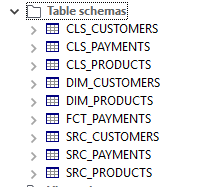
|  |
| --- |
| BI.Talend Data Integration Exit Task |

Solution Description.

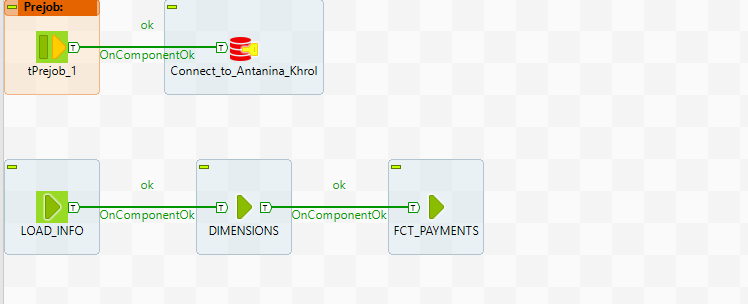
1. File DDL.sql. Scripts to create tables from SRC\_, CLS\_, DIM\_ levels. Also we create here sequences.
2. Create connection as context, retrieve tables from schema.



Pic.

1. Job ET\_03. General flow.

We can run this flow to get all ETL flow.

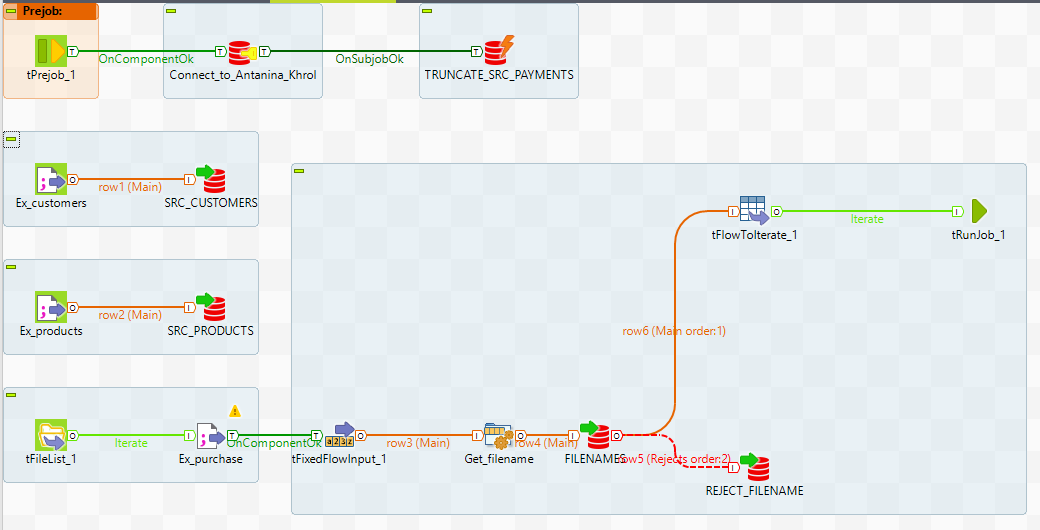


Pic.

1. Step Load info links to Job ET\_01.

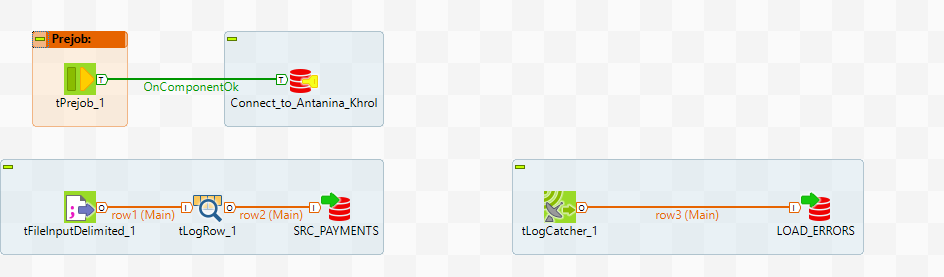
In this step we load data from Customers, Products files. Also the ability to load new incremental data from purchase\_yyyymmdd.csv is foreseen.

