Project Title: Cooking made easy!

Team 25(DexLab)

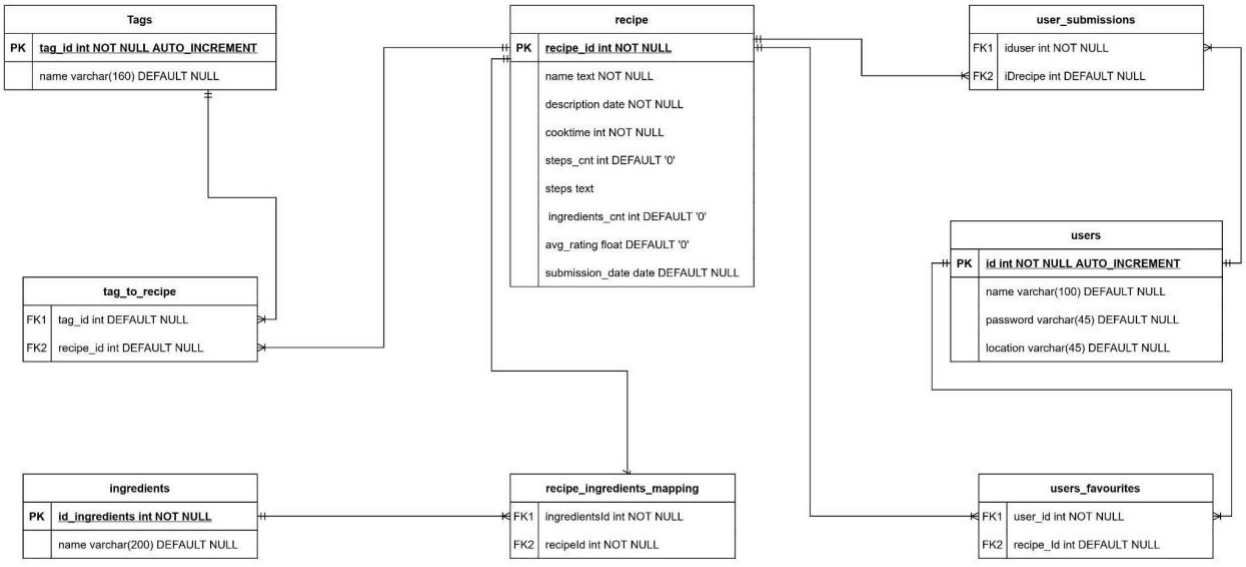
Vishwa Shrirame | vshriram@iu.edu

Kumar Saurabh | ksaurabh@iu.edu

Raghav Chegu Shyam Kumar | racheg@iu.edu

1. Database Design

ER Diagram



Cooking Made Easy ER Diagram

We have a total of 8 tables in our schema. They are as follows:-

Entities:

recipe

- The recipe table will have all the information of recipes
- Ease record will contain information such as recipe_id, name, description, number of steps, steps, ingredients count, average rating and submission date

- recipe_id will be the primary key

ingredients

- This table will contain all the unique ingredients in our database
- Each record will consist of ingredients_id, name
- Ingredients_id will be the primary key

recipe_ingredients_mapping

- This table will contain mapping between recipes and ingredients
- Each record will have recipe_id and ingredients_id as foreign key

Tags

- This table will contain all the unique tag related to recipes in our database
- Each record will consist of tag_id, name
- tag_id will be the primary key

tag_to_recipe

- This table will contain mapping between recipes and tags
- Each record will have recipe_id and tag_id as foreign key

users

- This table will contain all information of users.
- Each record will contain information such as user_id, name, name, location, password.
- user_id will be the primary key

users_favourites

- This table will contain mapping between users and their favourite recipes
- Each record will have recipe_Id and user_id as foreign key

users_submissions

- This table will contain mapping between users and their submission recipes
- Each record will have iDrecipe and iduser as foreign key

Relationship

recipe and recipe_ingredients_mapping – *Many: Many* relationship
User and recipe_ingredients_mapping – *Many: Many* relationship
Tags and tag_to_recipe – *Many: Many* relationship
recipe and tag_to_recipe – *Many: Many relationship*
user and users_favourites – *Many: Many relationship*
recipe and users_favourites – *Many: Many relationship*

user and users_submissions – *Many: Many relationship*
recipe and users_submissions – *Many: Many relationship*

So our ER Diagram explains the relationship between recipes and ingredients, recipes and tags, recipes and users. The users table is mapped to user favourites on a one-to-many relationship, whereas the recipes table is mapped to ingredients on one to one relationship.

