DIMENSIONS

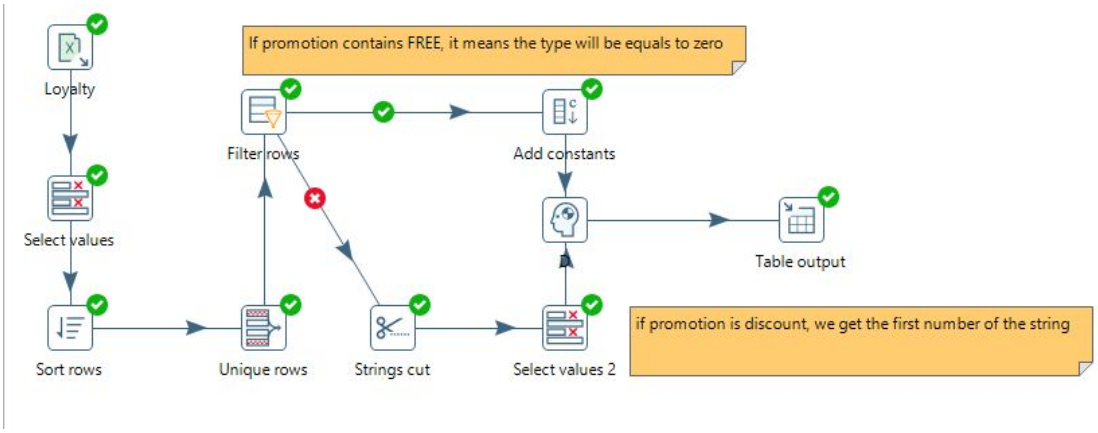
In all dimensions I followed the logic of getting the date from sources, getting the unique values from it and then storing it into the DB.

- dim_cinema transformation

- dim_format transformation

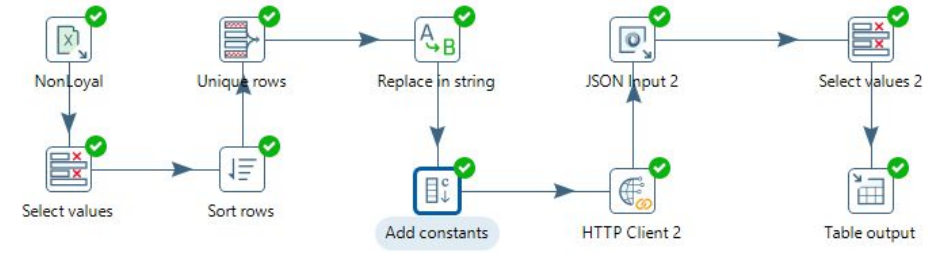
- dim_promotion transformation

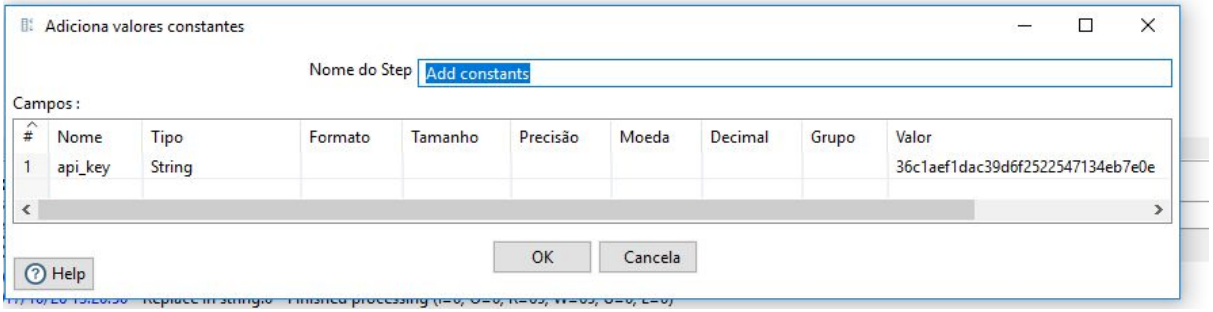
With promotion something else should be done, we have a string that describes the promotion. We can take the first character of the string and then we get the discount. But, we should also handle it when it means 1 free ticket. So I created a filter rows step (like if/else statement) that allowed me to split the data flow in two parts. After that I store it into the DB with type = 0 meaning that it is 1 free ticket promotion. It will be used in fact field calculations.

