UNIVERSITY OF PISA

COMPUTER ENGINEERING MASTER DEGREE

Large-Scale and Multi-Structured DataBases

WeFood

Professors:

Pietro Ducange

Alessio Schiavo

Group Members:

Giovanni Ligato

Cleto Pellegrino

Giuseppe Soriano

ACADEMIC YEAR 2023/2024

Index

1. Introduction

WeFood is a Social Network where users can share their recipes and provide feedbacks about other users' recipes through comments and star rankings. It manages in a completely automatic way the calories of the recipes, so the users do not have to worry about that when they post a new recipe. Using the search engine, users can discover new top rated recipes to amaze their friends, filtering by ingredients or calories. Furthermore, users can follow other users and get suggestions about new users to follow.

2. Requirements

Describing the requirements it is important to distinguish between functional and non-functional requirements.

2.1. Functional Requirements

The main functional requirements for *WeFood* can be organized by the actor that is involved in the use case.

1. **Unregistered User**:
   - 1.1. Browse *recent*[1] recipes;
   - 1.2. Sign Up.
2. **Registered User**:
   - 2.1. Log In;
   - 2.2. Log Out;
   - 2.3. Upload a Post (Recipe);
   - 2.4. Modify his/her Posts;
   - 2.5. Delete his/her Posts;
   - 2.6. Comment a Post;
   - 2.7. Modify his/her own comments;
   - 2.8. Delete his/her own comments or comments on his/her Posts;
   - 2.9. Evaluate by a star ranking a Post;
   - 2.10. Delete his/her own star rankings;
   - 2.11. View the Recipe of a Post;
   - 2.12. View the Total Calories of a Recipe;
   - 2.13. View the Steps of a Recipe;

[1] Time interval can be manipulated by the actor using a slider.

2.14. View the Ingredients of a Recipe;

2.15. View the Calories of an Ingredient;

2.16. Browse most *recent* Posts;

2.17. Browse most *recent* top rated Posts;

2.18. Browse most *recent* Posts by ingredients;

2.19. Browse most *recent* Posts by calories (minCalories and maxCalories);

2.20. View his/her own personal profile;

2.21. Modify his/her own personal profile (e.g. change Name, Surname, Password);

2.22. Delete his/her own personal profile;

2.23. Find a User by username;

2.24. View other Users' profiles;

2.25. Follow a User;

2.26. Unfollow a User;

2.27. View his/her Friends (i.e. the Users he/she follows and that follow him/her);

2.28. View his/her Followers;

2.29. View his/her Followed Users.

3. **Admin**:

3.1 Log In;

3.2 Log Out;

3.3. Browse all the Users;

3.4. Browse all the Posts;

3.5. Ban a User;

3.6. Unban a banned User;

3.7. Delete a Post;

3.8. Delete a Comment;

3.9. See statistics about the usage of *WeFood*;

3.10. Add a new Ingredient.

2.2. Non-Functional Requirements

The non-functional requirements for *WeFood* are as follows.

1. **Performance**: the overall system must be able to handle a request in less than `1.5` seconds, because the user experience would be negatively affected by a longer response time. Being a social network, it is necessary to have a good performance in order to provide a good user experience.
2. **Availability**: the system must be available 24/7 for allowing users to use it at any time.
3. **Security**: the system must be secure and protect users' data even from possible attacks. In particular, the information transmitted between client and server must be over HTTPS. Furthermore, the system must protect users' passwords by hashing them before the storing in the database.
4. **Reliability**: the system must be reliable and must not lose the information uploaded by the users. It must be caple of recovering from a crash and restore the data in a consistent state, exploiting the replicas of the database.
5. **Usability**: the GUI offered to the users must be easy to use and intuitive. Each user should be able to use the application without any training and in about 15 minutes.
6. The **Back-End** must be written in Java.

3. Design

After having defined the requirements, it is possible to proceed with the design of the system.

3.1. Use Case Diagram

Translating the requirements into a graphical representation we obtain the *UML Use Case Diagram* shown in Figure 1.

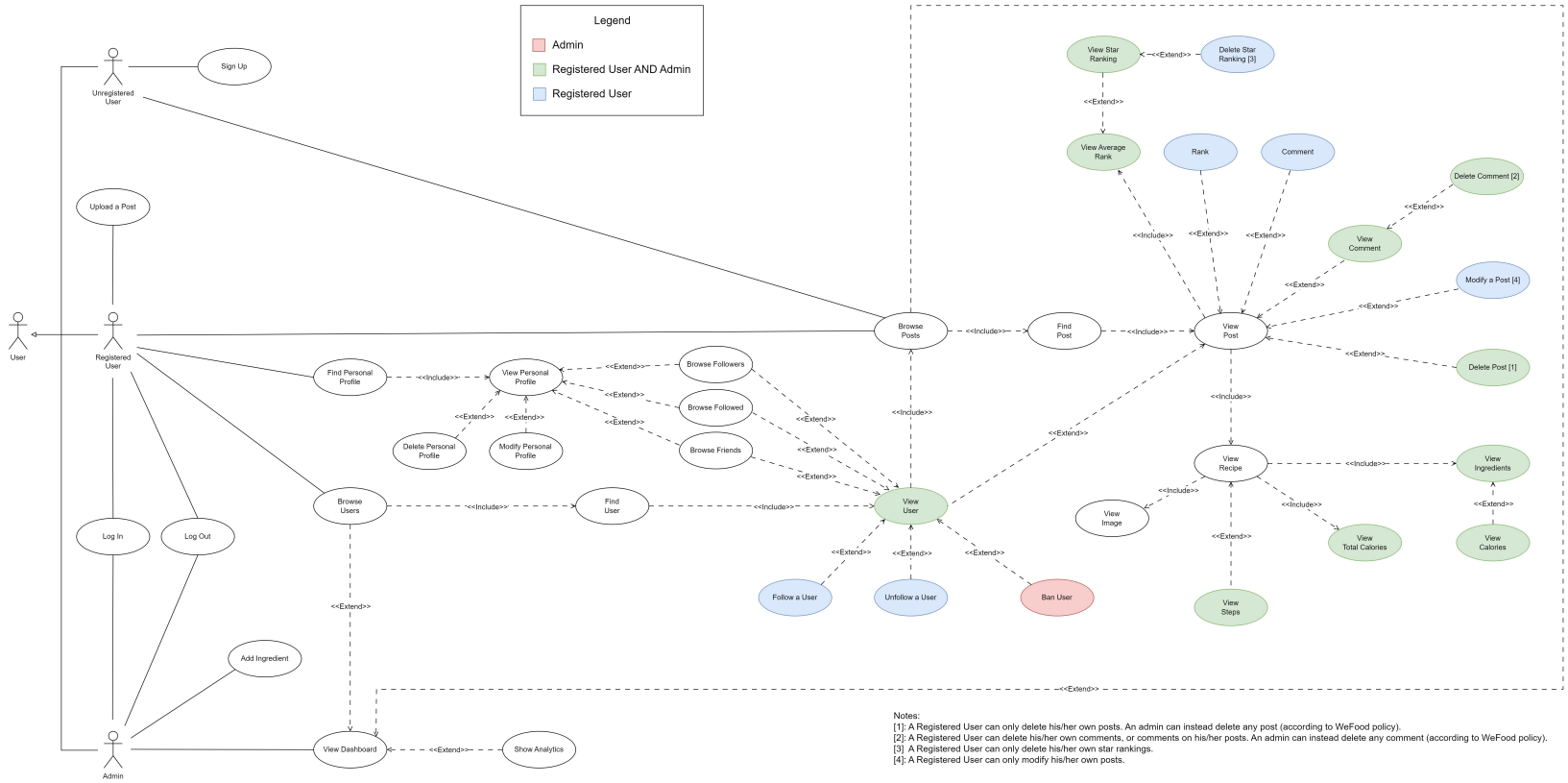

Figure 1: UML Use Case Diagram.

3.2. Class Diagram

The *UML Class Diagram* shown in Figure 2 represents the main entities of the system and their relationships.

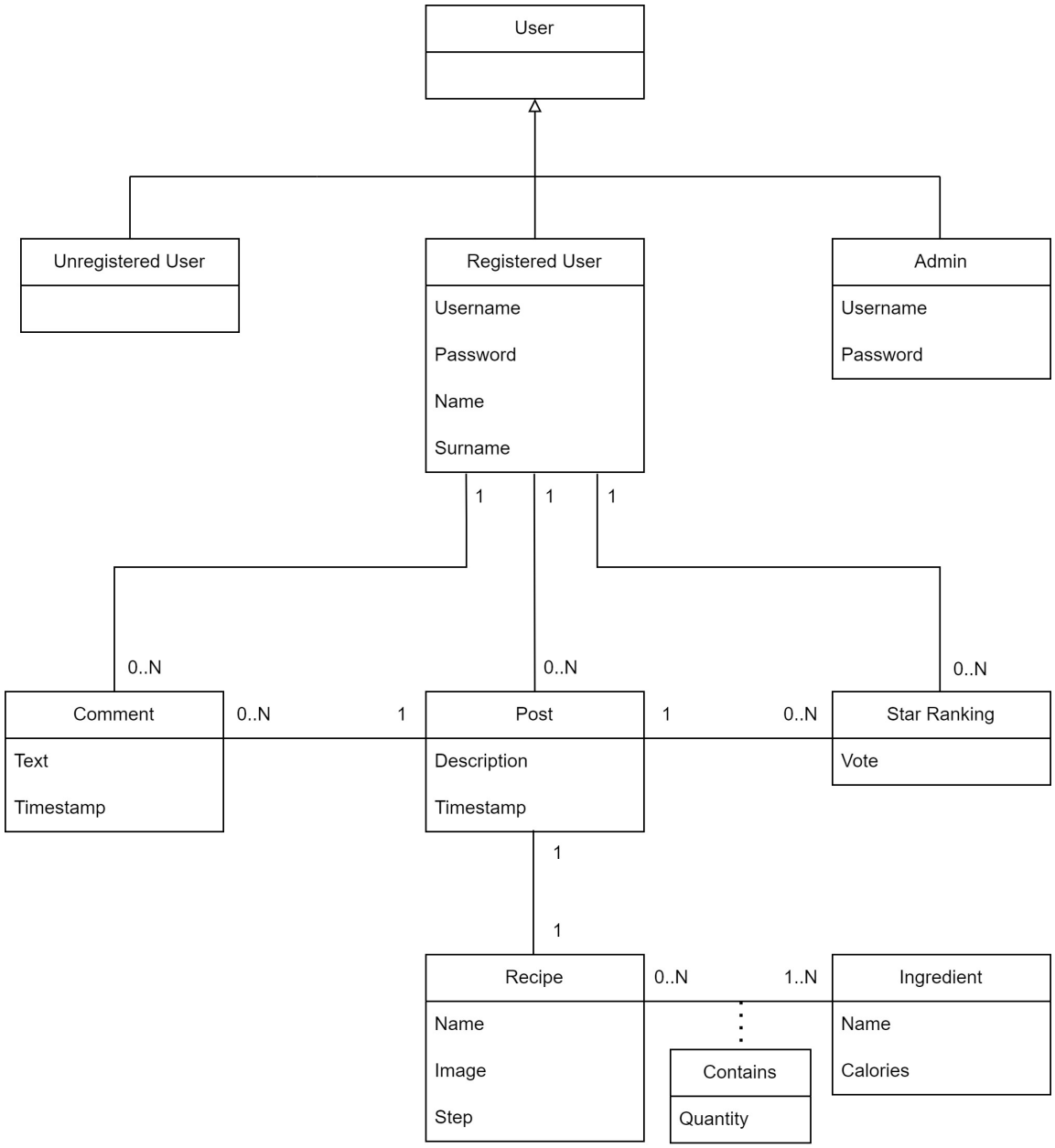

Figure 2: UML Class Diagram.

4. DataBases

Before cleaning and preparing the dataset needed to populate the databases, it is necessary to define the structure of the latter. In particular, two different databases will be used: a document DB and a graph DB.

4.1. Document DB

The entities managed by the document DB are the following.

- User (Unregistered User, Registered User, Admin);
- Post;
- Recipe;
- Comment;
- StarRanking;
- Ingredient.

4.1.1. Collections

The collections designed for storing the information inside the document DB are three: User, Post and Ingredient.