2. Describe Data types

recipes -

Attribute	**Data-type**
recipe_id	INT
name	TEXT
description	TEXT
cooktime	INT
steps_cnt	INT
steps	TEXT
ingredients_cnt	INT
avg_rating	FLOAT
submission_date	DATE

users -

Attribute	**Data-type**
id	INT
name	VARCHAR (100)
password	VARCHAR (45)
location	VARCHAR (45)

tags -

Attribute	**Data-type**
tag_id	INT
name	VARCHAR (160)

users_favourites -

Attribute	**Data-type**
user_id	INT
recipe_id	INT

ingredients -

Attribute	**Data-type**
id_ingredients	INT
name	VARCHAR (200)

user_submissions -

Attribute	**Data-type**
iduser	INT
IDrecipe	INT

tag_to_recipe -

Attribute	Data-type
tag_id	INT
recipe_id	INT

recipe_ingredients_mapping -

Attribute	Data-type
ingredients_id	INT
recipe_id	INT

3. Descirbe Constraints and Primary keys

Primary Keys

PRIMARY KEY	Description and Table Name
recipe_id	Unique ID of a recipe
user_id	Unique ID of a user
tag_id	Unique ID of a tag
ingredients_id	Unique ID of an ingredient

4. Code to create database

Create cooking_made_easy database:

```
1 •   Create schema cooking_made_easy;
2
3 •   use cooking_made_easy;
```

Create recipe table:

DDL for cooking_made_easy.recipes

```
1  CREATE TABLE `recipes` (
2    `recipe_id` int NOT NULL,
3    `name` text,
4    `description` text,
5    `cooktime` int DEFAULT '0',
6    `steps_cnt` int DEFAULT '0',
7    `steps` text,
8    `ingredients_cnt` int DEFAULT '0',
9    `avg_rating` float DEFAULT '0',
10   `submission_date` date DEFAULT NULL,
11   PRIMARY KEY (`recipe_id`)
12 ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb3
```

Create Tags table:

DDL for cooking_made_easy.Tags

```
CREATE TABLE `Tags` (
  `tag_id` int NOT NULL AUTO_INCREMENT,
  `name` varchar(160) DEFAULT NULL,
  PRIMARY KEY (`tag_id`)
) ENGINE=InnoDB AUTO_INCREMENT=17 DEFAULT CHARSET=utf8mb3
```

Create ingredients table:

DDL for cooking_made_easy.ingredients

```
CREATE TABLE `ingredients` (
  `id_ingredients` int NOT NULL,
  `name` varchar(200) DEFAULT NULL,
  PRIMARY KEY (`id_ingredients`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb3
```

Create users table:

DDL for cooking_made_easy.users

```
CREATE TABLE `users` (
  `id` int NOT NULL AUTO_INCREMENT,
  `name` varchar(100) DEFAULT NULL,
  `password` varchar(45) DEFAULT NULL,
  `location` varchar(45) DEFAULT NULL,
  PRIMARY KEY (`id`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb3
```

Create tag_to_recipe table:

DDL for cooking_made_easy.tag_to_recipe

```
CREATE TABLE `tag_to_recipe` (
  `tag_id` int DEFAULT NULL,
  `recipe_id` int DEFAULT NULL,
  KEY `tag_id_idx` (`tag_id`),
  KEY `recipe_id_idx` (`recipe_id`),
  CONSTRAINT `recipe_id` FOREIGN KEY (`recipe_id`) REFERENCES `recipes` (`recipe_id`) ON DELETE CASCADE ON UPDATE CASCADE,
  CONSTRAINT `tag_id` FOREIGN KEY (`tag_id`) REFERENCES `Tags` (`tag_id`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb3
```

