CS 689 Term Project Report
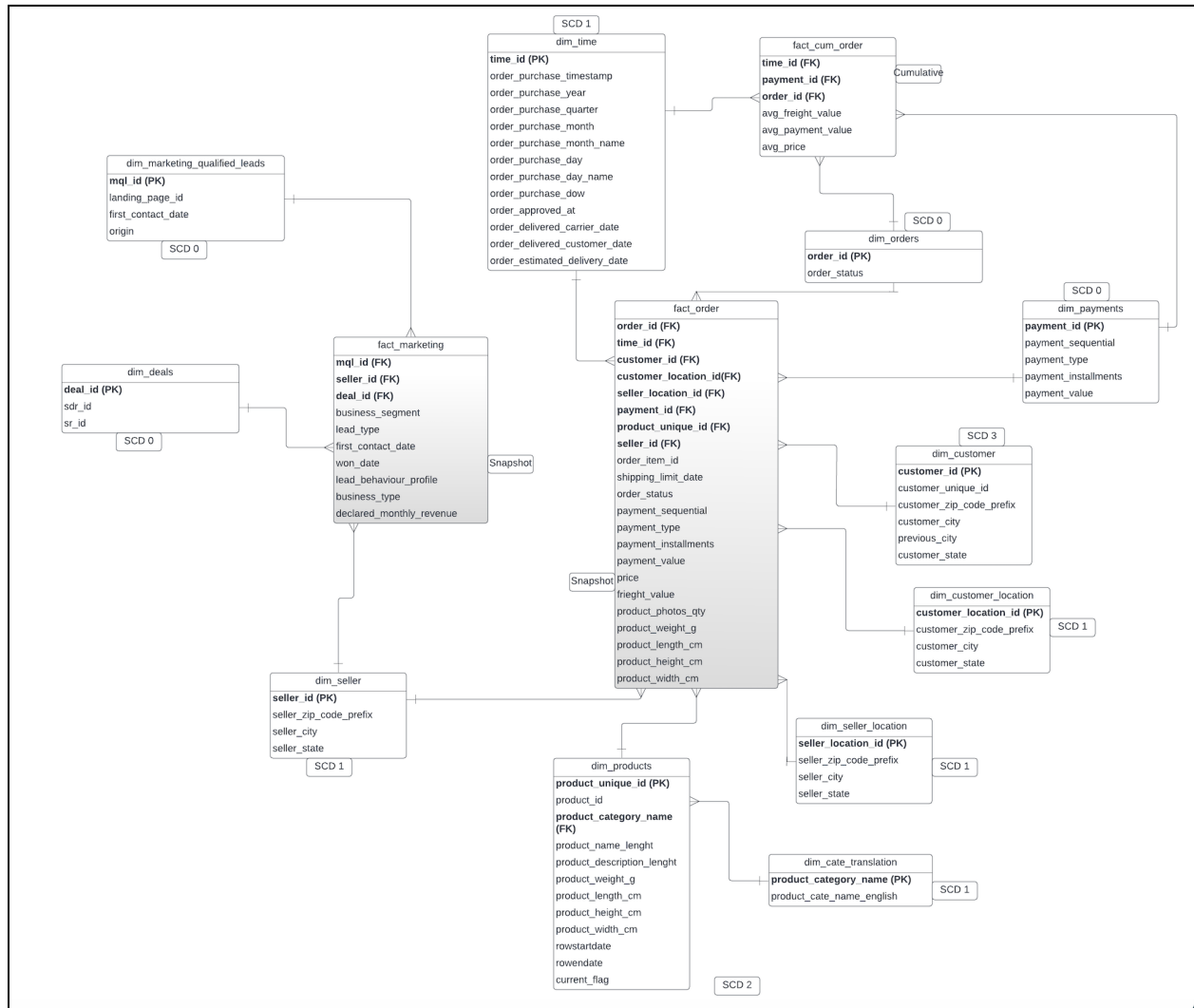
Topic: E-commerce Analysis

