

Assignment 1 — Tech Crunch

Heejae Roh (ID 002795339)

College of Professional Studies, Northeastern University

ALY 6030: Data Warehousing & SQL

Professor Venkata Duvvuri

November 15, 2023

Table of Contents

Introduction.....	3
Content.....	3
Problem 1. Normalization.....	3
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.....	3
2) Does the table satisfy 1NF? Why or why not?	3
3) Does the table satisfy 2NF? Why or why not?	4
4) Does the table satisfy 3NF? Why or why not?	4
5) Sketch a proposed Entity-Relationship diagram that would bring this dataset into 3NF.	
Problem 2. Case study (adapted from Comeau, Chapter 9).....	5
1) Once you run the sql code you'll notice that each table has data populated for two recipes, Chicken Marsala and Absolute Brownies.	5
2) Write only one SQL query that returns all information on only the two new recipes you inserted from all the tables you created in step 1 above.	8
3) Write a SELECT query that identifies the recipe name, category name, and ingredient name, and ingredient amount. No other variables should be included. .	9
Conclusion	10
References.....	11

Introduction

In this project, I will learn about table normalization in databases. Let's find out what 1NF (First Normal Form), 2NF (Second Normal Form), and 3NF (Third Normal Form) are using TechCrunch data. Secondly, I will add two recipes to the SQL 'recipes' database using MySQL Workbench. I will utilize various SQL queries and use the JOIN and ORDER BY functions.

Content

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.**

Fund_id is a primary key. A primary key is a unique identifier for each record in a table (herovired, 2023). In other words, a primary key is unique within a table and is used to identify each row. A table can have only one primary key. The techcrunch csv file contains: fund_id, company, numEmps, category, city, state, fundedDate, raisedAmt, raisedCurrency, round. There are a total of 10 columns, of which Fund_id is a unique number with different numbers from 1 to 523, and can be used to identify each row.

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

Yes. The table satisfies 1NF (First Normal Form).

1NF is a table where columns have unique records (ShareTheSameSky, 2021). Because every value contains only one value or item. In other words, there is no cell with multiple values connected by ',' (or other symbol). For example, looking at the 'company' column, each row

contains only one company. There is no table that contains two or more values, such as Twitter or Facebook. All columns are configured to satisfy this condition.

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

No. The table does not satisfies 2NF (Second Normal Form).

To satisfy 2NF, the table must be fully dependent to Primary Key. In this case the table has partial dependency. For example, city may be completely dependent on state. If the company does not include each branch, numEmps (which represent the number of employees), city, state can be fully dependent to 'company'. We can make new table whose primary key is company.

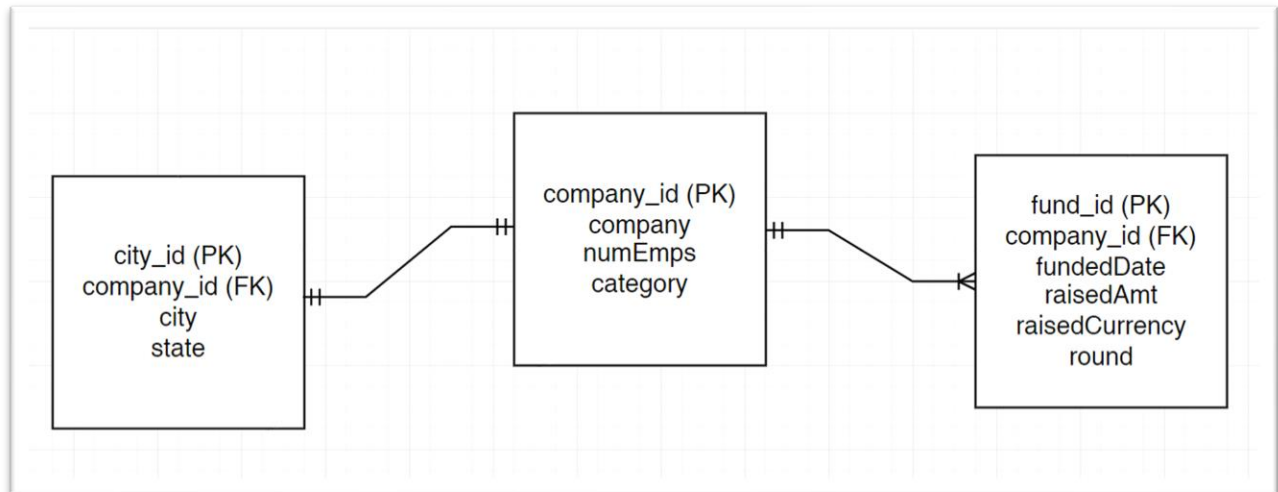
4) Does the table satisfy 3NF? Why or why not?

