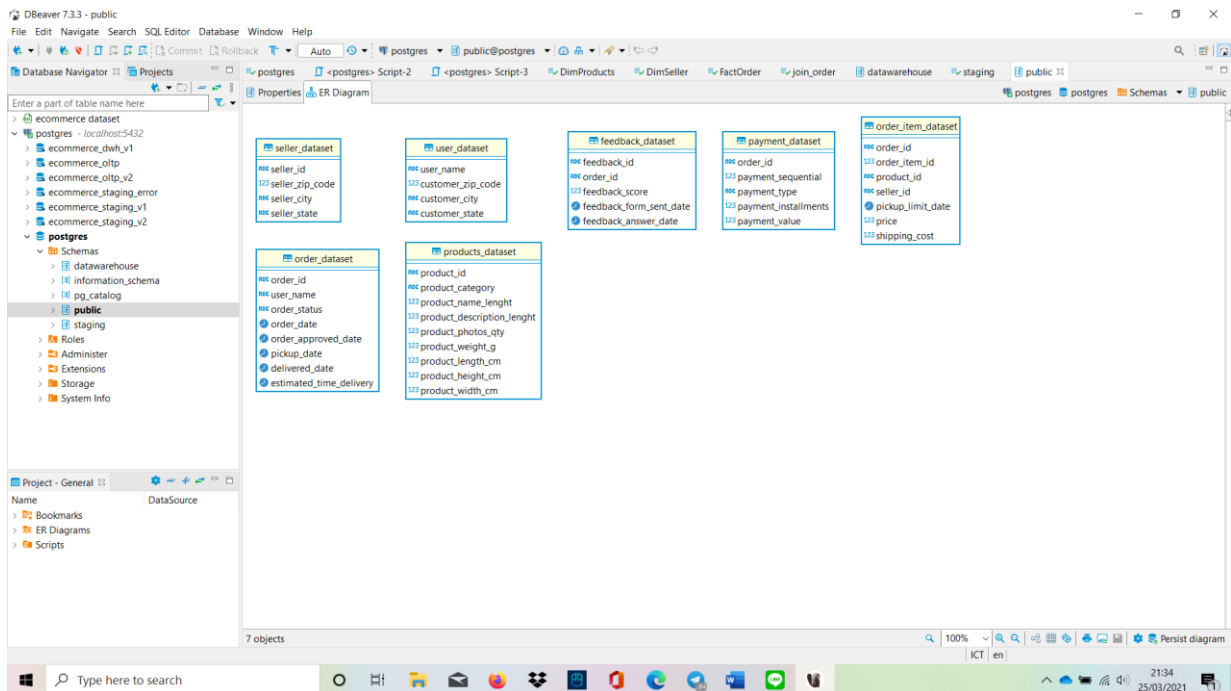


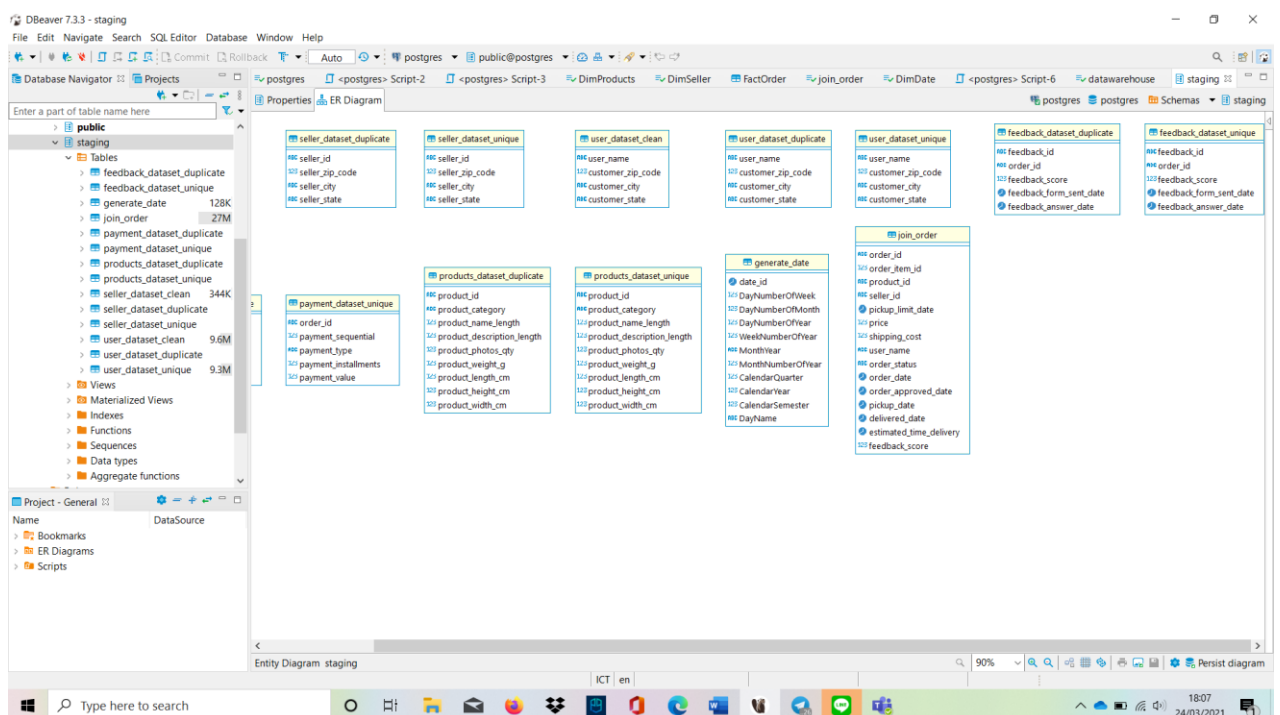
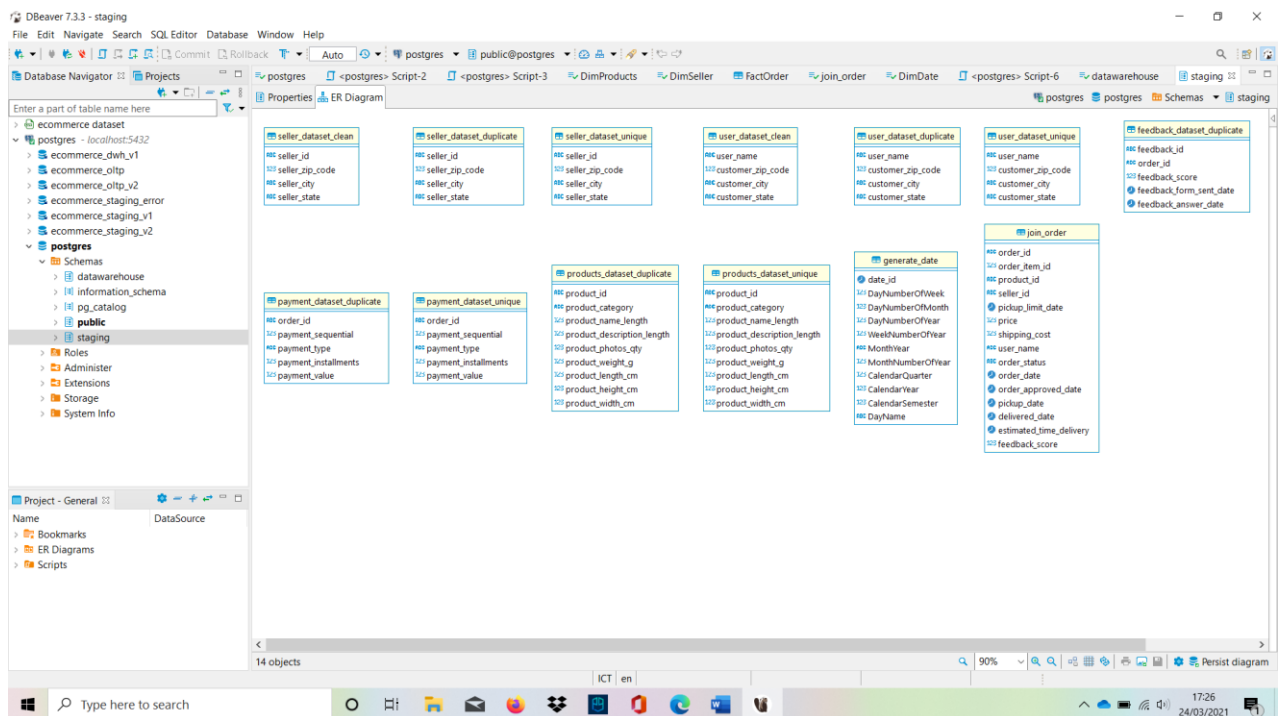
## PROGRESS 24 MARET 2021

Disini ada sedikit perubahan dari oltp, staging, dan datawarehouse yang terpisah-pisah databasenya, menjadi satu database di database **postgres** namun terdapat 3 schema, yaitu schema OLTP, STAGING, dan DATAWAREHOUSE

Di database OLTP, berisi csv dari kakak mentor, yang di load ke dalam **public** schema. Terdapat 6 table dalam schema ini, antara lain seller\_dataset, user\_dataset, feedback\_dataset, payment\_dataset, order\_item\_dataset, order\_dataset, dan products\_dataset.