- RISHABH SINGH
  SANSIT KABI

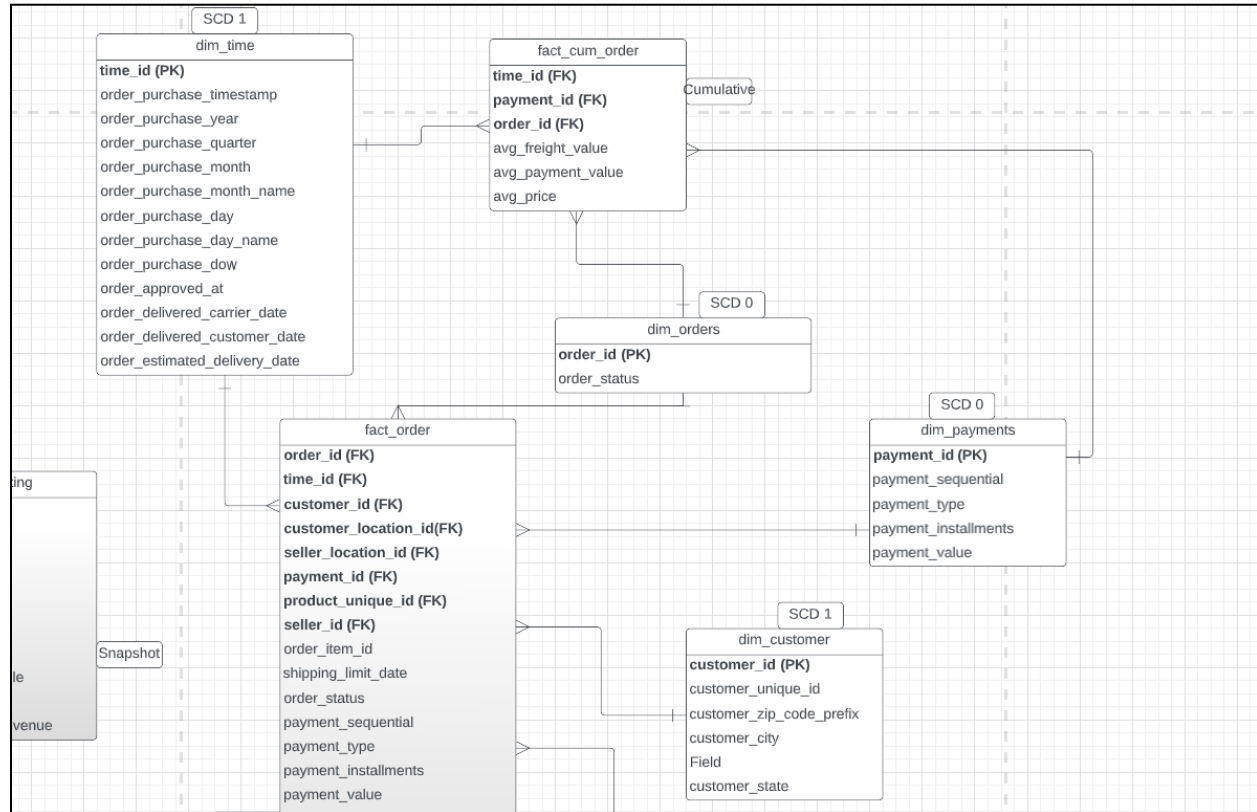# Table of contents

1. Updated Design
2. Changes Based on Feedback
   - Cumulative fact_cum_order
   - Complex Query with Window Function
   - SCD 2 Implementation
   - SCD 3 Implementation

Updated Design

dim_time — SCD 1
- time_id (PK)
- order_purchase_timestamp
- order_purchase_year
- order_purchase_quarter
- order_purchase_month
- order_purchase_month_name
- order_purchase_day
- order_purchase_day_name
- order_purchase_dow
- order_approved_at
- order_delivered_carrier_date
- order_delivered_customer_date
- order_estimated_delivery_date

fact_cum_order — Cumulative
- time_id (FK)
- payment_id (FK)
- order_id (FK)
- avg_freight_value
- avg_payment_value
- avg_price

dim_marketing_qualified_leads — SCD 0
- mql_id (PK)
- landing_page_id
- first_contact_date
- origin

dim_orders — SCD 0
- order_id (PK)
- order_status

dim_payments — SCD 0
- payment_id (PK)
- payment_sequential
- payment_type
- payment_installments
- payment_value

