Ονοματα: Παναγιωτης Σπεντζαρης,Γιωργος Μποτσας

ΑΜ:1071110,1070929

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

04 / 06/ 2022

Υπογραφή

04 / 06/ 2022

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

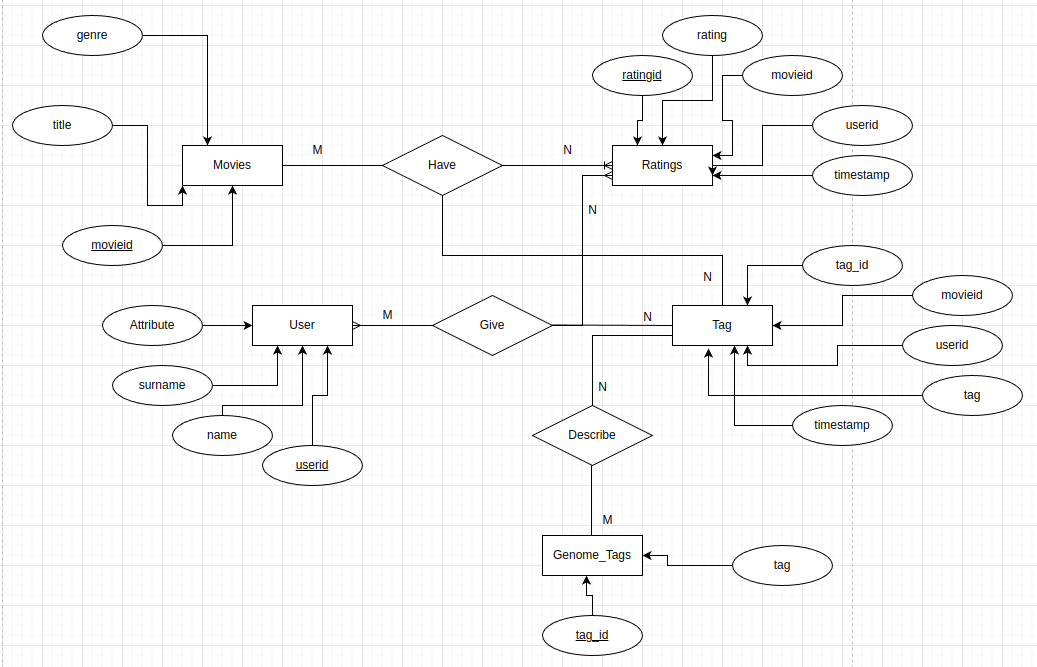
για το ερ. 1

Τεχνικά χαρακτηριστικά περιβάλλοντος λειτουργίας

|  |  |
| --- | --- |
| Χαρακτηριστικό | Τιμή |
| CPU model | Intel core i5 4th gen |
| CPU clock speed | 2.6GHz |
| Physical CPU cores | 4 |
| Logical CPU cores | 4 |
| RAM | 8gb ddr3 |
| Secondary Storage Type | HDD/ssd |

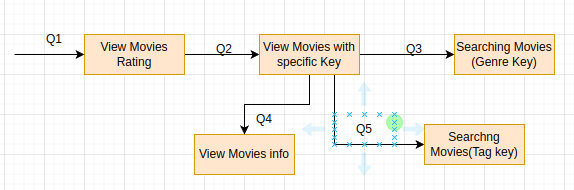
Ερώτημα 1: Σχεδιασμός ΒΔ

Παρακάτω παρατίθεται το εννοιολογικό μοντέλο της βάσης δεδομένων που δημιουργήσαμε ή αλλιώς του Keyspace με όνομα bigdata

