Miniproject 1 - DIS

Lars Andersen, Mathias Winde Pedersen & Søren Skibsted Als

September 29, 2014

D

We started this exercise with an overall plan, where we decided to load the data from the database, and then to transform this data and insert it into the dimension tables. After this, the fact table can be filled. We realised that the schema had to be changed, and as of such we moved the price from the dimension product, to the fact table, as price is what we measure. Additionally, we do not have incremental loads, but should be strived for in a larger and more permanent system.

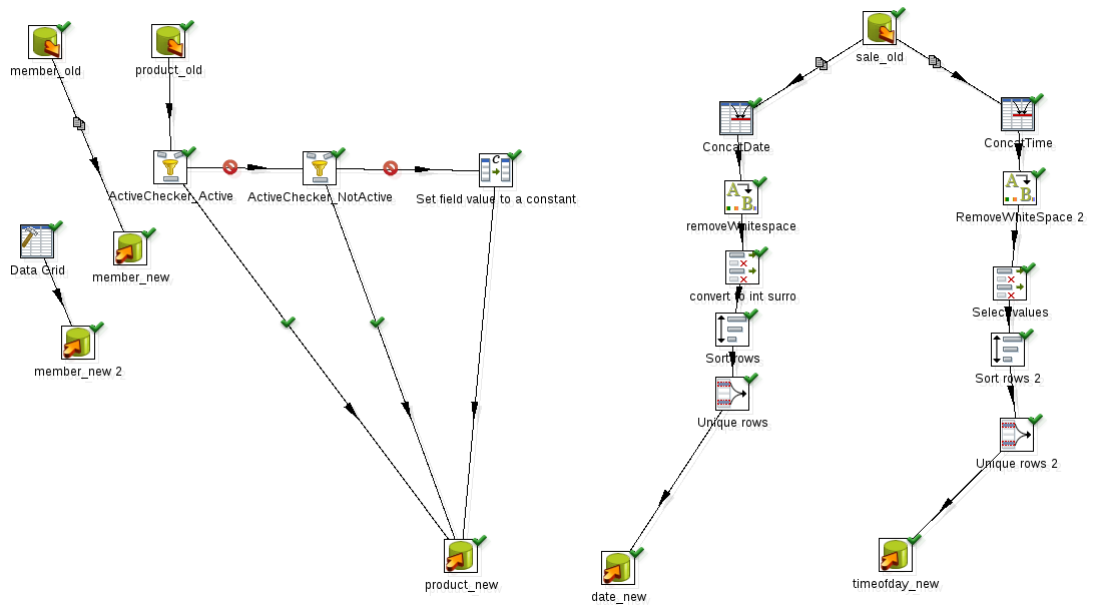In figure 1 a flow of how the dimensions are filled can be seen. As can be

Figure 1: Flow for filling of dimensions

seen, members are added straightforward, as the old and new database map one to one, with just an additional surrogate key. However, a dirty fix is that we add a member with memberid -1, in order to handle an error that occurred when filling the fact table (more on this in the description of the fact table flow).

The filling of the product is straight forward too, but we do some datacleaning, as some product were active but had a deactivation date, this data got deactivated instead.

For the filling of the date and timeofday dimension, it gets a bit more complicated. What we do there is to extract the information (year, month, day, hour, minute, second). When this is done, we generate a primary key by concatenating the information. So for date the primary key consists of year month day, and then for timeofday it is hour minute second.

In figure 2, the flow for filling the fact table can be seen. The figure depicts how the various surrogatekeys gets looked up in the dimension tables, as well as the price which is looked up in the old product table. After the lookups, the same generation of surrogate keys as in the filling of date and time dimensions
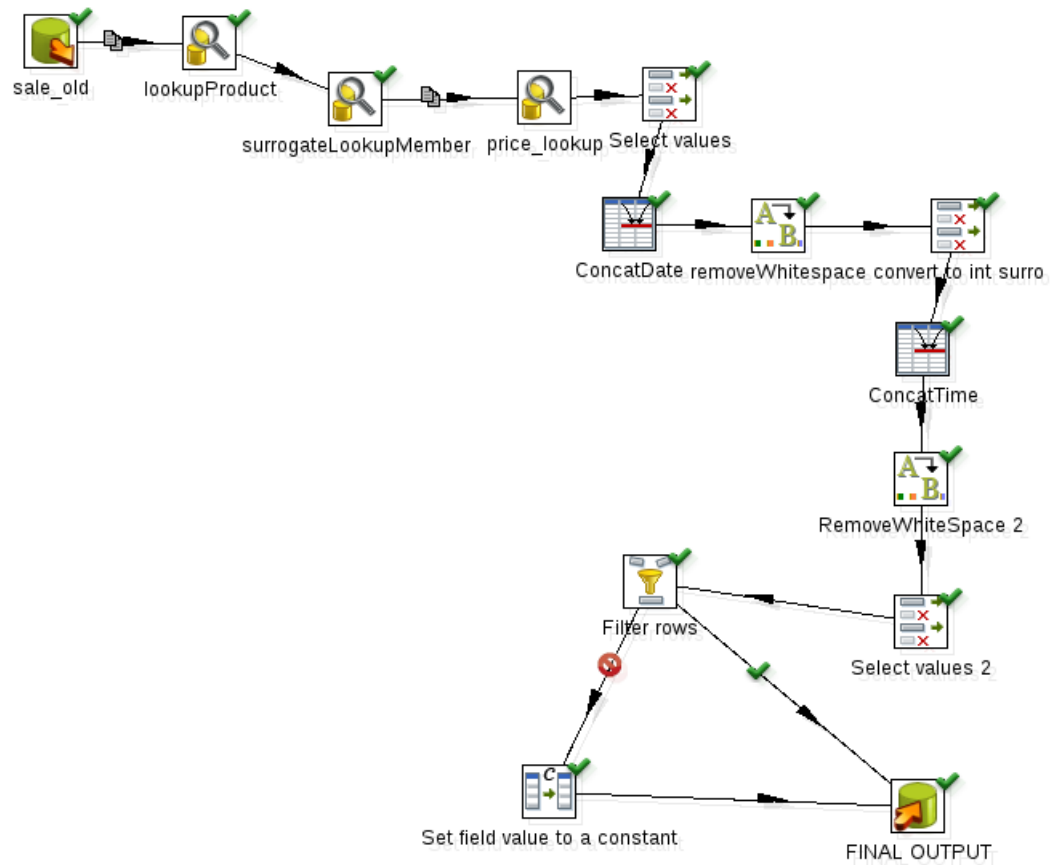
Figure 2: Flow for filling the fact table

are used, to get the same keys. Lastly, the mentioned erroneous memberids are filtered, such that members with an invalid memberid(0) gets changed to -1.