

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ ΣΤΑ ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ

Συστήματα Διαχείρισης Δεδομένων Μεγάλης Κλίμακας
(Χειμερινό Εξάμηνο 2023-2024)

Διδάσκων: Ιωάννης Κωτίδης (kotidis@aueb.gr)

Βοηθός: Χρυσόστομος Καπέτης (mkar@aueb.gr)

Οδηγίες

- Η εργασία είναι ΑΤΟΜΙΚΗ
- Η εργασία είναι ΥΠΟΧΡΕΩΤΙΚΗ και προσμετρά 25% στον τελικό βαθμό του μαθήματος.

Ανάθεση: **17-11-2023**

Παράδοση: **27-11-2023 Ώρα (23:55)**

Αποθήκες Δεδομένων

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

Η Ένωση Ευρωπαϊκών Ποδοσφαιρικών Ομοσπονδιών ενδιαφέρεται να αναπτύξει μια αποθήκη δεδομένων για την άντληση στατιστικών στοιχείων για τα πρωταθλήματα ποδοσφαίρου των ευρωπαϊκών χωρών. Οι απαιτήσεις της ομοσπονδίας εστιάζουν μεταξύ άλλων στην ανάλυση των στοιχείων των ποδοσφαιρικών αγώνων ανα πρωτάθλημα, σύλλογο και ποδοσφαιριστή. Εξυπακούεται ότι στην ανάλυση των δεδομένων θα πρέπει να ληφθεί υπόψη και ο παράγοντας του χρόνου έτσι ώστε, η ομοσπονδία να είναι σε θέση να παράγει αναφορές με στατιστικά στοιχεία ανα έτος, μήνα, και ημέρα.

Καλείστε να σχεδιάσετε και να υλοποιήσετε την παραπάνω αποθήκη δεδομένων προκειμένου να αυξήσετε την αποτελεσματικότητα της διεξαγωγής χρήσιμων στατιστικών στοιχείων. Στην συνέχεια να τροφοδοτήσετε την αποθήκη με τα δεδομένα του αρχείου **footballData.txt** και να εκτελέσετε ορισμένες επερωτήσεις για την παραγωγή χρήσιμων στατιστικών στοιχείων. Την αποθήκη δεδομένων θα την υλοποιήσετε με την χρήση του DBMS SQL Server. Για την παρουσίαση των αποτελεσμάτων θα αξιοποιήσετε τις δυνατότητες του Power BI.

Ακολουθεί αναλυτική περιγραφή των δεδομένων και των ζητούμενων της εργασίας.

Το αρχείο **footballdata.txt** περιέχει 98751 εγγραφές. Κάθε εγγραφή αποτελείται από 22 πεδία τα οποία διαχωρίζονται με τον χαρακτήρα "|" (pipe). Ακολουθεί η περιγραφή των πεδίων.

footballData.txt		
game_id	Integer	Κωδικός ποδοσφαιρικού αγώνα
game_date	date	Ημερομηνία διεξαγωγής του αγώνα
home_club_id	Integer	Κωδικός γηπεδούχου ομάδας
home_club_name	varchar(100)	Ονομασία γηπεδούχου ομάδας
home_club_goals	Integer	Αριθμός τερμάτων που πέτυχε η γηπεδούχος ομάδα
away_club_id	Integer	Κωδικός φιλοξενούμενης ομάδας
away_club_name	varchar(100)	Ονομασία φιλοξενούμενης ομάδας
away_club_goals	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).
competition_id	varchar(10)	Κωδικός πρωταθλήματος
competition_name	varchar(100)	Ονομασία πρωταθλήματος
country_name	varchar(100)	Χώρα διεξαγωγής του πρωταθλήματος.
player_club_id	Integer	Κωδικός ομάδας στην οποία αγωνίζεται ο παίκτης (ποδοσφαιριστής).
player_id	Integer	Κωδικός παίκτη.
player_name	varchar(100)	Όνομα παίκτη
yellow_cards	Int	Αριθμός κίτρινων καρτών με τις οποίες τιμωρήθηκε ο παίκτης στον αγώνα (0,1, 2).
red_cards	Int	Ένδειξη η οποία δείχνει αν ο παίκτης αποβλήθηκε από τον αγώνα με κόκκινη κάρτα (0=Δεν δέχθηκε κόκκινη κάρτα, 1=Αποβλήθηκε με κόκκινη κάρτα).
goals	Int	Αριθμός τερμάτων που πέτυχε ο παίκτης στον αγώνα
assists	Int	Αριθμός assists που έκανε ο παίκτης στον αγώνα (Assist=Μία assist χρεώνεται στον παίκτη που έδωσε την τελευταία πάσα στον σκόρερ).
minutes_played	Int	Λεπτά συμμετοχής του παίκτη στον αγώνα.
date_of_birth	Date	Ημερομηνία γέννησης του παίκτη
position	varchar(50)	Θέση στην οποία αγωνίζεται ο παίκτης (Goalkeeper, Defender, Midfield, Attack).

