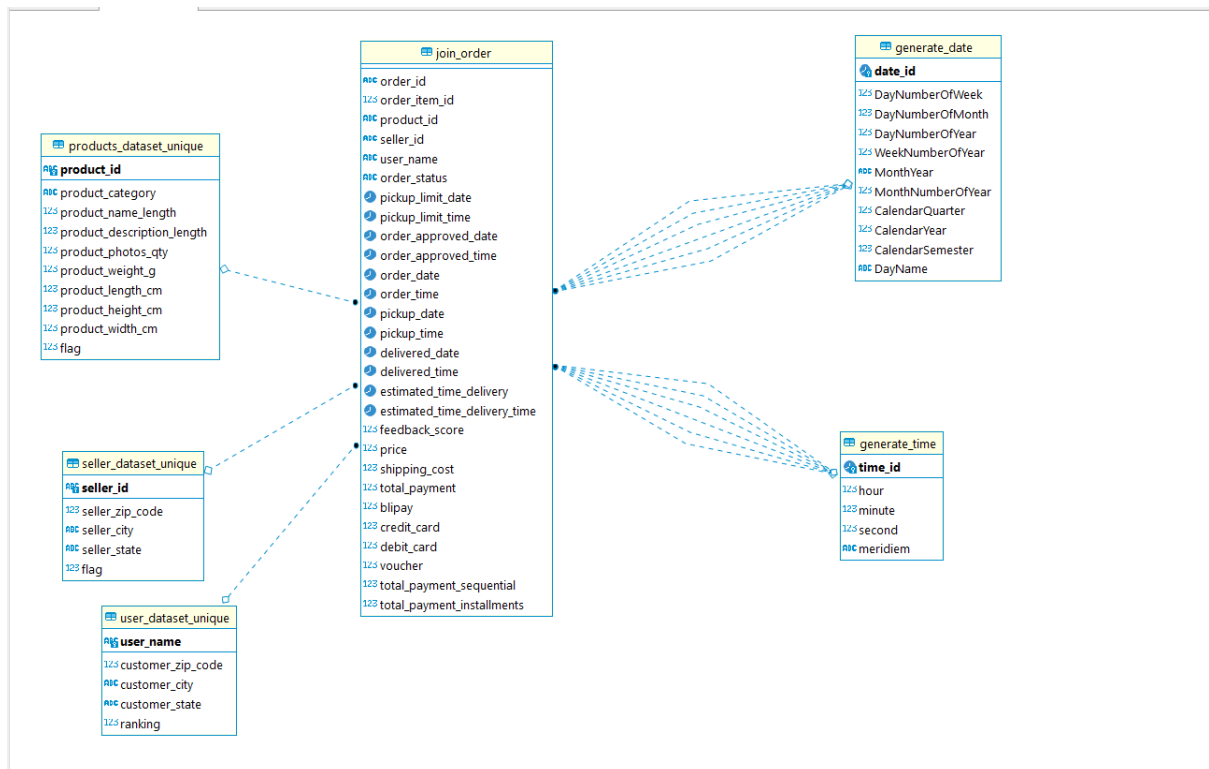


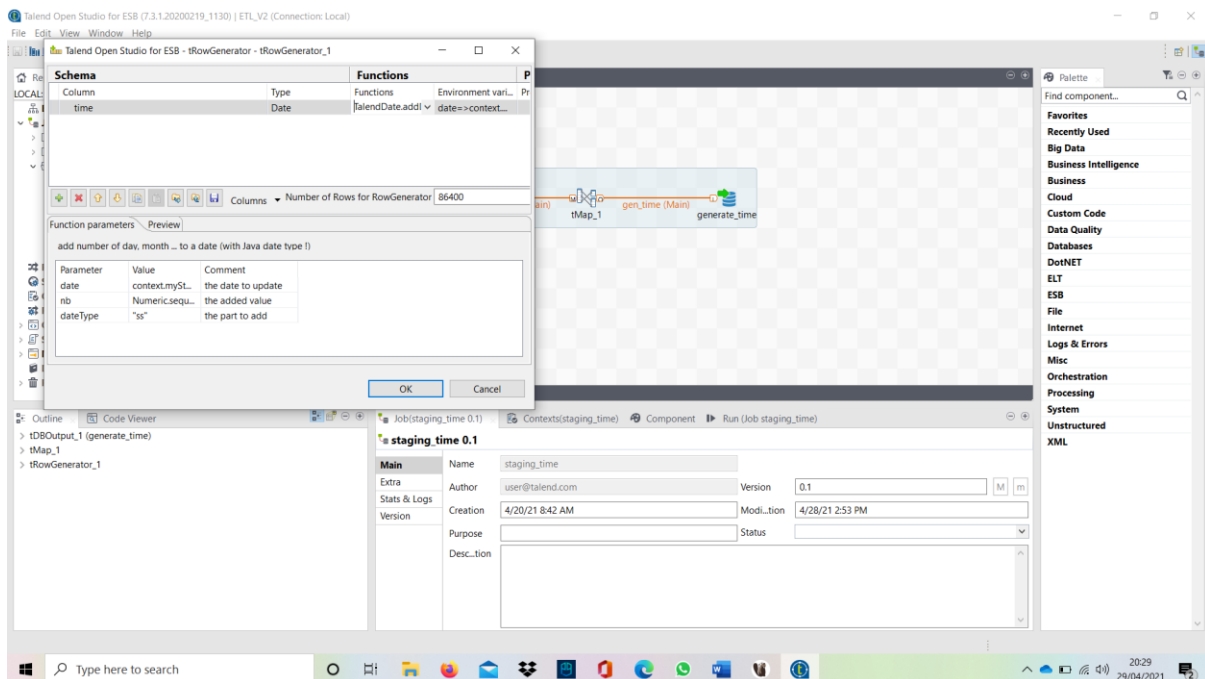
PROGRESS 29 APRIL 2021

DATE & TIME

Ada beberapa perubahan yang dilakukan pada progress kali ini. Yang pertama adalah pada date dan time. Pada staging terdapat table generate_date. Di generate_date terdapat date_id dengan tipe data date. Kemudian ada atribut lain seperti DayNumberOfWeek, DayNumberOfMonth, DayNumberOfYear, WeekNumberOfYear, MonthYear, MonthNumberOfYear, CalendarQuarter, CalendarYear, CalendarSemester, dan DayName. Jika sebelumnya saya generate date dari 2016-2018, saya tambah menjadi 2016-2020. Kemudian terdapat table generate_time yang saya generate sendiri dari talend. Atribut yang terdapat dalam table ini adalah time_id dengan tipe data time, kemudian atribut lainnya adalah hour, minute, second, dan meridiem (AM/PM).



