Συστήματα Διαχείρισης Δεδομένων Μεγάλου Όγκου

Εργαστηριακή Άσκηση 2022/23

|  |  |  |
| --- | --- | --- |
| **Όνομα** | **Επώνυμο** | **ΑΜ** |
| Πάρης | Σεργιάννης | 1067467 |
| Στυλιανός | Στυλιανάκης | 1059713 |

Βεβαιώνω ότι είμαι συγγραφέας της παρούσας εργασίας και ότι έχω αναφέρει ή παραπέμψει σε αυτήν, ρητά και συγκεκριμένα, όλες τις πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών, προτάσεων ή λέξεων, είτε αυτές μεταφέρονται επακριβώς (στο πρωτότυπο ή μεταφρασμένες) είτε παραφρασμένες. Επίσης βεβαιώνω ότι αυτή η εργασία προετοιμάστηκε από εμένα προσωπικά ειδικά για το συγκεκριμένο μάθημα/σεμινάριο/πρόγραμμα σπουδών.

Έχω ενημερωθεί ότι σύμφωνα με τον εσωτερικό κανονισμό λειτουργίας του Πανεπιστημίου Πατρών άρθρο 50§6, τυχόν προσπάθεια αντιγραφής ή εν γένει φαλκίδευσης της εξεταστικής και εκπαιδευτικής διαδικασίας από οιονδήποτε εξεταζόμενο, πέραν του μηδενισμού, συνιστά βαρύ πειθαρχικό παράπτωμα.

|  |  |
| --- | --- |
| Υπογραφή    16 / 2 / 2024 | Υπογραφή    16 / 2 / 2024 |

**Συνημμένα αρχεία κώδικα**

Μαζί με την παρούσα αναφορά υποβάλλουμε τα παρακάτω αρχεία κώδικα

|  |  |  |
| --- | --- | --- |
| Αρχείο | Αφορά το ερώτημα | Περιγραφή/Σχόλιο |
| main.py | 2,3,4 | Σε αυτο το αρχείο καλούμε τις συναρτήσεις από το db\_functions.py, για να εκτελέσουμε τα inserts και τα selects στην βάση. Επιλέγουμε mode αλλάζοντας την μεταβλητή  operation\_mode. |
| db\_functions.py | 2,3,4 | Περιέχει όλες τις συναρτήσεις που αφορούν τις βασικές λειτουργίες της βασης, οι οποίες καλούνται από την main.py. |
| query\_generation.py | 2, 4 | Περιέχει τα table schemas και χρησιμοποιείται για την κατασκευή ερωτημάτων. |
| data\_shortening.py | 2,3,4 | Script που χρησιμοποιείται για την κατασκευή csv με ένα υποσύνολο δεδομένων, με σκοπό την ταχύτερη υλοποίηση των ερωτημάτων. |

