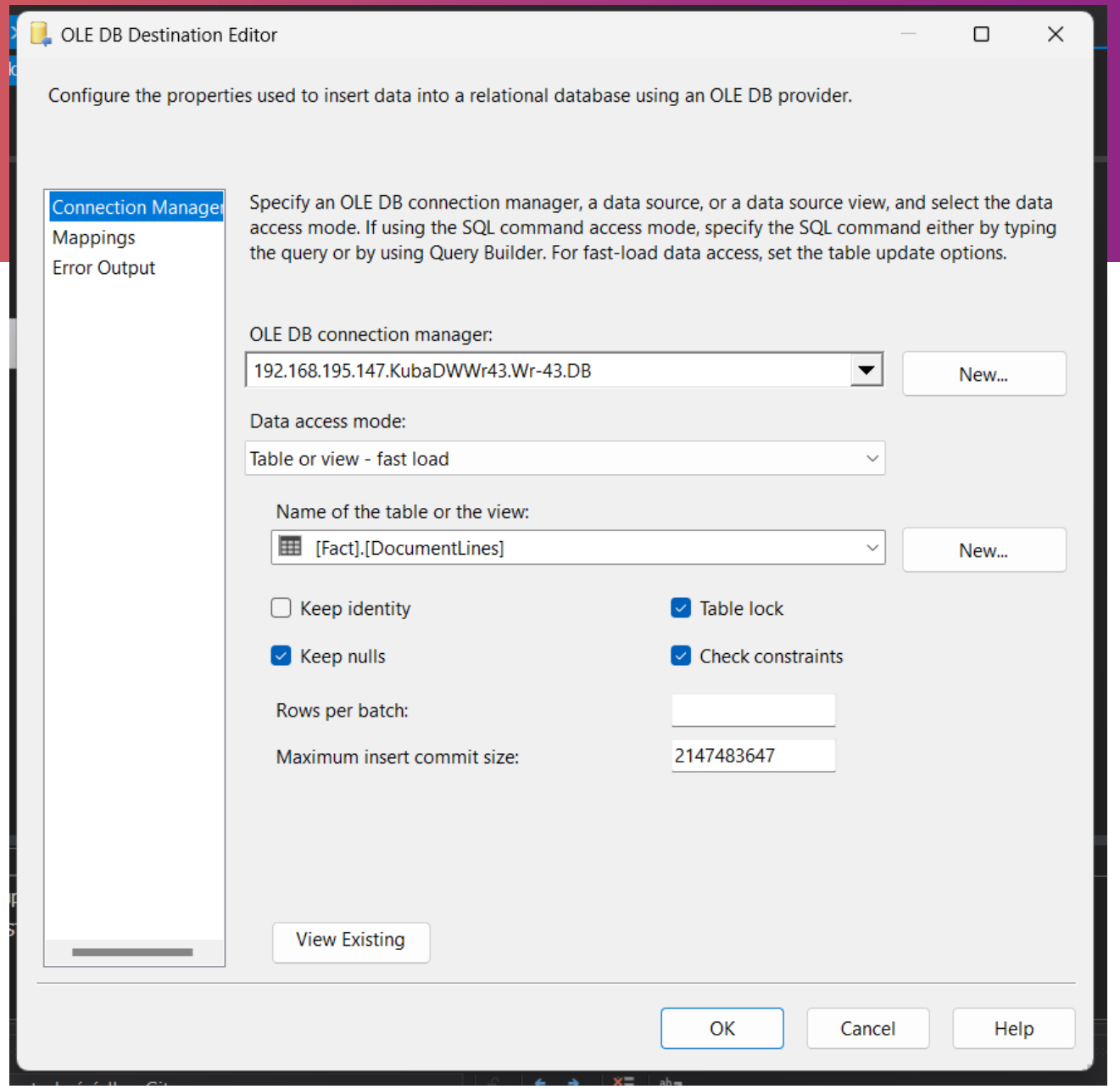

Fig. 9. Example of an upload from Staging to OLAP database

ETL Process

SSIS Job was supported by a *.txt log file for log documentation of the ETL Job.