Yang saya lakukan Ketika generate table generate_time adalah table ini menggunakan second untuk iterasi.



USER_DATASET

Kemudian untuk table user karena berdasarkan kesepakatan, kami menggunakan query untuk melakukan ranking dimana ranking pertama yang akan diambil.

create table user_dataset_fix as

```
SELECT *
```

```
FROM (SELECT distinct * ,
```

```
        RANK() OVER( PARTITION BY user_name
```

```
        ORDER BY  customer_zip_code DESC, customer_state DESC ,
        customer_city DESC ) AS ranking
```

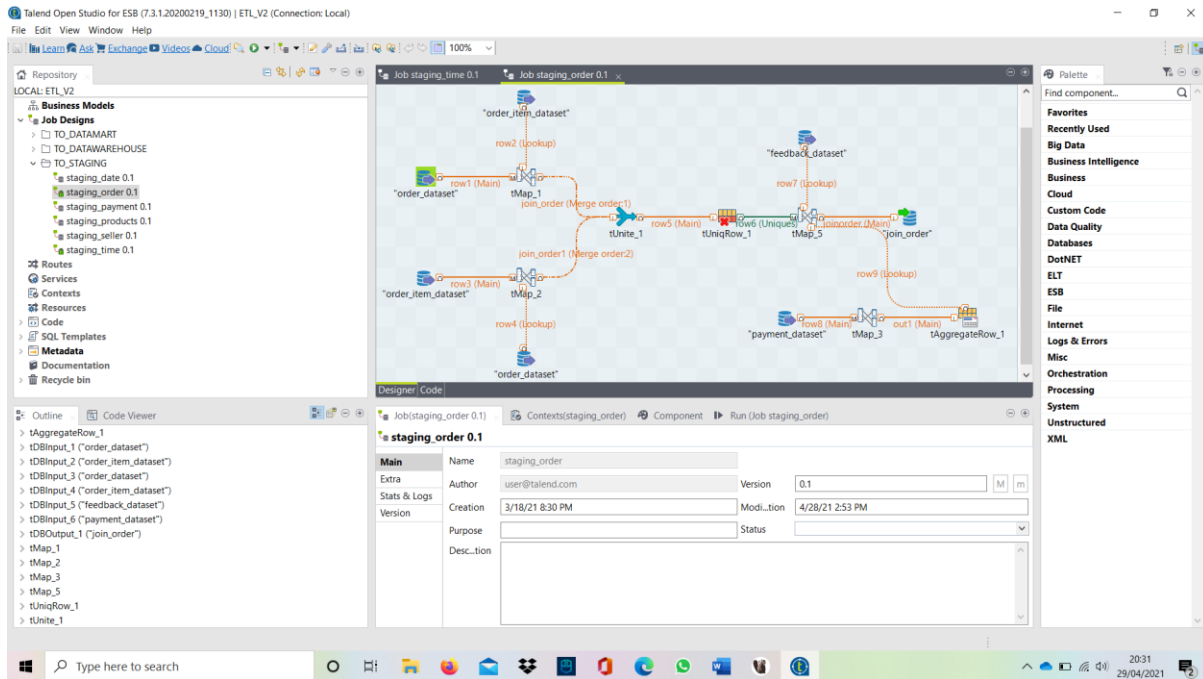
```
        FROM user_dataset_ori_csv) t
```

```
WHERE t.ranking = 1
```

```
order by customer_zip_code
```

JOIN_ORDER

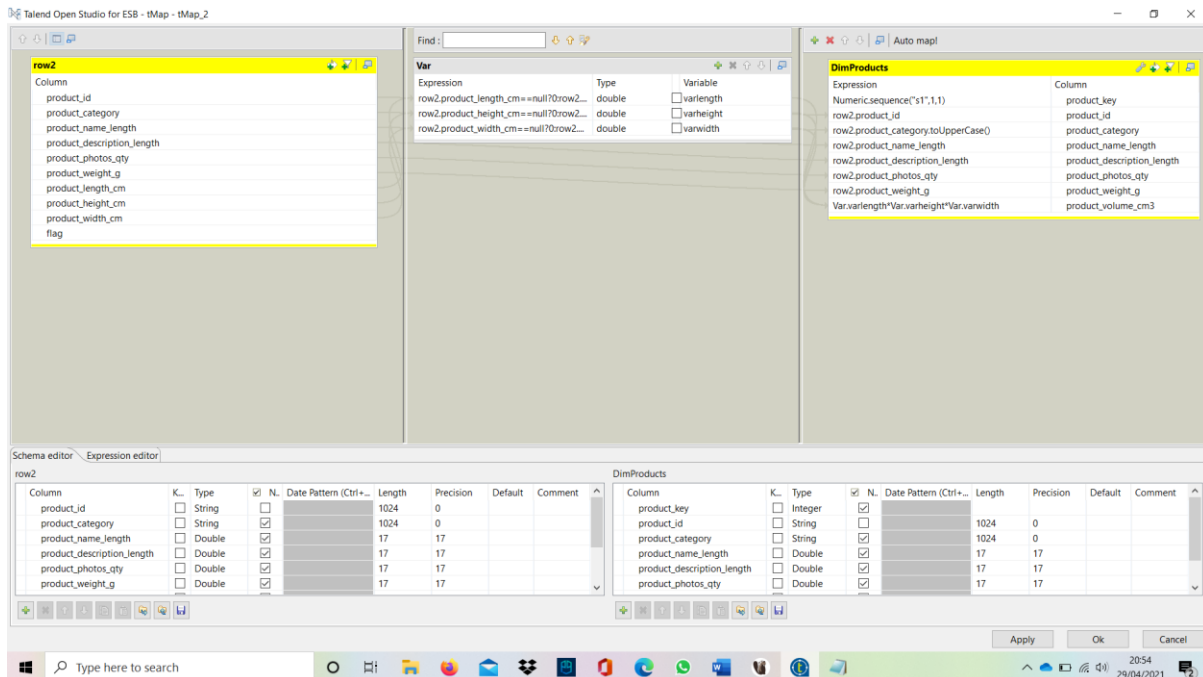
Kemudian pada join_order, disini outputnya akan mengarah ke schema staging. Jadi jika sebelumnya untuk mengenerate paymentnya pada job yang terpisah, agar lebih ringkas maka dijadikan satu pada job yang sama.