Kemudian pada database, di schema **staging** ada banyak table. Saya akan jelaskan tiap table dari kiri atas ke kanan bawah.



Yang pertama ada seller\_dataset\_clean. Seller\_dataset\_clean berasal dari seller\_dataset yang datanya di cleaning, dimana maksud cleaning disini adalah memastikan bahwa satu zip code hanya dimiliki oleh satu kota dan provinsi. Dicocokkan juga dengan data yang ada di user\_dataset\_clean.

Yang kedua ada seller\_dataset\_duplicate. Seller\_dataset\_duplicate berasal dari seller\_dataset\_clean yang diproses di Talend menggunakan tUniqrow. Seller\_dataset\_duplicate berisi data dengan seller\_id yang duplikat.

Yang ketiga ada seller\_dataset\_unique. Seller\_dataset\_unique berasal dari seller\_dataset\_clean yang diproses di Talend menggunakan tUniqrow. Seller\_dataset\_unique berisi data dengan seller\_id yang unik.

Yang keempat ada user\_dataset\_clean. User\_dataset\_clean berasal dari user\_dataset yang datanya di cleaning, dimana maksud cleaning disini adalah memastikan bahwa satu zip code hanya dimiliki oleh satu kota dan provinsi.

Yang kelima ada user\_dataset\_duplicate. User\_dataset\_duplicate berasal dari user\_dataset\_clean yang diproses di Talend menggunakan tUniqrow. User\_dataset\_duplicate berisi data dengan user\_name yang duplikat.

Yang keenam ada user\_dataset\_unique. User\_dataset\_unique berasal dari user\_dataset\_clean yang diproses di Talend menggunakan tUniqrow. User\_dataset\_unique berisi data dengan user\_name yang unik.

Yang ketujuh ada feedback\_dataset\_duplicate. Feedback\_dataset\_duplicate berasal dari feedback\_dataset yang diproses di Talend menggunakan tUniqrow. Feedback\_dataset\_duplicate berisi data dengan feedback\_id dan order\_id yang duplikat.

Yang kedelapan ada payment\_dataset\_duplicate. Payment\_dataset\_duplicate berasal dari payment\_dataset yang diproses di Talend menggunakan tUniqrow. Payment\_dataset\_duplicate berisi data dengan order\_id dan payment\_sequential yang duplikat.

Yang kesembilan ada payment\_dataset\_unique. Payment\_dataset\_unique berasal dari payment\_dataset yang diproses di Talend menggunakan tUniqrow. Payment\_dataset\_unique berisi data dengan order\_id dan payment\_sequential yang unik.

Yang kesepuluh ada products\_dataset\_duplicate. Products\_dataset\_duplicate berasal dari products\_dataset yang diproses di Talend menggunakan tUniqrow. Products\_dataset\_duplicate berisi data dengan product\_id yang duplikat.

Yang kesebelas ada products\_dataset\_unique. Products\_dataset\_unique berasal dari products\_dataset yang diproses di Talend menggunakan tUniqrow. Products\_dataset\_unique berisi data dengan products\_id yang duplikat.