The structure of the User collection is as follows.

```
[User]:
{
    _id: ObjectId('...'),
    type: "Admin", # Applicable only for Admin
    username: String, [UNIQUE]
    password: String,
    name: String, # Not Applicable for Admin
    surname: String, # Not Applicable for Admin
    posts: [{ [REDUNDANCY(1)]
                idPost: ObjectId('...'),
                name: String,
                image: String
    }, ...] # Not Applicable for Admin
}
```

This collection is used both to store information about the Registered Users and the Admins. The field `type` is used to distinguish between the two types of Users, and it is set only for the Admins because they will be in minority compared to the Registered Users. The field `username` is required for the authentication of the Users and it is unique. So there cannot be two Users with the same username. The `password` is used for the authentication as well, and it contains the password provided by the User at the moment of the registration. For security reasons, the password is hashed before being stored in the database. The fields `name` and `surname` are used to store the name and the surname of the Registered Users and hence they are not applicable for the Admins for which it is not necessary to know their names and surnames. Lastly, the field `posts` is used to link (i.e. document linking) the Registered Users with their Posts. In particular, it contains a list of

objects, each one containing the id of a Post and the `name` and the `image` of the Recipe of the Post. The fields `name` and `image` are redundant because they are already stored in the Post collection, but they are useful for the queries that involve the User collection because they avoid the need of joining the Post collection.

The structure of the Post collection is a little bit more complex because it manages the information about the Posts, the Recipes, the Comments and the StarRankings. The decision of storing a Post in his own collection instead of embedding it in the document of the User that created it (i.e. document embedding) is due to several reasons:

- a User can publish an unlimited number of Posts, and a Post can have an many Comments and StarRankings. So the size of the document of a User would have grown indefinitely and very quickly, and this would have lead to a degradation of the performance of the system;
- the Posts are the main entities of the system and they are used in many queries. So having them in a separate collection ensures better performance by limiting the size of the individual document and taking advantage of tailored indexes.

```
[Post]:
{
    _id: ObjectId('...'),
    idUser: ObjectId('...'),
    username: String, [REDUNDANCY(2)]
    description: String,
    timestamp: Long,
    recipe: {
                name: String,
                image: String,
                steps: [String, ...],
                totalCalories: Double, [REDUNDANCY(3)]
                ingredients: [{
                                name: String,
                                quantity: Double
                }, ...]
    },
    starRankings: [{
                        idUser: ObjectId('...'),
                        username: String, [REDUNDANCY(4)]
                        vote: Double
    }, ...],
    avgStarRanking: Double, [REDUNDANCY(5)]
    comments: [{
                idUser: ObjectId('...'),
                username: String, [REDUNDANCY(6)]
                text: String,
                timestamp: Long
    }, ...]
}
```

Here the field `idUser` is used to link the Post with the Registered User that created it and `username` contains his/her username. The field `description` is used to store the description of the Post provided by the User. The field `timestamp` is used to store the timestamp of when the Post was uploaded. The field `recipe` is used to store the information about the Recipe contained in the Post. In particular, it contains the `name` and the `image` of the Recipe, the `steps` of the Recipe, the `totalCalories` of the Recipe and the list of `ingredients` of the Recipe together with the respective quantities in grams. It is important to notice that `image` does not contain the whole image, but just the URL of the image. The latter is stored or online or locally in the server. The field `starRankings` is used to store the information about the star rankings of the Post. In particular, it contains a list of objects, each one containing the id and the `username` of the Registered User that provided the star ranking and the `vote`, that represents the vote expressed by the Registerd User. The field `avgStarRanking` is used to store the average of the star rankings of the Post. The field `comments` is used to store the information about the comments of the Post. It contains a list of objects, each one containing the id and the `username` of the RegisteUser that provided the comment, the `text` and the `timestamp` of the comment.

The last collection is the Ingredient collection, that is used to store the information about the Ingredients. The structure of the Ingredient collection is very simple because it contains only the `name` and the `calories` per 100 grams of the Ingredient. The `name` is unique because does not make sense to have two Ingredients with the same name.

```
[Ingredient]:
{
    _id: ObjectId('...'),
    name: String, [UNIQUE]
    calories: Double
}
```

4.2. Graph DB

The entities that are managed by the graph DB are just: User (Registered User), Recipe and Ingredient.

In the following two sections, a comprehensive analysis of the nodes and relationships within the Graph DB schema, as illustrated in Figure 3, will be presented.

4.2.1. Nodes

The nodes designed for storing the information inside the graph DB are three, one for each entity.

The User node is used to store the information about the Registered Users. Each node contains the `_id` of the Registered User that is stored in the User collection of the document DB and his/her `username`.

```
(User):
    - _id: String
    - username: String [REDUNDANCY(7)]
```

The Recipe node is used to store the information about the Recipes. Each node contains the `_id` of the Post that contains the Recipe, which is stored withing the Post collection of the document DB.

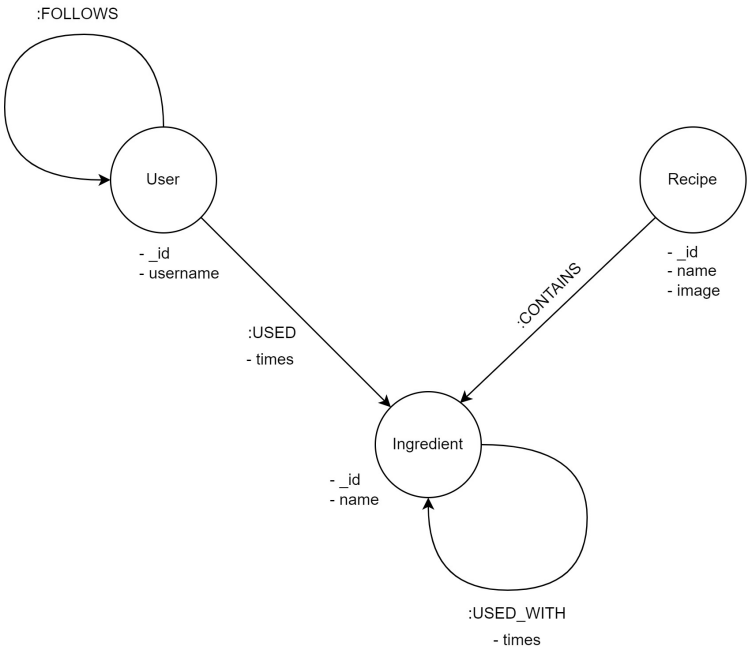

Figure 3: Graph DB schema.

Additionally, it includes the Recipe's `name` along with its corresponding `image`.

```
(Recipe):
    - _id: String
    - name: String [REDUNDANCY(8)]
    - image: String [REDUNDANCY(9)]
```

The Ingredient node is used to store the information about the Ingredients. Each node contains the `_id` of the Ingredient that is stored in the Ingredient collection of the document DB and the `name` of the Ingredient.

```
(Ingredient):
    - _id: String
    - name: String [REDUNDANCY(10)]
```

4.2.2. Relationships

Between the described nodes are possible several relationships that are formally defined as follows.

`(User)-[:FOLLOWS]->(User)`

This relationship allows Users to follow other Users. Two Users become friends when they follow each other.

`(User)-[:USED]->(Ingredient)`

```
(times: int) [REDUNDANCY(11)]
```

This relationship, instead, allows to quickly retrieve the Ingredients that have been used by the Users in their Recipes. The `times` attribute, in addition to being used for counting the number of times that an Ingredient has been used by a User, is used to keep track of the fact that the relationship with the Ingredient still can exist in other Recipes after the deletion of a Recipe by a User. Only when `times` becomes 0 the relationship can be deleted.

```
(Ingredient)-[:USED_WITH]->(Ingredient)
             (times: int) [REDUNDANCY(12)]
            [BIDIRECTIONAL]
```

This relationship allows to quickly retrieve the Ingredients that have been used together in the Users' Recipes. Here the `times` attribute is used for counting the number of times that two Ingredients have been used together. This relationship is bidirectional because if an Ingredient `A` has been used with an Ingredient `B`, then also the Ingredient `B` has been used with the Ingredient `A`.

```
(Recipe)-[:CONTAINS]->(Ingredient)
```

This last relationship allows to retrieve the Ingredients that are contained in a Recipe.

4.3. Redundancies

The proposed database models have redundancies, denoted as `[REDUNDANCY(n)]`. This means that the information they contain can be obtained through alternative means, often involving more intricate operations than a straightforward retrieval of the redundant value. These redundancies were cautiously introduced to enhance the system's reading performance and subsequently reduce response times. However, to maintain *data consistency*, writing operations are necessary to keep the redundancies updated. While reading operations are more frequent for the redundancies, the writing operations are generally less frequent. This approach is deemed more convenient for optimizing overall system performance. Redundancies also allow to avoid the need for joins, particularly when dealing with inter-database connections. A detailed explanation of the reasons justifying the introduction of the redundancies is provided in Table 1.

Table 1: Redundancies introduced into the database models.

(1) `DocumentDB:User:posts` **Reason**: To avoid joins. **Original/Raw Value**: `DocumentDB:Post`
(2) `DocumentDB:Post:username` **Reason**: To avoid joins. **Original/Raw Value**: `DocumentDB:User:username`
(3) `DocumentDB:Post:recipe:totalCalories` **Reason**: To avoid joins and to avoid computing the total calories of a Recipe every time a Post is shown.

Original/Raw Value: It is possible to compute the total calories of a Recipe by summing the calories of the Ingredients contained in the Recipe In particular the precise formula is the following: $\sum_i \left(quantity_i \cdot \frac{calories100g_i}{100} \right)$ where $quantity_i$ is the quantity of the i-th Ingredient contained in the Recipe and $calories100g_i$ is the amount of calories contained in 100 grams of the i-th Ingredient that can be retrieved from the `Ingredient` collection.

---

(4) `DocumentDB:Post:starRankings:username`
Reason: To avoid joins.
Original/Raw Value: `DocumentDB:User:username`

---

(5) `DocumentDB:Post:avgStarRanking`
Reason: To avoid computing the average star ranking of a Post every time is shown.
Original/Raw Value: It is possible to compute the average star ranking of a Post by averaging the values contained in `DocumentDB:Post:starRankings:vote`

---

(6) `DocumentDB:Post:comments:username`
Reason: To avoid joins.
Original/Raw Value: `DocumentDB:User:username`

---

(7) `GraphDB:(User):username`
Reason: To avoid joins with the DocumentDB.
Original/Raw Value: `DocumentDB:User:username`

---

(8) `GraphDB:(Recipe):name`
Reason: To avoid joins with the DocumentDB.
Original/Raw Value: `DocumentDB:Post:recipe:name`

---

(9) `GraphDB:(Recipe):image`
Reason: To avoid joins with the DocumentDB.
Original/Raw Value: `DocumentDB:Post:recipe:image`

---

(10) `GraphDB:(Ingredient):name`
Reason: To avoid joins with the DocumentDB.
Original/Raw Value: `DocumentDB:Ingredient:name`

---

(11) `GraphDB:(User)-[:USED]->(Ingredient):times`
Reason: To avoid computing the total number of times that a User used an Ingredient.
Original/Raw Value: It is possible to compute the total number of times that a User used an Ingredient by counting the number of times that the User used that Ingredient in his/her Recipes (information that can be retrieved from the DocumentDB).

---

(12) `GraphDB:(Ingredient)-[:USED_WITH]->(Ingredient):times`
Reason: To avoid computing the number of times that an Ingredient is used with another one.
Original/Raw Value: It is possible to compute the number of times that an Ingredient is used with another one by counting the number of times that all the Users used these two Ingredients together in their Recipes (information that can be retrieved from the DocumentDB).

---

5. Dataset

To populate the databases with a substantial volume of realistic data, datasets sourced from Kaggle were employed.

5.1. Raw Dataset

The initial raw datasets are related to the main functionalities of *WeFood*. In particular, datasets about recipes and ingredients were found.

- Calories per 100 grams in Food Items `[1]`
- Recipes and Interactions `[2]`
- Recipes and Reviews `[3]`

Contained in these datasets there are *almost* all the information needed to populate the databases. Indeed, in around 1 GB of raw data it is possible to find 2225 food items, over 500,000 recipes and 1,400,000 reviews. After a careful analysis, however, it was found that something was missing.

1. Personal Information about the Users: only the username of the Users was available (`AuthorName` in `recipes.csv` of `[3]`).
2. The quantity of each ingredient in the recipes: `RecipeIngredientQuantities` in `recipes.csv` of `[3]` contains some numbers that could be useful for this purpose, but they are not clear and there is no documentation about them. Indeed there is no way to understand if they are the quantities of the ingredients in grams or in other units of measure (e.g. just to have an idea there numbers like the following: 1, 1/2, 3, 5, etc). Furthermore, there are a lot of `NA` values in this column.

Everything else is in the datasets, and need only to be cleaned and appropriately merged to obtain the structure needed for the population of the databases.

5.2. Cleaning Process

There is the need to clean the datasets because in them there are plenty of information that are not useful for the purposes of *WeFood* and would result only in a waste of space. For achieving this goal, the datasets were analyzed in detail and the information that were not useful were discarded. The cleaning process was performed using Python.

Below a separate description of the cleaning process for each entity identified in the design phase is provided.