The used filenames is written in table Filenames. There is the Primary key in this table. If it’s possible to make insert we work with this file. If it’s impossible, we put the filename in the table REJECT\_FILENAME.



Pic.

1. The component RunJob goes to Job ET\_02.

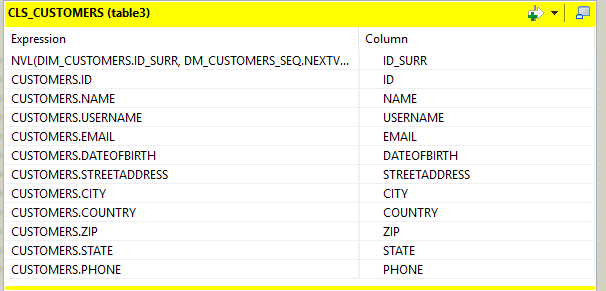


Pic.

In this step we write data from all files to SRC\_PAYMENTS and log errors.

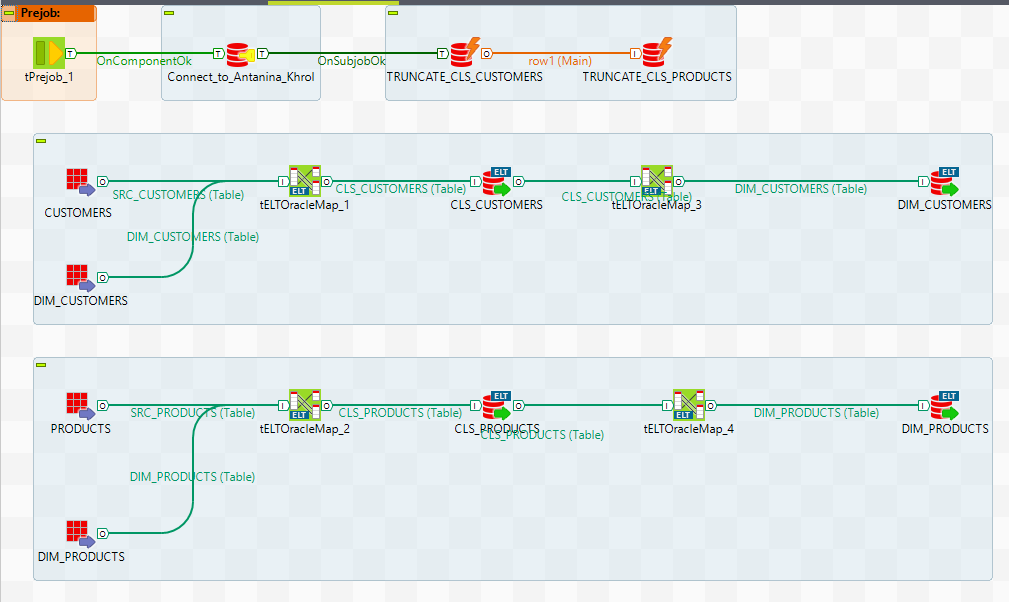
1. Block RunJob DIMENSIONS from the main flow goes to Job ET\_05.

In this step we work with tables in CLS\_ level. The main purpose is to create and insert the right ID\_SURR. If the row already in dimension table – we get the ID\_SURR from this table. If not, we create ID\_SURR using sequence.



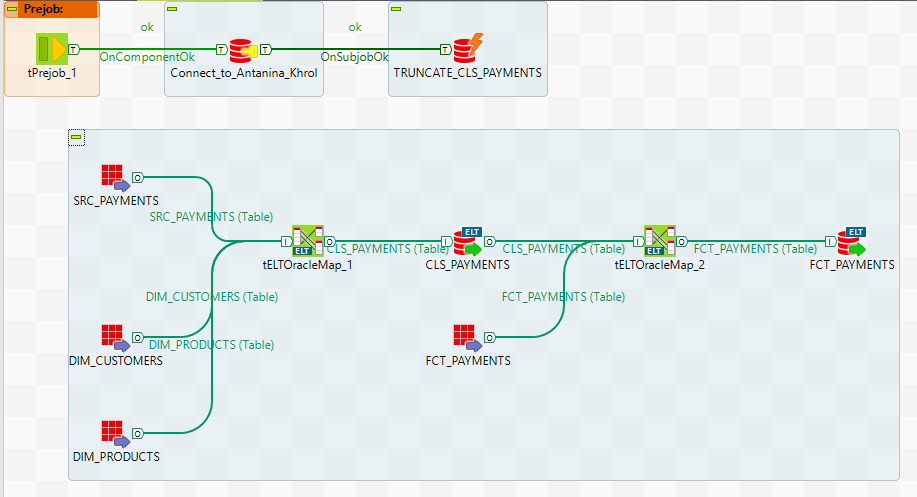
Pic.