# Τεχνικά χαρακτηριστικά περιβάλλοντος λειτουργίας

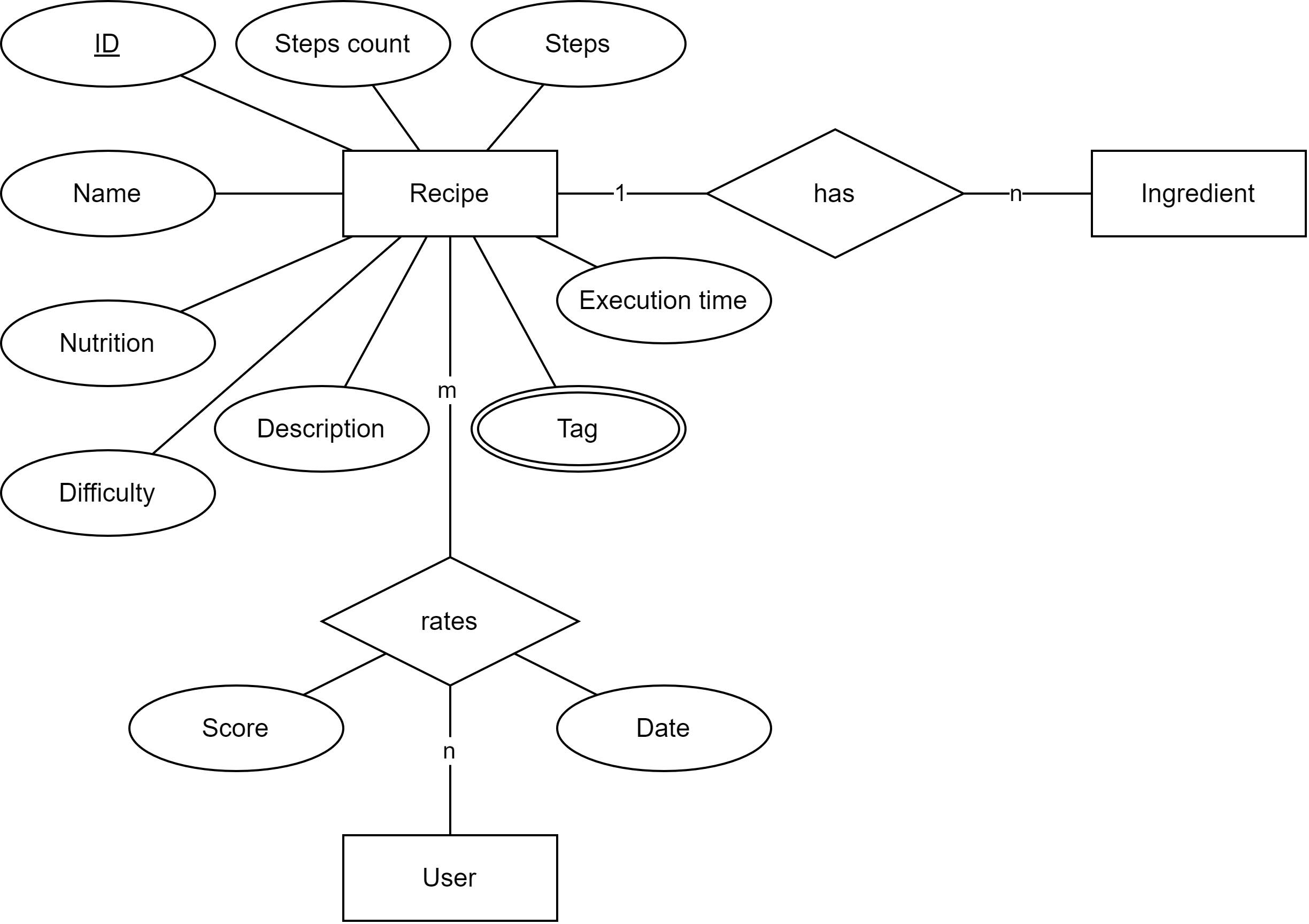
Για την υλοποίηση της εργασίας χρησιμοποιήθηκε αποκλειστικά το Astra DB.