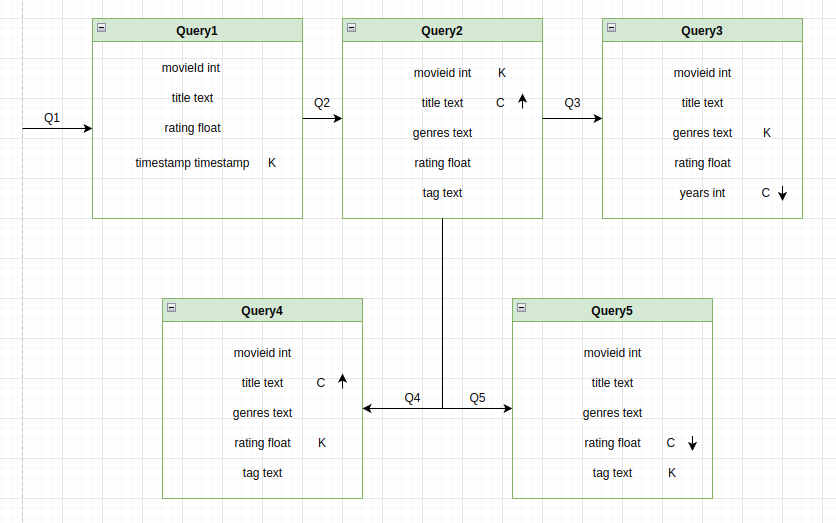


Όπως βλέπουμε αποτελείται από 5 πίνακες η δημιουργία των οποίων προήλθε από την επιλογή καταλλήλων δεδομένων από τα csv αρχεία που μας δόθηκαν.

Στην συνέχεια παρατίθεται το workflow διάγραμμα του οποίου η δημιουργία προήλθε από τα ερωτήματα που θέλουμε να εκτελέσουμε στο συγκεκριμένο keyspace.

****

Τέλος το chebotko διάγραμμα



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]

Η δημιουργία του Keyspace με όνομα bigdata γίνεται με την εντολή:

*CREATE KEYSPACE bigdata WITH replication = {'class':'SimpleStrategy', 'replication\_factor' : 3};*

Στην συνέχεια γίνεται η δημιουργία των πινάκων που θα χρησιμοποιήσουμε για να εξάγουμε τα δεδομένα:

CREATE TABLE query11(

Unnamed int,

movieId int,

title text,

rating float,

timestamp timestamp ,

PRIMARY KEY(timestamp,rating));

CREATE TABLE query2(

Unnamed int,

movieId int,

title text,

genres text,

rating float,

tag text,

PRIMARY KEY(movieId,title));

CREATE TABLE query3(

Unnamed int,

movieId int,

title text,

genres text,

rating float,

years int,

PRIMARY KEY(genres,years));

CREATE TABLE query4(

Unnamed int,

movieId int,

title text,

genres text,

rating float,

tag text,

PRIMARY KEY((rating),title));

CREATE TABLE query5(

Unnamed int,

movieId int,

title text,

genres text,

rating float,

tag text,

PRIMARY KEY((tag),rating));

SCREENSHOT μετα από εκτέλεση select:

Εικόνα που περιέχει κείμενο, εσωτερικό, στιγμιότυπο οθόνης, οθόνη

Περιγραφή που δημιουργήθηκε αυτόματα

Ερώτημα 3: Απαντήσεις ερωτημάτων

Ερώτημα Απάντηση:

Εμφάνιση των 30 ταινιών με την

υψηλότερη μέση βαθμολογία μεταξύ

01/01/2015 και 15/01/2015:

Εικόνα που περιέχει κείμενο, εσωτερικό, κλείσιμο

Περιγραφή που δημιουργήθηκε αυτόματα

Εμφάνιση όλων των λεπτομερειών για την

ταινία Jumanji (κατηγορία, μέση

βαθμολογία, top-5 ετικέτες):

Εικόνα που περιέχει κείμενο

Περιγραφή που δημιουργήθηκε αυτόματα

Εμφάνιση των ταινιών της κατηγορίας

“adventure” ταξινομημένες ως προς το έτος

παραγωγής

Εικόνα που περιέχει κείμενο

Περιγραφή που δημιουργήθηκε αυτόματα

Εμφάνιση των ταινιών που περιέχουν τη

λέξη “star”(Χρησιμοποιηθηκε η εντολη indexing: CREATE CUSTOM INDEX employee\_firstname\_idx ON query4 (title) USING 'org.apache.cassandra.index.sasi.SASIIndex' WITH OPTIONS = {'mode': 'CONTAINS', 'analyzer\_class': 'org.apache.cassandra.index.sasi.analyzer.StandardAnalyzer', 'case\_sensitive': 'false'};)

Εικόνα που περιέχει κείμενο

Περιγραφή που δημιουργήθηκε αυτόματα

Εμφάνιση των 20 ταινιών με την

υψηλότερη μέση βαθμολογία για την

ετικέτα “comedy”:

Εικόνα που περιέχει κείμενο

Περιγραφή που δημιουργήθηκε αυτόματα

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]