

NYU Tandon School of Engineering
CS-6083 Sections A and INET, Fall 2024
Prof Frankl

If you opt for doing a mini-project for the end of semester assessment, choose one of these three options: Option 1 deals with enhancing the design of the course project ; Options 2 and 3 deal with analyzing performance of executing a given query (Option 2) or a bulk insertion (Option 3) with and without indexes on a real-world dataset, which is provided.

Mini-project Requirements, Submission, and Presentation:

You may consult your course notes, text-book, and any published material. *You may discuss the general ideas of how to approach these with others, but you should develop your solutions individually and present them individually.*

You should make a short (about 5 min) video in which you explain what you did, show your results, and mention any lessons learned. At the start of the video, you should introduce yourself with your camera on and show your ID close enough to the camera that it's readable.

After that, you'll probably want to switch to screen-sharing as you explain your work. You can do this using zoom or NYU stream and then share the link with Prof Frankl. In addition to the link to the recording, you should hand in the documents you use in your presentation (and relevant scratch work, if any) via Gradescope.

Mini Project Options 1: Enhance Welcome Home Design

In this mini-project option, you'll add to the design of the WelcomeHome database application that you developed in the course project, compare ER design to normalization approaches, and discuss plans for some of the queries.

WelcomeHome is expanding to many cities around the country. There are now branches. Each branch has a branchID and some additional information, such as its address. Within each branch, the (roomNum, shelfNum) is unique, but the same (roomNum, shelfNum) can occur in different branches. In addition, a person is designated as the manager of each branch. (Ideally this should be a staff member, but you may allow it to be any Person in your design, and assume this will be enforced by the application code.) This expansion of WelcomeHome affects several aspects of the design. In this problem, you'll consider them from two points of view and compare.

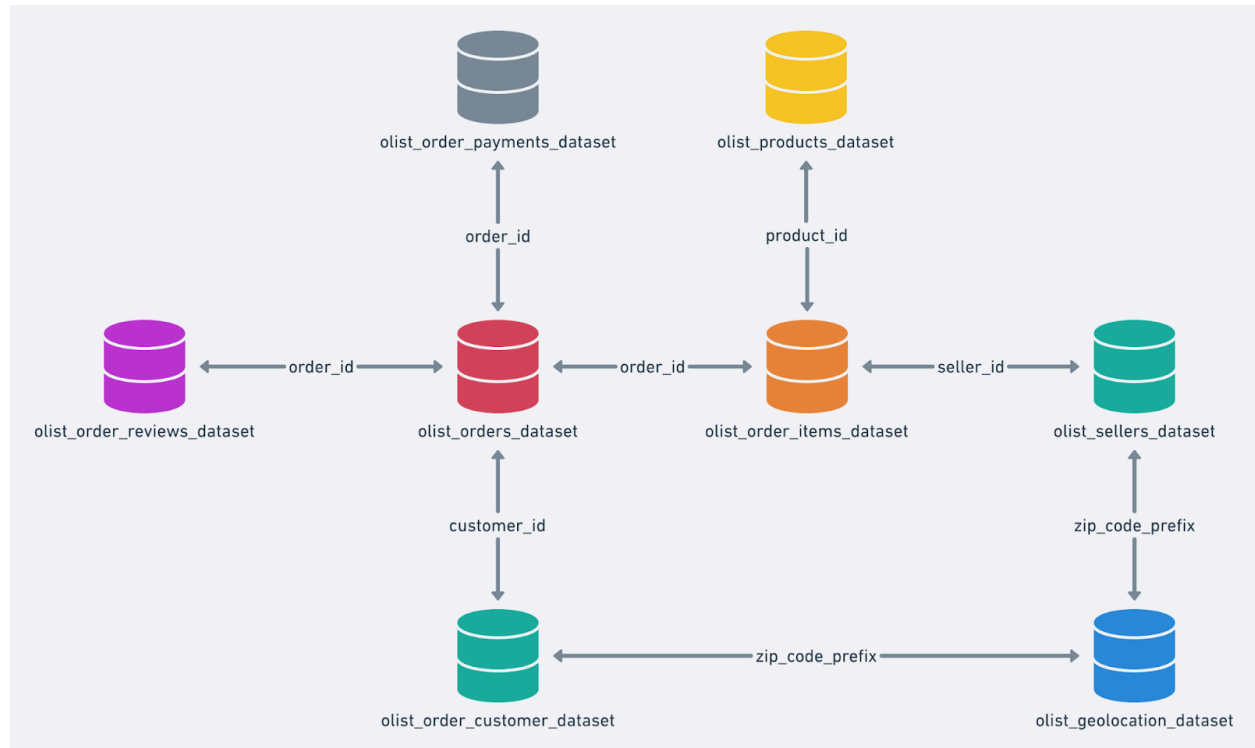
1. Add all of the attributes of the branch to the Location schema; identify at least one non-trivial functional dependency that results from this change, and decompose the schema into BCNF. Also add any schemas or attributes and constraints needed to record who is the manager of which branch.
2. Starting with the original ER diagram, modify to incorporate an entity set representing Branches. You should convert at least one of the existing entity sets to a weak entity set, identify its discriminator, and add the identifying relationship set. Also add any entity sets, relationship sets (with cardinality/participation constraints), or attributes needed to record who is the manager of which branch.
3. Derive a schema from the ER diagram in (b) and compare it to your schema from (a)
4. Briefly discuss other changes that you might want to make, for example to associate people with branches and decisions you'd need to make (or to discuss with the stake-holders) about this.