# Ερώτημα 1: Σχεδιασμός ΒΔ

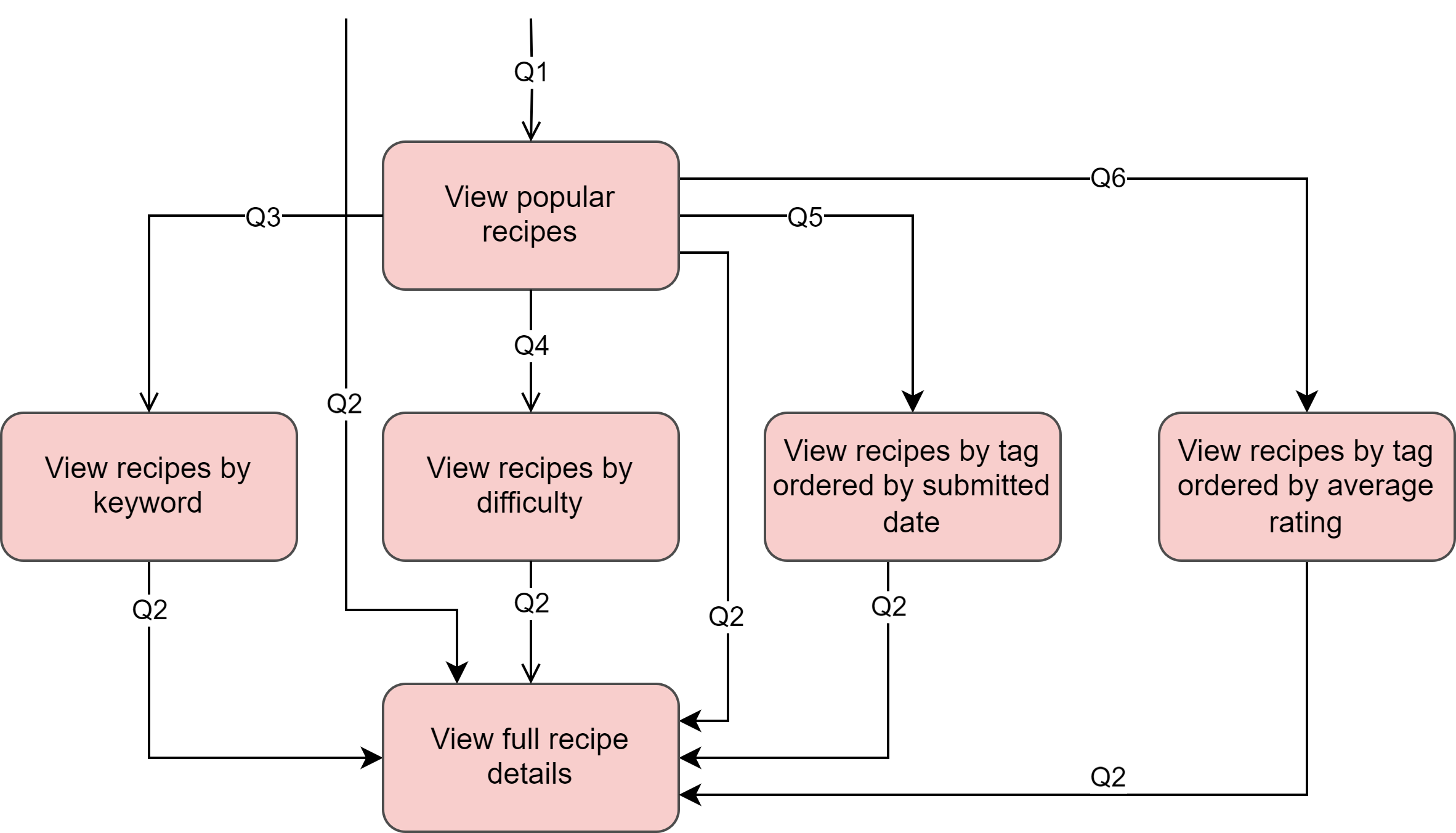
### Απαιτήσεις χρήστη:

1. Εμφάνιση των συνταγών που είναι δημοφιλείς (έχουν καλή βαθμολογία) εντός ενός χρονικού διαστήματος (π.χ. τους τελευταίους 3 μήνες) – αυτό μπορεί να αποτελεί κάλλιστα την αρχική οθόνη «προτάσεων» προς το χρήστη.
2. Να αναζητήσει την/τις συνταγές που περιέχουν κάποιες λέξεις – κλειδιά στον τίτλο.
3. Να αναζητήσει συνταγές με βάση την κατηγορία δυσκολίας (π.χ. χρόνος παρασκευής ή πλήθος βημάτων) και να τις λάβει με βάση κάποια ταξινόμηση (π.χ. μέση βαθμολογία.
4. Να δει τις λεπτομερείς πληροφορίες για κάποια συνταγή (κατηγορία δυσκολίας, μέση βαθμολογία, ετικέτες, υλικά, διαδικασία παρασκευής).
5. Να δει τις top-n συνταγές που σχετίζονται με κάποια ετικέτα.