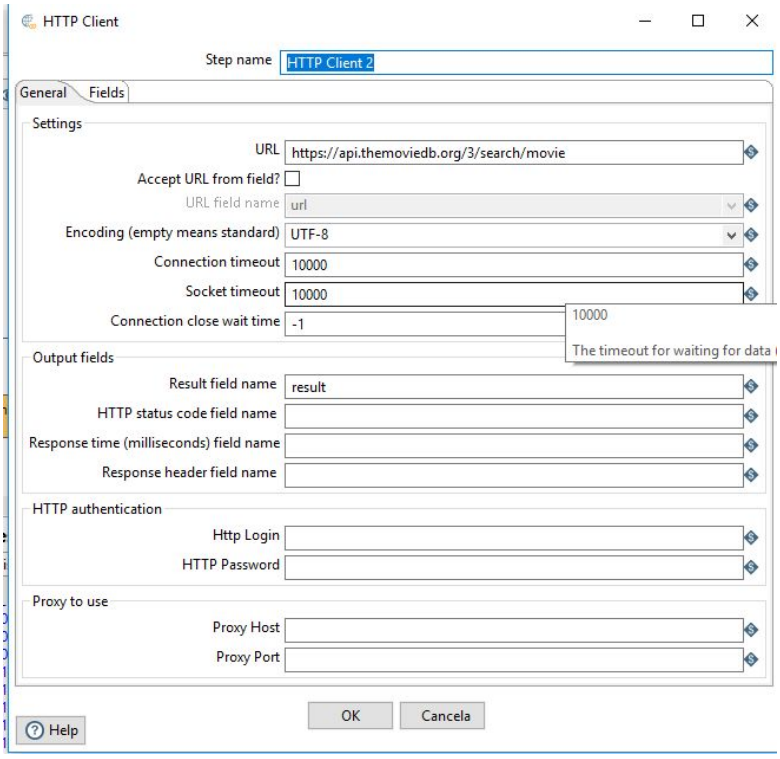


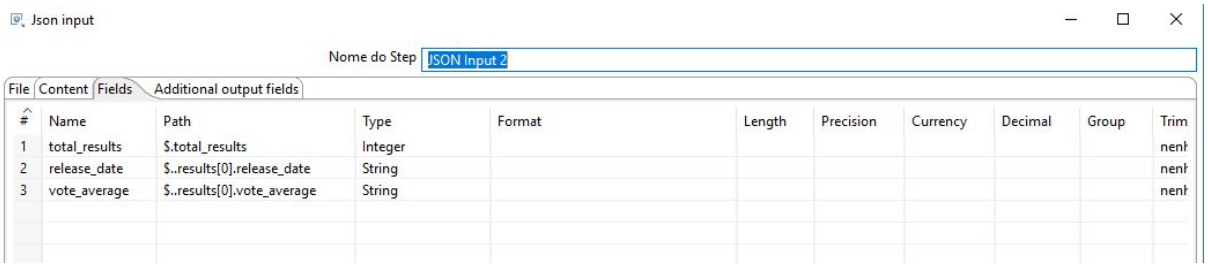
- dim_movies transformation

First we get the unique values from the source, then we add some constants column to hold the api_key information and the query parameter. After that, we create a HTTP client step where we inform the base API URL. Finally, we get a JSON result, using the JSON input step, we can handle this JSON and get only what we need. Then we store the data into DB









#	Name	Path	Type	Format	Length	Precision	Currency	Decimal	Group	Trim
1	total_results	$.total_results	Integer							nenh
2	release_date	$..results[0].release_date	String							nenh
3	vote_average	$..results[0].vote_average	String							nenh

FACT

In order to get the fact table, we must remove all the string information and replace them by foreign key from dimension tables. But before that, we must do some adjustments in our data. Join the data from our two sources (loyal and nonLoyal customers). They have different columns, so we must correct it.

Handles the loyalty program customers. We must get the information from promotions and get how much they have paid (total paid) and how many tickets were sold. To get the total paid we must just the data from our ticket price source. Also we have to make the calculations considering type 0 as 1 free tickets, and the rest following the formula described below.

After all calculations are done, some of the dimensions are missing in our Non Loyalty Program customers, so we must add those field with null values. After that we can finally join the two streams (as shown in Union Stream steps). After that, we can use Stream lookup step to get the id from dimension tables we created before. Finally we remove all non id_fields and store the data into DB.