

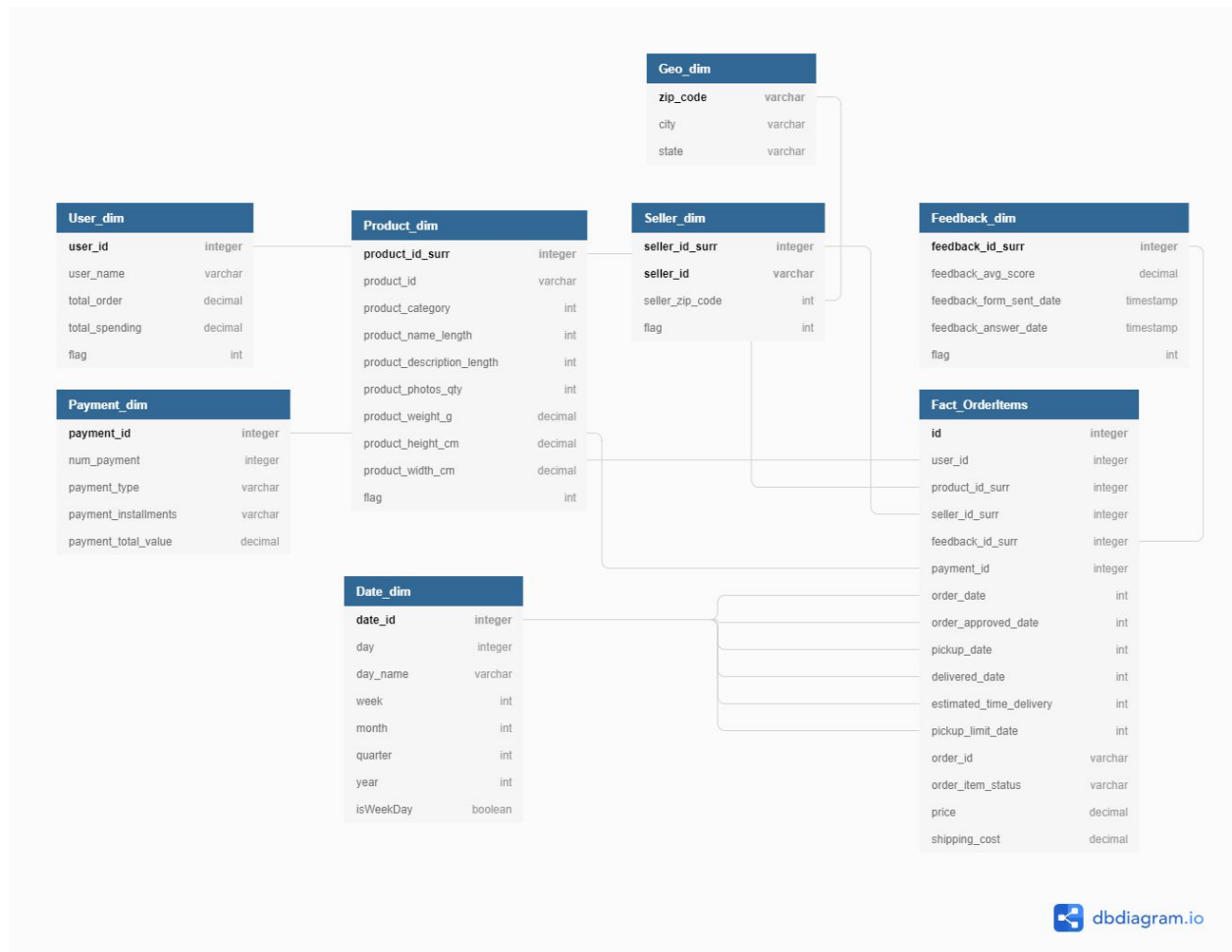
## **DataWarehouse Diagram**

### **Business Need / Requirement**

Berikut merupakan beberapa pertanyaan bisnis yang mendasari pembuatan struktur skema data warehouse :