After this step we merge info into DIM\_ tables. ID\_SURR is Primary key.



Pic.

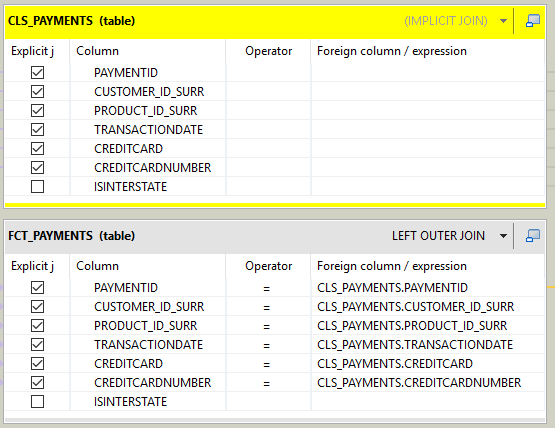
1. Block RunJob FCT\_PAYMENTS from the main flow goes to Job ET\_06.



Pic.

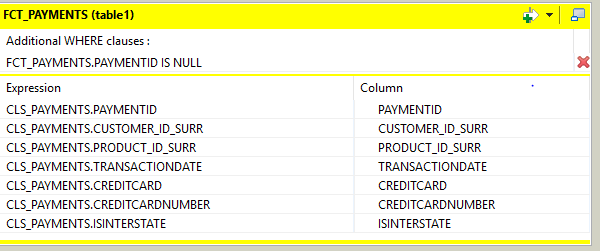
The main purposes of this step are: 1. Create new column isInterState. 2. Find new id for table CLS\_PAYMENTS (ID\_SURR). 3. Find in table CLS\_PAYMENTS data that is absent in table FCT\_PAYMENTS.

For the last goal we make join for all columns



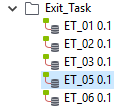
Pic.

and take row with Where clause.

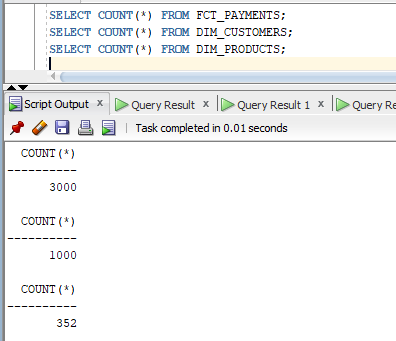


Pic.

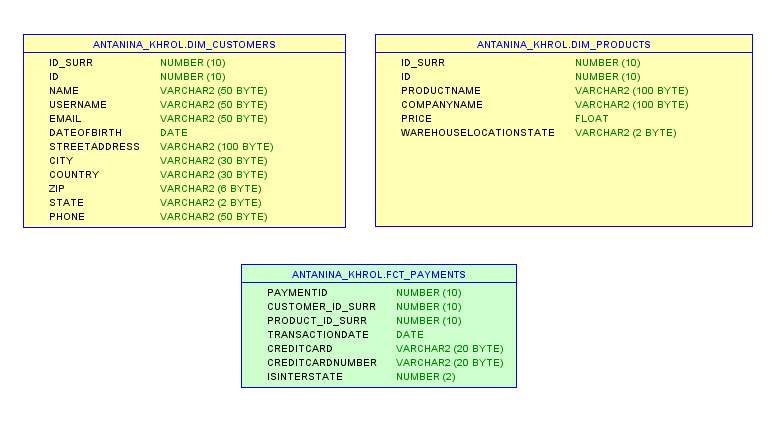
1. So, the project consists of 5 Jobs



Pic.



Pic.



Pic.