Ingredient: The starting point was: `ingr_map.pickle` in `[2]`. Here there are lots of fields useful for machine learning related tasks, but not for the *WeFood* purposes. Only fields strictly needed for linking recipes with `[1]` have been kept. The final result is the following. To facilitate referencing in the subsequent merging process, each intermediate product generated during the cleaning process will be assigned a distinct name, starting with the following.

```
Ingredient_A: {
    "raw_ingr":"pretzels",
    "replaced":"pretzel",
```

```
    "id":5711
}
```

The first field `raw_ingr` contains the original text of the ingredient, the one inserted by the user in the recipe. The second field `replaced` contains the simplified representation of the ingredient and the last field `id` contains the unique identifier of the ingredient.

At this point, the file `calories.csv` in `[1]` was analyzed. In this file in addition to the calories per 100 grams of each ingredient there are also Food Categories associated to them. These categories were really useful because they allowed to devise a plan for dealing with the lack of the quantity of each ingredient in the recipes in a simple but effective way. The idea was the following:

1. to associate to each `FoodCategory` two quantities, `quantity_min` and `quantity_max`, that are a realistic representation of the quantities used in real life for that specific `FoodCategory` (this labour intensive work was done with the support of *ChatGPT*);
2. to generate a random quantity for each ingredient in each recipe in the range `[quantity_min, quantity_max]`.

This solution does not provide a precise quantity for each ingredient in each recipe, but it is a good approximation that won't produce unrealistic results. This means that there won't be a recipe where there are 500 grams of `salt`, because the maximum quantity of `salt` that can be used in a recipe is 10 grams (i.e. `Herbs&Spices`, the `FoodCategory` of `salt`, has `quantity_max` equal to 10 grams).

After having removed some duplicates from `calories.csv`, based on the `FoodItem` field, the fields that were not useful were discarded. An example can be of help for understanding the final structure of the file.

```
Ingredient_B: {
    "FoodCategory":"Pastries,Breads&Rolls",
    "FoodItem":"Pretzel",
    "Cals_per100grams":"338 cal",
    "quantity_min":100,
    "quantity_max":500
}
```

Recipe: After the conversion of `RAW_recipes.csv` in `json` to have a more readable format, the file was analyzed in detail. Here there were lots of fields that could be discarded. The structure after the cleaning is, as before, better described by an example object.

```
Recipe_A: {
    "name":"pretzel crust",
    "id":194491,
    "contributor_id":356062,
    "submitted":"2006-11-07",
    "steps":"['preheat oven to 350', 'crush pretzels in a blender', 'add sugar
    ↪  and butter and mix well', 'press into a 9 inch pie plate', 'bake at 350
    ↪  for 8 minutes and cool', 'add desired filling']",
    "description":"recipe for a basic pretzel crust i found in a magazine. i
    ↪  don't think i saw it posted here yet.",
```

```
    "ingredients":"['pretzels', 'sugar', 'butter']"
}
```

Here the `name` is the name of the Recipe, `id` is the unique identifier of the Recipe and will be useful for linking the recipe with the interactions (i.e. comments and star rankings) of the dataset [2]. The `contributor_id` is the unique identifier of the User that created the Recipe. The `submitted` field is the timestamp of when the Recipe was uploaded. The `steps` field contains the steps of the Recipe, the `description` field contains the description that will be used for the Post that contains the Recipe and the `ingredients` field contains the list of the ingredients of the Recipe. Observing carefully the `steps` and the `ingredients` it is clear that they must be transformed into an array of strings because at the moment they are just strings.

From `recipes.csv` of [3], instead, it is possible to retrieve the URLs of the images of the Recipes. Thus only the fields `RecipeId` and `Images` are retained. Note that not all the Recipes have an image, and some of them have more than one image. Where no image is available, a default image will be applied. Viceversa, if multiple images are present, only the first image will be utilized.

```
Recipe_B: {
    "RecipeId":194491,
    "Image":"https:\/\/img.sndimg.com\/food\/image\/upload\/
    ↪  w_555,h_416,c_fit,fl_progressive,q_95\/v1\/img\/recipes
    ↪  \/19\/44\/91\/3xjO4aXTeiZpOajAsBRX_0S9A6246.jpg"
}
```

Comment: In `RAW_interactions.csv` of [2] the reviews (i.e. comments of *WeFood*) and the ratings (i.e. star rankings of *WeFood*) are stored together, because to each review there is associated a rating. In *WeFood* it is not the same, because a Registered User can leave a comment without providing a star ranking and viceversa. So the first step was to separate the two types of interactions.

```
Comment_A: {
    "user_id":430471,
    "recipe_id":194491,
    "timestamp":"2007-04-09"",
    "text":"This tasted great, however, I couldn't get it to hold together good.
    ↪  Either I didn't crush the pretzels small enough or I should have used a
    ↪  little more butter. But either way it was easy to make and tasted great.
    ↪  I used it as a base for a cheesecake pudding with fresh strawberries on
    ↪  the top."
}
```

Star Ranking: As previously mentioned, also the star rankings were stored in `RAW_interactions.csv` of [2]. For this reason the cleaning process was similar as the one used for the comments.

```
StarRanking_A: {
    "user_id":254614,
    "recipe_id":194491,
    "vote":4
}
```

Post: Posts are an abstraction of the Recipes that was introduced in *WeFood*. In the datasets there

is no information about them, so they will be created from scratch in the next merging process.

User: As previously noted, another lack of the datasets was the absence of personal information about the Users. Indeed, only the username of the Users was available. For not having an inconsistent situation where the Users have a name and a surname that are not coherent with their username, the username was dropped as well. In this way, it was possible to generate all the information needed for the Users using the Python library `Faker`. In particular, the name and the surname of the Users were generated randomly, and the username was computed accordingly using the following structure:

```
username = f"{name.lower()}_{surname.lower()}_{num}"
```

Where `num` is a random number in the range `[1, 99]`. All this is made paying attention to the fact that the username must be unique. By employing this approach, it was possible to create a User for each `contributor_id` of the Recipes (i.e. the Users who uploaded at least one Recipe). Users who did not upload any Recipe (i.e. who only appear in interactions) were excluded as a sufficient number of users was already available. The passwords (currently in plain) were generated randomly as well.

```
User_A: {
    "contributor_id":356062,
    "name":"Justin",
    "surname":"Alexander",
    "username":"justin_alexander_34",
    "password":"e6FMX30hGu"
}
```

5.3. Merging Process

After having cleaned the datasets, and having generated the missing information, it was possible to proceed with the merging process. This step was necessary to recreate the structure needed for the population of the databases. Also the merging process was performed using Python and Jupyter Notebook.

Because the merging process aims to produce the final structure of the documents that will be stored in the document DB, the latter will follow a different partitioning compared to the one used in the cleaning process. Here the partitioning will be based on the collections of the document DB.

Ingredient: Obtaining the final structure for the `Ingredient` collection is straightforward if `Ingredient_B` is considered. Indeed, it is sufficient to:

- rename `FoodItem` to `name`;
- rename `Cals_per100grams` to `calories`;
- take the float value of `calories`;
- discard `FoodCategory`, `quantity_min` and `quantity_max`.

```
{
    name: "Pretzel",
    calories: 338.0
}
```

The `_id` field will be automatically generated at the moment of the insertion in the database.

Post: Being the main collection of *WeFood*, the `Post` collection was also the one that takes more time to be merged. For this reason it is necessary to describe a step at a time for not complicating too much the explanation.

1. Merge of `Recipe_A` and `Recipe_B` on `id` and `RecipeId` respectively for including the URLs of the images inside the Recipes. In this way `Recipe_AB` is obtained.
2. The subsequent task involves converting the `ingredients` array of strings within `Recipe_A` into an array of objects. Each of these objects will include in the final structure the ingredient's `name` and its corresponding `quantity` expressed in grams. It is crucial to emphasize that the `name` of the ingredient must match an entry in the Ingredient collection. For achieving this:
    a. Firstly, `Ingredient_A` and `Ingredient_B` are matched on `replaced` and `FoodItem` respectively. For maximizing the number of matches, the matching was performed with the support of a Python Library called `fuzzywuzzy` which merges strings by likelyhood using *Levenshtein Distance*, which is a metric used in information theory, linguistics, and computer science for measuring the difference between two sequences. Thanks to this approach, no `Ingredient_A` was left unmatched with an `Ingredient_B`. An example of the matching is the following.

```
Ingredient_AB: {
    "raw_ingr":"pretzels",
    "replaced":"pretzel",
    "id":5711,
    "FoodCategory":"Pastries,Breads&Rolls",
    "FoodItem":"Pretzel",
    "Cals_per100grams":"338 cal",
    "quantity_min":100,
    "quantity_max":500
}
```

    b. By noticing that the `raw_ingr` field of `Ingredient_AB` contains the name of the ingredients as they are inserted by the users in the recipes, it possible to do a further match (this time using exact equality) between the latter and the strings contained in `ingredients` of `Recipe_AB`.

    c. At this point, `ingredients` will be an array of objects of the type of `Ingredient_AB`. For each of this object, it is possible to add a field `quantity` that is a random number in the range `[quantity_min, quantity_max]`. This is the quantity of the ingredient in grams that will be used in the recipe. After removing the no longer needed fields and making other minor modifications, here's the new structure.

```
Recipe_C: {
    "name": "pretzel crust",
    "id": 194491,
    "contributor_id": 356062,
    "submitted": "2006-11-07",
    "steps": [
        "preheat oven to 350",
        "crush pretzels in a blender",
```

```
            "add sugar and butter and mix well",
            "press into a 9 inch pie plate",
            "bake at 350 for 8 minutes and cool",
            "add desired filling"
        ],
        "description": "recipe for a basic pretzel crust i found in a
        ↪  magazine. i don't think i saw it posted here yet.",
        "ingredients": [
            {
                "foodItem": "Pretzel",
                "Cals_per100grams": 338.0,
                "quantity": 176
            },
            {
                "foodItem": "Sugar",
                "Cals_per100grams": 405.0,
                "quantity": 102
            },
            {
                "foodItem": "Butter",
                "Cals_per100grams": 720.0,
                "quantity": 43
            }
        ],
        "Image": "https://img.sndimg.com/food/image/upload/
        ↪  w_555,h_416,c_fit,fl_progressive,q_95/v1/img/recipes/
        ↪  19/44/91/3xjO4aXTeiZpOajAsBRX_0S9A6246.jpg"
    }
```

d. Now, exploiting the field `Cals_per100grams`, which has not been discarded yet, it is possible to compute the `totalCalories` of the recipe. This is done by executing the following operations for each recipe.

```
totalCalories = 0
for ingredient in ingredients:
    totalCalories += ingredient["quantity"] *
↪  (ingredient["Cals_per100grams"]/100)
```

e. In the end, the desired structure for `ingredients` was obtained.

```
Recipe_D: {
    "name": "pretzel crust",
    "id": 194491,
    "contributor_id": 356062,
    "submitted": "2006-11-07",
    "steps": [
        "preheat oven to 350",
        "crush pretzels in a blender",
```

```
            "add sugar and butter and mix well",
            "press into a 9 inch pie plate",
            "bake at 350 for 8 minutes and cool",
            "add desired filling"
        ],
        "description": "recipe for a basic pretzel crust i found in a
        ↪  magazine. i don't think i saw it posted here yet.",
        "ingredients": [
            {
                "foodItem": "Pretzel",
                "quantity": 176
            },
            {
                "foodItem": "Sugar",
                "quantity": 102
            },
            {
                "foodItem": "Butter",
                "quantity": 43
            }
        ],
        "totalCalories": 1317.58,
        "Image": "https://img.sndimg.com/food/image/upload/
        ↪  w_555,h_416,c_fit,fl_progressive,q_95/v1/img/recipes/
        ↪  19/44/91/3xjO4aXTeiZpOajAsBRX_0S9A6246.jpg"
    }
```


3. Creation of the Posts involves relocating fields unrelated to the Recipes. Furthermore, the Posts themselves contain the Recipes.


```
Post_A: {
    "id": 194491,
    "idUser": 356062,
    "description": "recipe for a basic pretzel crust i found in a magazine. i
    ↪  don't think i saw it posted here yet.",
    "timestamp": "2006-11-07",
    "recipe": {
        "name": "pretzel crust",
        "image": "https://img.sndimg.com/food/image/upload/
        ↪  w_555,h_416,c_fit,fl_progressive,q_95/v1/img/recipes/
        ↪  19/44/91/3xjO4aXTeiZpOajAsBRX_0S9A6246.jpg",
        "steps": [
            "preheat oven to 350",
            "crush pretzels in a blender",
            "add sugar and butter and mix well",
            "press into a 9 inch pie plate",
            "bake at 350 for 8 minutes and cool",
```

```
            "add desired filling"
        ],
        "ingredients": [
            {
                "foodItem": "Pretzel",
                "quantity": 176
            },
            {
                "foodItem": "Sugar",
                "quantity": 102
            },
            {
                "foodItem": "Butter",
                "quantity": 43
            }
        ],
        "totalCalories": 1317.58
    }
}
```