- Perlu melihat jumlah order setiap harian, mingguan, bulanan, atau tahunan dari :
  - a. Tiap seller
  - b. Tiap segmen customer
  - c. Tiap daerah
- Perbandingan antara jumlah order untuk variasi periode waktu yang berbeda.  
Variasi dapat berupa :
  - Libur vs Tidak Libur
  - Akhir bulan vs Tanggal Biasa
  - Per bulan
- Perbandingan antara jumlah order untuk waktu yang berbeda (per jam setiap harinya)
- Mengetahui trend peningkatan order dari sisi mingguan, bulanan ataupun tahun
- Jenis - jenis produk paling populer per segmen waktu
- Bagaimana segmentasi produk (produk mana yang sering dibeli dengan produk lainnya).
- Bagaimana kualitas customer dari e-commerce (lifetime Valuenya)
- Bagaimanakah trend rating / feedback per order dari sisi mingguan, bulanan ataupun tahun
- Bagaimanakah trend rating / feedback tiap seller dari sisi mingguan, bulanan ataupun tahun
- Bagaimanakah revenue seller dari sisi mingguan, bulanan ataupun tahun

Dari business *need / requirement* tersebut, beserta dengan data yang available, saya menyusun struktur data warehouse dengan skema sebagai berikut :



## Star Schema Design

### Penjabaran Struktur :

#### Dimension Table

Berikut penjabaran tabel dimensi dari perancangan data warehouse :

- Product\_dim

Menggunakan cara SCD tipe II : Adds a new row to a dimension table.

Attribute	Constraint	Deskripsi
prod_id_surr	PK	Surrogate key untuk dimensi user untuk suatu order
product_id		unique product identifier
product_category		root category of product
product_name_length		number of characters of the product name

product_description_length		number of characters of the product description
product_photos_qty		number of product photos
product_weight_g		product weight measure in grams
product_length_cm		product length measure in centimeters
product_height_cm		product height measure in centimeters
product_width_cm		product width measure in centimeters
flag		Flag untuk menandakan latest update pada tabel dimensi

- User\_dim

Menggunakan cara SCD tipe II : Adds a new row to a dimension table.

Attribute	Constraint	Deskripsi
user_id	PK	Surrogate key untuk dimensi user untuk suatu order
user_name		unique identifier of a user
total_order		Aggregate value of total order from this user
total_spending		Aggregate value of user spending
flag		Flag untuk menandakan latest update pada tabel dimensi

- Seller\_dim

Menggunakan cara SCD tipe II : Adds a new row to a dimension table.

Attribute	Constraint	Deskripsi
seller_id_sur r	PK	Surrogate key untuk dimensi seller untuk suatu order
seller_id		seller unique identifier
seller_zip_c ode		zip code
flag		Flag untuk menandakan latest update pada tabel dimensi

- Feedback\_dim

Menggunakan cara SCD tipe I : Overwrites an attribute in a dimension table.

Attribute	Constraint	Deskripsi
feedback_id_surr	PK	Surrogate key untuk dimensi feedback untuk suatu order
feedback_avg_score		Rata - rata skor feedback terhadap suatu order (menjadi 1 dengan order per item)
feedback_form_sent_date		timestamp that survey was sent to the customer.
feedback_form_answer_date		feedback answered timestamp
flag		Flag untuk menandakan latest update pada tabel dimensi

- Payment\_dim

Menggunakan cara SCD tipe II : Adds a new row to a dimension table.

Attribute	Constraint	Deskripsi
payment_id	PK	Surrogate key untuk dimensi payment untuk suatu order
num_payemnt		Jumlah pembayaran yang dilakukan
payment_type		method of payment chosen by the customer
payment_installment		number of installments chosen by the customer
payment_value		total order amount
flag		Flag untuk menandakan latest update pada tabel dimensi

- Date\_dim

Menggunakan cara SCD tipe I : Overwrites an attribute in a dimension table.

Attribute	Constraint	Deskripsi
date_id	PK	Surrogate key untuk mengidentifikasi 1 satuan waktu
day		Tanggal

day_name		Nama hari . cth: Monday, Tuesday, etc
week		Minggu ke - ...
month		Bulan ke ...
quarter		Quarter ke .. (etc : 1,2,3,4)
year		Tahun
isWeekend		Apakah akhir pekan (Jumat - Minggu)

#### - Geo\_dim

Menggunakan cara SCD tipe I : Overwrites an attribute in a dimension table.