No. The table does not satisfies 3NF (Third Normal Form).

To satisfy 3NF, A relation that is in First and Second Normal Form and in which no non-primary-key attribute is transitively dependent on the primary key (Upadhyay, M). All its attributes are not only fully functionally dependent on the primary key but also non-transitively dependent, meaning no non-prime attribute depends on another non-prime attribute. In other words, there should be no functional dependency between any two non-key attributes (Jay, 2023).

In order to satisfy 3NF, it must be non-transitively dependent, meaning no non-prime attribute depends on another non-prime attribute. There is transitive dependency between attributes like company and others. For example, city is dependent on state. city is dependent on company. numEmps is dependent on company. We can manage the table separately to satisfy 2NF and 3NF.

- 5) Sketch a proposed Entity-Relationship diagram that would bring this dataset into 3NF. If you answered “yes” to (4), for example, your ERD would just be the raw data table with no changes. If however your ERD requires multiple tables to be in 3NF, you should draw all relationships between them and indicate their type (one-to-one, one-to-many, etc.)



One company_id can have one or multiple fund_ids (one-to-many). One company_id can have one city_id (one-to-one).

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

- 1) Once you run the sql code you'll notice that each table has data populated for two recipes, Chicken Marsala and Absolute Brownies.

```

98 • SELECT *
99 FROM recipe_main;
100

```

recipe_id	rec_title	category_id	recipe_desc	prep_time	cook_time	servings	difficulty	directions
1	Chicken Marsala	1	Chicken and mushrooms	20	20	4	2	Flatten chicken breasts to 1/4 inch. Place flour in...
2	Absolute Brownies	2	Rich, chocolate brownies	25	35	16	2	Preheat oven to 350 degrees F (175 degrees C...

Guideline:

- Use the INSERT INTO statement to insert new information about **two (2)**
completely new recipes of your choosing into the database.
1. -- Lasagna and Korean short ribs are added in 'recipe_main' column INSERT INTO
`recipe_main` VALUES (3,'Easy Lasagna',3,'Cheese and meat sauce',10,30,8,1,'poon
meat sauce on the bottom of a lightly greased casserole dish. Add 4 boiled lasagna
noodles (or use no-boil or fresh noodles). Spread 1/3 of the ricotta cheese mixture on
top. Add 1.5 cups of meat sauce. Top with mozzarella cheese. Cover and bake at 375°
for 30 minutes. Remove cover and bake for 15 more minutes. Broil at the end if
desired. Let it rest for 15 minutes prior to serving with Garlic Bread With Cheese.');
 2. INSERT INTO `recipe_main` VALUES (4,'Korean short ribs',4,'Beef short ribs and
soy sauce',5,10,4,1,'Before cooking your short ribs in the air fryer, make sure to
defrost them. This is easy to do in a bowl of cold water for 10 – 15 minutes (do not
open the plastic bag if using the water to defrost). The time you cook your short ribs
in the air fryer will depend on your air fryer model and how thick the meat is. Cooked
at 400 degrees F, they should be ready in 15 minutes. Korean short ribs cook very
quickly, so keep an eye on them so they don't overcook.');
- You can make up your own recipes or copy them from a website. **Be sure to
insert values into each of the tables generated by the recipe.sql file:**
recipe_main, rec_ingredients, ingredients, and categories.
3. -- Add ingredients for Lasagna INSERT INTO `ingredients` VALUES (20, 'Ricotta
Cheese'), (21, 'Meat Sauce'), (22, 'Mozzarella Cheese');