4. By considering `Comment_A` and `StarRanking_A`, it is possible to create the `comments` and `starRankings` arrays of the Posts. Specifically, each Comment or StarRanking is added to the corresponding Post based on the `recipe_id` field.


```
Post_B: {
    "id": 194491,
    "idUser": 356062,
    "description": "recipe for a basic pretzel crust i found in a magazine. i
    ↪  don't think i saw it posted here yet.",
    "timestamp": "2006-11-07",
    "recipe": {
        "name": "pretzel crust",
        "image": "https://img.sndimg.com/food/image/upload/
        ↪  w_555,h_416,c_fit,fl_progressive,q_95/v1/img/recipes/
        ↪  19/44/91/3xjO4aXTeiZpOajAsBRX_OS9A6246.jpg",
        "steps": [
            "preheat oven to 350",
            "crush pretzels in a blender",
            "add sugar and butter and mix well",
            "press into a 9 inch pie plate",
            "bake at 350 for 8 minutes and cool",
            "add desired filling"
        ],
        "ingredients": [
            {
                "foodItem": "Pretzel",
                "quantity": 176
```

```
            },
            {
                "foodItem": "Sugar",
                "quantity": 102
            },
            {
                "foodItem": "Butter",
                "quantity": 43
            }
        ],
        "totalCalories": 1317.58
    },
    "comments": [
        {
            "user_id": 430471,
            "timestamp": 1176076800000,
            "text": "This tasted great, however, I couldn't get it to hold
            ↪  together good.  Either I didn't crush the pretzels small
            ↪  enough or I should have used a little more butter. But either
            ↪  way it was easy to make and tasted great.  I used it as a
            ↪  base for a cheesecake pudding with fresh strawberries on the
            ↪  top."
        },
        {
            "user_id": 254614,
            "timestamp": 1182902400000,
            "text": "You have to crush the pretzels fine. There is a definite
            ↪  taste of salt, sugar and butter in the crust. It was great
            ↪  with a pudding pie filling but I would not make it with a
            ↪  fruit filling.You want the crust flavor to be a part of the
            ↪  dessert. Add waxed paper or non stick foil to press into pie
            ↪  pan, works very well. Thanks for posting."
        }
    ],
    "starRankings": [
        {
            "user_id": 430471,
            "vote": 3
        },
        {
            "user_id": 254614,
            "vote": 4
        }
    ]
}
```


5. To reconstruct the collection's structure there isn't much left to do. Indeed, it is only necessary

to convert the `timestamp` of creation of the Post in `Long`, to add the `avgStarRanking` field and to include the `username` of the Users, available in `User_A`, whenever their `id` appears.

```
Post_C: {
    "id": 194491,
    "idUser": 356062,
    "username": "justin_alexander_34",
    "description": "recipe for a basic pretzel crust i found in a magazine. i
    ↪  don't think i saw it posted here yet.",
    "timestamp": 1162857600000,
    "recipe": {
        "name": "pretzel crust",
        "image": "https://img.sndimg.com/food/image/upload/
        ↪  w_555,h_416,c_fit,fl_progressive,q_95/v1/img/recipes/
        ↪  19/44/91/3xjO4aXTeiZpOajAsBRX_OS9A6246.jpg",
        "steps": [
            "preheat oven to 350",
            "crush pretzels in a blender",
            "add sugar and butter and mix well",
            "press into a 9 inch pie plate",
            "bake at 350 for 8 minutes and cool",
            "add desired filling"
        ],
        "ingredients": [
            {
                "quantity": 176,
                "name": "Pretzel"
            },
            {
                "quantity": 102,
                "name": "Sugar"
            },
            {
                "quantity": 43,
                "name": "Butter"
            }
        ],
        "totalCalories": 1317.58
    },
    "comments": [
        {
            "timestamp": 1176076800000,
```

```
                "text": "This tasted great, however, I couldn't get it to hold
                ↪  together good.  Either I didn't crush the pretzels small
                ↪  enough or I should have used a little more butter. But either
                ↪  way it was easy to make and tasted great.  I used it as a
                ↪  base for a cheesecake pudding with fresh strawberries on the
                ↪  top.",
                "idUser": 430471,
                "username": "bonnie_vincent_27"
            },
            {
                "timestamp": 1182902400000,
                "text": "You have to crush the pretzels fine. There is a definite
                ↪  taste of salt, sugar and butter in the crust. It was great
                ↪  with a pudding pie filling but I would not make it with a
                ↪  fruit filling.You want the crust flavor to be a part of the
                ↪  dessert. Add waxed paper or non stick foil to press into pie
                ↪  pan, works very well. Thanks for posting.",
                "idUser": 254614,
                "username": "christopher_bass_52"
            }
        ],
        "avgStarRanking": 3.5,
        "starRankings": [
            {
                "vote": 3,
                "idUser": 430471,
                "username": "bonnie_vincent_27"
            },
            {
                "vote": 4,
                "idUser": 254614,
                "username": "christopher_bass_52"
            }
        ]
    }
```


User: In `User_A`, the only thing missing for having the User collection is the array field `posts` which contains a simplified representation of the Posts uploaded by the User. By employing the `idUser` in `Post_C`, all the user's posts can be retrieved, and the necessary fields can be selected.


```
User_B: {
    "contributor_id": 356062,
    "name": "Justin",
    "surname": "Alexander",
    "username": "justin_alexander_34",
    "password": "e6FMX30hGu",
    "posts": [
```

```
        {
            "idPost": 234229,
            "name": "layered ice cream candy cake",
            "image": "https://img.sndimg.com/food/image/upload/
            ↪  w_555,h_416,c_fit,fl_progressive,q_95/v1/img/recipes/
            ↪  23/42/29/picztrpPR.jpg"
        },
        {
            "idPost": 194491,
            "name": "pretzel crust",
            "image": "https://img.sndimg.com/food/image/upload/
            ↪  w_555,h_416,c_fit,fl_progressive,q_95/v1/img/recipes/
            ↪  19/44/91/3xjO4aXTeiZpOajAsBRX_0S9A6246.jpg"
        },
        {
            "idPost": 226341,
            "name": "quesadilla combos",
            "image": "https://img.sndimg.com/food/image/upload/
            ↪  w_555,h_416,c_fit,fl_progressive,q_95/v1/img/recipes/
            ↪  22/63/41/Xc9MqI3jSBmO7sx9cUFN_quesadilla-combos_0511.jpg"
        },
        {
            "idPost": 282971,
            "name": "raspberry lime rugalach",
            "image": "https://img.sndimg.com/food/image/upload/
            ↪  w_555,h_416,c_fit,fl_progressive,q_95/v1/img/recipes/
            ↪  28/29/71/picYumCGj.jpg"
        },
        ...
    ]
}
```


5.4. Population

After completing the preceding steps, the datasets have been appropriately cleansed, merged, and are now prepared for being imported into the databases. The file dimensions are as follows:

- Post collection: approximately 700MB;
- User: 69MB;
- Ingredient: 266KB.

However, specific procedures are needed to populate the two types of databases.

5.4.1. Document DB

It's true that all the necessary files for populating the documentDB have been generated, but there are some points that may need further clarifications.

- Firstly, in cases where a field is null or empty (e.g. a Post with no associated image, comment, or star ranking), within the context of a NoSQL database, there is no necessity for these fields to be assigned the null value to reduce unnecessary memory usage. Consequently, if there is missing information in a field, that field will simply be absent from the structure.
- Secondly, the `_id` fields are not yet incorporated into the aforementioned structures, while conversely, the old *ids* from the datasets still persist. This is due to the intention to utilize the `_id` values provided by MongoDB (i.e. the documentDB employed in the implementation). To preserve the connection between documents and facilitate the substitution, a straightforward yet effective procedure was implemented. Documents were imported into their respective MongoDB collections, and subsequently, a *JSON export* was performed, leading to the insertion of MongoDB's `_id` within all the documents. Utilizing the old *ids*, the linkage between documents was established, and a substitution was executed by assigning the new MongoDB `_id` whenever the old *id* was encountered. Subsequently, all collections were re-imported. In this way, the linkage was now based on the MongoDB `_id`, and no problems were encountered at all.

5.4.2. Graph DB

Populating Neo4j (i.e. the adopted graphDB) proved to be a bit more challenging. Unlike MongoDB, the import process from JSON is not as straightforward. Moreover, the structure defined earlier was specific to the MongoDB collections. To address this, Python scripts were developed. After establishing a connection with the Neo4j DBMS using the Neo4j driver for Python, these scripts initiate the creation of all the nodes before establishing the relationships between them. This process relies on both *Cypher* queries and the information found in the JSON documents to construct the entire graph database from scratch.

6. Queries

Here are all the queries required to access the databases and implement *WeFood* functionalities. They are grouped into basic CRUD operations and more intricate aggregations or query suggestions.

6.1. CRUD operations

The set of fundamental operations includes creating, reading, updating, and deleting data within the databases.

6.1.1. Create

Creation operations are:

1. `Create a new User`: a new User signs up to *WeFood*;
2. `Create a new Post / Recipe`: a User uploads a new Recipe;
3. `Create a new Comment`: a User comments a Post;
4. `Create a new StarRanking`: a User rates a Post;
5. `Create a new Ingredient`: an Admin adds a new ingredient;
6. `Create a new following relationship`: a User follows another User;
7. `Create a new Recipe-Ingredient relationship`: a new recipe is created and contains an ingredient;
8. `Create a new User-Ingredient relationship`: a User uses an ingredient;
9. `Create a new Ingredient-Ingredient relationship`: an ingredient is used with another ingredient.

MongoDB

1. `Create a new user`:

```
db.User.insertOne({
    username: String,
    password: [HASHEDSTRING],
    name: String,
    surname: String
})
```

2. `Create a new Post`:

```
db.Post.insertOne({
    idUser: ObjectId("..."),
    username: String,
    description: String,
    timestamp: Long,
    recipe: {
        name: String,
        image: String,
        steps: [String, ...],
        totalCalories: Double,
        ingredients: [{
```

```
            name: String,
            quantity: Double,
        }, ...]
    }
})
```

3. Create a new Comment:

```
db.Post.updateOne({
    _id: ObjectId("...")
}, {
    $push: {
        comments: {
            idUser: ObjectId("..."),
            username: String,
            text: String,
            timestamp: Long
        }
    }
})
```

4. Create a new StarRanking:

```
db.Post.updateOne({
    _id: ObjectId("...")
}, {
    $push: {
        starRankings: {
            idUser: ObjectId("..."),
            username: String,
            vote: Double
        }
    }
})
```

5. Create a new Ingredient:

```
db.Ingredient.insertOne({
    name: String,
    calories: Double
})
```

Neo4j

1. Create a new User:

```
CREATE (u:User {
    _id: String,
    username: String
})
```

2. Create a new Recipe:

```
CREATE (r:Recipe {
    _id: String,
    name: String,
    image: String
})
```

5. Create a new Ingredient:

```
CREATE (i:Ingredient {
    _id: String,
    name: String
})
```

6. Create a new following relationship:

```
MATCH (u1:User {username: String}), (u2:User {username: String})
MERGE (u1)-[:FOLLOWS]->(u2)
```

7. Create a new Recipe-Ingredient relationship:

```
MATCH (r:Recipe {_id: String}), (i:Ingredient {name: String})
CREATE (r)-[:CONTAINS]->(i)
```

8. Create a new User-Ingredient relationship:

```
MATCH (u:User {username: String}), (i:Ingredient {name: String})
MERGE (u)-[r:USED]->(i) ON CREATE SET r.times = 1 ON MATCH SET r.times =
↪  r.times + 1
```

9. Create a new Ingredient-Ingredient relationship:

```
MATCH (i1:Ingredient {name: String}), (i2:Ingredient {name: String})
MERGE (i1)-[r:USED_WITH]->(i2) ON CREATE SET r.times = 1 ON MATCH SET r.times
↪  = r.times + 1
```

6.1.2. Read

Reading operations are:

1. Find User by username;
2. Find User Page by username;
3. Get all the Ingredients;
4. Find Ingredient by name;
5. Find Most Recent Top Rated Posts;
6. Find Most Recent Top Rated Posts by set of ingredients;
7. Find Most Recent Posts by minCalories and maxCalories;
8. Find Post by Recipe name;
9. Find Post by _id;
10. Find Banned Users;
11. Find Users Followed by a User;

12. `Find Followers of a User`;
13. `Find Friends of a User`: a User's friends are the Users that follow him/her and that he/she follows;
14. `Find Recipes by set of ingredients`;

MongoDB

1. `Find User by username`:

```
db.User.find({
    username: String
}, {
    posts: 0
})
```

2. `Find User Page by username`:

```
db.User.find({
    username: String
}, {
    username: 1,
    posts: 1,
    deleted: 1
})
```