fact_marketing — Snapshot
- mql_id (FK)
- seller_id (FK)
- deal_id (FK)
- business_segment
- lead_type
- first_contact_date
- won_date
- lead_behaviour_profile
- business_type
- declared_monthly_revenue

fact_order — Snapshot
- order_id (FK)
- time_id (FK)
- customer_id (FK)
- customer_location_id(FK)
- seller_location_id (FK)
- payment_id (FK)
- product_unique_id (FK)
- seller_id (FK)
- order_item_id
- shipping_limit_date
- order_status
- payment_sequential
- payment_type
- payment_installments
- payment_value
- price
- frieght_value
- product_photos_qty
- product_weight_g
- product_length_cm
- product_height_cm
- product_width_cm

dim_customer — SCD 3
- customer_id (PK)
- customer_unique_id
- customer_zip_code_prefix
- customer_city
- previous_city
- customer_state

dim_deals — SCD 0
- deal_id (PK)
- sdr_id
- sr_id

dim_customer_location — SCD 1
- customer_location_id (PK)
- customer_zip_code_prefix
- customer_city
- customer_state

dim_seller — SCD 1
- seller_id (PK)
- seller_zip_code_prefix
- seller_city
- seller_state

dim_seller_location — SCD 1
- seller_location_id (PK)
- seller_zip_code_prefix
- seller_city
- seller_state

dim_products — SCD 2
- product_unique_id (PK)
- product_id
- product_category_name (FK)
- product_name_lenght
- product_description_lenght
- product_weight_g
- product_length_cm
- product_height_cm
- product_width_cm
- rowstartdate
- rowendate
- current_flag

dim_cate_translation — SCD 1
- product_category_name (PK)
- product_cate_name_english

Changes Based on Feedback

- ## Cumulative fact_cum_order table

- ## SQL code

```
--Fact_cummulative
create table fact_cum_order(
    time_id int,
    payment_id int,
    order_id character varying(300),
    avg_freight_value float(25),
    avg_payment_value float(25),
    avg_price float(25)
)

alter table fact_cum_order
add constraint fk2_time foreign key(time_id)
references dim_time(time_id);

alter table fact_cum_order
add constraint fk2_payment foreign key(payment_id)
references dim_payments(payment_id);

alter table fact_cum_order
add constraint fk2_order foreign key(order_id)
references dim_order(order_id);
```

```sql
--inserting values in fact_cum_order
insert into fact_cum_order(time_id,payment_id,order_id,avg_freight_value,avg_payment_value,avg_price)
    select dim_time.time_id, dim_payments.payment_id,temp_order.order_id,
            avg(temp_order.freight_value) as avg_freight_value,
            avg(dim_payments.payment_value) as avg_payment_value,
            avg(temp_order.price) as avg_price
    from temp_order
    join dim_time on temp_order.order_purchase_timestamp = dim_time.order_purchase_timestamp
    join dim_payments on temp_order.payment_sequential = dim_payments.payment_sequential and
                temp_order.payment_type = dim_payments.payment_type and
                temp_order.payment_installments = dim_payments.payment_installments and
                temp_order.payment_value = dim_payments.payment_value
    group by 1,2,3
    order by 3

--data check
select * from fact_cum_order limit 5;
select count(*) from fact_cum_order;
```

```sql
--data check
select * from fact_cum_order limit 5;
select count(*) from fact_cum_order;
```

Data Output   Messages   Notifications

	time_id integer	payment_id integer	order_id character varying (300)	avg_freight_value double precision	avg_payment_value double precision	avg_price double precision
1	50876	237	00010242fe8c5a6d1ba2dd792cb162...	13.29	72.19	58.9
2	88111	1946	00018f77f2f0320c557190d7a144bdd3	19.93	259.83	239.9
3	94211	58560	000229ec398224ef6ca0657da4fc703e	17.87	216.87	199
4	70181	19656	00024acbcdf0a6daa1e931b038114c...	12.79	25.78	12.99
5	97827	417	00042b26cf59d7ce69dfabb4e55b4fd9	18.14	218.04	199.9

