****

#### ALY6030-Mod1: Assignment 1- Tech crunch

#### Integrated Data Warehousing and SQL

Instructor: [Kayal Chandrasekaran](https://northeastern.instructure.com/courses/140410/users/121784)

2nd June 2023

Submitted By: Srishti Singh

**Problem 1. Normalization**

**1) What is a good choice for a primary key here? In contrast, give an example of an attribute (or composite) that would *not* be a valid primary key.**

**Answer:** The "fund\_id" column is a good choice for a primary key in this dataset because it uniquely identifies each funding event and can serve as a unique identifier for each row. To confirm that it satisfies the definition of a primary key, you can sort the data by the "fund\_id" column in Excel and ensure that there are no duplicate values.

On the other hand, the "company" column would not be a valid primary key because it can have duplicate values. Since multiple funding events can be associated with the same company, it does not uniquely identify each row in the table.

**2) Does the table satisfy 1NF? Why or why not?**

The table satisfies 1NF (First Normal Form) because each column contains atomic values, meaning that they do not contain multiple values or repeating groups. Each column represents a single attribute, fulfilling the requirements of 1NF.

**3) Does the table satisfy 2NF? Why or why not?**

In the dataset, the "numEmps" column represents the number of employees, and it is associated with the "company" column, not the "fund\_id" column (primary key). This implies that the number of employees is dependent on the company identification rather than the funding event identification.

In the context of normalization, if a non-key column (in this case, "numEmps") depends on only a part of the primary key (in this case, "fund\_id"), it indicates a partial dependency. To satisfy 2NF, the dataset should be decomposed into separate tables to remove partial dependencies.

Therefore, based on the provided dataset, it appears that the statement about **the table not satisfying 2NF** due to the dependency of the number of employees on the company ID rather than the fund ID is accurate.

**4) Does the table satisfy 3NF? Why or why not?**

The table doesn’t satisfy 3NF (Third Normal Form) as it doesn’t satisfy 2NF and there are transitive dependencies between non-key attributes. There are transitive dependencies in the given table, meaning that each non-key attribute doesn’t depend only on the primary key. For example: columns like number of employees seem to be dependent on company rather than fund \_id.

5)

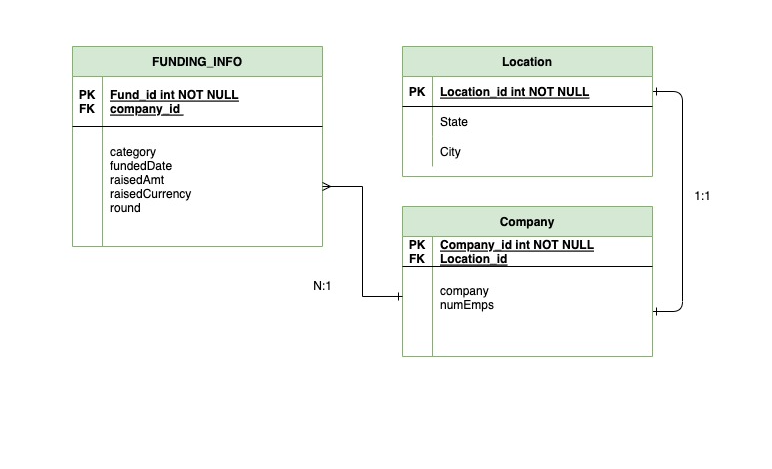
**To attain the 3NF in the provided Dataset, we can define the relationships between the tables as follows:**

**One-to-One Relationship:**

Location table and Company table: Each company is associated with one location, and each location is linked to one company. This is a one-to-one relationship because each company can have only one location, and each location can be associated with only one company.