4. -- Add rec_ingredients for Lasagna INSERT INTO `rec_ingredients` VALUES (20, 3, 4, 20), (21, 3, 3, 21), (22, 3, 4, 22);
5. -- Add ingredients for Korean short ribs INSERT INTO `ingredients` VALUES (23, 'Beef short ribs'), (24, 'Soy Sauce');
6. -- Add rec_ingredients for Korean short ribs INSERT INTO `rec_ingredients` VALUES (23, 4, 5, 23), (24, 4, 6, 24);
7. -- insert categories INSERT INTO `categories` VALUES (3,'Entree'),(4,'Main');
 - Use Chicken Marsala and Absolute Brownies as your guides for the problem as the instructions follow similar steps as what was already done in the code, but remember you need to do the same thing for two new recipes.

```

107 -- Lasagna and Korean short ribs are added in 'recipe_main' column
108 • INSERT INTO `recipe_main` VALUES (3,'Easy Lasagna',3,'Cheese and meat sauce',10,30,8,1,'poon meat sauce on the bottom of a lightly greased casserole dish. Add 4 boiled lasagna
109 • INSERT INTO `recipe_main` VALUES (4,'Korean short ribs',4,'Beef short ribs and soy sauce',5,10,4,1,'Before cooking your short ribs in the air fryer, make sure to defrost them.
110
111 -- Add ingredients for Lasagna
112 • INSERT INTO `ingredients`
113 VALUES
114 (20, 'Ricotta Cheese'),
115 (21, 'Meat Sauce'),
116 (22, 'Mozzarella Cheese');
117
118 -- Add rec_ingredients for Lasagna
119 • INSERT INTO `rec_ingredients`
120 VALUES
121 (20, 3, 4, 20), |
122 (21, 3, 3, 21),
123 (22, 3, 4, 22);
124
125 -- Add ingredients for Korean short ribs
126 • INSERT INTO `ingredients`
127 VALUES
128 (23, 'Beef short ribs'),
129 (24, 'Soy Sauce');
130
131 -- Add rec_ingredients for Korean short ribs
132 • INSERT INTO `rec_ingredients`
133 VALUES
134 (23, 4, 5, 23),
135 (24, 4, 6, 24);
136
137 -- insert categories
138 • INSERT INTO `categories` VALUES (3,'Entree'),(4,'Main');
```

2) Write only one SQL query that returns all information on only the two new recipes you inserted from all the tables you created in step 1 above.

Guideline:

- Your query should show all relevant information from all four of the tables from step 1.
- Don't worry that the output table may duplicate rows, for example the recipe_main table will duplicate rows for each of the ingredients you enter which is ok. The idea is to get you comfortable with joins.
- Show the results of your query as screenshots pasted in below:

8. SELECT * FROM recipes.recipe_main AS m

INNER JOIN recipes.categories AS c ON m.category_id=c.category_id INNER

JOIN recipes.rec_ingredients AS r ON m.recipe_id=r.recipe_id INNER JOIN

recipes.ingredients AS i ON r.ingredient_id=i.ingredient_id

WHERE m.recipe_id=3 OR m.recipe_id=4;

```
140 -- all the information on only the two new recipes
```

```
141 • SELECT *
```

```
142
```

```
143 FROM recipes.recipe_main AS m
```

```
144 INNER JOIN recipes.categories AS c
```

```
145 ON m.category_id=c.category_id
```

```
146 INNER JOIN recipes.rec_ingredients AS r
```

```
147 ON m.recipe_id=r.recipe_id
```

```
148 INNER JOIN recipes.ingredients AS i
```

```
149 ON r.ingredient_id=i.ingredient_id
```

```
150 WHERE m.recipe_id=3 OR m.recipe_id=4;
```

```
151
```

recipe_id	rec_title	category_id	recipe_desc	prep_time	cook_time	servings	difficulty	directions	category_id	category_name	rec_ingredient_id	recipe_id	amount	ingredient_id	ingredient_id	ingred_name
3	Easy Lasagna	3	Cheese and meat sauce	10	30	8	1	poon meat sauce on the bottom of a lightly gre...	3	Entree	20	3	4.00	20	20	Ricotta Cheese
3	Easy Lasagna	3	Cheese and meat sauce	10	30	8	1	poon meat sauce on the bottom of a lightly gre...	3	Entree	21	3	3.00	21	21	Meat Sauce
3	Easy Lasagna	3	Cheese and meat sauce	10	30	8	1	poon meat sauce on the bottom of a lightly gre...	3	Entree	22	3	4.00	22	22	Mozzarella Cheese
4	Korean short ribs	4	Beef short ribs and soy sauce	5	10	4	1	Before cooking your short ribs in the air fryer, ...	4	Main	23	4	5.00	23	23	Beef short ribs
4	Korean short ribs	4	Beef short ribs and soy sauce	5	10	4	1	Before cooking your short ribs in the air fryer, ...	4	Main	24	4	6.00	24	24	Soy Sauce

- 3) Write a SELECT query that identifies the recipe name, category name, and ingredient name, and ingredient amount. No other variables should be included.**

Guideline:

- Your output should be sorted first by descending category name, then by ascending recipe name, followed by descending ingredient name.
- Recall that string variables are sorted alphabetically (ascending only) when you check your query results.