- Complex Query with Window Function

```
692
693  -- Window function rank()--ranking payment_value from cities with customers
694
695  select dim_customer_location.customer_city,fact_order.customer_id,dim_payments.payment_value,
696         rank() over(order by avg(dim_payments.payment_value) desc)city_rank
697         from fact_order
698         join dim_customer_location on fact_order.customer_location_id = dim_customer_location.customer_location_id
699         join dim_payments on fact_order.payment_id = dim_payments.payment_id
700         group by 1,2,3
701
702
```

Data Output   Messages   Notifications

	customer_city character varying (100)	customer_id character varying (300)	payment_value double precision	city_rank bigint
1	rio de janeiro	1617b1357756262bfa56ab541c47bc16	13664.08	1
2	vila velha	ec5b2ba62e574342386871631fafd3fc	7274.88	2
3	campo grande	c6e2731c5b391845f6800c97401a43a9	6929.31	3
4	vitoria	f48d464a0baaea338cb25f816991ab1f	6922.21	4
5	marilia	3fd6777bbce08a352fddd04e4a7cc8f6	6726.66	5
6	divinopolis	05455dfa7cd02f13d132aa7a6a9729c6	6081.54	6
7	araruama	df55c14d1476a9a3467f131269c2477f	4950.34	7

- SCD 2 Implementation

```
--- scd2 trigger
--procedure for change
create or replace function new_date()returns trigger
 LANGUAGE plpgsql as
$$
begin
     update dim_products
         set current_flag = 'N',
             rowstartdate = product_staging.change_date
               from product_staging
             where product_staging.product_id = dim_products.product_id and
                   dim_products.current_flag = 'Y';
    insert into dim_products (product_id,product_category_name,product_name_lenght,product_description_lenght
                        ,product_photos_qty,product_weight_g,product_length_cm,product_height_cm
                        ,product_width_cm,product_category_name_english,current_flag, rowstartdate, rowenddate)
     select distinct product_id,product_category_name,product_name_lenght,product_description_lenght
                        ,product_photos_qty,product_weight_g,product_length_cm,product_height_cm
                        ,product_width_cm,product_category_name_english, 'Y', change_date, '9999-12-31'::date
     from product_staging where product_staging.change_date !='9999-12-31';
     return null;
end$$;
```

```
5
6   --trigger
7   create or replace trigger scd_type
8       after insert
9       on product_staging
0       for each row
1       execute procedure new_date();
2
3
4   -- inserting values in product_staging
5
6   insert into product_staging(product_id,product_category_name,product_name_lenght,product_description_lenght
7                               ,product_photos_qty,product_weight_g,product_length_cm,product_height_cm
8                               ,product_width_cm,product_category_name_english,Change_date)
9   values('03b29ca1beec2e03e15a3b980c1505ed','perfumaria',34,250,2,800,27,5,20,
0          'perfumery',current_timestamp);
1          -- changing
2
```

- After insert in product_staging

```
743
744  -- the above trigger should create a new record with the history saved
745  select * from product_staging where product_id = '03b29ca1beec2e03e15a3b980c1505ed'
746  select * from dim_products where product_id = '03b29ca1beec2e03e15a3b980c1505ed'
747
```

Data Output   Messages   Notifications

	product_id character varying (300)	product_category_name character varying (100)	product_name_lenght double precision	product_description_lenght double precision	product_photos_qty double precision	product_weight_g double precision	product_length_cm double precision	product_height_cm double precision	product_width_cm double precision	product_category_name_e character varying (100)
1	03b29ca1beec2e03e15a3b980c1505ed	perfumaria	43	656	3	100	18	7	11	perfumery
2	03b29ca1beec2e03e15a3b980c1505ed	perfumaria	34	250	2	800	27	5	20	perfumery

```
743
744  -- the above trigger should create a new record with the history saved
745  select * from product_staging where product_id = '03b29ca1beec2e03e15a3b980c1505ed'
746  select * from dim_products where product_id = '03b29ca1beec2e03e15a3b980c1505ed'
747
```

Data Output   Messages   Notifications

	product_id character varying (300)	product_category_name character varying (100)	product_name_lenght double precision	product_description_lenght double precision	product_photos_qty double precision	product_weight_g double precision	product_length_cm double precision	product_height_cm double precision	product_width_cm double precision	product_category_name_e character varying (100)
1	03b29ca1beec2e03e15a3b980c1505ed	perfumaria	43	656	3	100	18	7	11	perfumery
2	03b29ca1beec2e03e15a3b980c1505ed	perfumaria	34	250	2	800	27	5	20	perfumery