HANDLING NULL

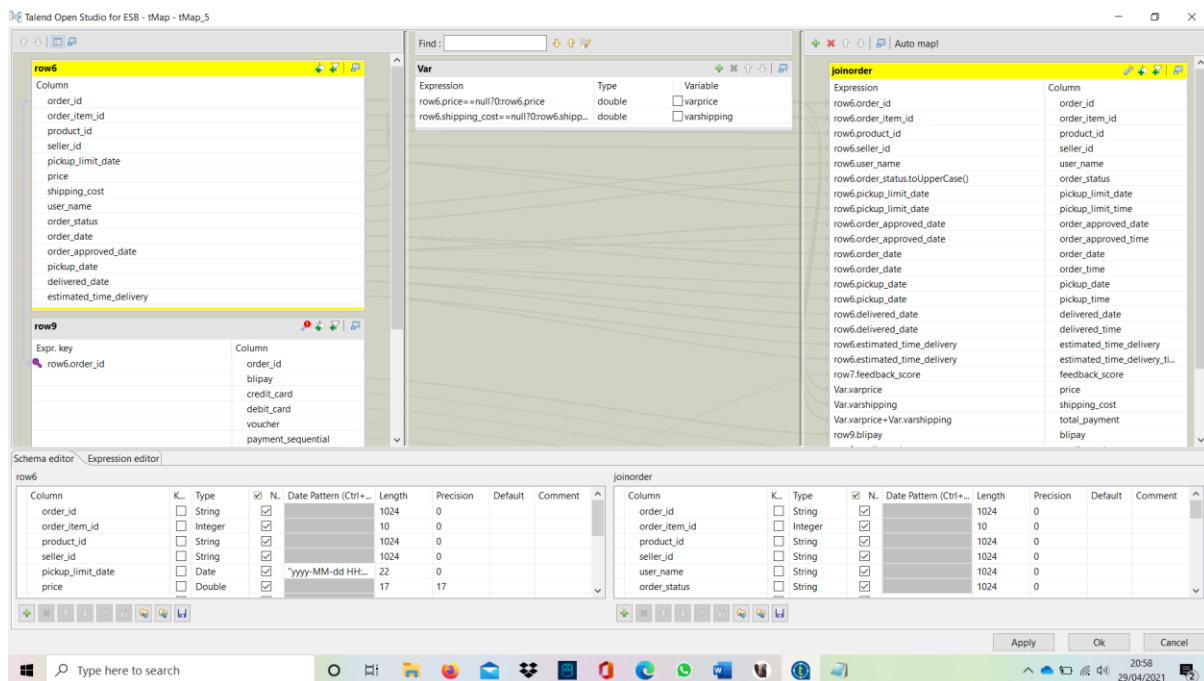
Products

Kemudian saya mengeksplor fitur yang baru saya eksplor di talend. Pada dimensi produk saya ingin membuat volume. Namun ini akan error jika ada row yang memiliki nilai null. Kemudian saya menggunakan var untuk menangkap nilai null dari length, height, dan width. Nilai Null tersebut kemudian diubah menjadi 0. Kemudian saya kalikan length, height, dan width untuk mendapatkan volume.