Ζήτημα Πρώτο

Να δημιουργήσετε το λογικό σχήμα της αποθήκης δεδομένων και να το τροφοδοτήσετε με τα απαραίτητα δεδομένα. Συγκεκριμένα:

1. Να δημιουργήσετε μία βάση δεδομένων με όνομα **footballDW (Football Data Warehouse)**. Στη συνέχεια να δημιουργήσετε τον πίνακα **footballData** στον οποίο να φορτώσετε τα δεδομένα του αρχείου **footballData.txt** χρησιμοποιώντας την παρακάτω εντολή:

```
BULK INSERT footballData
FROM 'C:\MY_DATA\footballData.txt' !!!ΠΡΟΣΟΧΗ: Προσαρμόστε το PATH
WITH (FIRSTROW =2, FIELDTERMINATOR= '|', ROWTERMINATOR = '\n');
```

2. Να υλοποιήσετε το λογικό σχήμα της αποθήκης δεδομένων το οποίο θα πρέπει να έχει την μορφή αστέρα (Star Schema).
3. Να γράψετε κατάλληλες εντολές σε γλώσσα SQL, οι οποίες θα τροφοδοτούν το σχήμα της αποθήκης με τα απαραίτητα στοιχεία από τον πίνακα **footballData**.
4. Να αναπαραστήσετε διαγραμματικά το σχήμα της αποθήκης χρησιμοποιώντας την επιλογή "Database diagrams" του SQL Server Management Studio.

Η δημιουργία του λογικού σχήματος και η τροφοδότηση της αποθήκης με τα δεδομένα θα γίνουν με την εκτέλεση ενός **SQL script** το οποίο θα πρέπει να γράψετε.

Ζήτημα Δεύτερο

Χρησιμοποιώντας την αποθήκη δεδομένων που δημιουργήσατε στο προηγούμενο ζήτημα, να γράψετε και να εκτελέσετε επερωτήσεις σε γλώσσα SQL, οι οποίες να απαντούν στα ακόλουθα ερωτήματα (απαιτήσεις) της Ένωσης Ευρωπαϊκών Ποδοσφαιρικών Ομοσπονδιών:

1. 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.

2. Εμφανίστε έναν κατάλογο με τα παρακάτω στοιχεία για κάθε ομάδα:
 - Το όνομα του πρωταθλήματος στο οποίο συμμετέχει η ομάδα
 - Το όνομα της ομάδας
 - Τον αριθμό των γκολ που πέτυχε η ομάδα συνολικά σε όλους τους αγώνες του πρωταθλήματος.
 - Τον συνολικό αριθμό των κίτρινων καρτών που δόθηκαν σε όλους τους παίκτες της ομάδας.
 - Τον συνολικό αριθμό των κόκκινων καρτών που δόθηκαν σε όλους τους παίκτες της ομάδας.

Ο κατάλογος να είναι ταξινομημένος σε αύξουσα διάταξη με βάση το όνομα του πρωταθλήματος και το όνομα της ομάδας.

3. 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 να είναι ταξινομημένος σε αύξουσα διάταξη με βάση το όνομα του πρωταθλήματος.

4. Εμφανίστε έναν κατάλογο με τα ονόματα των κορυφαίων σκόρερ κάθε ομάδας του ελληνικού πρωταθλήματος. 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. Γράψτε μια επερώτηση σε γλώσσα SQL το αποτέλεσμα της οποίας είναι η δημιουργία ενός κύβου το αποτέλεσμα της οποίας είναι η δημιουργία ενός κύβου (data cube), κάθε κελί του οποίου περιέχει τον συνολικό αριθμό των γκολ που σημειώθηκαν στο πρωτάθλημα της Αγγλίας (premier-league) ανά ομάδα, έτος και μήνα.

Ζήτημα 3

1. Δημιουργήστε μια αναφορά (report) με το power BI με τα παρακάτω:
 - 1.1 Κατάλληλο γράφημα στο οποίο θα απεικονίζονται τα αποτελέσματα της τρίτης και της τέταρτης επερώτησης που γράψατε στο προηγούμενο ζήτημα.
 - 1.2 Κατάλληλο γράφημα ή/και πίνακα για την αναπαράσταση των δεδομένων του κύβου που ζητείται στο προηγούμενο ζήτημα.
2. Εξερευνήστε τα δεδομένα της αποθήκης και δημιουργήστε ένα καλαίσθητο dashboard με κατάλληλα γραφήματα και άλλα στοιχεία παρουσίασης στο οποίο να απεικονίζονται οι πλέον χρήσιμες κατά την γνώμη σας πληροφορίες για το πρωτάθλημα της αρεσκείας σας.

ΠΑΡΑΔΟΤΕΑ

Τα παραδοτέα της εργασίας σας θα είναι τα εξής:

1. Ένα αρχείο pdf με όνομα AM.pdf το οποίο θα περιέχει:

Παραδοτέα πρώτου ζητήματος

- Τον κώδικα (εντολές SQL) για την δημιουργία του λογικού σχήματος της αποθήκης δεδομένων και την εισαγωγή των εγγραφών στους αντίστοιχους πίνακες.
- Το διάγραμμα του σχήματος αστέρα της αποθήκης δεδομένων.

Παραδοτέα δεύτερου ζητήματος

- Τον κώδικα με τις επερωτήσεις SQL του δεύτερου ζητήματος.

Παραδοτέα τρίτου ζητήματος

- Την αναφορά (report) που δημιουργήσατε στο PowerBI
- Το dashboard που δημιουργήσατε

2. Το αρχείο .rbix που δημιουργήσατε με το power bi. Το όνομα του αρχείου να είναι της μορφής: AM.rbix

Τα δύο παραπάνω αρχεία (PDF, και ZIP) μπορείτε να τα τοποθετήσετε σε έναν φάκελο τον οποίο στην συνέχεια θα τον συμπιέσετε σε ένα αρχείο της μορφής AM.zip (όπου AM είναι ο αριθμός μητρώου σας) και θα το ανεβάσετε στο eclass.