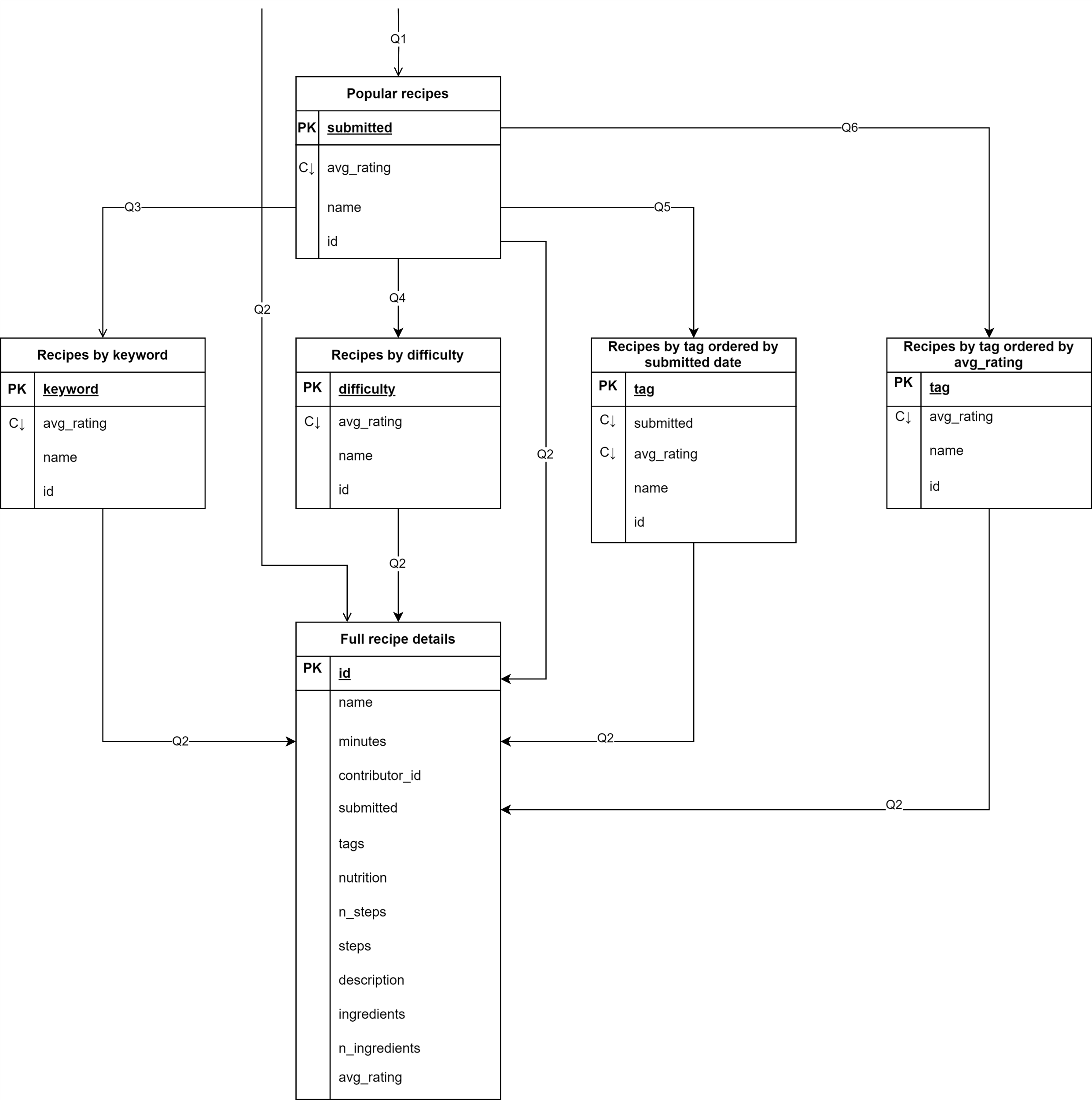
### Εννοιολογικό μοντέλο



### Application workflow



### Chebotko diagram



# Ερώτημα 2: Ερωτήματα DDL

[επαναλαμβάνετε τον παρακάτω πίνακα για κάθε keyspace στη ΒΔ σας]