Create user_favourite and user_submission tables:

DDL for cooking_made_easy.users_favourites

```
CREATE TABLE `users_favourites` (
  `user_id` int NOT NULL,
  `recipe_Id` int DEFAULT NULL,
  KEY `recipe_Id_idx` (`recipe_Id`),
  KEY `user_id_idx` (`user_id`),
  CONSTRAINT `id_recipe` FOREIGN KEY (`recipe_Id`) REFERENCES `recipes` (`recipe_id`) ON DELETE CASCADE ON UPDATE CASCADE,
  CONSTRAINT `user_id` FOREIGN KEY (`user_id`) REFERENCES `users` (`id`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb3
```

DDL for cooking_made_easy.user_submissions

```
CREATE TABLE `user_submissions` (
  `iduser` int NOT NULL,
  `iDrecipe` int DEFAULT NULL,
  KEY `iDrecipe_idx` (`iDrecipe`),
  KEY `iduser_idx` (`iduser`),
  CONSTRAINT `iDrecipe` FOREIGN KEY (`iDrecipe`) REFERENCES `recipes` (`recipe_id`) ON DELETE CASCADE ON UPDATE CASCADE,
  CONSTRAINT `iduser` FOREIGN KEY (`iduser`) REFERENCES `users` (`id`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb3
```

For data insertion we have cleaned our data and have made changes in the python notebook and exported it in csv after that it was imported in the database using command line.

1. **Loading data for processing:**

```
# Load Data from csv

raw_recipes = pd.read_csv('data/RAW_recipes.csv')
raw_interactions = pd.read_csv('data/RAW_interactions.csv')
```

[3] ✓ 7.4s

2. **Processing data and cleaning for csv formation for recipes table**

```
# Creating dataframe with all required variables

x = raw_interactions.groupby('recipe_id')['rating'].mean()
x = pd.DataFrame(x)
x.reset_index(inplace=True)
recipes_final = raw_recipes.merge(x, how = 'inner', left_on='id', right_on='recipe_id')
```

[9] ✓ 0.2s

```
# Creating recipes dataframe to be exported to the recipes table in the database

recipes_df = pd.DataFrame()
recipes_df['recipe_id'] = recipes_final['id']
recipes_df['name'] = recipes_final['name']
recipes_df['description'] = recipes_final['description']
recipes_df['cooktime'] = recipes_final['minutes']
recipes_df['steps_cnt'] = recipes_final['n_steps']
recipes_df['steps'] = recipes_final['steps']
recipes_df['ingredients_cnt'] = recipes_final['n_ingredients']
recipes_df['avg_rating'] = recipes_final['rating_x']
recipes_df['submission_date'] = recipes_final['submitted']
```

[17] ✓ 11.3s

```
from pathlib import Path
filepath = Path(
    '/Users/saurabh/workspace/Spring_2022/ADT/Cooking-made-easy/data/recipes_new.csv')
filepath.parent.mkdir(parents=True, exist_ok=True)
recipes_df.to_csv(filepath)
```

[20] ✓ 3.4s

3. **Formatting data for Tags and tags_recipe_mapping table:**

```
# Creating tags dataframe

unique_tags = calc_unique(recipes_final['tags'])

main_tag = []
for tag in unique_tags:
    if ('less' in tag or 'more' in tag) and '-' in tag:
        main_tag.append(tag)

other_tags = ['potluck','weeknight','dinner-party','appetizers', 'for-1-or-2',
 'for-large-groups', 'time-to-make', 'to-go', 'toddler-friendly', 'for-large-groups-holiday-event']
main_tag.extend(other_tags)

main_tags = {'tag_id': list(range(1,len(main_tag)+1)),'name': main_tag}
tags_df = pd.DataFrame(main_tags)
```

