REPORT

on the

U2M8.ETL Overview – Extraction

U2M9.LW.ETL Overview - Transportation

Alina Sadovskaya

LAB8\_2. ETL Extraction - BASIC

2.1. Task 01: Extraction Description

Full Extraction

The data is extracted completely from the source system. Because this extraction reflects all the data currently available on the source system, there's no need to keep track of changes to the data source since the last successful extraction. The source data will be provided as-is and no additional logical information (for example, timestamps) is necessary on the source site.

#### Offline Extraction

The data is not extracted directly from the source system but is staged explicitly outside the original source system. The data already has an existing structure (for example, redo logs, archive logs or transportable tablespaces) or was created by an extraction routine.

LAB9\_2. ETL Transportation - BASIC

2.1. Task 01: Transportation Description

### Transportation Using Flat Files

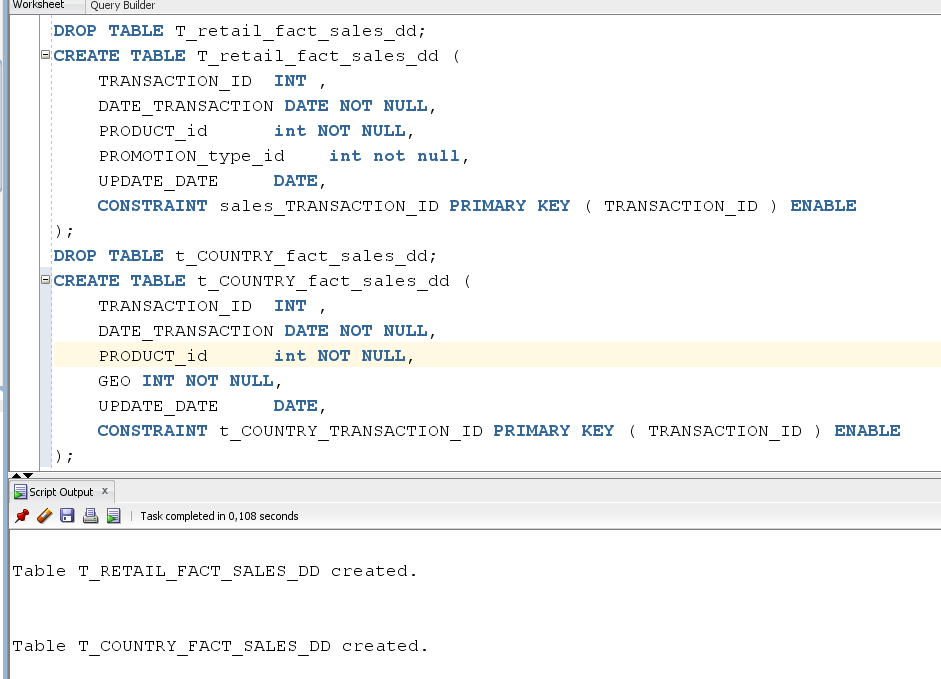
The most common method for transporting data is by the transfer of flat files, using mechanisms such as FTP or other remote file system access protocols. Data is unloaded or exported from the source system into flat files and is then transported to the target platform using FTP or similar mechanisms.

Because source systems and data warehouses often use different operating systems and database systems, using flat files is often the simplest way to exchange data between heterogeneous systems with minimal transformations. However, even when transporting data between homogeneous systems, flat files are often the most efficient and most easy-to-manage mechanism for data transfer.

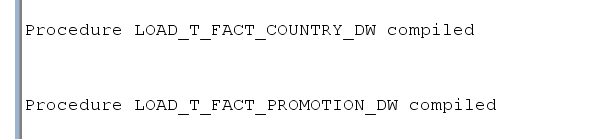
3. ETL Extraction – Example of Loading FCT\_\*

3.1. Task 02: Prepare Table of Facts to DW Layer

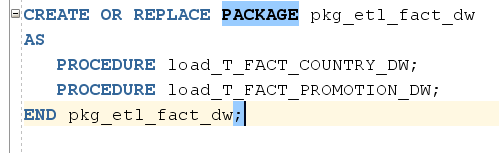
Back in lab 4, we cleared transactions and converted natural keys to primary keys in them. now all we have to do is group the data into the corresponding fact tables:



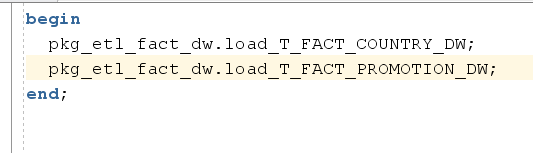
Now let's create a procedure for filling them in:

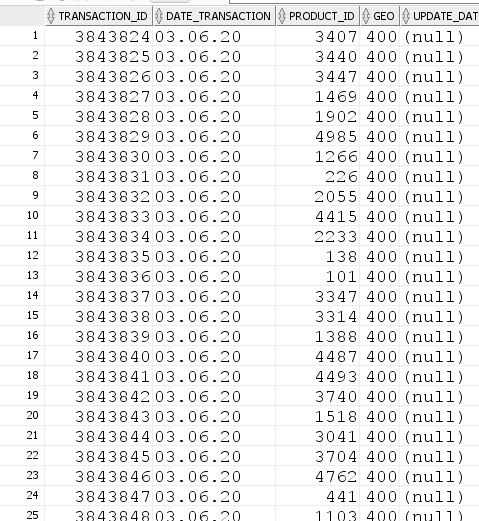
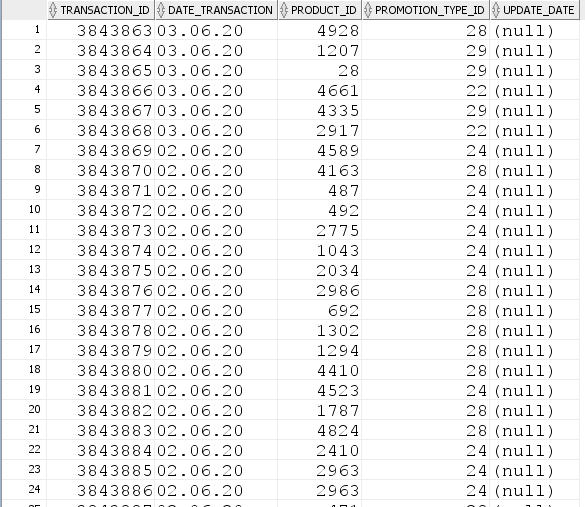


Combine the created procedures into a single package:



Сreate a script to run them:



Let's look at what the data looks like in our tables:

T\_Country\_fact\_sales\_dd T\_Retail\_fact\_sales\_dd