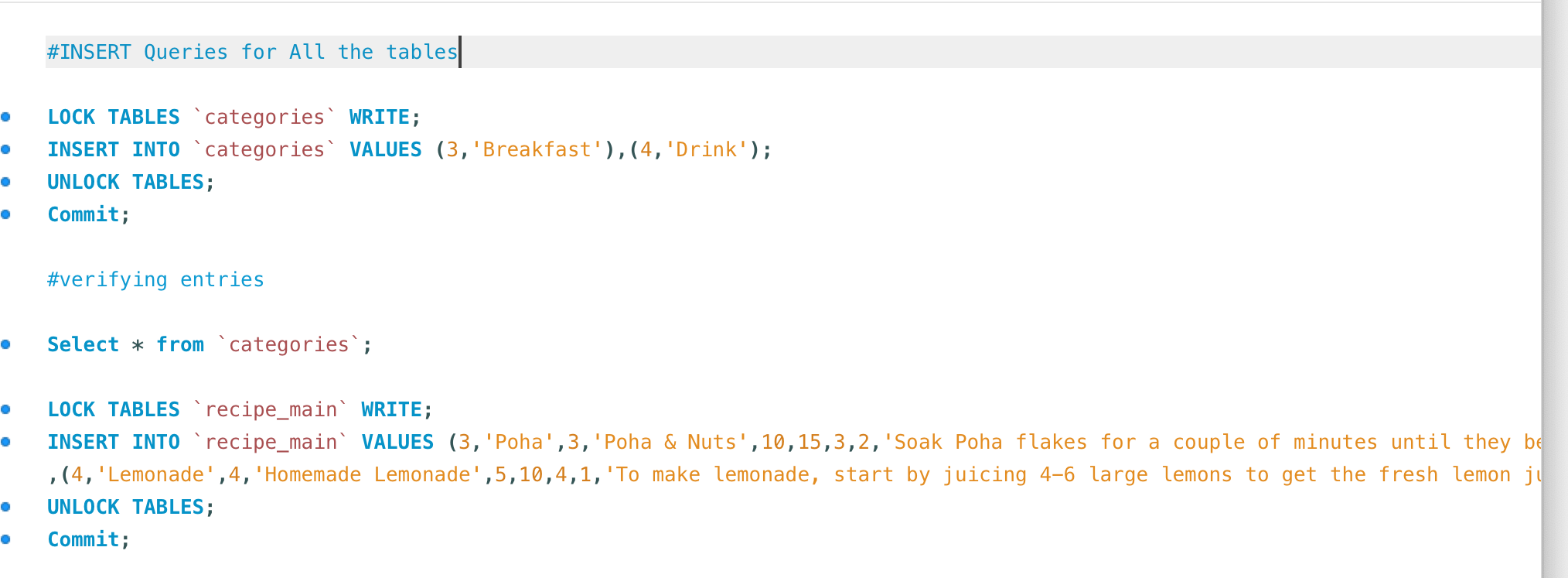
**One-to-Many Relationship:**

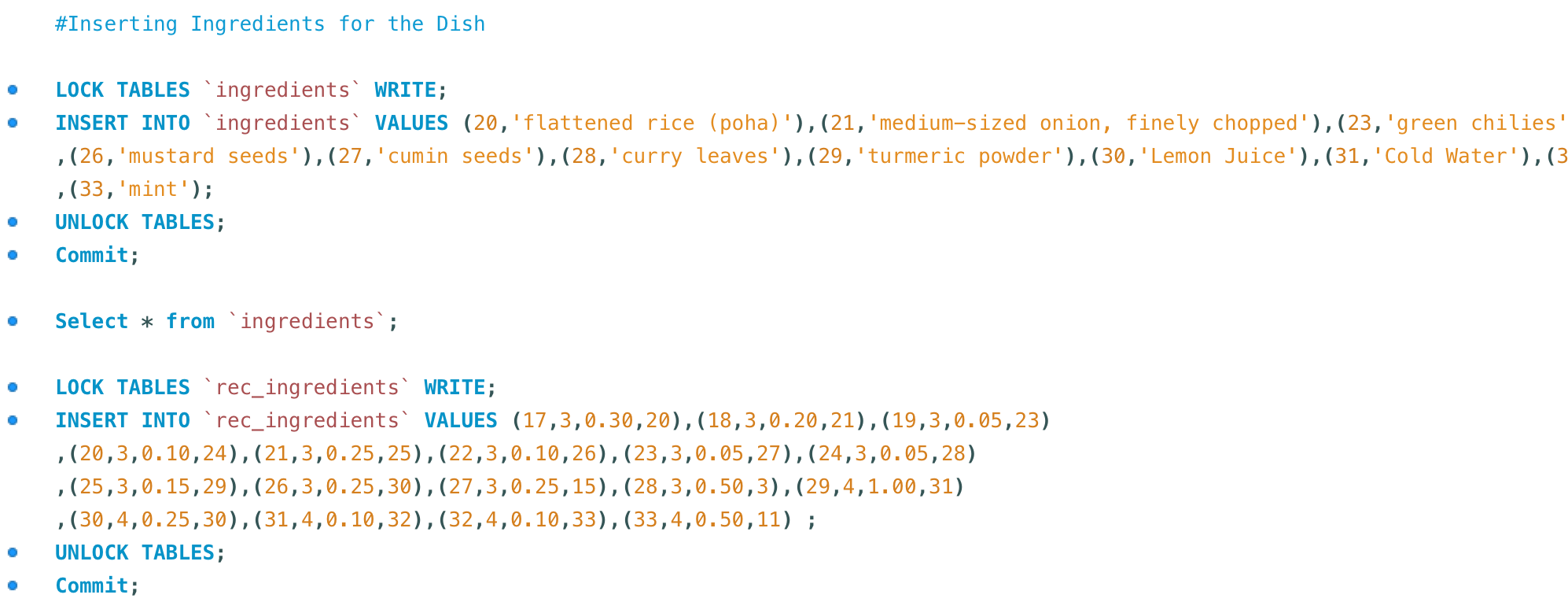
Company table and Fund table: A company can have multiple funds, but each fund is associated with only one company. This is a one-to-many relationship because one company can have multiple funds, but each fund can be linked to only one company.