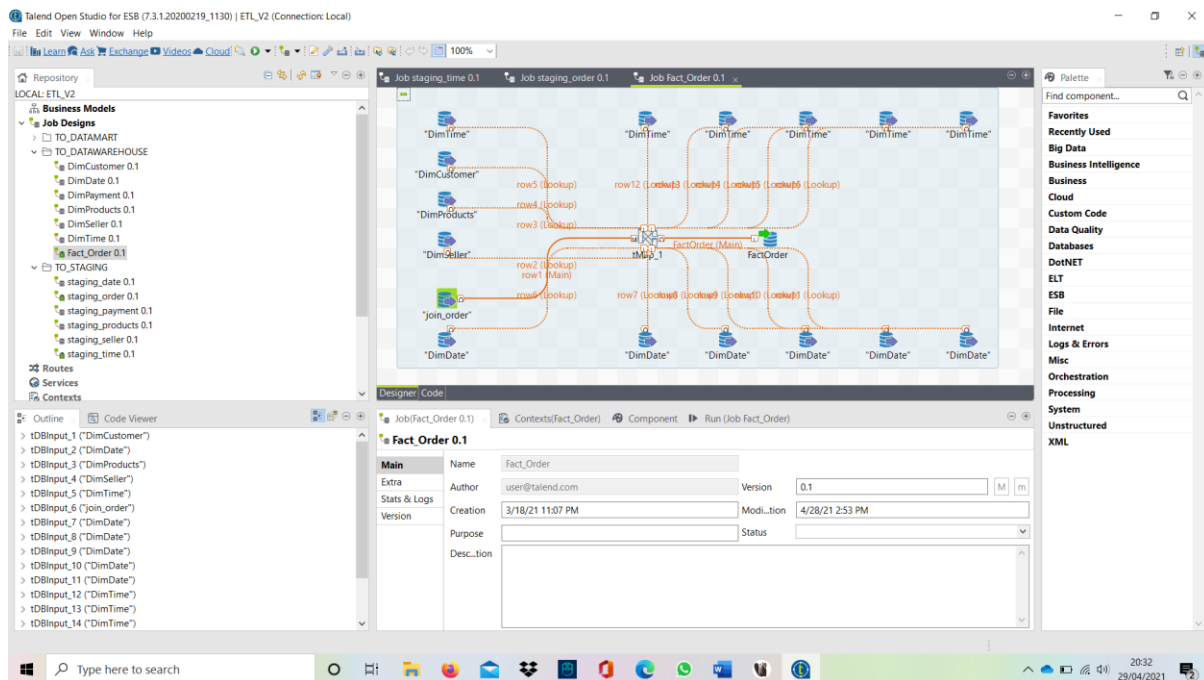
Join_order

Hal ini saya lakukan juga pada tMap join_order, dimana saya ingin membuat total_payment tiap order item dengan menjumlahkan price dengan shipping_cost. Sebelum dijumlahkan, null dari price dan shipping cost kemudian ditangkap dan diubah jadi 0. Saya juga menggunakan variable ini untuk mengisi price dan shipping cost di join_order. Pada table join_order juga ada perubahan lagi dimana saya memisahkan antara date dan time. Sehingga terdapat pickup_limit_date, pickup_limit_time, order_approved_date, order_approved_time, order_date, order_time, pickup_date, pickup_time, delivered_date, delivered_time, estimated_time_delivery, estimated_time_delivery_time. Sehingga nantinya pada FactOrder, atribut-atribut tersebut akan berisi surrogate key dari dimensi date atau time.



FACTORDER

Kemudian pada FactOrder yang akan masuk datawarehouse. Disini ada perubahan terutama pada DimDate dan DimTime. FactOrder disini jika kemarin berisi tanggal, sekarang berisikan key dari table Time atau Date.



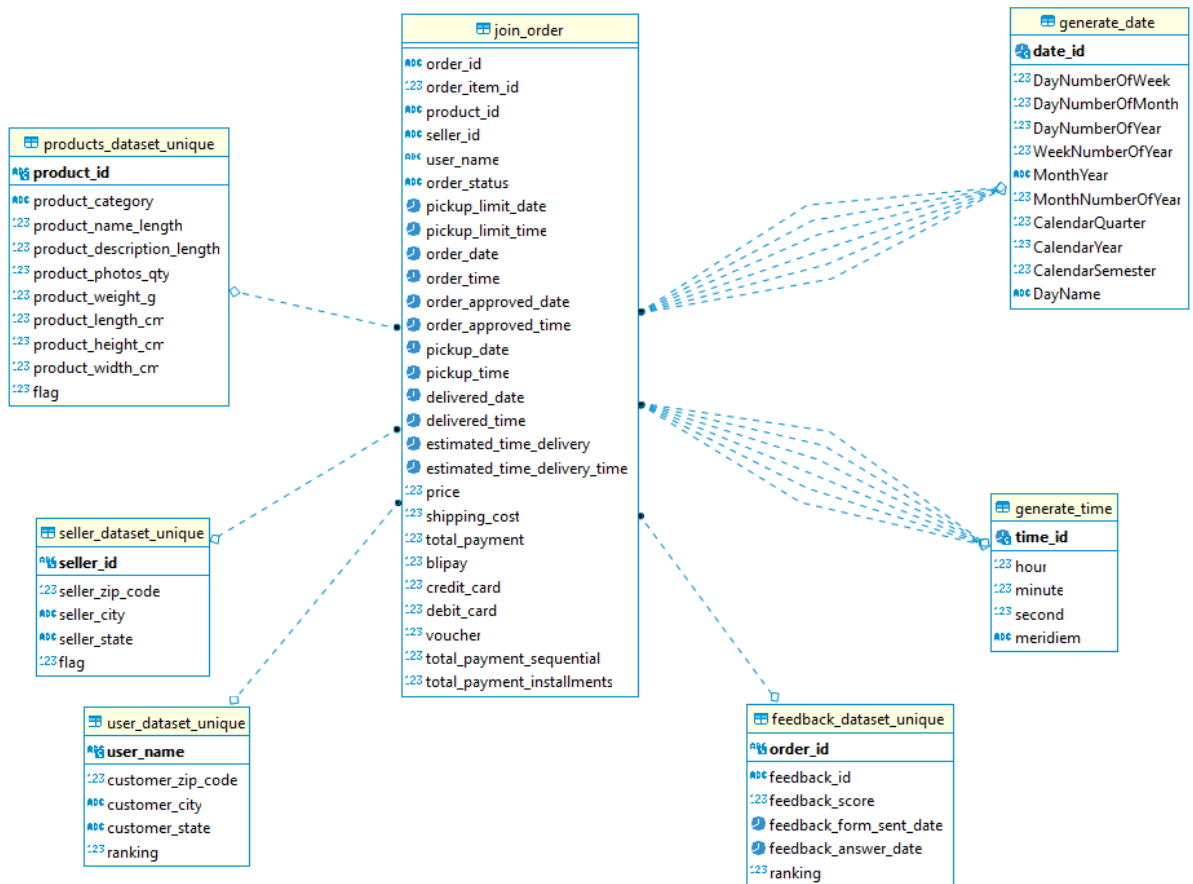
Karena beberapa perubahan ini, desain DataWarehouse juga ikut berubah.

