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  \_\_\_ / \_\_\_ / 2023 | Υπογραφή  \_\_\_ / \_\_\_ / 2023 |

**Συνημμένα αρχεία κώδικα**

Μαζί με την παρούσα αναφορά υποβάλλουμε τα παρακάτω αρχεία κώδικα

|  |  |  |
| --- | --- | --- |
| Αρχείο | Αφορά το ερώτημα | Περιγραφή/Σχόλιο |
| *Erotima1.py* | *1* | *Περιέχει όλα τα ερωτήματα για το ερ. 1* |
|  |  |  |
|  |  |  |

# Τεχνικά χαρακτηριστικά περιβάλλοντος λειτουργίας

*[Τεχνικά χαρακτηριστικά φυσικού Η/Υ που χρησιμοποιήθηκε για την εργασία, αν χρησιμοποιήθηκε μόνο το Astra DB μπορείτε απλά να αναφέρετε αυτό αντί για τον πίνακα]*

Για την υλοποίηση της εργασίας χρησιμοποιήθηκε αποκλειστικά το Astra DB.

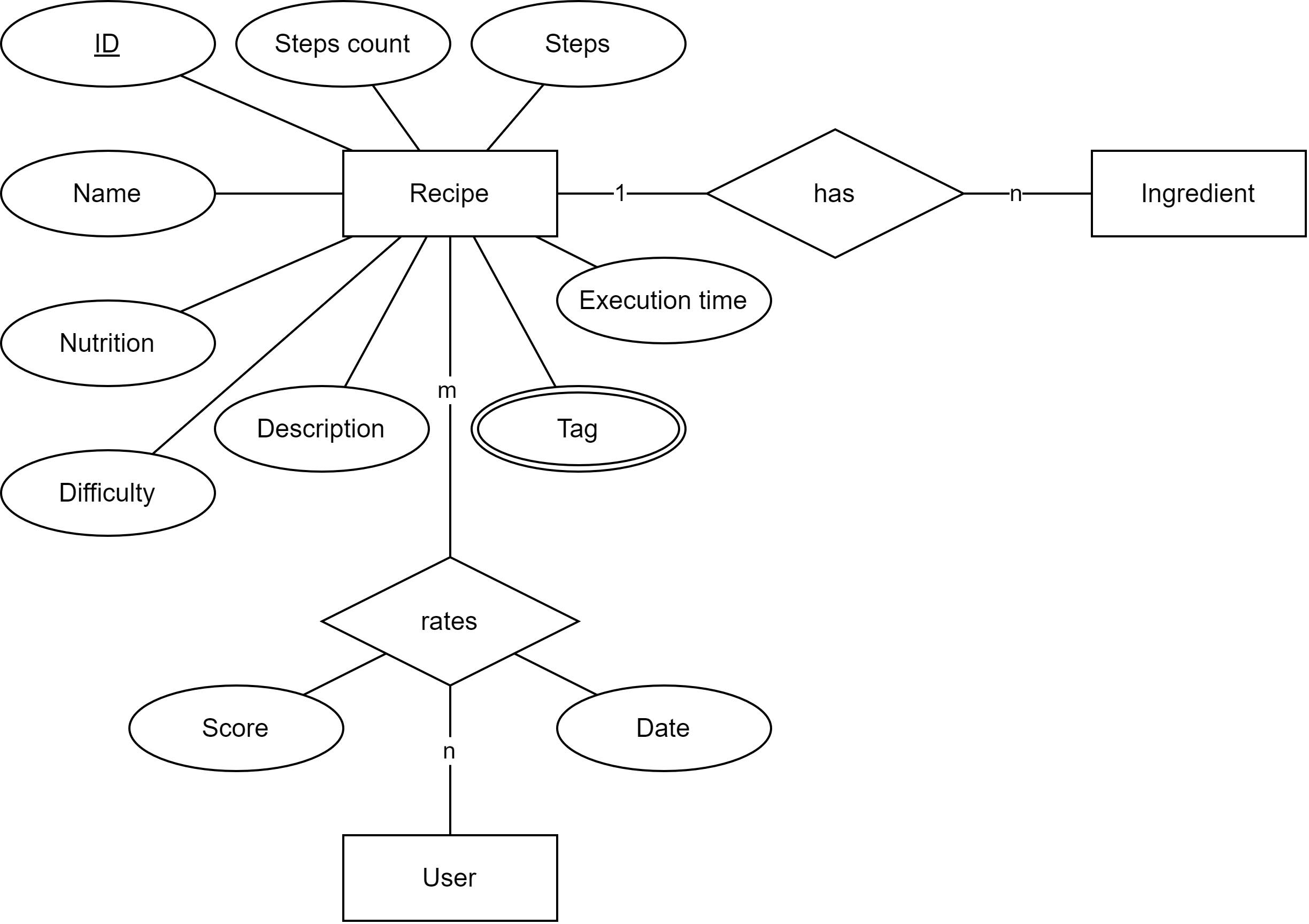
# Ερώτημα 1: Σχεδιασμός ΒΔ

*[δώστε το εννοιολογικό μοντέλο, το application workflow και το Chebotko diagram μαζί με τυχόν επεξηγήσεις που θέλετε να γράψετε για να εξηγήσετε τη φιλοσοφία του καθενός και να το περιγράψετε]*

### 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 ταινίες 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** | *[δώστε το όνομα του keyspace προς δημιουργία]* |
| **DDL statement** | *[δώστε το DDL statement για τη δημιουργία του keyspace]* |
| **Screenshot** | *[δώστε ένα screenshot που δείχνει δεδομένα μέσα στο keyspace με ως αποτέλεσμα του ερωτήματος SELECT \* FROM [keyspace] LIMIT 5]* |

# 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*