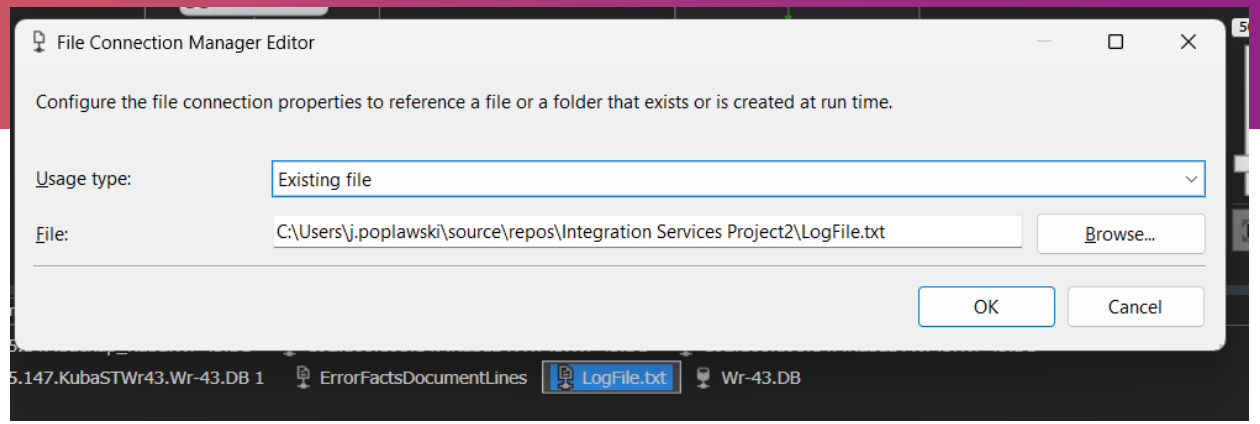

Fig. 10. Log file connection

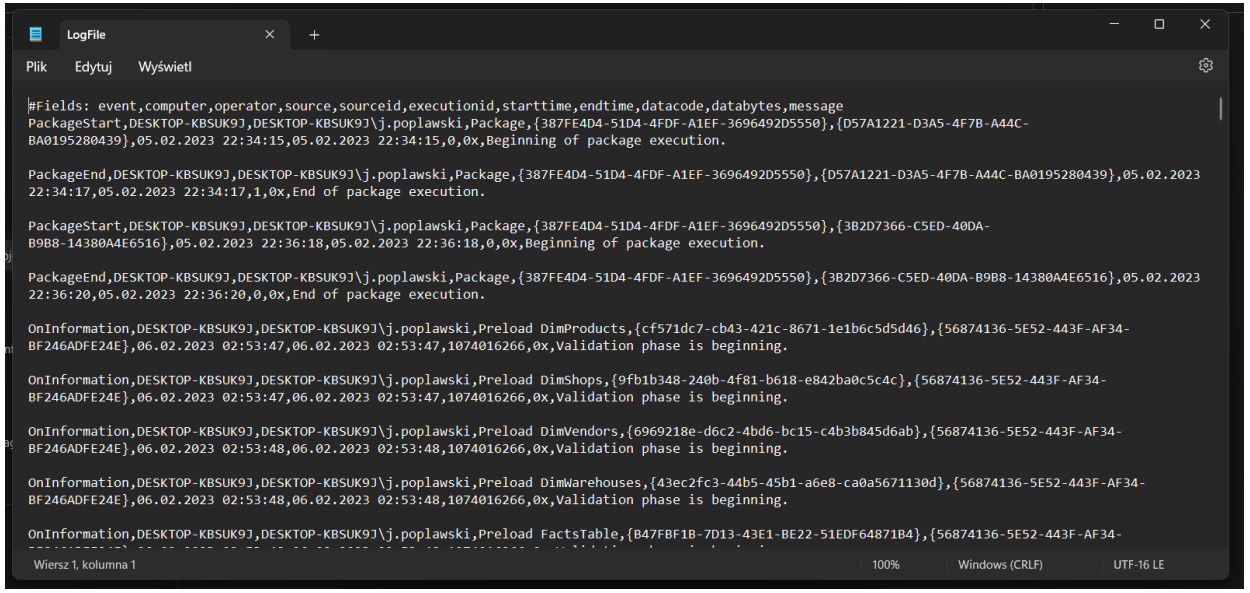

Fig. 11. Overview of the log file


FOR
THE
FUTURE
softserve