9. SELECT m.rec_title, c.category_name, i.ingred_name, r.amount

FROM recipes.recipe_main AS m

INNER JOIN recipes.categories AS c

ON m.category_id=c.category_id

INNER JOIN recipes.rec_ingredients AS r

ON m.recipe_id=r.recipe_id

INNER JOIN recipes.ingredients AS i

ON r.ingredient_id=i.ingredient_id

ORDER BY 2 DESC, 1 ASC, 3 DESC;

```

152  -- specific columns in our table
153  • SELECT m.rec_title, c.category_name, i.ingred_name, r.amount
154  FROM recipes.recipe_main AS m
155  INNER JOIN recipes.categories AS c
156  ON m.category_id=c.category_id
157  INNER JOIN recipes.rec_ingredients AS r
158  ON m.recipe_id=r.recipe_id
159  INNER JOIN recipes.ingredients AS i
160  ON r.ingredient_id=i.ingredient_id
161  ORDER BY 2 DESC, 1 ASC, 3 DESC;

```

Result Grid				
Filter Rows: <input type="text"/> Export: Wrap Cell Content:				
	rec_title	category_name	ingred_name	amount
▶	Korean short ribs	Main	Soy Sauce	6.00
	Korean short ribs	Main	Beef short ribs	5.00
	Chicken Marsala	Entree	Sliced mushrooms	2.00
	Chicken Marsala	Entree	Rosemary, dried and crushed	0.25
	Chicken Marsala	Entree	Parsley, minced	2.00
	Chicken Marsala	Entree	Parmesan cheese, grated	2.00
	Chicken Marsala	Entree	Olive oil	2.00
	Chicken Marsala	Entree	Marsala wine	0.75
	Chicken Marsala	Entree	Flour	2.00
	Chicken Marsala	Entree	Chicken breast halves, boneless	4.00
	Chicken Marsala	Entree	Butter	2.00
	Easy Lasagna	Entree	Ricotta Cheese	4.00
	Easy Lasagna	Entree	Mozzarella Cheese	4.00
	Easy Lasagna	Entree	Meat Sauce	3.00
	Absolute Brownies	Desserts	White sugar	1.00
	Absolute Brownies	Desserts	Vanilla extract	1.00
	Absolute Brownies	Desserts	Unsweetened cocoa powder	0.33
	Absolute Brownies	Desserts	Salt	0.25
	Absolute Brownies	Desserts	Olive oil	0.50
	Absolute Brownies	Desserts	Flour	0.50
	Absolute Brownies	Desserts	Egg(s)	2.00

Conclusion

In this project, I learned how to normalize data such as 1NF, 2NF, and 3NF using a single CSV file. Normalization is a key data organization method for building a database. In the 'recipes' section, I learned how to add data using SQL and how to extract the necessary information using JOIN. Through JOIN, I can extract and check all the necessary data from the normalized data at once.

References

Herovired. (2023, August 8). Difference between primary key and foreign key. Retrieved from

<https://herovired.com/learning-hub/blogs/difference-between-primary-key-and-foreign-key>

ShareTheSameSky. (2021). ELI5: The difference between 3NF, 2NF, and 1NF. Reddit.

Retrieved from

https://www.reddit.com/r/explainlikeimfive/comments/q1897c/eli5_the_difference_between_3nf_2nf_and_1nf/

Upadhyay, M. (2023, November 6). Second Normal Form (2NF). GeeksforGeeks. Retrieved

from <https://www.geeksforgeeks.org/second-normal-form-2nf/>

Upadhyay, M. (2019, July 31). Third Normal Form (3NF). GeeksforGeeks. Retrieved from

<https://www.geeksforgeeks.org/third-normal-form-3nf/>

Jay, R. (2023). What is first, second, or third normal form. Retrieved from

https://northeastern.instructure.com/courses/160692/discussion_topics/2059663

Stephanie. (2022, April 27). Easy Lasagna Recipe. Retrieved from

<https://thecozycook.com/easy-lasagna-recipe/>

Tatiana. (2022, March 23). Trader Joe's Korean Short Ribs Air Fryer. Retrieved from

<https://thetopmeal.com/trader-joes-korean-short-ribs-air-fryer/>

Jayaram, P. (2018, November 23). Overview of the SQL ORDER BY clause. Retrieved from

<https://www.sqlshack.com/overview-of-the-sql-order-by-clause/>