

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ ΣΤΑ ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ

Συστήματα Διαχείρισης Δεδομένων Μεγάλης Κλίμακας
(Χειμερινό Εξάμηνο 2025-2026)

Διδάσκων: Ιωάννης Κωτίδης (kotidis@aueb.gr)

Βοηθός: Χρυσόστομος Καπέτης (mkar@aueb.gr)

Οδηγίες

- Η εργασία είναι ΟΜΑΔΙΚΗ (ομάδες δύο ατόμων)
- Η εργασία είναι **ΥΠΟΧΡΕΩΤΙΚΗ** και **προσμετρά 25%** στον τελικό βαθμό του μαθήματος.

Ανάθεση: **11/12/2025**

Παράδοση: **11/01/2026 (23:55)**

Neo4J Graph Database

Το αρχείο **FootballData.csv** περιέχει τα στοιχεία των ποδοσφαιρικών αγώνων δεκατεσσάρων πρωταθλημάτων ευρωπαϊκών χωρών την περίοδο 2022-2023.

Ακολουθεί αναλυτική περιγραφή των δεδομένων και των ζητούμενων της εργασίας.

Το αρχείο **FootballData.csv** περιέχει 98751 εγγραφές. Κάθε εγγραφή αποτελείται από 19 πεδία τα οποία διαχωρίζονται με τον χαρακτήρα ";" (pipe). Ακολουθεί η περιγραφή των πεδίων.

footballData.txt		
GameID	Integer	Κωδικός ποδοσφαιρικού αγώνα
GameDate	date	Ημερομηνία διεξαγωγής του αγώνα
HomeTeamName	String	Ονομασία γηπεδούχου ομάδας.
AwayTeamName	String	Ονομασία φιλοξενούμενης ομάδας
HomeGoals	Integer	Αριθμός τερμάτων που πέτυχε η γηπεδούχος ομάδα
AwayGoals	Integer	Αριθμός τερμάτων που πέτυχε η φιλοξενούμενη ομάδα
Winner	Integer	Κωδικός που δηλώνει το νικητή του αγώνα (1=Νικητής η γηπεδούχος Ομάδα, 2=Νικητής η φιλοξενούμενη ομάδα, 0=Ο αγώνας έληξε ισοπαλία).
CompetitionName	String	Ονομασία πρωταθλήματος.
Country	String	Χώρα διεξαγωγής του πρωταθλήματος.
PlayerID	Integer	Μοναδικός Κωδικός για κάθε παίκτη.
PlayerName	String	Όνομα παίκτη

Position	String	Θέση στην οποία αγωνίζεται ο παίκτης (Goalkeeper, Defender, Midfield, Attack).
BirthDate	date	Ημερομηνία γέννησης του παίκτη.
PlayerTeamName	String	Η ονομασία της ομάδας στην οποία αγωνίστηκε ο παίκτης στον συγκεκριμένο αγώνα.
Minutes	Integer	Λεπτά συμμετοχής του παίκτη στον αγώνα.
Goals	Integer	Αριθμός τερμάτων που πέτυχε ο παίκτης στον αγώνα.
Assists	Integer	Αριθμός assists που έκανε ο παίκτης στον αγώνα (Assist=Μία assist χρεώνεται στον παίκτη που έδωσε την τελευταία πάσα στον σκόρερ).
YellowCards	Integer	Αριθμός κίτρινων καρτών με τις οποίες τιμωρήθηκε ο παίκτης στον αγώνα (0,1, 2).
RedCards	Integer	Ένδειξη η οποία δείχνει αν ο παίκτης αποβλήθηκε από τον αγώνα με κόκκινη κάρτα (0=Δεν δέχθηκε κόκκινη κάρτα, 1=Αποβλήθηκε με κόκκινη κάρτα).