- Changes in dim_products

```
743
744  -- the above trigger should create a new record with the history saved
745  select * from product_staging where product_id = '03b29ca1beec2e03e15a3b980c1505ed'
746  select * from dim_products where product_id = '03b29ca1beec2e03e15a3b980c1505ed'
747
```

Data Output   Messages   Notifications

	product_unique_id [PK] integer	product_id character varying (300)	product_category_name character varying (100)	product_name_lenght double precision	product_description_lenght double precision	product_photos_qty double precision	product_weight_g double precision	product_length_cm double precision	product_height_cm double precision	product_width_cm double precision	p c
1	129318	03b29ca1beec2e03e15a3b980c1505ed	perfumaria	34	250	2	800	27	5	20	p
2	129315	03b29ca1beec2e03e15a3b980c1505ed	perfumaria	43	656	3	100	18	7	11	p

```
        'perfumery',current_timestamp);
        -- changing


-- the above trigger should create a new record with the history saved
select * from product_staging where product_id = '03b29ca1beec2e03e15a3b980c1505ed'
select * from dim_products where product_id = '03b29ca1beec2e03e15a3b980c1505ed'
```

Data Output   Messages   Notifications

...e	product_name_lenght double precision	product_description_lenght double precision	product_photos_qty double precision	product_weight_g double precision	product_length_cm double precision	product_height_cm double precision	product_width_cm double precision	product_category_name_english character varying (100)	current_flag character (1)	rowstartdate date	rowenddate date
	34	250	2	800	27	5	20	perfumery	Y	2022-12-13	9999-12-31
	43	656	3	100	18	7	11	perfumery	N	9999-12-31	9999-12-31

- SCD 3 Implementation

```
-- scd 3 on dim_customer

alter table order_staging
add change_date date default '9999-12-31';
alter table dim_customer
add present_city character varying(100) default 'Unknown' ;

--procedure for change
create or replace function scd_3newtrigg()returns trigger
 LANGUAGE plpgsql as
$$
begin
      update dim_customer
          set previous_city = dim_customer.customer_city,
             present_city = order_staging.customer_city
          from order_staging
            where order_staging.customer_id = dim_customer.customer_id and
                   order_staging.change_date != '9999-12-31';
      return null;
end$$;
```

```
--trigger
create or replace trigger scd_3
    after update
    on order_staging
    for each row
    execute procedure scd_3newtrigg();

--changing in order_staging
update order_staging
set customer_city = 'san paulo',change_date = current_timestamp
    where customer_id = 'd67b6cca5a87299f711a6961f579fe67'


-- the above trigger should create value in previous_city with the history saved
select * from order_staging where customer_id = 'd67b6cca5a87299f711a6961f579fe67'
select * from dim_customer where customer_id = 'd67b6cca5a87299f711a6961f579fe67'
```

–update on order_staging

```
788
789  -- the above trigger should create value in previous_city with the history saved
790  select * from order_staging where customer_id = 'd67b6cca5a87299f711a6961f579fe67'
791  select * from dim_customer where customer_id = 'd67b6cca5a87299f711a6961f579fe67'
```

Data Output   Messages   Notifications

ght_value ble precision	customer_unique_id character varying (300)	customer_zip_code_prefix character varying (100)	customer_city character varying (100)	customer_state character varying (100)	seller_zip_code_prefix character varying (100)	seller_city character varying (100)	seller_state character varying (100)	payment_sequential double precision	p
14.28	aea90564d6f09ae11bf936f55ed49d72	82030	san paulo	PR	9350	maua	SP	1	

– update on dim_customer

customer_id [PK] character varying (300)	customer_unique_id character varying (300)	customer_zip_code_prefix character varying (100)	customer_city character varying (100)	customer_state character varying (100)	previous_city character varying (100)	present_city character varying (100)
d67b6cca5a87299f711a6961f579fe67	aea90564d6f09ae11bf936f55ed49d72	82030	curitiba	PR	curitiba	san paulo