[22] ✓ 6.6s

```
filepath = Path(
    '/Users/saurabh/workspace/Spring_2022/ADT/Cooking-made-easy/data/tags.csv')
filepath.parent.mkdir(parents=True, exist_ok=True)
tags_df.to_csv(filepath)
```

[31] ✓ 0.5s

```
# Creating tag_to_recipe dataframe

tag_recipe = recipes_final[['id', 'tags']]

tags_arr = []
tags_id = []

for index, row in tag_recipe.iterrows():
    row['tags'] = row['tags'].replace("[","")
    row['tags'] = row['tags'].replace("]","")
    row['tags'] = row['tags'].replace("'","")
    row['tags'] = row['tags'].replace(" ","")

    tag = [j for j in row['tags'].split(',') if j in list(tags_df['name'])]

    tags_arr.append(tag)
    tags_id.append(row['id'])

tag_fin_recipe = pd.DataFrame({'tag_id': tags_arr, 'recipe_id': tags_id})
tag_fin_recipe = tag_fin_recipe.explode('tag_id')
tag_to_recipe = tag_fin_recipe.merge(tags_df, how = 'inner', left_on='tag_id', right_on='name')
```

[36] ✓ 33.6s

```
filepath = Path(
    '/Users/saurabh/workspace/Spring_2022/ADT/Cooking-made-easy/data/tag_to_recipe.csv')
filepath.parent.mkdir(parents=True, exist_ok=True)
tag_to_recipe.to_csv(filepath)
```

[38] ✓ 1.5s

4. **Formatting the indregdients and recipe_ingredients_mapping tables:**

```
# Creating ingredients dataframe

unique_ingredients = list(calc_unique(recipes_final['ingredients']))

ingredients = {'id_ingredients': list(range(1,len(unique_ingredients)+1)),'name': unique_ingredients}
ingredients_df = pd.DataFrame(ingredients)
```

[32] ✓ 4.1s

```
filepath = Path(
    '/Users/saurabh/workspace/Spring_2022/ADT/Cooking-made-easy/data/ingredients.csv')
filepath.parent.mkdir(parents=True, exist_ok=True)
ingredients_df.to_csv(filepath)
```

[34] ✓ 0.1s

```
ind_data = ind_data[['ingredients_name', 'recipe_id']]
ind_data
```

[84] ✓ 0.2s

	ingredients_name	recipe_id
0	blueberries	38
0	granulated sugar	38
0	vanilla yogurt	38
0	lemon juice	38
1	saffron	39
...	...	...
231636	blue cheese	537716
231636	cheez whiz	537716
231636	rolls	537716
231636	green onion	537716
231636	parsley	537716

2096582 rows × 2 columns

```
filepath = Path(
    '/Users/saurabh/workspace/Spring_2022/ADT/Cooking-made-easy/data/ingredients_map.csv')
filepath.parent.mkdir(parents=True, exist_ok=True)
ind_data.to_csv(filepath, index=False)
```

[85] ✓ 2.2s

5. **Explain how data is imported**

We have used MySQL CLI to import data from CSV file to the table in the readify database. Set global local_infile to 1 to use load command for importing csv data to table

```
(base) saurabh@Kumars-MacBook-Pro data %
(base) saurabh@Kumars-MacBook-Pro data % mysql --local_infile=1 -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 547
Server version: 8.0.28 MySQL Community Server - GPL

Copyright (c) 2000, 2022, Oracle and/or its affiliates.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql>
```