3. `Get all the Ingredients`:

```
db.Ingredient.find({}, {
    _id: 0
})
```

4. `Find Ingredient by name`:

```
db.Ingredient.find({
    name: String
}, {
    _id: 0
})
```

5. `Find Most Recent Top Rated Posts`:

```
db.Post.find({
    timestamp: {
        $gte: Long
    }
}, {
    "recipe.name": 1,
    "recipe.image": 1
}).sort({
    avgStarRanking: -1
}).limit(limit)
```

6. Find Most Recent Top Rated Posts by set of ingredients:

```
db.Post.find({
    timestamp: {
        $gte: Long
    },
    "recipe.ingredients.name": {
        $all: [String, ...]
    }
}, {
    "recipe.name": 1,
    "recipe.image": 1
}).sort({
    avgStarRanking: -1
}).limit(limit)
```

7. Find Most Recent Posts by minCalories and maxCalories:

```
db.Post.find({
    timestamp: {
        $gte: Long
    },
    "recipe.totalCalories": {
        $gte: minCalories,
        $lte: maxCalories
    }
}, {
    "recipe.name": 1,
    "recipe.image": 1
}).sort({
    timestamp: -1
}).limit(limit)
```

8. Find Posts by Recipe name:

```
db.Post.find({
    "recipe.name": {
        $regex: String,
        $options: "i"
    }
}, {
    "recipe.name": 1,
    "recipe.image": 1
}).limit(10)
```

9. Find Post by _id:

```
db.Post.find({
    _id: ObjectId("..."),
}, {
```

```
    _id: 0,
    idUser: 0,
    "starRankings.idUser": 0,
    "comments.idUser": 0
})
```

10. Find Banned Users:

```
db.User.find({
    deleted: true,
    name: { $exists: true }
})
```

Neo4j

11. Find Users Followed by a User:

```
MATCH (u1:User {username: String})-[:FOLLOWS]->(u2:User)
RETURN u2
```

12. Find Followers of a User:

```
MATCH (u1:User)-[:FOLLOWS]->(u2:User {username: String})
RETURN u1
```

13. Find Friends of a User:

```
MATCH (u1:User {username: String})-[:FOLLOWS]->(u2:User)-[:FOLLOWS]->(u1)
RETURN u2
```

14. Find Recipes by set of ingredients:

```
MATCH (r:Recipe)-[:CONTAINS]->(i:Ingredient)
WHERE i.name IN [String, ...]
WITH r, COLLECT(i.name) AS ingredients
WHERE ALL(ingredient IN [String, ...]
          WHERE ingredient IN ingredients)
RETURN r
LIMIT 15
```

6.1.3. Update

Update operations are:

1. Update User's information: A User can update his/her password, name or surname;
2. Update Post: A User can update the description of a Post;
3. Update Comment: A User can update the text of a Comment;

MongoDB

1. Update User's information:

```
db.User.updateOne({
    _id: ObjectId("...")
```

```
}, {
    $set: {
        password: [HASHEDSTRING],
        name: String,
        surname: String
    }
})
```

2. Update Post:

```
db.Post.updateOne({
    _id: ObjectId("...")
}, {
    $set: {
        description: String
    }
})
```

3. Update Comment:

```
db.Post.updateOne({
    _id: ObjectId("..."),
    comments: {
        $elemMatch: {
            idUser: ObjectId("..."),
            timestamp: Long
        }
    }
}, {
    $set: {
        "comments.$.text": String
    }
})
```

6.1.4. Delete

Deletion operations are:

1. `Delete User`: Users have the option to delete their own profiles, bearing in mind that all *non-personal information* will be retained for statistical purposes. Once a profile is deleted, re-registration using the previous username is not permitted;
2. `Delete Post`;
3. `Delete Post from User`;
4. `Delete Comment`;
5. `Delete StarRanking`;
6. `Delete Recipe`;
7. `Delete following relationship`;
8. `Delete / Decrement User-Ingredient relationship`;

Deletions not allowed:

- `Delete Ingredient`: it is not possible to delete an Ingredient because otherwise all the Recipes that contain it would be inconsistent;
- `Delete Ingredient-Ingredient relationship`: this relationship is neither removed nor decremented when a Recipe is deleted, for statistical purposes.

MongoDB

1. `Delete User`: It is important to first mark the User as deleted before proceeding to erase his/her personal information.

```
db.User.updateOne({
    _id: ObjectId("...")
}, {
    $unset: {
        password: "",
        name: "",
        surname: "",
        posts: ""
    },
    $set: {
        deleted: true
    }
})
```

2. `Delete Post`:

```
db.Post.deleteOne({
    _id: ObjectId("...")
})
```

3. `Delete Post from User`:

```
db.User.updateOne({
    _id: ObjectId("...")
}, {
    $pull: {
        posts: {
            idPost: ObjectId("...")
        }
    }
})
```

4. `Delete Comment`:

```
db.Post.updateOne({
    _id: ObjectId("...")
}, {
    $pull: {
        comments: {
```

```
            username: String,
            timestamp: Timestamp
        }
    }
})
```

5. Delete StarRanking:

```
db.Post.updateOne({
    _id: ObjectId("...")
}, {
    $pull: {
        starRankings: {
            idUser: ObjectId("...")
        }
    }
})
```

Neo4j

1. Delete User:

```
MATCH (u:User {username: String})
DETACH DELETE u
```

6. Delete Recipe:

```
MATCH (r:Recipe {_id: String})
DETACH DELETE r
```

7. Delete following relationship:

```
MATCH (u1:User {username: String})-[r:FOLLOWS]->(u2:User {username: String})
DELETE r
```

8. Delete / Decrement User-Ingredient relationship:

```
MATCH (u:User {username: String})-[r:USED]->(i:Ingredient {name: String})
SET r.times = r.times - 1
WITH r
WHERE r.times = 0
DELETE r
```

6.2. Suggestions and Aggregations

In this section, more relevant queries are presented, categorized into two sub-sections: suggestions and aggregations. Suggestions queries propose new information to users, based on their preferences and the preferences of their friends. Aggregations queries, instead, offer statistical insights into the stored data.

6.2.1. Suggestions

1. Show Most / Least used Ingredients;
2. Show Most used Ingredients by a User;
3. Suggest users to follow: a User is suggested to follow the friends of his/her friends;
4. Suggest most popular combination of ingredients;
5. Suggest new ingredients based on friends' usage;
6. Suggest most followed users;
7. Find Users by Ingredient usage: find Users who have employed a particular ingredient most frequently.

Neo4j

1. Show Most / Least used Ingredients:

```
MATCH (u:User)-[r:USED]->(i:Ingredient)
RETURN i, SUM(r.times) AS times
ORDER BY times DESC
LIMIT 5

MATCH (u:User)-[r:USED]->(i:Ingredient)
RETURN i, SUM(r.times) AS times
ORDER BY times ASC
LIMIT 5
```

2. Show Most used Ingredients by a User:

```
MATCH (u:User {username: String})-[r:USED]->(i:Ingredient)
RETURN i, r.times AS times
ORDER BY times DESC
LIMIT 5
```

3. Suggest users to follow:

```
MATCH (u1:User {username:
↪  String})-[:FOLLOWS]->(u2:User)-[:FOLLOWS]->(u3:User)
WHERE (u2)-[:FOLLOWS]->(u1)
AND NOT (u1)-[:FOLLOWS]->(u3)
RETURN u3
LIMIT 10
```

4. Suggest most popular combination of ingredients:

```
MATCH (i1:Ingredient {name: String})-[r:USED_WITH]->(i2:Ingredient)
RETURN i1, i2, r.times AS times
ORDER BY times DESC
LIMIT 5
```

5. Suggest new ingredients based on friends' usage:

```
MATCH (u1:User {username:
↪  String})-[:FOLLOWS]->(u2:User)-[r:USED]->(i:Ingredient)
```

```
WHERE (u2)-[:FOLLOWS]->(u1)
AND NOT (u1)-[:USED]->(i)
RETURN i, r.times AS times
ORDER BY times DESC
LIMIT 5
```

6. Suggest most followed users:

```
MATCH (u1:User)-[:FOLLOWS]->(u2:User)
RETURN u2, COUNT(u1) AS followers
ORDER BY followers DESC
LIMIT 5
```

7. Find Users by Ingredient usage:

```
MATCH (u:User)-[r:USED]->(i:Ingredient {name: String})
RETURN u, i, r.times AS times
ORDER BY times DESC
LIMIT 10
```

6.2.2. Aggregations

(#1): Compute the *ratio of interactions* and the *average* avgStarRanking by distinguishing between posts with and without images (i.e. no field `image` inside `recipe`). For the Posts with images, the ratio of interactions are computed as follows:

$$ratioOfComments = \frac{TotNumberOfComments}{TotNumberOfPosts}$$

$$ratioOfStarRankings = \frac{TotNumberOfStarRankings}{TotNumberOfPosts}$$

here $TotNumberOfComments$, $TotNumberOfStarRankings$ and $TotNumberOfPosts$ are computed by considering only the Posts with images. Similarly, the same calculations are performed for Posts without images.

```
db.Post.aggregate([
    {
        $project: {
            _id: 1,
            hasImage: {
                $cond: {
                    if: {
                         $eq: [{ $type: "$recipe.image" }, "missing"]
                    },
                    then: false,
                    else: true
                }
            },
            comments: {
```

```
                $size: {
                    $ifNull: ["$comments", []]
                }
            },
            starRankings: {
                $size: {
                    $ifNull: ["$starRankings", []]
                }
            },
            avgStarRanking: {
                $ifNull: ["$avgStarRanking", 0]
            }
        }
    },
    {
        $group: {
            _id: "$hasImage",
            numberOfPosts: {
                $sum: 1
            },
            totalComments: {
                $sum: "$comments"
            },
            totalStarRankings: {
                $sum: "$starRankings"
            },
            avgOfAvgStarRanking: {
                $avg: "$avgStarRanking"
            }
        }
    },
    {
        $project: {
            _id: 0,
            hasImage: "$_id",
            ratioOfComments: {
                $divide: ["$totalComments", "$numberOfPosts"]
            },
            ratioOfStarRankings: {
                $divide: ["$totalStarRankings", "$numberOfPosts"]
            },
            avgOfAvgStarRanking: 1
        }
    }
])
```

(#2): Given a User, show the *number* of Comments and StarRankings he/she has done and the

average of this StarRankings.

```
db.Post.aggregate([
    {
        $match: {
            $or: [
                {"comments.username": String},
                {"starRankings.username": String}
            ]
        }
    },
    {
        $project: {
            filteredComments: {
                $filter: {
                    input: "$comments",
                    as: "comment",
                    cond: {$eq: ["$$comment.username", String]}
                }
            },
            filteredStarRankings: {
                $filter: {
                    input: "$starRankings",
                    as: "starRanking",
                    cond: {$eq: ["$$starRanking.username", String]}
                }
            }
        }
    },
    {
        $group: {
            _id: null,
            avgOfStarRankings: {
                $avg: {$sum: "$filteredStarRankings.vote"}
            },
            numberOfStarRankings: {
                $sum: {$size: "$filteredStarRankings"}
            },
            numberOfComments: {
                $sum: {$size: "$filteredComments"}
            }
        }
    },
    {
        $project: {
            _id: 0,
            numberOfComments: 1,
```

```
            numberOfStarRankings: 1,
            avgOfStarRankings: 1
        }
    }
])
```

(#3): After filtering the Recipes by name, retrieve the *average* amount of calories of the first 10 Recipes ordered by descending avgStarRanking.

```
db.Post.aggregate([
    {
        $match: {
            "recipe.name": { $regex: String, $options: "i" }
        }
    },
    {
        $sort: {
            avgStarRanking: -1
        }
    },
    {
        $limit: 10
    },
    {
        $group: {
            _id: null,
            avgOfTotalCalories: {
                $avg: "$recipe.totalCalories"
            }
        }
    },
    {
        $project: {
            _id: 0,
            avgOfTotalCalories: 1
        }
    }
])
```

(#4): Given a User, show the *average* `totalCalories` of the Recipes published by him/her.

```
db.Post.aggregate([
    {
        $match: {
            username: String
        }
    },
    {   $project: {
```

```
            recipeCalories: "$recipe.totalCalories"
        }
    },
    {   $group: {
            _id: null,
            avgCalories: {
                $avg: "$recipeCalories"
            }
        }
    }
])
```

7. DataBases Deployment

As previously mentioned, MongoDB and Neo4j are the chosen databases for the deployment phase, serving as the documentDB and graphDB, respectively. The subsequent discussion will focus on the precautions taken when deploying these databases in a real-world setting.

7.1. MongoDB

MongoDB is the first database to be deployed, and it has been set up in a cluster with *three* machines to improve fault tolerance.

