Data loading

Load all the five sets of files provided into corresponding snowflake tables. (note: order_product_prior and order_products_train should be loaded into a table called order_product)

ER diagram

Draw an ER diagram for the five tables you created, you can use online tools like Lucidchart (https://www.lucidchart.com/pages/)

SQL design

1. Create a table called **order_products_prior** by using the last SQL query you created from the previous assignment. It should be similar to below (note you need to replace the s3 bucket name “imba” to yours own bucket name):

```
CREATE TABLE order_products_prior AS
     (SELECT a.*,
          b.product_id,
          b.add_to_cart_order,
          b.reordered
     FROM   orders a
          JOIN order_products b
          ON a.order_id = b.order_id
     WHERE  a.eval_set = 'prior')
```

2. Create a SQL query (user_features_1). Based on table **orders**, for each user, calculate the max order_number, the sum of days_since_prior_order and the average of days_since_prior_order.

3. Create a SQL query (user_features_2). Similar to above, based on table **order_products_prior**, for each user calculate the total number of products, total number of distinct products, and user reorder ratio(number of reordered = 1 divided by number of order_number > 1)

4. Create a SQL query (up_features). Based on table **order_products_prior**, for each user and product, calculate the total number of orders, minimum order_number, maximum order_number and average add_to_cart_order.

5. Create a SQL query (prd_features). Based on table **order_products_prior**, first write a sql query to calculate the sequence of product purchase for each user, and name it product_seq_time (For example, if a user first time purchase a product A, mark it as 1. If it’s the second time a user purchases a product A, mark it as 2). Below are some examples:

User_id	Order_number	Product_id	Product_seq_time	….
123	1	Egg	1	
123	1	Apple	1	
123	1	Fish	1	
123	2	Egg	2	
123	2	Mop	1	
123	2	Banana	1	
123	3	Egg	3	
123	3	Orange	1	
123	3	Fish	2	
123	3	Salad	1	

Then on top of this query, for each product, calculate the count, sum of reordered, count of product_seq_time = 1 and count of product_seq_time = 2.