Use load data local in file command to import data from csv file to the tables in the database

```
mysql>
mysql> LOAD DATA LOCAL INFILE  '/Users/saurabh/workspace/Spring_2022/ADT/Cooking-made-easy/data/ingredients.csv' INTO TABLE ingredients FIELDS TERMINATED BY ',' ENCLOSED BY '"' LINES TERMINATED BY '\n' IG
NORE 1 ROWS;
Query OK, 14758 rows affected, 14758 warnings (0.05 sec)
Records: 14758  Deleted: 0  Skipped: 0  Warnings: 14758

mysql>
mysql>
mysql> LOAD DATA LOCAL INFILE  '/Users/saurabh/workspace/Spring_2022/ADT/Cooking-made-easy/data/tags.csv' INTO TABLE Tags FIELDS TERMINATED BY ',' ENCLOSED BY '"' LINES TERMINATED BY '\n' IGNORE 1 ROWS;
Query OK, 16 rows affected, 18 warnings (0.00 sec)
Records: 17  Deleted: 0  Skipped: 1  Warnings: 18

mysql>
mysql>
mysql> LOAD DATA LOCAL INFILE  '/Users/saurabh/workspace/Spring_2022/ADT/Cooking-made-easy/data/tag_to_recipe.csv' INTO TABLE tag_to_recipe FIELDS TERMINATED BY ',' ENCLOSED BY '"' LINES TERMINATED BY '\n
' IGNORE 1 ROWS;
Query OK, 0 rows affected, 65535 warnings (21.31 sec)
Records: 664580  Deleted: 0  Skipped: 664580  Warnings: 664580

mysql>
mysql>
mysql> LOAD DATA LOCAL INFILE  '/Users/saurabh/workspace/Spring_2022/ADT/Cooking-made-easy/data/ingredients_map.csv' INTO TABLE recipe_ingredients_mapping FIELDS TERMINATED BY ',' ENCLOSED BY '"' LINES TE
RMINATED BY '\n' IGNORE 1 ROWS;
Query OK, 2052107 rows affected, 44475 warnings (8.06 sec)
Records: 2096582  Deleted: 0  Skipped: 44475  Warnings: 44475

mysql>
```

Data dictionary -

recipes -

Table Name	Attribute	Description	Data_type	nullable	Example
recipes	recipe_id	recipe ID - PRIMARY KEY	INT	N	1
recipes	name	recipe name	TEXT	Y	Briyani
recipes	description	User-provided description	TEXT	Y	Veg Briyani
recipes	cooktime	Minutes to prepare recipe	INT	Y	30
recipes	steps_cnt	Number of steps in recipe	INT	Y	1
recipes	steps	Text for recipe steps, in order	TEXT	Y	Cook in Handi
recipes	ingredients_cnt	Number of ingredients	INT	Y	10
recipes	avg_rating	Average rating for that recipe	FLOAT	Y	4.3
recipes	submission_date	Date recipe was submitted	DATE	Y	2020-01-01

users

Table Name	Attribute	Description	Data_type	nullable	Example
users	id	User id - PRIMARY KEY	INT	N	1
users	name	User name	VARCHAR(45)	Y	Samar Deep
users	password	User password	VARCHAR(45)	Y	gurudatta
users	location	User location	VARCHAR(45)	Y	Bloomington

tags -

Table Name	Attribute	Description	Data_type	nullable	Example
tags	tag_id	Tag id - PRIMARY KEY	INT	N	1
tags	name	Tag name	VARCHAR(160)	N	Under 30 min

users_favourites -

Table Name	Attribute	Description	Data_type	nullable	Example
users_favourites	user_id	User id - PRIMARY KEY	INT	N	1
users_favourites	recipe_id	Recipe id	INT	Y	2

ingredients -

Table Name	Attribute	Description	Data_type	nullable	Example
ingredients	id_ingredients	Ingredient id - PRIMARY KEY	INT	N	1
ingredients	name	Ingredient name	VARCHAR(200)	Y	Milk

user_submissions -

Table Name	Attribute	Description	Data_type	nullable	Example
user_submissions	iduser	User id - PRIMARY KEY	INT	N	100
user_submissions	IDrecipe	Recipe id	INT	Y	32

tag_to_recipe -

Table Name	Attribute	Description	Data_type	nullable	Example
tag_to_recipe	tag_id	Tag id - PRIMARY KEY	INT	Y	45
tag_to_recipe	recipe_id	Recipe id	INT	Y	29