PROGRESS 30 APRIL 2021

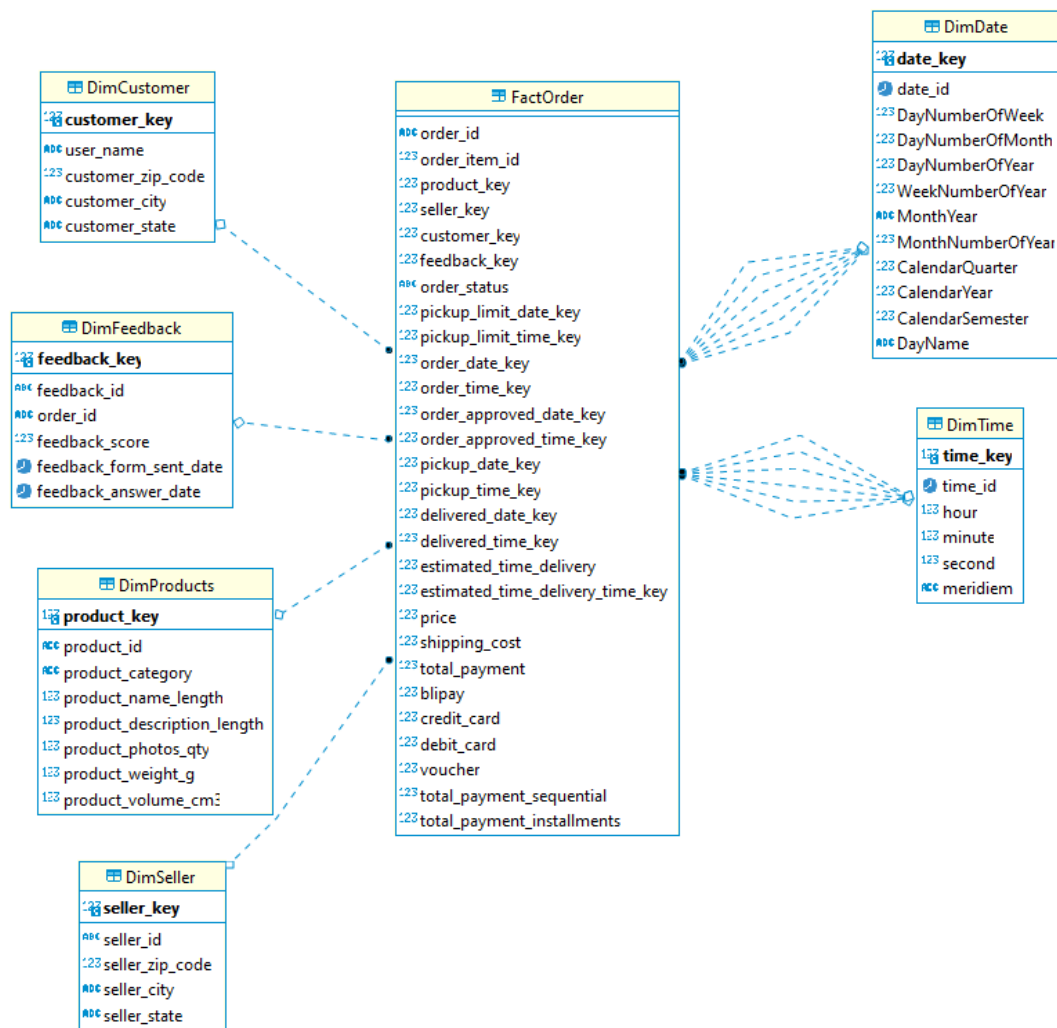
Progress pada 30 April adalah jika sebelumnya feedback_score merupakan bagian dari FactOrder, kali ini saya memisahkan menjadi dimensi sendiri yaitu DimFeedback. Cara saya mendapatkan DimFeedback adalah dari feedback_dataset, saya olah menggunakan query rank. Partitionnya berdasarkan order_id dan diranking berdasarkan feedback_answer_date terbaru ke terlama. Sehingga semakin tinggi rankingnya, maka semakin baru feedback tersebut. Query yang digunakan adalah sebagai berikut:

```
create table staging.feedback_dataset_unique as
SELECT *
FROM (SELECT distinct * ,
        RANK() OVER( PARTITION BY order_id
                     ORDER BY feedback_answer_date DESC) AS ranking
        FROM feedback_dataset) t
WHERE t.ranking = 1
order by feedback_answer_date
```

Kemudian feedback_dataset_unique tersebut terhubung dengan order_id yang ada di join_order. Kemudian pada DimFeedback terdapat surrogate_key yang bernama feedback_key. Feedback_key kemudian dimasukkan ke dalam FactOrder, fact table yang ada di datawarehouse. Tampilannya akan seperti berikut pada staging.



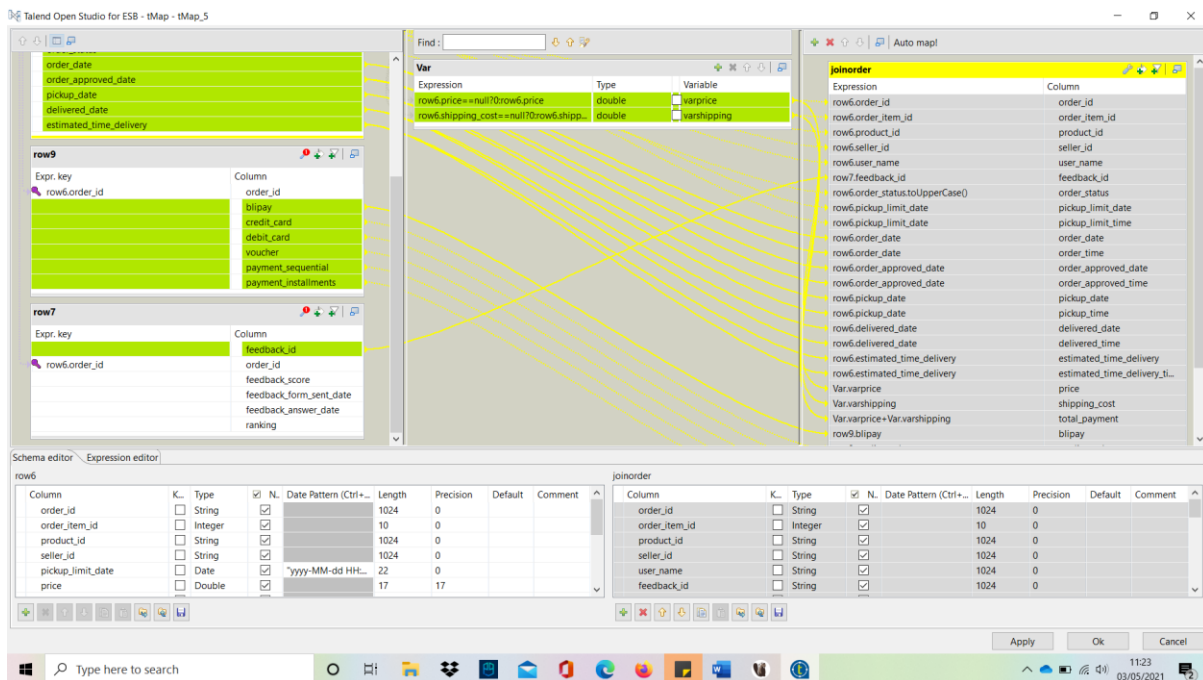
Kemudian berikut tampilan pada saat di datawarehouse