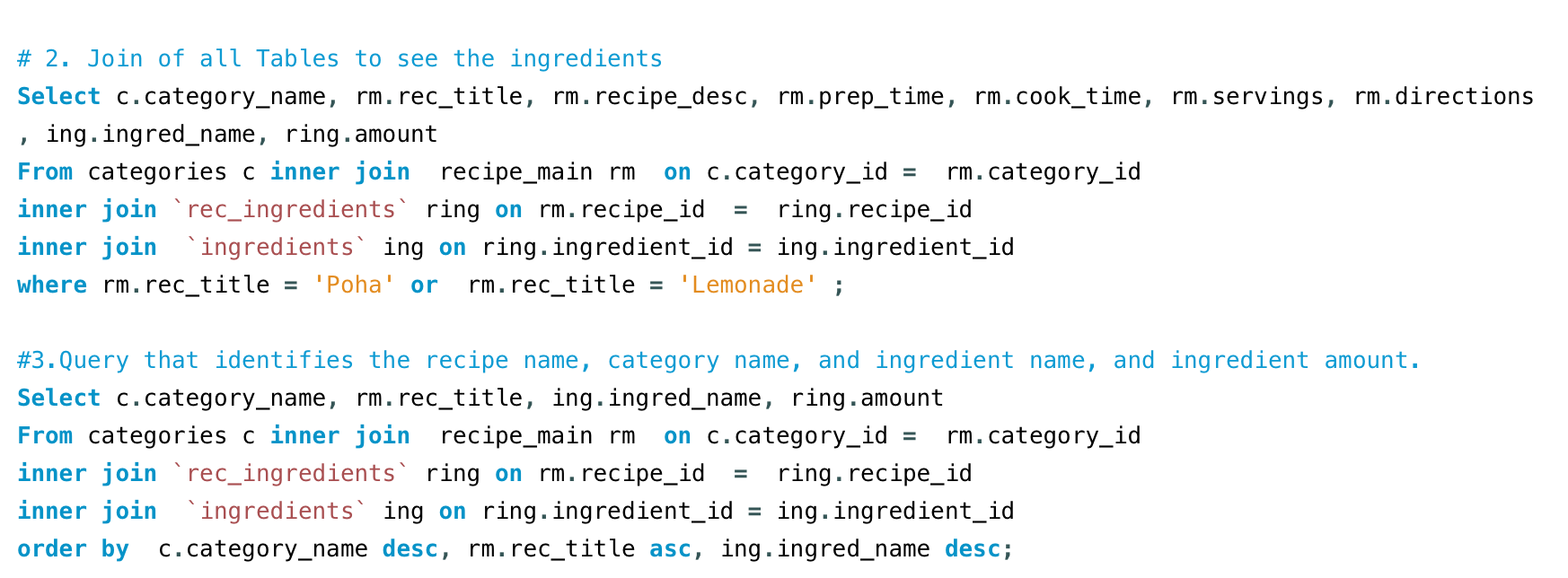
**Problem 2. Case study (adapted from Comeau, Chapter 9)**

