



Πανεπιστήμιο Πειραιώς, Τμήμα Πληροφορικής
ΠΜΣ «Κυβερνοασφάλεια και Επιστήμη Δεδομένων»,
2024-2025

Big Data Management

Διδάσκοντες: Νίκος Πελέκης, Γιώργος Παπαστεφανάτος
Εργαστηριακοί βοηθοί: Γ. Αλεξίου, Σ. Μαρούλης

1^η Εργαστηριακή Άσκηση (ατομική)

Αντικείμενο της άσκησης

Σκοπός της άσκησης είναι η εξοικείωση με NoSQL συστήματα διαχείρισης βάσεων δεδομένων και πιο συγκεκριμένα με τη MongoDB. Το σύνολο δεδομένων που θα χρησιμοποιηθεί προέρχεται από το TripAdvisor και περιλαμβάνει πληροφορίες για ευρωπαϊκά εστιατόρια. Συγκεκριμένα, περιλαμβάνει 1.083.397 ευρωπαϊκά εστιατόρια, με αναλυτικά στοιχεία όπως τοποθεσία, μέση βαθμολογία, αριθμός κριτικών, ώρες λειτουργίας, τύποι κουζίνας, και άλλα. Μπορείτε να κατεβάσετε τα δεδομένα σε μορφή JSON από [εδώ](#).

Μεθοδολογία υλοποίησης της άσκησης

Ερώτημα 1. Εισαγωγή Δεδομένων

Εισάγετε το αρχείο με τα δεδομένα στη βάση, χρησιμοποιώντας είτε κατάλληλες εντολές (hint: `mongoimport`), είτε το εργαλείο MongoDB Compass. Το παραδοτέο του ερωτήματος θα είναι οι εντολές που χρησιμοποιήθηκαν ή screenshots από τα βήματα που ακολουθήθηκαν στο MongoDB Compass, καθώς και screenshots που θα παρουσιάζουν δείγμα των δεδομένων στη βάση.

Ερώτημα 2. Ανάκτηση δεδομένων από τη ΒΔ

Θα πρέπει να υλοποιήσετε τα παρακάτω queries:

1. Εντοπίστε τα 5 κορυφαία εστιατόρια στην Αθήνα που προσφέρουν vegan επιλογές και σερβίρουν πρωινό (πεδίο `meals`). Η κατάταξη βασίζεται στον μέσο όρο αξιολόγησης (`avg_rating`), με την προϋπόθεση ότι έχουν περισσότερες από 100 συνολικές αξιολογήσεις (`total_reviews_count`). Εμφανίστε τα ονόματα, τις διευθύνσεις, τη μέση αξιολόγηση και τον συνολικό αριθμό αξιολογήσεων,

ταξινομημένα πρώτα κατά μέσο όρο αξιολόγησης και στη συνέχεια κατά συνολικό αριθμό αξιολογήσεων.

2. Εντοπίστε τους 10 πιο δημοφιλείς τύπους κουζίνας (πεδίο `cuisines`) στην Αθήνα που προσφέρονται από εστιατόρια με `vegan` επιλογές. Εμφανίστε το όνομα κάθε τύπου κουζίνας, τον αριθμό των εστιατορίων που τον προσφέρουν και τον μέσο όρο του αριθμού αξιολογήσεων αυτών των εστιατορίων.
3. Βρείτε τις 5 πόλεις εκτός Ελλάδας με τον μεγαλύτερο αριθμό εστιατορίων που προσφέρουν ελληνική κουζίνα. Εμφανίστε τα ονόματα των πόλεων και τον αριθμό των εστιατορίων που προσφέρουν ελληνική κουζίνα σε κάθε πόλη.
4. Εντοπίστε την πόλη (ανεξαρτήτως χώρας) με τον μεγαλύτερο αριθμό εστιατορίων που έχουν μέση αξιολόγηση (`avg_rating`) μεγαλύτερη από 4.5 και περισσότερες από 1000 συνολικές αξιολογήσεις (`total_reviews_count`). Εμφανίστε το όνομα της πόλης και τον αριθμό αυτών των εστιατορίων.
5. Εμφανίστε όλες τις περιφέρειες της Ελλάδας (πεδίο `region`) και τον αριθμό των εστιατορίων σε κάθε περιφέρεια με μέση αξιολόγηση (`avg_rating`) μεγαλύτερη από 4.5. Ταξινομήστε τις περιφέρειες κατά φθίνουσα σειρά με βάση τον αριθμό των εστιατορίων.

Για τα παραπάνω queries, χρησιμοποιήστε τις μεθόδους **find()** και **aggregate()** της MongoDB. Το παραδοτέο του ερωτήματος θα είναι τα queries που απαντούν τις παραπάνω ερωτήσεις στη βάση καθώς και screenshots των αποτελεσμάτων.

Ερώτημα 3. Optimization

Δημιουργήστε κατάλληλα ευρετήρια για την επιτάχυνση των ερωτημάτων του Ερωτήματος 2. Αξιολογήστε την απόδοση των queries πριν και μετά την εφαρμογή των ευρετηρίων (χρησιμοποιώντας τη μέθοδο **explain()**). Καταγράψτε και συγκρίνετε τα αποτελέσματα μέσω screenshots των πλάνων εκτέλεσης και σχολιάστε τις διαφορές.

Στην αναφορά σας, συμπεριλάβετε τις εντολές δημιουργίας των ευρετηρίων και αιτιολογήστε την επιλογή σας. Εξηγήστε γιατί επιλέξατε συγκεκριμένα πεδία για την ευρετηρίαση, λαμβάνοντας υπόψη στοιχεία όπως π.χ. τους τελεστές που χρησιμοποιήθηκαν στα ερωτήματα και την πληθικότητα των τιμών στα επιλεγμένα πεδία.

Ημερομηνία και τρόπος παράδοσης

Η αναφορά σας για την εργασία να αποσταλεί σε μορφή **PDF** στο email stavmars@athenarc.gr, μέχρι και τις 7/1/2025. Παρακαλείσθε να έχετε την παρακάτω πρόταση ως θέμα στο email που θα στείλετε: **“CDS110 - Εργασία MongoDB”**.