7.1.1. ReplicaSet

To deploy MongoDB across different machines, a *ReplicaSet* was necessary. A ReplicaSet is a collection of `mongod` instances organized hierarchically to ensure *redundancy and high availability*. The ReplicaSet, named `lsmdb`, comprises a Primary node with a priority of `2` and two Secondary nodes with priorities of `1.5` and `1`. The Primary node is the sole member capable of receiving write operations. Once a write operation is executed on the Primary node, the Secondary nodes replicate the same operation in their datasets. Configuring the ReplicaSet settings allows specifying the desired level of acknowledgment for a write operation, known as the *write concern*. In the proposed scenario of *WeFood*, having a social network that handles non-sensitive data, waiting for all nodes to acknowledge the operation is unnecessary. In the event of a primary node that crash immediately after a write operation, resulting in the failure to replicate to secondary nodes, the data loss is acceptable. Users can simply redo the operation (after the secondary nodes have elected a new primary node) without encountering dangerous or critical consequences. Therefore, a write concern of `w: 1` is adopted, allowing the ReplicaSet to return control once *one* node has acknowledged the write operation.

Additional configurable options include `j` and `wtimeout`.

- The `j` option determines when a node acknowledges writes, either after applying the write operation in memory or after writing to the *on-disk journal*. The default, when `w: 1`, is as if it were specified as false, acknowledging writes after the write in memory.
- The `wtimeout` specifies the time limit for an operation, and the default is `0`, meaning the write operation will block indefinitely if the level of write concern is unachievable. This situation will be handled in the code by setting a *client-level* `wtimeout` of `5000ms` (5 seconds), causing an exception to be thrown if the write operation is not completed within this time frame.

Regarding reading operations, all members of the replica set *can* accept read operations, although by *default*, applications direct reads to the primary member. In the case of a social network like *WeFood*, read operations can be directed to secondary nodes, even if they are not as updated as the primary node, as MongoDB asynchronously updates data to the secondary nodes. Thus, to ensure the lowest response time for read operations, the *read concern* is set to `nearest` at the *client-level*, meaning read operations will be performed on the nearest node (i.e. the node with the lowest latency).

7.1.2. Sharding

When it comes to *Sharding*, it is crucial to assess before the potential benefits it can bring. Sharding is the horizontally partitioning of data across multiple servers, that can help in enhancing scalability and performance. However, in certain situations, opting for sharding may prove impractical or undesirable. For instance, in the current implementation of *WeFood*, sharding the Post collection based on a specific field, such as `timestamp`, could result in latency issues when querying with unrelated filters (e.g., by `totalCalories` or by `avgStarRanking`), leading to an inefficient process.

To elaborate a little further, if the application was designed with predefined *Categories* for Recipes, sharding the Post collection based on the `category` field might achieve balanced load distribution among shards. However, this approach was *intentionally avoided* to offer users the flexibility to explore diverse recipes without constraints. Users indeed, in the current implementation, can search for Recipes using *various filters*, discovering Recipes *beyond* fixed categories.

Similarly, the decision not to implement the Sharding for the User collection is justified by the fact that it is realistic to expect that the User collection will not grow as much as the Post collection. For this reason, the complexity of implementing and managing the Sharding for the User collection is deemed unnecessary.

In summary, while sharding can significantly enhance database performance, its appropriateness is strictly tied to the specific demands of the application. In this case, considerations regarding uncorrelated filters, diverse exploration, and distinct growth rates between Users and Posts *influence* the decision not to adopt a sharding approach.

7.1.3. Indexes and Constraints

Considering the different collections stored in MongoDB, the following indexes and constraints have been implemented:

- `Ingredient`:
  - `name`: basic index (ascending order) *and* unique constraint.
- `Post`:
  - `timestamp`: basic index (ascending order);
  - `recipe.totalCalories`: basic index (ascending order).
- `User`:
  - `username`: basic index (ascending order) *and* unique constraint.

Specifically:

```
--> Ingredient:name
```

```
db.Ingredient.createIndex( { "name": 1 }, { unique: true } )
```

Index	nReturned	executionTimeMillis	totalKeysExamined	totalDocsExamined
No	1	4	0	1911
Yes	1	0	1	1

Operation: Find Ingredient by `name`.

Reason: It becomes evident that the inclusion of an index on the `name` field can speed up the find operation. Furthermore, the drawbacks of adding this index are negligible, given the infrequency of writing operations on the Ingredient collection: only the admin has the authority to introduce new ingredients, and the expectation is that such additions will occur rarely. On the other hand, the reading operations like the one that has been analyzed are expected to be frequent among the Users because all the social network is based on Recipes and so on Ingredients. So, adding an index on the `name` field will improve the User experience.

```
--> Post:timestamp
```

```
db.Post.createIndex( { "timestamp": 1 } )
```

Index	nReturned	executionTimeMillis	totalKeysExamined	totalDocsExamined
No	100	109	0	231323
Yes	100	3	100	100

Operation: Find the 100 most recent Posts.

```
--> Post:recipe.totalCalories
```

```
db.Post.createIndex( { "recipe.totalCalories": 1 } )
```

Index	nReturned	executionTimeMillis	totalKeysExamined	totalDocsExamined
No	100	144	0	231323
Yes	100	7	100	100

Operation: Find 100 Posts with totalCalories between `minCalories` and `maxCalories`.

Reasons: Given the substantial volume of the Post collection, it is necessary to define specific indexes, as outlined above. This approach ensures that operations frequently executed by users of *WeFood* yield significant benefits. Since a larger number of users browse posts compared to those creating new posts, the presence of these indexes is justified, despite the potential slowdown in write operations required to maintain them.

It's worth noting a subtle aspect: the decision not to define an index on the `avgStarRanking` field. This choice is thoughtful because this field is updated every time a user rates a post. Introducing an index on `avgStarRanking` would necessitate frequent updates with every rating, resulting in more write operations than the potential benefits conferred by the index.

In contrast, the fields `totalCalories` and `timestamp` remain the same after the initial Post creation. As a result, the indexes defined on them experience updates only once. This distinction is crucial for optimizing the overall performance of the system.

```
--> User:username
```

```
db.User.createIndex( { "username": 1 }, { unique: true } )
```

Index	nReturned	executionTimeMillis	totalKeysExamined	totalDocsExamined
No	1	78	0	27901
Yes	1	2	1	1

Operation: Find User by `username`.

Reason: The inclusion of an index on the `username` field for the Users, as illustrated in the table, proves highly beneficial in significantly reducing the execution time of queries involving `username` lookup. This optimization is particularly valuable during the login phase and when Users or administrators need to access a User Profile. Examining potential drawbacks in this scenario, it becomes evident that the index does not require frequent updates. Specifically, updates are unnecessary after the creation (i.e., sign up) of a new user, as users cannot change their usernames. Similarly, updates are not required when a user decides to delete their account from the platform. This is because the `username` information is retained even after account deletion, preventing other users from signing up with the same `username`.

7.2. Neo4j

The deployment of Neo4j was more straightforward compared to the previous setup. This ease was attributed to the deployment on a *single machine*, whereas setting up a cluster on multiple machines with replicas would have required the enterprise edition of Neo4j.

7.2.1. Indexes

The indexes implemented in Neo4j are:

- `Ingredient`:
  - `name`: text index.
- `Recipe`:
  - `_id`: text index.
- `User`:
  - `username`: text index.

More in detail:

`--> Ingredient:name`

```
CREATE TEXT INDEX ingredient_index FOR (i:Ingredient) ON (i.name);
```

Index	Total DB hits	ms
No	3826	74
Yes	5	5

Operation: Find Ingredient by `name`.

Reason: As for MongoDB, introducing an index on the `name` field of the Ingredients helps to improve the User experience. Indeed, all the suggestions regarding the Ingredients will receive a boost after the introduction of this index.

--> Recipe:_id

```
CREATE TEXT INDEX recipe_index FOR (r:Recipe) ON (r._id);
```

Index	Total DB hits	ms
No	462651	202
Yes	6	14

Operation: Find Recipe by `_id`.

Reason: The introduction of such an index is justified by the fact that this `_id` corresponds to the `_id` of the Post that contains the Recipe that is stored in MongoDB. By doing this, the process of finding a Recipe by `_id` is optimized as it is in MongoDB, where `_id` is by default indexed, being the primary key of the collection. Improvement in the performance of the system will be evident both in the creation phase and in the deletion phase of the Recipes, during which a particular Recipe is searched by `_id` among all the Recipes stored in the graph DB.

--> User:username

```
CREATE TEXT INDEX user_index FOR (u:User) ON (u.username);
```

Index	Total DB hits	ms
No	55808	65
Yes	5	12

Operation: Find User by `username`.

Reason: locating a specific User by `username` is a prerequisite for various features in *WeFood*. These operations, along with those involving *suggestions*, are also connected to the *friendship system* provided to Users. Updating this index is not a relevant issue since Users cannot modify their usernames. The index only needs updates in the rare event of Users deleting their profiles. Considering the infrequency of such occurrences, the introduction of this index is expected to bring about more benefits than drawbacks.

7.3. Consistency, Availability and Partition Tolerance

The trio of properties highlighted in the title represents the fundamental characteristics of a distributed system. However, as the *CAP theorem* states, achieving all three simultaneously in a functional system is deemed impossible. Aligned with the predetermined Non-Functional Requirements (Section **2.2.**), it becomes necessary to prioritize the *Availability* and *Partition Tolerance* of the system, allowing for a measured relaxation in *Consistency constraints*. This strategic approach allows to the primary system actors, the Users, to persist in utilizing the application even if certain displayed information is *not entirely up-to-date*. In practical terms, this might translate to scenarios such as not immediately viewing the latest Recipes (e.g., during a network partition) or encountering non-updated suggestions after the upload of a new Recipe (e.g., if Neo4j is temporarily down). Nevertheless, Users are assured of *eventually* accessing the most recent information. The timeline for this "eventually" remains uncertain, as consistency updates may

employ diverse approaches, and in the event of significant faults, human operator intervention may be necessary.

This is the main idea behind the design of the system. Consequently, the intentional alignment of the design with the intersection of Availability and Partition Tolerance in the CAP theorem places considerable emphasis on achieving eventual consistency. The subsequent paragraphs will delve into a detailed discussion of the system's consistency management, focusing particularly on the *Intra-Database* and *Inter-Database* consistency.

7.4. Intra-Database Consistency

Ensuring consistency within a database is crucial for maintaining the integrity of stored data, particularly in cases of *redundancies*, as evident in the *WeFood* Database Model. MongoDB demands particular attention for the management of redundancy updates, while Neo4j, lacking redundancies among nodes, requires careful consideration only for *inter-database consistency*, that will be discussed later on.

Turning attention to MongoDB collections, a deliberate decision has been made to prioritize updating the Post collection *first*, in scenarios involving multiple operations for redundancy updates. This choice is justified by the need to *consistently update* the Post collection to avoid compromising the User experience, especially if subsequent operations encounter issues. For instance:

- Creation or Deletion of a Post: the Post collection is updated first, followed by the update of redundancy in the User collection;
- Deletion of a User Account: after removing all Posts associated with the User in the Post collection, the User collection is updated by eliminating all the non-personal information.

When ranking a Post, the entire process is contained within the Post collection. Specifically, when a User adds a new vote to a Post, a *practical approximation* is employed to update the `avgStarRanking` field. Instead of re-evaluating all elements in the `starRankings` array, the server follows this sequence:

1. when the User views the post, the server has already transmitted all star rankings;
2. upon receiving the new vote and the Post with all star rankings, the server first adds the new star ranking to the `starRankings` array field in the Post collection;
3. subsequently, the server calculates the updated average star ranking by considering the User's provided copy of the starRankings array, along with the new vote;
4. the `avgStarRanking` field is then directly updated on the Post collection.

While recognizing that this approach may result in a temporarily stale `avgStarRanking` due to the local client copy *and* potential votes occurring after the User viewed the Post, it is understood that this discrepancy is more likely during the initial wave of votes but diminishes over time. As time progresses, the `avgStarRanking` is expected to *converge* to the correct value, ensuring accuracy as the upload time recedes into the past. This same approach applies to the deletion of a vote, with the understanding that the observations regarding the staleness of `avgStarRanking` remain pertinent.

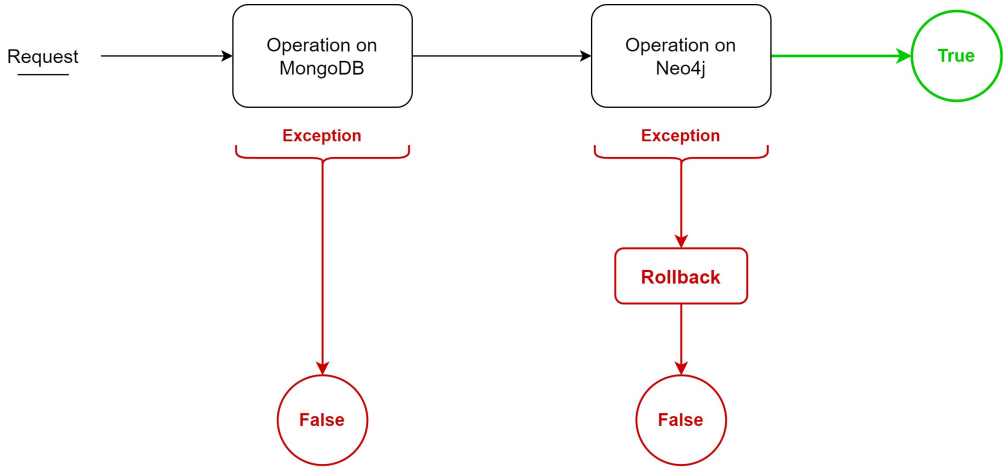

Figure 4: Strict Consistency scheme.