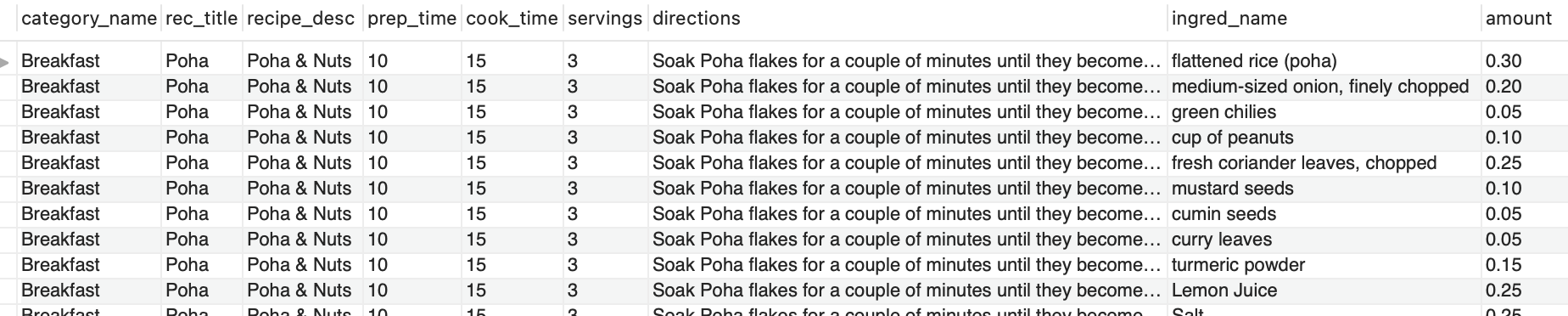
1. **Inserting entries into table for dishes and drinks i.e., Poha and lemonade**



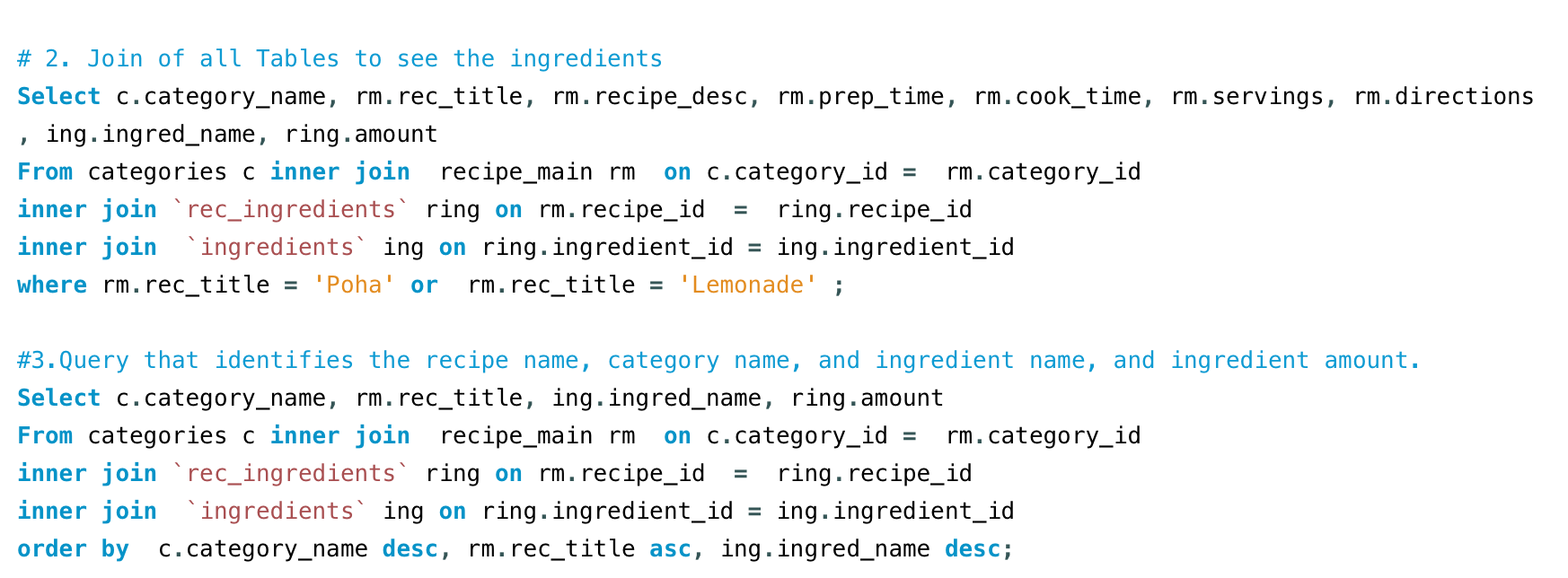


1. Query showing all relevant information from all four of the tables





1. **Final Query Result**

****

****