|  |  |
| --- | --- |
| **Keyspace** | recipes.popular\_recipes |
| **DDL statement** | CREATE TABLE recipes.popular\_recipes (  id int,  submitted date,  avg\_rating float,  name text,  PRIMARY KEY (submitted, avg\_rating, id)  ) WITH CLUSTERING ORDER BY (avg\_rating DESC, id ASC); |
| **Screenshot** |  |

|  |  |
| --- | --- |
| **Keyspace** | recipes.recipes\_difficulty |
| **DDL statement** | CREATE TABLE recipes.recipes\_difficulty (  difficulty text,  id int,  avg\_rating float,  name text,  PRIMARY KEY (difficulty, avg\_rating, name)  ) WITH CLUSTERING ORDER BY (avg\_rating DESC, name ASC); |
| **Screenshot** |  |

|  |  |
| --- | --- |
| **Keyspace** | recipes.recipes\_tag\_rating |
| **DDL statement** | CREATE TABLE recipes.recipes\_tag\_rating (  tag text,  id int,  avg\_rating float,  name text,  PRIMARY KEY (tag, avg\_rating, id)  ) WITH CLUSTERING ORDER BY (avg\_rating DESC, id ASC); |
| **Screenshot** |  |

|  |  |
| --- | --- |
| **Keyspace** | recipes.recipes\_tag\_submitted |
| **DDL statement** | CREATE TABLE main.recipes\_tag\_submitted (  tag text,  submitted date,  avg\_rating float,  id int,  name text,  PRIMARY KEY (tag, submitted, avg\_rating, id)  ) WITH CLUSTERING ORDER BY (submitted DESC, avg\_rating DESC, id ASC); |
| **Screenshot** |  |

|  |  |
| --- | --- |
| **Keyspace** | recipes.recipes\_keywords |
| **DDL statement** | CREATE TABLE main.recipes\_keywords (  id int,  keywords set<text>,  avg\_rating float,  name text,  PRIMARY KEY (id, avg\_rating, name)  ) WITH CLUSTERING ORDER BY (avg\_rating DESC, name ASC); |
| **Screenshot** |  |

|  |  |
| --- | --- |
| **Keyspace** | recipes.recipes\_details |
| **DDL statement** | CREATE TABLE recipes.recipes\_details (  id int,  name text,  minutes int,  contributor\_id int,  submitted date,  tags set<text>,  nutrition list<float>,  n\_steps smallint,  steps list<text>,  description text,  ingredients set<text>,  n\_ingredients smallint,  avg\_rating float,  difficulty text,  keywords set<text>,  PRIMARY KEY (name)  ); |
| **Screenshot** |  |

# Ερώτημα 3: Απαντήσεις ερωτημάτων

*[Μην παραθέσετε στο έντυπο όλες τις επιστρεφόμενες εγγραφές! Να καταγράψετε μόνο αυτές που αναφέρει το πρότυπο.]*