7.5. Inter-Database Consistency

A more complex issue than the previous one arises when it comes to maintaining consistency across different databases. The type of consistency guaranteed depends on the nature of the operation being executed.

1. For *operations necessitating inter-database consistency and performed by actors other than Registered Users* (i.e. the primary actors of the system), **strict consistency** is implemented. This includes:

   - *Registration of an Unregistered User*: prioritizing the verification of the setup's accuracy precedes the User's ability to utilize the application;
   - *Creation of an Ingredient by the Admin*: the emphasis lies on confirming the success of the process rather than providing a swift response.

   In cases of strict consistency, the databases are rolled back to their previous state if an operation fails. The depicted scheme (Figure 4) involves executing the operation first in MongoDB. If unsuccessful, no modifications are made, and a `False` response is returned. Upon success, the same operation is performed in Neo4j, introducing a minor delay for the User. If issues arise, MongoDB is rolled back to maintain consistency between the two databases, resulting in a `False` return. Successful execution yields a `True` return, signifying consistency between the databases.

2. Conversely, when dealing with Registered Users who require high availability and quick responses, even delicate operations demanding inter-database consistency are managed with **eventual consistency**. These operations include:

   - *Uploading or Deleting a Post*;
   - *Deleting a Registered User Account*.

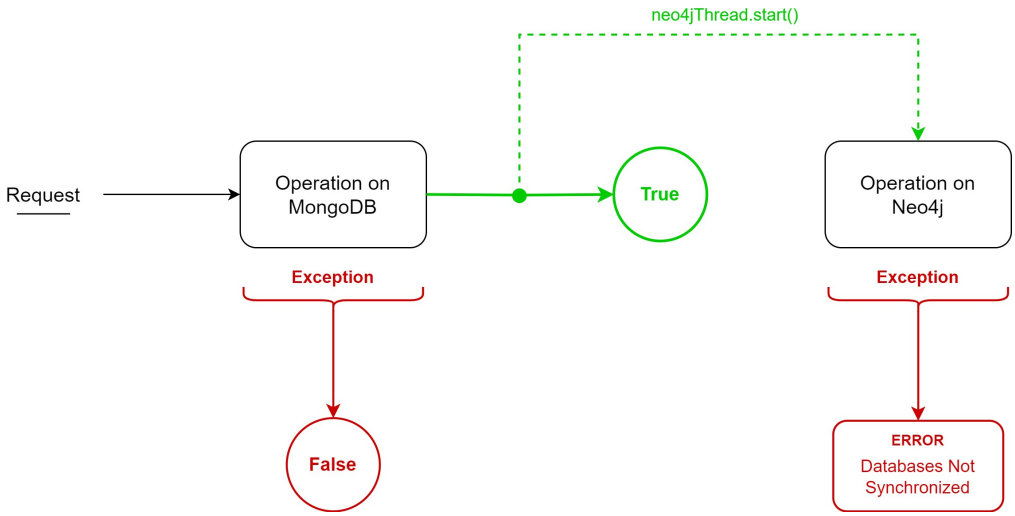

Figure 5: Eventual Consistency scheme.

Eventual consistency (Figure 5) is then employed in these scenarios. The operation begins in MongoDB, and if unsuccessful, no further action is taken before returning `False`. If MongoDB operation succeeds, an immediate `True` response is provided, eliminating the need for waiting. Subsequently, a Thread is initiated to handle consistency asynchronously with Neo4j. Since no critical information is stored in Neo4j, Users can continue using the application even if the two databases are temporarily inconsistent. If the Thread encounters issues, no action is taken, prioritizing system availability over consistency. Log messages or automatic triggers can be scheduled for future consistency recovery. Users can seamlessly use the application, and if Neo4j is down, they can still access features unrelated to Neo4j. Human operators, notified by log messages, are responsible for rectifying the situation if necessary. If both MongoDB and Neo4j operations succeed, no further intervention is required.

---

aggiungere link github e rendere pubblico il repository + inserire credenziali utente e admin in manuale utente

8. Implementation

8.1. System Architecture - Frameworks and components

The architecture of your application is composed by two main component: -Client -Server

Firstly, let's take a closer look at the Client component. The client serves as the interface through which users interact with your application. It plays a crucial role in initiating communication with the server by sending HTTP requests. The client component is responsible for creating a seamless and intuitive user experience, encapsulating the presentation and user interface logic.

Moving on to the Server component, it serves as the backbone of your application, handling the requests received from the client. The server adopts a Model-View-Controller (MVC) architecture, a widely adopted design pattern that promotes a modular and organized approach to software development. In this context, the Model represents the data and business logic and the Controller manages the flow of information. Is it common, like in this case, to no have the view, because we make use of APIs to manage the comunication between client and server.

The server's role is not only to process the incoming requests but also to efficiently manage the data and comunicate with our databases. The use of MVC helps to maintain a separation of concerns, making the codebase more modular, scalable, and easier to maintain. This design pattern contributes to the overall robustness and flexibility of the server-side architecture.

A key characteristic of your communication model is its adherence to RESTful principles. REST, or Representational State Transfer, is an architectural style that emphasizes a stateless and standardized approach to communication between systems. In your case, all information essential for communication is encapsulated within the body of the HTTP requests, formatted in JSON (JavaScript Object Notation).

This RESTful communication approach offers advantages such as scalability, flexibility, and simplicity. By embracing a stateless communication model, your application becomes more resilient, making it easier to scale horizontally as the user base grows.

In summary, the architecture of your application revolves around a well-defined interaction between the client and server components. The client, responsible for user interaction, initiates communication through HTTP requests, while the server, structured with an MVC architecture, efficiently processes these requests and manages the application's data and business logic. The adoption of RESTful communication, with information conveyed in JSON format, adds a layer of standardization and efficiency to the overall system, contributing to a robust and scalable application design.

8.1.1. Server

Our server is structured in this way: In the deepest part of our code, there are the two Base classes, found in the repository/base package, which act as a driver to handle query in a "coherent" way. In particular the BaseMongoDB class act as a string parser, reading all the strings provided by the classes above it and building step by step the Java queries each time the parser find a new component. The parser is also able to understand the content of the document to perform different queries. This is done so that from the classes above it will be possible to send queries exaclty how a human would write the queries in the mongo shell, so that each time a new query shoul be tested or

implemented, is not necessary to build the method from scratch, but it will be necessary just to send the string to the Base class and all the steps to create the query will be done inside the class.

Going above we have all the interfaces, which provides the structure of all the classes found in the mongodb and neo4j packages, which will implement all the queries as strings (this is possible for the reasons described before).

Going even further above, we have the DAOs and the DTOs. DAO classes will call all the methods found in the respective classes of the mongodb/neo4j packages. DTOs are more interesting in our case. They will provided a rapresentation of the classes found in the model package which is elaborated with respect to the original one. For instance the PostDTO rapresents what the user sees before clicking on a post, like Instagram feed page, before seeing comments or informations about the post, every user sees just the image. The image (and in our case the recipe name), is our PageDTO.

In the service package we find all the classes which, in some cases, also handle the consistency in the two databases and inside the single database. In general four cases of consistency management can be found: User creation consistency: If a user is created in MongoDB, we try the creation in Neo4j. If it fails in Neo4j, we display server-side that the databases are not synchronized. Ingredient creation consistency: The same as described in the case of the user. Post consistency: During the creation of a post, first of all we try the creation in MongoDB in the Post collection. After this we update the redunducies in the User collection. If the latter fails, an inconsistency in MongoDB is displayed server-side (we know the updated version of the database can be found in the Post collection). After all this steps, we have to create nodes and the relations in Neo4j. This are the Recipe node, the ingredient-ingredient relations, the user-ingredient relation and the recipe-ingredient relation. We try as the first step, to create the node. If this fails, we handle the case of the failed operation in Neo4j (described at the end). If this goes through, we start the creation of the relations. If one of this creations fails, we try to delete the recipe node and we display that Neo4j remains consistent but with a failed operation. If the ingredient-ingredient operation fails, the deletion of the recipe-ingredient relation is not handled separately. In fact the deletion of the recipe node is done in a way so that also all the relations are delete (DETACH DELETE). If the creation of user-ingredient fails, the ingredient-ingredient relations are not deleted, because utilized just for statistical purposes, and the user will never directly see that his informations are not precise. But this is not a problem since the most important thing is to show to the user some “good” informations that makes him stay active in the social network. If the creation in Neo4j of all the previous described steps fails, we try a rollback in MongoDB. We try to remove the post from the User collection. If the operation fails, we display that the databases are not consistent. Otherwise, we try to delete the Post also in the Post collection. If the operation fails, we display that MongoDB is not consistent and that the databases are not synchronized. Otherwise we only display that the operations was not completed successfuly (databases in this case are synchronized and consistent).

```
During the deletion of a post, we start from deleting informations from MongoDB, in particular from
```

```
There are several techniques to handle the inconsistency (eventual consistency) wich can remain from
```

Last but not least, controller package and the apidto package manage all the comunication with the client and provide a possibility for the client to interact with all the classes described before.

BaseMongoDB: The Driver to deal with MongoDB queries The BaseMongoDB class (`it.unipi.lsmsdb.wefood.repository.base.BaseMongoDB`) is an abstract Java class providing a comprehensive suite of functionalities to interact with MongoDB. It estabilishes connections, executes queries and handles data manipulation tasks like insterions, updates and deletions. The code is structured to support various MongoDB operations through a unified interface, offering a unique method that takes as input parameter the query in the MongoShell format.

Connection handling At the core, the class manages a MongoDB connection using the MongoClient instance. It employs the Singleton pattern to ensure that only one instance of MongoClient exists. This approach prevents unnecessary multiple connections to the database. The connection details, such as host addresses and database name, are configured as static final strings. In details, these are how parameters have been setted:

- `private static final String MONGODB_DATABASE = "WeFood";`
- `private static final String WRITE_CONCERN = "1";` This parameter is set to "1", indicating that the write operation will be considered successful as soon as the data is written to the primary node in the MongoDB cluster. This level of write concern balances between performance and data safety, ensuring that each write operation is acknowledged by the primary node, thus offering a moderate level of data durability without the overhead of waiting for multiple nodes to acknowledge.
- `private static final String WTIMEOUT = "5000";` The write timeout is configured to 5000 milliseconds (5 seconds). This setting specifies the maximum amount of time the server will wait for a write operation to be acknowledged. If the operation is not acknowledged within this timeframe, it will result in a timeout exception. This timeout value helps in maintaining a balance between application responsiveness and waiting for database operations, ensuring that the application does not hang indefinitely on slow write operations.
- `private static final String READ_PREFERENCE = "nearest";` The read preference has been set to "nearest", which directs the server to read from the nearest member (either primary or secondary) of the MongoDB cluster based on network latency. This setting is crucial for achieving low-latency read operations, as it ensures that the application reads data from the geographically closest node, thereby reducing network latency and improving the overall read performance.

Error handling Throughout the class, there's a consistent approach to error handling. The methods throw exceptions like MongoException, IllegalArgumentException, and IllegalStateException to signal failures during database operations. This exception-based approach ensures robust error reporting and handling. Error handling is entrusted to calling methods.

Query execution The central component of query execution in the `BaseMongoDB` class is the method `executeQuery`, designed to handle various MongoDB operations. This method receives a single string input, `mongosh_string`, which is crucial as it contains all the information necessary to determine the type of operation to perform and its specific parameters. The input string, `mongosh_string`, follows a format resembling a MongoDB shell command. It includes the collection name, the operation to be performed, and the parameters for that operation. Example format: `db.collection.operation({param1: value1, ...})`. The method starts by dissecting

the `mongosh_string` to extract essential components: the collection name and the operation details. The operation name (e.g., `find`, `insertOne`) is isolated, which guides the method to invoke the corresponding operation-specific method. The parameters for the operation are extracted from the remaining part of the `mongosh_string`. This part of the string represents the operation's arguments and is in JSON format. Depending on the operation, the parameters extracted from `mongosh_string` may need further processing. For example, in a `find` operation, the query parameters might need to be split into individual components like `find`, `project`, `sort`, and `limit`. This is achieved through various parsing methods, regular expressions, or JSON manipulations.