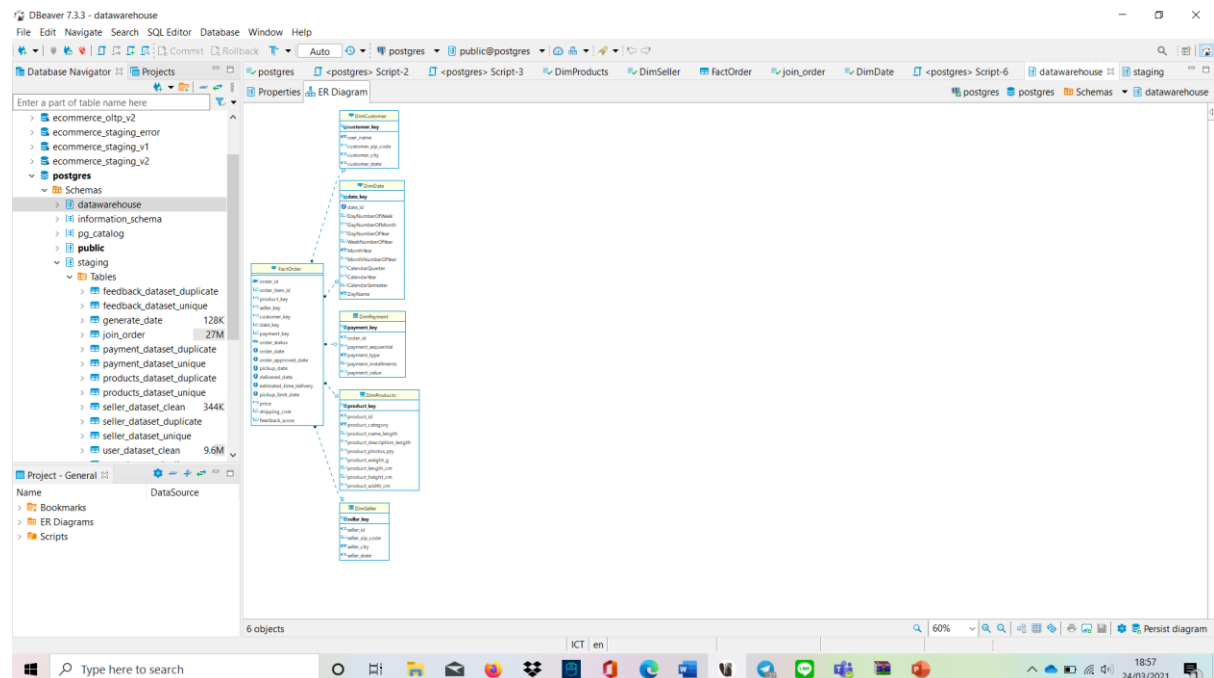
Yang keduabelas ada generate\_date. Generate\_date dibuat dengan di Talend. Caranya adalah dengan membuat context di job, ada myStartDate dan myEndDate. Disini saya lebih menggunakan myStartDate. Tanggal paling awal adalah 2016-09-04 dan paling akhir adalah 2018-10-17. Disini saya set 2016-01-01 sampai 2018-12-31. Kemudian menggunakan tRowGenerator, dimulai dari context.myStartDate, kemudian untuk urutan menggunakan Numeric.sequence("s1",1,1) -1, agar dari tanggal 2016-01-01 juga muncul. Kemudian date type adalah dd. Kemudian saya membuat berbagai macam kolom. Yang pertama ada day\_number\_of\_week, yang merupakan minggu ke berapa pada tanggal tersebut (Week 1-4). Kemudian ada day\_number\_of\_month, yang merupakan hari ke berapa pada suatu tanggal di bulan tersebut(1-31). Kemudian ada day\_number\_of\_year, yang merupakan hari ke berapa pada suatu tanggal di tahun tersebut(1-366). Month\_year merupakan bulan dalam bentuk string tanggal tersebut(January – December). Terdapat month\_number\_of\_year, yaitu bulan ke berapa pada tanggal tersebut (1-12). Calendar\_quarter merupakan kuartal ke berapa tanggal tersebut (1-4). Calendar\_year merupakan tahun berapa pada tanggal tersebut (2016-2018). Calendar\_semester merupakan semester ke berapa pada tanggal tersebut (1-2). Day name merupakan nama hari pada tanggal tersebut (Monday-Sunday).

Yang ketigabelas ada join\_order. Join\_order merupakan gabungan dari order\_dataset dan order\_item\_dataset. Kedua table tersebut di outer join. Kemudian menggunakan tUniqrow, dan memilih order\_id dan order\_item\_dataset yang unique. Kemudian menggunakan feedback\_dataset\_unique dan user\_dataset\_unique sebagai lookup.

Yang terakhir ada feedback\_dataset\_unique. Feedback\_dataset\_unique berasal dari feedback\_dataset yang diproses di Talend menggunakan tUniqrow. Feedback\_dataset\_unique berisi data dengan feedback\_id dan order\_id yang unik.

Lanjut pada schema **datawarehouse**, berasal dari schema **staging**.

Disini terdapat 6 table, antara lain FactOrder, DimCustomer, DimDate, DimPayment, DimProducts, dan DimSeller.



Yang pertama ada table FactOrder. FactOrder berasal dari table join\_order yang diproses di Talend menggunakan tMap.

Yang kedua ada table DimCustomer. DimCustomer berasal dari table user\_dataset\_unique yang ditambah surrogate key.

Yang ketiga ada table DimDate. DimDate berasal dari table generate\_date yang ditambah surrogate key.

Yang keempat ada table DimPayment. DimPayment berasal dari table payment\_dataset\_unique yang diolah datanya menggunakan talend, dengan mereplace payment type yang mengandung “\_” untuk direplace dengan “.”.

Yang kelima ada table DimProducts. DimProducts berasal dari table products\_dataset\_unique yang diolah datanya menggunakan talend, dengan mereplace product category yang mengandung “\_” untuk direplace dengan “ ”.

Yang terakhir ada table DimSeller. DimSeller berasal dari table seller\_dataset\_unique yang ditambah surrogate key.