Attribute	Constraint	Deskripsi
zip_code	PK	zip_code
city		city
state		state

Sebagai tambahan, terdapat dimensi tambahan yang mungkin dapat diterapkan pada data warehouse. Dimensi ini ditambah sebagai penambah informasi untuk warehouse, namun tidak dapat diolah pada proses transformation akibat kurangnya metadata.

#### - User\_dest\_dim

Menggunakan cara SCD tipe I : Overwrites an attribute in a dimension table.

Attribute	Constraint	Deskripsi
user_name	PK	Unique identifier of user
zip_code	PK	zip_code
num_order	-	Number of order of certain user for this zip_code

### Fact Table

#### **Fact\_OrderItems**

Bentuk *granularity* : Each product in each orders

Dimensi yang digunakan :

- Product\_dim → Produk apa yang diorder
- User\_dim → Siapa yang mengorder produk dalam order

- Seller\_dim → Seller mana user mengorder produk
- Feedback\_dim → Feedback order dari order\_items (jika ada)
- Payment\_dim → Pembayaran terhadap order
- Date\_dim → Kapan order dilakukan

Attribute	Constraint	Deskripsi
id	PK	Surrogate untuk fact table
prod_id_surr	FK	Reference for Product_dim
user_id	FK	Reference for User_dim
seller_id_surr	FK	Reference for Seller_dim
feedback_id_surr	FK	Reference for Feedback_dim
payment_surr_key	FK	Reference for Payment_dim
order_date	FK	Reference for Date_dim
order_approved_date	FK	Reference for Date_dim
pickup_date	FK	Timestamp when it was handled to the logistic partner
delivered_date	FK	actual order delivery date to the customer
estimated_time_delivery	FK	estimated delivery date that was informed to customer at the purchase moment
pickup_limit_date	FK	Shows the seller limit date for handling the order over to the logistic partner.
order_id		Order identifier serial number
order_item_status		Order status (processing, delivered, shipped, etc)
Price		item price
shipping_cost		shipping cost

## **ETL Process Guideline**

### **<Node Process>**

#### **Tahapan Eksplorasi**

Berdasarkan eksplorasi yang dilakukan dengan memanfaatkan Pandas terhadap data source utama (flat files), maka diperoleh beberapa hal penting sebagai berikut :

#### **Poin - poin Penting :**

Dari eksplorasi tersebut, beberapa poin penting yang mendasari tahapan Extract & Transforming sebagai berikut :

#### **Tahapan Extraction: -**

Ekstraksi telah dilakukan oleh pemberi proyek

#### **Tahapan Transforming :**

**<on progress>**

#### **Tahapan Loading :**

**<on progress>**

Secara lengkap, berikut merupakan seluruh struktur data warehouse yang saya buat :

<Gambar lengkap skema (external data source (flat files) → Staging Area → Data Warehouse → Per Subject Analytics>  
<on progress>

### Reference

- Creating date time dimension table  
<https://stackoverflow.com/questions/47150709/how-to-create-a-calendar-table-date-dimension-in-pandas>
- How to update data warehouse using SCD II method  
<https://stackoverflow.com/questions/54481196/what-is-the-process-of-updating-record-in-a-dimension-table-following-scd-method>
- Postgres Datawarehousing Guide  
[https://wiki.postgresql.org/images/3/38/PGDay2009-EN-Datawarehousing\\_with\\_PostgreSQL.pdf](https://wiki.postgresql.org/images/3/38/PGDay2009-EN-Datawarehousing_with_PostgreSQL.pdf)
- ETL with pygrametl  
<http://chrthomsen.github.io/pygrametl/>
- Data Warehousing Guide  
[https://www.ibm.com/support/knowledgecenter/SSGU8G\\_12.1.0/com.ibm.whse.doc/ids\\_ddi\\_360.htm](https://www.ibm.com/support/knowledgecenter/SSGU8G_12.1.0/com.ibm.whse.doc/ids_ddi_360.htm)
- Postgres Trigger SCD Implementation  
<https://dannybeachnau.com/2020/02/05/creating-a-slowly-changing-dimension-in-postgres/>