FIND Once the executeQuery method has isolated `operationDoc` for the find operation, it transitions into a tailored query execution process. Parsing this string involves converting the query parameters into a BSON format compatible with the MongoDB Java driver. The query criteria and any specified projection details are transformed into Document objects. Similarly, sorting instructions and the query limit are extracted and formatted appropriately. Once the query components are parsed and formatted, the find method executes the query against the specified MongoDB collection. The results are then collected and returned, providing a seamless bridge between the input string format and the MongoDB query execution.

AGGREGATE Handling the aggregate operation follows a similar pattern of precision and adaptation as the find operation, yet it caters to the more complex nature of aggregation in MongoDB. After the executeQuery method identifies and isolates the `operationDoc` specific to the aggregate operation, the next step involves interpreting and structuring this string for execution. The `operationDoc` for an aggregation contains a sequence of MongoDB aggregation pipeline stages, represented as an array of JSON objects. The primary task is to parse this string into a series of Bson objects, each corresponding to a stage in the MongoDB aggregation pipeline. The parsing process ensures that stages as `$match`, `$group`, `$sort` and others are accurately converted from their string representation into Bson objects, which are the format required by the MongoDB Java driver. Once the stages are parsed and organized, the aggregate method proceeds to execute this pipeline against the specified MongoDB collection. The result of this aggregation is a list of Document objects, each representing a record in the final aggregated output.

INSERT ONE The `operationDoc` contains the data for the document to be inserted in a JSON-like format. The core task here is to parse this data string into a Document object. Once the data is parsed into a Document, the insertOne method is invoked. The operation results in an `InsertOneResult`, which includes information about the success of the operation (the `_id` of the inserted document).

UPDATE ONE For the updateOne operation in the BaseMongoDB class, the process is tailored to update a single document in a MongoDB collection. The critical step here is to parse the `operationDoc` into two main components: the filter criteria and the update details. The filter criteria determine which document in the collection will be updated, while the update details specify how the document should be modified.

- The filter part of `operationDoc` is converted into a Document object. This conversion is essential to match the BSON format expected by the MongoDB Java driver.
- The update portion, potentially containing various update operators (`$set`, `$unset`, `$push`, `$pull`), is also parsed into a BSON format. Each operator and its corresponding data need to be accurately represented to ensure the update is performed correctly.

With these components structured correctly, the updateOne method proceeds to execute the update operation. This involves calling the appropriate method from the MongoDB Java driver, passing the filter criteria and update details. The operation results in an `UpdateResult`, which includes information about the success of the operation (the number of documents updated).

DELETE ONE When the executeQuery method identifies a `deleteOne` operation, it shifts its focus to handling the `operationDoc`, which in this case contains the criteria for selecting the document to be deleted. This conversion is achieved by parsing the string into a Document object. With the deletion criteria correctly formatted, the deleteOne method executes the delete operation using the MongoDB Java driver. The result of this operation is a `DeleteResult` object, which provides information about the outcome (the number of documents deleted).

8.1.1.2. Models

More details on the classes and their attributes are as follows.

Admin:

- username: `String`
- password: `String` (hashed)

RegisteredUser:

- username: `String`
- password: `String` (hashed)
- name: `String`
- surname: `String`

Post:

- user: `RegisteredUser`
- description: `String`
- timestamp: `Date`
- comments: `List<Comment>`
- starRankings: `List<StarRanking>`
- recipe: `Recipe`

Comment:

- user: `RegisteredUser`
- text: `String`
- timestamp: `Date`

StarRanking:

- user: `RegisteredUser`
- vote: `Double`

Recipe:

- name: `String`
- image: `String`
- steps: `List<String>`
- ingredients: `Map<Ingredient, Double>`

Ingredient:

- name: `String`
- calories: `Double`

DI QUESTO HO SCRITTO SOPRA NELLA DESCRIZIONE DEL DRIVER, EVENTUALMENTE CONTROLLARE Fare riferimento a deployment database private static final String MONGODB_DATABASE = “WeFood”; private static final String WRITE_CONCERN = “1”; private static final String WTIMEOUT = “5000”; private static final String READ_PREFERENCE = “nearest”; private static final String mongoString = String.format(“mongodb://%s/%s/?w=%s&wtimeout=%s&readPreference= MONGODB_HOST, MONGODB_DATABASE, WRITE_CONCERN, WTIMEOUT, READ_PREFERENCE);

Here in the server try to describe also the Driver that we have implemented.

8.1.2. Client

For the Client implementation, the project was build using HTML, CSS and TypeScript. HTML, or HyperText Markup Language, is the standard markup language for creating web pages.

CSS, or Cascading Style Sheets, complements HTML by providing styling and layout.

Now, let’s move on to AngularJS. AngularJS is a JavaScript framework developed by Google, designed to make both the development and testing of web applications easier. It extends HTML with new attributes and binds data to HTML with expressions, making it a powerful tool for creating dynamic and interactive user interfaces.

In AngularJS, our application is organized into various components. Components are the building blocks of the user interface, each encapsulating a specific part of the application’s functionality.

Additionally, AngularJS employs services, which are reusable, injectable objects that perform specific functions across the application. Services are ideal for encapsulating shared functionality, such as data retrieval or business logic, and promoting modularity in our codebase.

Implementation Example for Login-Home and Home Functionalities:

The central component in this implementation is the HomeComponent, serving as the foundational element of the home page. Its corresponding template, home.component.html, delineates the page structure, encompassing a navigation bar, a central content area, and a footer. Noteworthy is the seamless integration of the LoggingPopupComponent, a critical component responsible for user authentication.

When a user initiates the “LOGIN” button in the navigation bar, the openPopup() method within the HomeComponent sets the showLoginPopup property to true. This action triggers the rendering of the LoggingPopupComponent through the *ngIf directive in the template. The popup is not merely a static UI element; it actively facilitates user authentication.

The LoggingPopupComponent is intricately linked with the RegisteredUserService. Upon a successful login attempt, this component emits the closePopup event. Subsequently, the HomeComponent responds by capturing this event, interpreting it as a signal to close the popup, and simultaneously redirecting the user to the registered user feed.

In the background, the PostService assumes a pivotal role, managing a range of post-related functionalities. From uploading and modifying posts to deletion operations, this service communicates with the backend server via HTTP requests. The browseMostRecentTopRatedPosts method stands out, fetching posts based on parameters such as recency and popularity, thereby addressing the dynamic content requirements of the application.

Adding another layer of complexity is the browse-most-recent-top-rated-post-by-ingredients component. This specialized component, along with its template (browse-most-recent-top-rated-post-by-ingredients.component.html), offers users a dedicated space for post filtering based on ingredient names. It includes input fields for dynamic entry and sliders to fine-tune temporal and post limit criteria.

While navigating the application's source code, a consistent theme of managing asynchronous operations becomes apparent. The isLoading property in HomeComponent ensures a seamless transition by displaying a loading spinner during data retrieval. Thoughtful error handling mechanisms are implemented to safeguard the user experience against unforeseen issues.

Reflecting on the unit tests, especially those for the BrowseMostRecentTopRatedPostByIngredientsComponent, underscores a commitment to quality assurance. These tests serve as vigilant checks, validating the successful creation of components and reinforcing the application's reliability.

In the client we can find actor classes, which inside have the main shell with all the commands for the users. Also we have methods wich guides the user through the steps to complete an operation an this methods also call the classes found in httprequest. The latters sends an http request to the server and if the status code is 200 (ok), a conversion from the response body to the desired object is perfomed. This objects can be found in model, dto, or apidto. The print of the object is done at the end inside the method called before by the Printer or Java print (System.out.println).

8.2. Future Works

I have a login api in java spring and i want that other apis are accessible only after the login is performed DIRE in breve come si doveva fare per rendere API non pubbliche public class SecurityConfig extends WebSecurityConfigurerAdapter { e poi dire come si è fatto e perchè, per semplificare l'implementazione....

Some possible future works that can be done to improve our application are the following: Firstly, one crucial step would be to implement Hypertext Transfer Protocol Secure (HTTPS), a secure communication protocol, to ensure a protected and encrypted experience for all users within the social network.

Then the utilization of additional replicas in Neo4j could be a way to procede. Presently, this is constrained by licensing limitations. By doing so, we can enhance the scalability and fault tolerance of our application.

Addressing the challenge of eventual consistency, as previously discussed during the server presentation, implementing one of the viable strategies for managing eventual consistency will contribute to a more reliable user experience.

Furthermore, to expand the feature set of our social network, it is necessary to introduce additional functionalities. To do so, optimizing our parser to handle these new queries will be fundamental to

achieve a more versatile and feature-rich social networking platform.

In conclusion, the future trajectory of our application involves not only ensuring security through HTTPS implementation but also overcoming licensing barriers to explore multiple replicas in Neo4j. Additionally, addressing issues related to eventual consistency and expanding the functional scope with new queries in MongoDB are pivotal steps in ensuring the sustained growth and improvement of our social network.

---

9. User Manual

This guide outlines the general approach to navigating and utilizing the features of our social network application. After this introduction, you'll find comprehensive instructions on interacting with the user interface, tailored for registered users, as well as an overview for non-registered users on accessing limited functionalities. A table presents all possible commands that users can input into the interface along with their corresponding outcomes.

For non-registered users, browsing recent posts sorted by upload date is permitted. However, all other operations, except for registration, are restricted until users complete the registration process. Upon registration, users are greeted with an empty personal profile page, zero followers/followed, and zero posts. The interface displays a personal shell prompting users to insert commands. The subsequent table details various commands and their corresponding results.

Command	Operation
`login`	To login
`logout`	To logout
`findIngredientByName`	To find an ingredient by name
`findIngredientsUsedWithIngredient`	To show suggestions about ingredients based on ingredient combinations
`findNewIngredientsBasedOnFriendsUsage`	To show suggestions about ingredients based on friends
`findUsersToFollowBasedOnUserFriends`	To show suggestions on new followers based on friends
`findMostFollowedUsers`	To show suggestions about the most followed users
`findUsersByIngredientUsage`	To show suggestions about users based on ingredients usage
`findMostUsedIngredientByUser`	To find the most used ingredient
`findMostLeastUsedIngredient`	To show an overview about ingredient usage
`uploadPost`	To upload a post
`modifyPost`	To modify a post
`deletePost`	To delete a post
`browseMostRecentTopRatedPosts`	To browse the most recent and top rated posts
`browseMostRecentTopRatedPostByIngredients`	To browse the most recent and top rated posts by ingredients
`browseMostRecentPostsByCalories`	To browse the most recent posts by calories
`findPostByRecipeName`	To find a post by recipe name
`averageTotalCaloriesByUser`	Statistics about calories
`findRecipeByIngredients`	To find a recipe by ingredients
`modifyPersonalInformation`	To modify personal informations
`deleteUser`	To delete your personal profile
`followUser`	To follow a user
`unfollowUser`	To unfollow a user
`findFriends`	To find your friends
`findFollowers`	To find your followers

Command	Operation
`findFollowed`	To find your followed users
`exit`	To exit from the site

After entering a command, the user will receive on-screen guidance to complete the operation. Accuracy in providing information is crucial; any inaccuracies will result in a failed operation, necessitating a restart. Notably, there is no back button available, meaning completed operations cannot be reversed. A pop-up message informs users of the success or failure of an operation, redirecting them to the main shell.

For post browsing, a folder is dynamically created/deleted upon issuing the relevant command, and the folder's path is displayed on the screen.

The admin of the social network is pre-registered, and credentials are communicated through various channels (e.g., voice, messages). The admin interacts with a personal shell using specific commands. Similar to regular users, the admin lacks a back button, and once an operation is correctly completed, it cannot be reversed.

The following table outlines admin-specific commands:

Command	Operation
`login`	To login
`logout`	To logout
`createIngredient`	To create a new ingredient
`banUser`	To ban a user from the site
`unbanUser`	To unban a user
`findIngredient`	To find information about a specific ingredient
`getAllIngredients`	To retrieve all information about ingredients
`findIngredientsUsedWithIngredient`	To find combinations of ingredients from a starting ingredient
`mostPopularCombinationOfIngredients`	Statistics about the most popular combinations of ingredients
`exit`	To close the application

This detailed guide aims to provide users, whether regular or admin, with a clear understanding of the social network's functionalities and the corresponding commands to interact effectively.

Da fare più mettere screen

10. References

[1] Calories in Food Items (per 100 grams) - https://www.kaggle.com/datasets/kkhandekar/calories-in-food-items-per-100-grams - Accessed: December 2023.

[2] Food.com Recipes and Interactions - https://www.kaggle.com/datasets/shuyangli94/food-com-recipes-and-user-interactions?select=PP_recipes.csv - Accessed: December 2023.

[3] Food.com - Recipes and Reviews - https://www.kaggle.com/datasets/irkaal/foodcom-recipes-and-reviews?select=reviews.csv - Accessed: December 2023.