|  |  |
| --- | --- |
| **Ερώτημα** | **Απάντηση** |
| Εμφάνιση των 30 συνταγών με την υψηλότερη μέση βαθμολογία μεταξύ 01/01/2012 και 31/05/2012 | *[παραθέστε τις 5 πρώτες μόνο]* |
| Εμφάνιση όλων των λεπτομερειών για την ταινία «chic greek salad» (κατηγορία δυσκολίας, διατροφικές αξίες, βήματα, περιγραφή, μέση βαθμολογία) | *[όλες τις λεπτομέρειες]* |
| Εμφάνιση των ταινιών της κατηγορίας «εύκολη» ταξινομημένες ως προς τη μέση βαθμολογία τους | *[παραθέστε τις 5 πρώτες μόνο]* |
| Εμφάνιση των συνταγών που περιέχουν την ετικέτα “slow-cooker” με ταξινόμηση ανά ημερομηνία προσθήκης (πιο πρόσφατες πρώτα) | *[παραθέστε τις 5 πρώτες μόνο]* |
| Εμφάνιση των 20 συνταγών με την υψηλότερη μέση βαθμολογία για την ετικέτα “cocktail”. | *[παραθέστε τις 5 πρώτες μόνο]* |

# Ερώτημα 4Α: Χρόνοι εισαγωγής δεδομένων

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Επίπεδο write consistency** | | |
| **ALL** | **QUORUM** | **ONE** |
| **[Keyspace 1]** | *[χρόνος εκτέλεσης]* | *[χρόνος εκτέλεσης]* | *[χρόνος εκτέλεσης]* |
| **[Keyspace 2]** | *[χρόνος εκτέλεσης]* | *[χρόνος εκτέλεσης]* | *[χρόνος εκτέλεσης]* |
| … | *…* | *…* | *…* |
| **[Keyspace n]** | *[χρόνος εκτέλεσης]* | *[χρόνος εκτέλεσης]* | *[χρόνος εκτέλεσης]* |
| **Μέσος όρος** |  |  |  |

# Ερώτημα 4Β: Χρόνοι ανάκτησης δεδομένων

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Επίπεδο write consistency** | | |
| **ALL** | **QUORUM** | **ONE** |
| **Ερώτημα 1** | *[μ.ο. για τις 10 επαναλήψεις]* | *[μ.ο. για τις 10 επαναλήψεις]* | *[μ.ο. για τις 10 επαναλήψεις]* |
| **Ερώτημα 2** | *[μ.ο. για τις 10 επαναλήψεις]* | *[μ.ο. για τις 10 επαναλήψεις]* | *[μ.ο. για τις 10 επαναλήψεις]* |
| **Ερώτημα 3** | *[μ.ο. για τις 10 επαναλήψεις]* | *[μ.ο. για τις 10 επαναλήψεις]* | *[μ.ο. για τις 10 επαναλήψεις]* |
| **Ερώτημα 4** | *[μ.ο. για τις 10 επαναλήψεις]* | *[μ.ο. για τις 10 επαναλήψεις]* | *[μ.ο. για τις 10 επαναλήψεις]* |
| **Ερώτημα 5** | *[μ.ο. για τις 10 επαναλήψεις]* | *[μ.ο. για τις 10 επαναλήψεις]* | *[μ.ο. για τις 10 επαναλήψεις]* |
| **Μέσος όρος** |  |  |  |

# Ερώτημα 4Γ: Σχολιασμός αποτελεσμάτων

*[Συνοψίστε τα αποτελέσματα των χρόνων εισαγωγής δεδομένων και ανάκτησης δεδομένων με κατάλληλες γραφικές παραστάσεις (δύο) και σχολιάστε τα ευρήματά σας – γιατί παρατηρούνται οι όποιες διαφορές στο χρόνο εκτέλεσης; Σε ποια στοιχεία της αρχιτεκτονικής της ΒΔ και του θεωρήματος CAP οφείλονται;]*

# Βιβλιογραφία

*[πηγές που χρησιμοποιήσατε για την εργασία]*

*Fatma Abdelhedi, Amal Ait Brahim, Faten Atigui, Gilles Zurfluh. Big Data and Knowledge Management: How to Implement Conceptual Models in NoSQL Systems?. 8th International Conference on*

*Knowledge Management and Information Sharing (IC3K 2016), Nov 2016, Porto, Portugal. pp.235-240, 10.5220/0006082302350240ff. hal-03193283f*

A graph of a bar chart

Description automatically generated with medium confidence

A graph of multiple colored bars

Description automatically generated with medium confidence