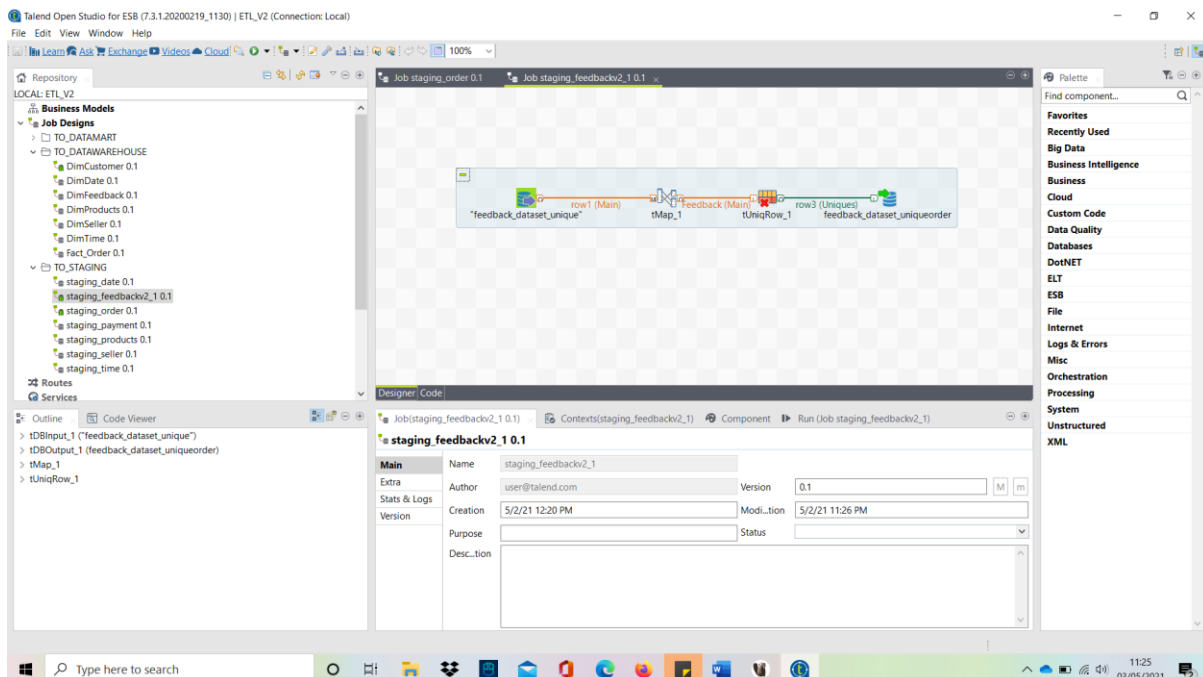
Kemudian untuk besok saya berencana untuk mengubah sedikit DimTime sehingga memiliki atribut DayNight.