Για τα παραπάνω δεδομένα ισχύουν τα εξής:

- Τα ονόματα των πρωταθλημάτων είναι μοναδικά
- Το όνομα κάθε ομάδας είναι μοναδικό.
- Στο πρωτάθλημα μιας χώρας αγωνίζονται μόνο ομάδες από αυτή την χώρα.
- Ένας παίκτης μπορεί να έχει αγωνιστεί το πολύ σε δύο ομάδες του ίδιου ή διαφορετικών πρωταθλημάτων (μεταγραφή).
- Η θέση (position) χαρακτηρίζει τον παίκτη (αμυντικός, επιθετικός κ.λπ.) και είναι η ίδια σε όλους τους αγώνες που συμμετείχε.

Ζήτημα Πρώτο

Καλείστε να μοντελοποιήσετε τα δεδομένα ως property graph σχεδιάζοντας τις κατάλληλες οντότητες (nodes) με σωστές ετικέτες (labels) , τους τύπους των σχέσεων (edges) και τις ιδιότητες (properties) των οντοτήτων και των σχέσεων. Κατά την μοντελοποίηση θα πρέπει να αποφύγετε την επανάληψη της ίδιας πληροφορίας σε κόμβους και σχέσεις, και να διασφαλίσετε ότι δεν δημιουργούνται σχέσεις μεταξύ κόμβων όταν αυτές δεν δικαιολογούνται εννοιολογικά από το πρόβλημα. Μελετήστε προσεκτικά τα ερωτήματα που καλείστε να απαντήσετε ώστε να διασφαλίσετε ότι το σχήμα του γράφου που θα σχεδιάσετε υποστηρίζει την αποδοτική εκτέλεση αυτών των ερωτημάτων σε γλώσσα Cypher.

Στη συνέχεια θα πρέπει να υλοποιήσετε το γράφο που σχεδιάσατε στο Neo4J και να φορτώσετε τα δεδομένα από το αρχείο FootballData.csv. Για τον σκοπό αυτό θα πρέπει να γράψετε ένα script σε γλώσσα Cypher.

Ζήτημα Δεύτερο

Κατόπιν της υλοποίησης της βάσης δεδομένων καλείστε να γράψετε κατάλληλες εντολές σε γλώσσα Cypher προκειμένου να απαντήσετε τα παρακάτω ερωτήματα:

1. Εμφανίστε έναν κατάλογο με τα παρακάτω στοιχεία για κάθε πρωτάθλημα:

- Ονομασία πρωταθλήματος
- Αριθμός αγώνων που κέρδισε η γηπεδούχος ομάδα
- Αριθμός αγών που κέρδισε η φιλοξενούμενη ομάδα
- Αριθμός αγώνων έληξαν ισόπαλοι.

Ο κατάλογος θα πρέπει να είναι ταξινομημένος σε αύξουσα διάταξη με βάση το όνομα του πρωταθλήματος.

2. Εμφανίστε έναν κατάλογο με τα αποτελέσματα των αγώνων του ελληνικού πρωταθλήματος με την παρακάτω μορφή:

Ημερομηνία Αγώνα, Γηπεδούχος Ομάδα, Φιλοξενούμενη Ομάδα, Γκολ Γηπεδούχου Ομάδας, Γκολ φιλοξενούμενης ομάδας.

Ο κατάλογος θα είναι ταξινομημένος με την ημερομηνία διεξαγωγής των αγώνων σε αύξουσα διάταξη.

3. Εμφανίστε έναν κατάλογο με τα ονόματα των παικτών που μέσα στην συγκεκριμένη σεζόν αγωνίστηκαν σε δύο ομάδες εκ των οποίων η μία συμμετέχει στο ελληνικό πρωτάθλημα και η άλλη σε οποιοδήποτε άλλο πρωτάθλημα (εκτός από το ελληνικό). Πρόκειται για παίκτες που πήραν μεταγραφή από κάποια ομάδα του ελληνικού πρωταθλήματος σε μια ομάδα ενός ξένου πρωταθλήματος ή το αντίστροφο. Ο κατάλογος πρέπει να περιέχει τις ακόλουθες πληροφορίες:

- Όνομα παίκτη
- Ονομασία ελληνικής ομάδας
- Ονομασία ξένης ομάδας
- Ονομασία ξένου πρωταθλήματος.

