

1. Data Lake process:

- layer 1 : contains all query from database sources.
- layer 2 : contains all transformation query from layer 1
- layer 3 : contains finalize query from layer 2

Data Warehouse process:

- layer 4 : contains query for dimension table and fact tables
- layer 5 : contains query for flat table and data mart.

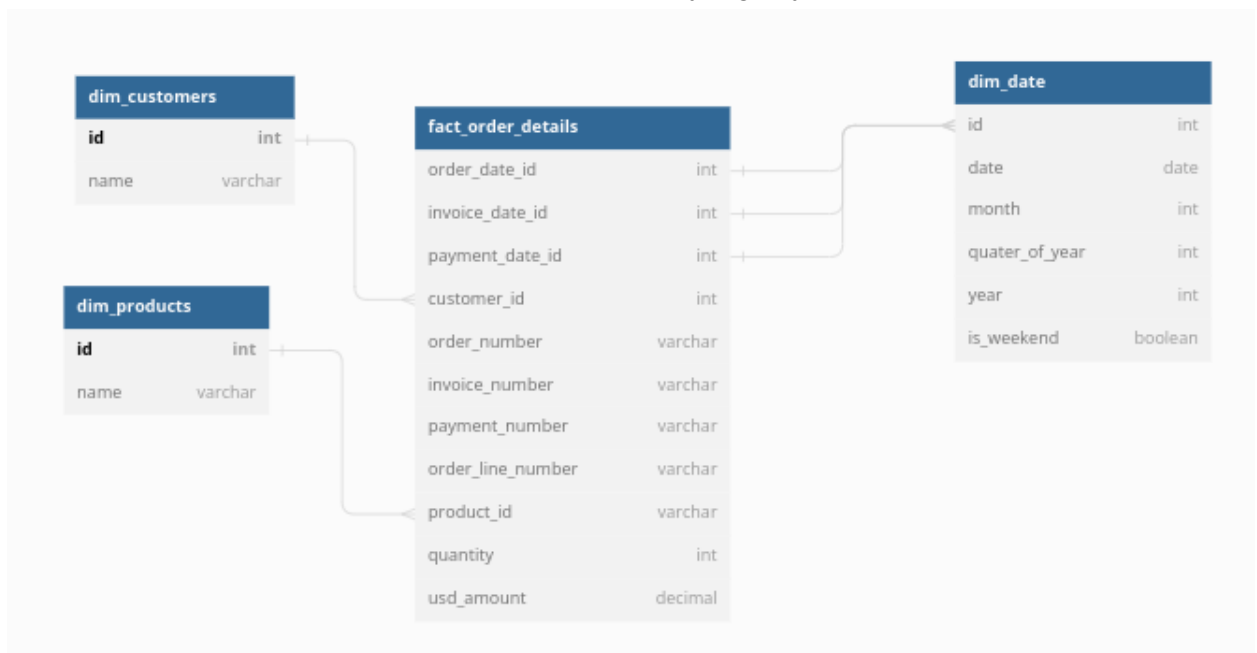
2. Proses data quality check dilakukan diantara layer kedua dan layer ketiga.

File quality check terdapat pada folder layer ketiga. Quality check tersebut melakukan pengecekan data dengan menghitung total data yang masuk berdasarkan tanggal ingestion dilakukan. Data hasil quality check dapat disimpan kedalam sebuah table yang mana table tersebut dapat digunakan untuk monitoring dan alerting.

3. Query contoh penggunaan data warehouse terdapat pada layer kelima. Query data mart tersebut bertujuan untuk memberikan report penjualan dari berbagai time frame yang berbeda. Berikut merupakan contoh sederhana penggunaan data warehouse:

- a. mart_monthly_order_reports.sql
- b. mart_quarterly_order_reports.sql
- c. mart_yearly_order_reports.sql

4. Berikut merupakan desain data warehouse alternatif yang saya buat.



Data warehouse terdiri dari table-table berikut.

1. Fact Table: *fact_order_details*
2. Dimension Tables:
 - a. *dim_customers*
 - b. *dim_products*
 - c. *dim_dates*

Data warehouse menerapkan start schema dimana *fact_order_details* menjadi central dari *dim_customers*, *dim_products*, dan *dim_date*. Dari dimension dan fact table tersebut

akan dibuat menjadi satu table besar berupa flat_orders yang terdapat pada layer kelima.

Perbedaan dari skema data warehouse sebelumnya adalah pada skema alternatif ini pengguna layer kelima bisa lebih leluasa dalam melakukan pengolahan data seperti agregasi. Hasil data olahan user yang menggunakan flat table sebagai source juga bisa langsung disimpan kedalam bentuk table atau views di BigQuery. Flat table bersifat menampung segala data transaksional dan semua atribut dalam satu table sehingga cenderung memiliki data yang cukup besar.

PROS	CONS
<p>User layer kelima dapat dengan leluasa melakukan pengolahan data seperti agregasi.</p> <p>Dependency terhadap data engineer berkurang karena user layer kelima dapat membuat data mart dengan flat table sebagai source sesuai dengan use case yang dihadapi atau langsung menggunakan flat table tersebut sebagai source untuk BI tools.</p>	<p>Flat table cenderung memiliki data yang besar (baik jumlah row maupun jumlah column) sehingga membutuhkan tingkat efisiensi query yang baik apabila ingin mengoptimalkan cost query.</p>