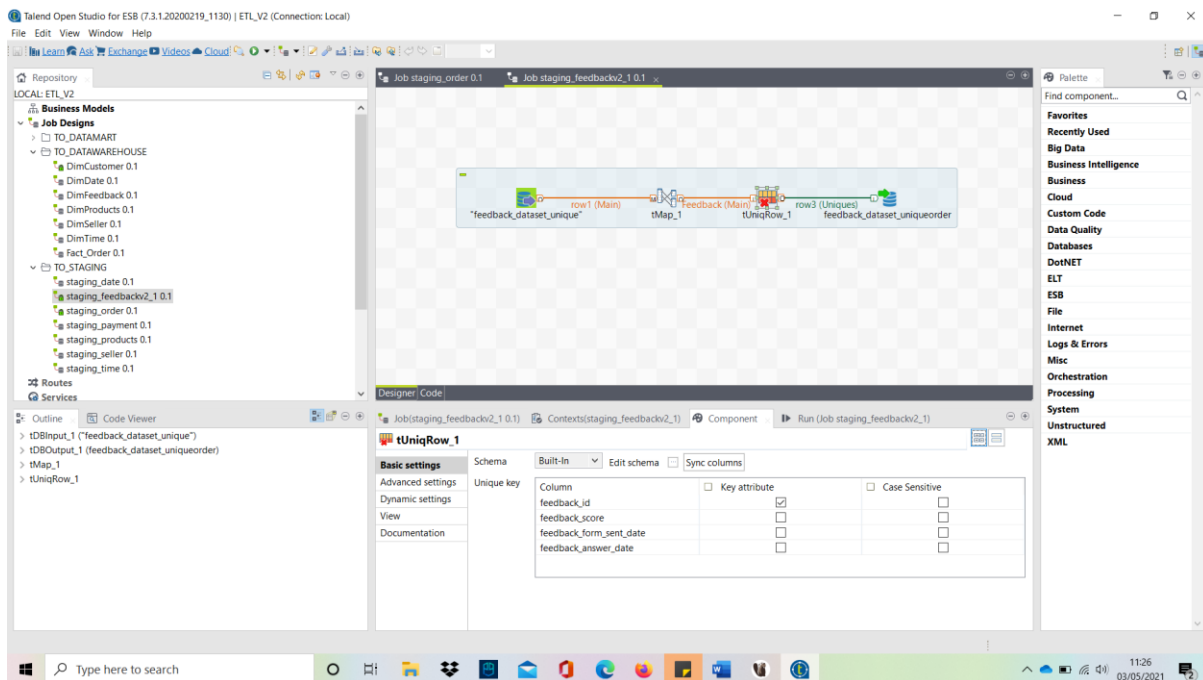
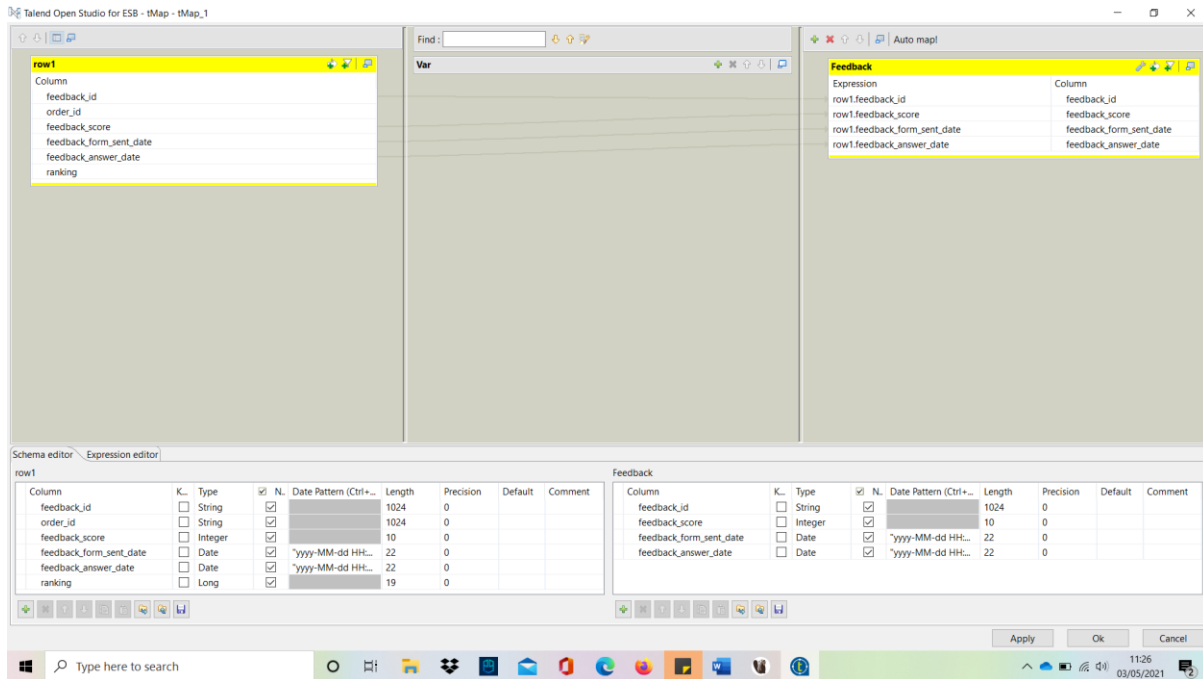
PROGRESS 2 MEI 2021

Di hari ini saya membetulkan table dari feedback. Jadi perubahan yang saya lakukan adalah pada table tersebut saya menghilangkan atribut order_id. Langkah yang saya lakukan adalah setelah terbentuk feedback_dataset_unique, saya left join dengan join_order yang merupakan gabungan dari order_dataset dan order_item_dataset, kemudian saya ambil feedback_id dari table feedback_dataset_unique untuk menjadi join_order.

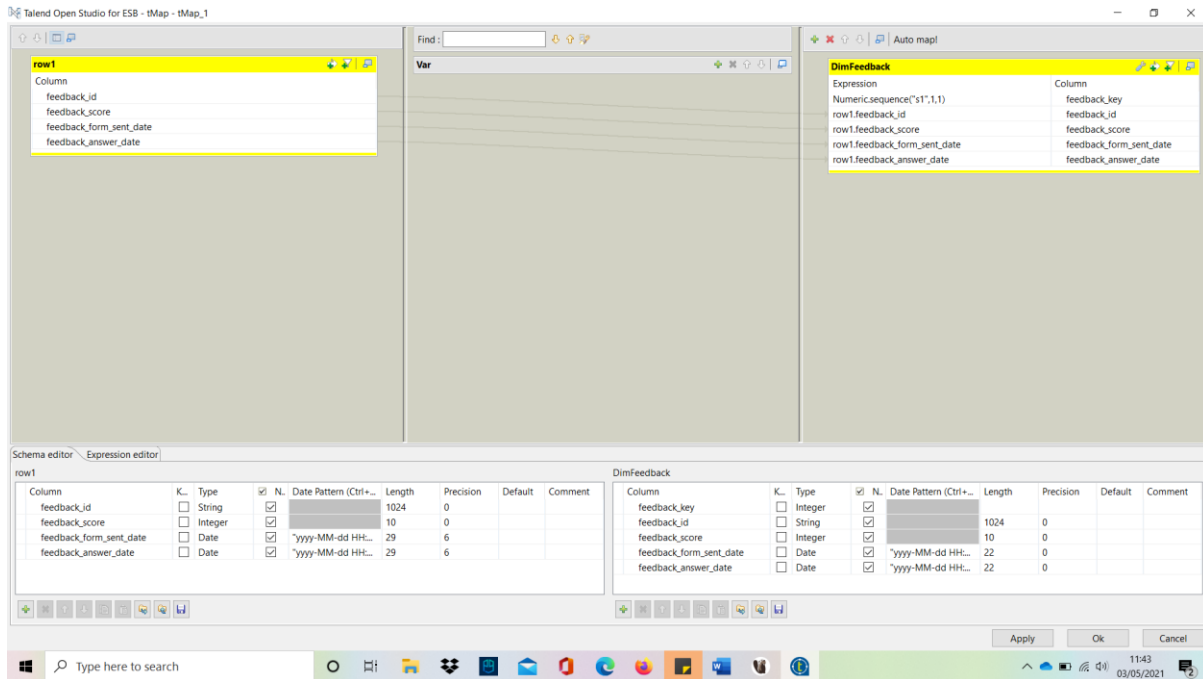


Kemudian saya membuat table `feedback_dataset_uniqueorder` dimana hanya akan berisi `feedback_id`, `feedback_score`, `feedback_form_sent_date`, dan `feedback_answer_date`. Kemudian dengan menggunakan `tUniqRow`, saya filter sehingga hanya menghasilkan `feedback_id` yang unik yang nantinya sebagai keynya. Jumlahnya terdapat 98854 row.