Ο κατάλογος να είναι ταξινομημένος σε αύξουσα διάταξη με βάση την ονομασία της ελληνικής ομάδας.

4. Εμφανίστε έναν κατάλογο με το όνομα του κορυφαίου σκόρερ κάθε ομάδας του πρωταθλήματος της αρεσκείας σας. Ο κατάλογος θα περιέχει τα παρακάτω στοιχεία:

- Όνομα ομάδας,
- Όνομα πρώτου σκόρερ
- Συνολικός αριθμός των γκολ που πέτυχε ο πρώτος σκόρερ της ομάδας.

Ο κατάλογος να είναι ταξινομημένος με τον αριθμό των γκολ σε φθίνουσα διάταξη.

5. Εμφανίστε έναν κατάλογο με τους 10 πιο «σκληρούς» αμυντικούς του πρωταθλήματος της αρέσκειας σας. Ένας αμυντικός χαρακτηρίζεται «σκληρός» από τον αριθμό των κόκκινων και των κίτρινων καρτών που έχει δεχτεί συνολικά σε όλα τα παιχνίδια στα οποία συμμετείχε.

Ο κατάλογος θα πρέπει να περιέχει τις παρακάτω πληροφορίες:

- Όνομα παίκτη
- Ονομασία ομάδας
- Αριθμός κόκκινων καρτών
- Αριθμός κίτρινων καρτών

Ο κατάλογος πρέπει να είναι ταξινομημένος με τον αριθμό των κόκκινων και των κίτρινων καρτών σε φθίνουσα διάταξη.

Για τα ερωτήματα που ακολουθούν θα πρέπει να **δημιουργήσετε** και να **αξιοποιήσετε** την σχέση **TEAMMATES** η οποία συνδέει δύο παίκτες που **έχουν αγωνιστεί μαζί στην ίδια ομάδα στον ίδιο αγώνα**. Η σχέση **TEAMMATES** θα πρέπει να διαθέτει μια ιδιότητα (Property) η οποία θα αποθηκεύει τον αριθμό των αγώνων στους οποίους οι δύο παίκτες αγωνίστηκαν μαζί.

6. Για μια ομάδα της αρέσκειας σας εμφανίστε το **βασικό επιθετικό δίδυμο**, δηλαδή τα ονόματα των δύο επιθετικών παικτών της ομάδας που αγωνίστηκαν μαζί στους περισσότερους αγώνες.
7. Για μια ομάδα της αρέσκειας σας εμφανίστε τους τρεις πιο **κομβικούς παίκτες**, δηλαδή τους τρεις παίκτες που έχουν αγωνιστεί με τους περισσότερους συμπαίκτες.

ΠΑΡΑΔΟΤΕΑ

Το παραδοτέο της εργασίας σας θα είναι ένα αρχείο AM.zip (όπου AM είναι ο Αριθμός Μητρώου) που θα ανεβάσετε στο eclass και θα περιέχει:

1. Μία αναφορά σε μορφή **.pdf** η οποία θα περιλαμβάνει:
 - Τη διαγραμματική αναπαράσταση του γράφου και την περιγραφή των κόμβων και των σχέσεων.
 - Τον κώδικα Cypher που γράψατε για να δημιουργήσετε τον γράφο και να φορτώσετε τα δεδομένα.
 - Τον κώδικα Cypher για τα ζητούμενα ερωτήματα μαζί με τα αντίστοιχα αποτελέσματα.
2. **CreateFootballGraph.cy**: Ένα script με τις εντολές της γλώσσας Cypher για την δημιουργία του γράφου και το φόρτωμα των δεδομένων.
3. **Queties.cy**: Ένα script με τις ερωτήσεις σε γλώσσα Cypher.