Mini Project Options 2 and 3: Performance Analysis Options

Dataset Overview

The Brazilian E-commerce Public Dataset by Olist provides real-world e-commerce transaction data from Brazil's largest department store, containing 100,000 orders from 2016 to 2018.

Dataset Structure



Dataset Setup Instructions

1. Data Acquisition

1. Download the dataset from Kaggle: [Brazilian E-commerce Public Dataset](#)
2. Extract the ZIP file to get the CSV files

2. Database Setup

Choose either MySQL or PostgreSQL for this project. Create the required tables using the table definitions present in the CSV File headers with required foreign keys

3. Data Import

MySQL: Use MySQL Workbench's Table Data Import Wizard

- Right-click on your schema
- Select "Table Data Import Wizard"
- Choose the corresponding CSV file
- Map columns correctly

PostgreSQL: Use pgAdmin's Import/Export data tool

- Right-click on your table
- Select "Import/Export"
- Configure CSV import settings

Option 2 : Query Performance Optimization

Problem Statement

The e-commerce platform is experiencing performance issues with their customer order history feature, affecting user experience and system stability.

Current Issues

1. Slow page load times
2. System unresponsiveness during peak hours

Baseline Query :

```
SELECT
o.order_id,
o.order_status,
o.order_purchase_timestamp,
p.product_id,
p.product_category_name,
oi.price,
oi.freight_value
FROM
olist_orders_dataset o
JOIN olist_order_items_dataset oi ON o.order_id = oi.order_id
```

```
JOIN olist_products_dataset p ON oi.product_id = p.product_id
JOIN olist_customers_dataset c ON o.customer_id = c.customer_id
WHERE
c.customer_unique_id = '8d50f5eadf50201ccdcedfb9e2ac8455' -- sample customer id
ORDER BY o.order_purchase_timestamp DESC;
```

a. Document Current Performance

- Run EXPLAIN ANALYZE on the baseline query
- Record:
 - Execution time
 - Number of rows processed
- Save the execution plan

b. Optimize Query

- Identify bottlenecks
- Create appropriate indexes
- Measure new performance using EXPLAIN ANALYZE on same query

Option 3 : Write Performance Analysis

1. Setup Test Environment

- Create a backup of the olist_order_items_dataset table to preserve original data
- Create two indexes on order_id and product_id column

2. Test Write Performance With Indexes

- Clear all existing data (Truncate) from the olist_order_items_dataset table
- Run EXPLAIN ANALYZE and perform a bulk insert using the backup data
- Record:
 - Execution time
 - Table size including indexes

3. Test Write Performance Without Indexes

- Drop all indexes from the olist_order_items_dataset table

- Clear the table data again
- Run EXPLAIN ANALYZE and perform the same bulk insert using the backup data
- Record :
 - Execution time
 - Final table size

d. Justify the difference in the execution time with and without indexes

Submission Guidelines: Record a video demonstrating and explaining the execution.