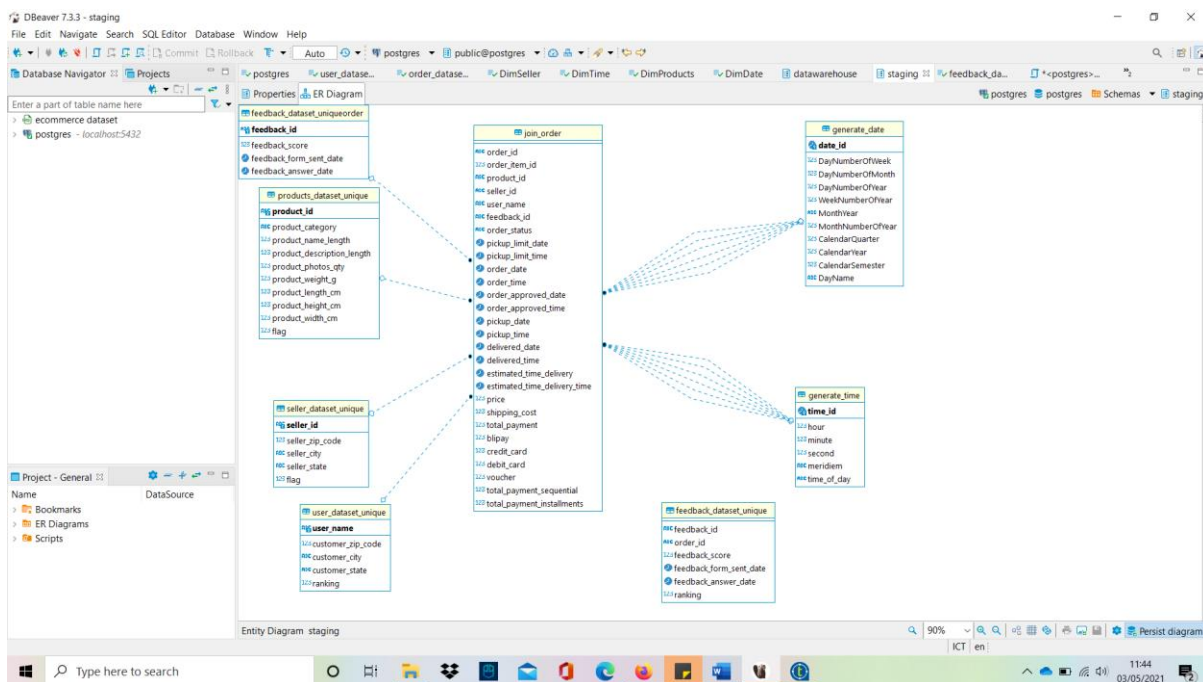


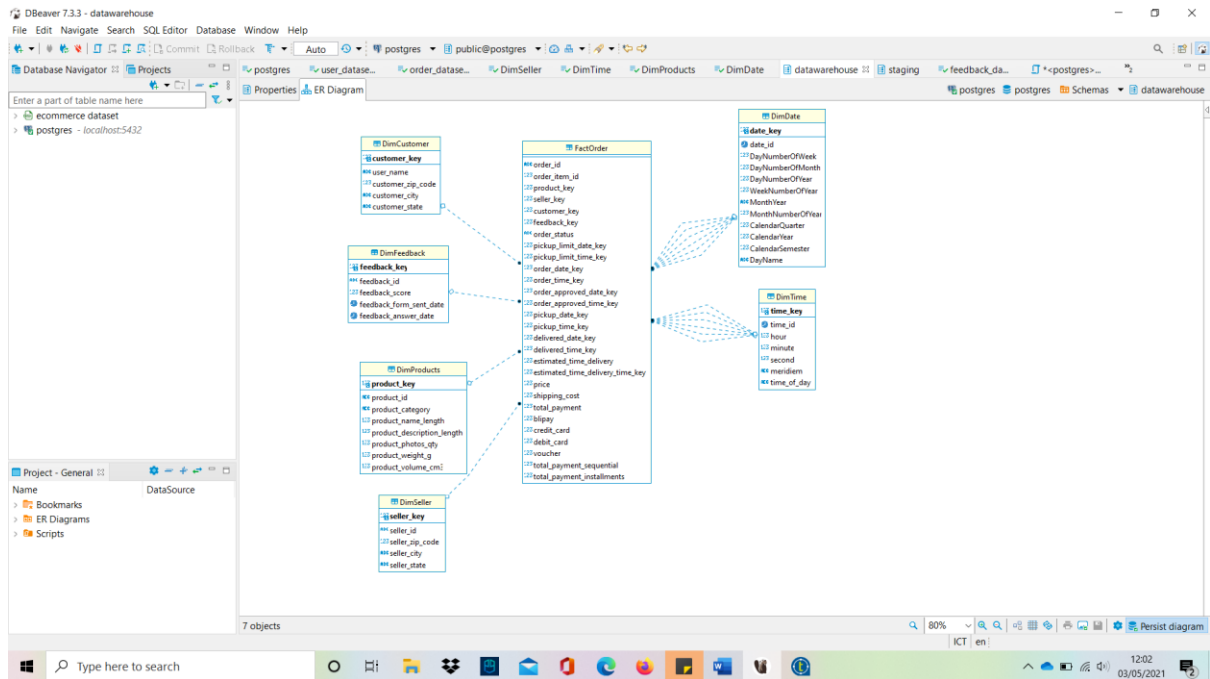


Kemudian pada dimensi feedback, saya berikan surrogate key.



Sehingga hasilnya akan terlihat pada gambar berikut. Disini bisa dilihat jika saya tidak menghapus feedback_dataset_unique karena disini saya membutuhkan informasi mengenai order_id pada table feedback yang sudah di deduct.





Berikut perubahan desain DataWarehouse yang terbaru.

