

ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ

Διδάσκων: Σ. Σκιαδόπουλος

Ακ. έτος 2022-23

Εργασία εξαμήνου Παράδοση: Δείτε τα ερωτήματα

## Γενικά για την εργασία

1. Η εργασία εκπονείται από ομάδες των 2 ατόμων αυστηρά: εργασίες με ένα, τρία και περισσότερα άτομα δεν βαθμολογούνται.
2. Για να ορίσετε την ομάδα σας χρησιμοποιήστε το [wiki](#) του μαθήματος στο [eclass](#). Η προθεσμία να δηλώσετε την ομάδα σας είναι η **12/12/2022**. Ομάδες μετά την ημερομηνία αυτή δεν θα γίνονται δεκτές.
3. Αν δεν μπορείτε να βρείτε συνεργάτη συμπληρώστε το όνομά σας στο τέλος της λίστας στην ενότητα «συνοικέσια». Ο σκοπός της ενότητας «συνοικέσια» είναι να σας βοηθήσει να βρείτε συνεργάτη. Οποιαδήποτε στιγμή, πριν από την λήξης της προθεσμίας, βρείτε συνεργάτη μπορείτε να μεταφερθείτε από τα «συνοικέσια» στις ομάδες (θυμηθείτε να σβήσετε τα ονόματά σας από τα «συνοικέσια»).

Σε κάθε περίπτωση, η ευθύνη του σχηματισμού και της λειτουργίας της ομάδας επαφίεται στα μέλη της και όχι στους διδάσκοντες.

4. Για το Παραδοτέο 1 δεν θα χρειαστεί να συνδεθείτε σε κάποιο σύστημα διαχείρισης βάσεων δεδομένων.
5. Για να υλοποιήσετε τα Παραδοτέα 2 και 3, με βάση τις ομάδες που δηλώσατε, θα δημιουργηθούν κωδικοί πρόσβασης στο [hilon.dit.uop.gr](#) για το λειτουργικό σύστημα και το σύστημα διαχείρισης βάσεων δεδομένων.
6. Τα Παραδοτέα 2 και 3 θα αναπτυχθούν και θα βαθμολογηθούν στη πλατφόρμα εργασίας που είναι εγκατεστημένη στο [hilon.dit.uop.gr](#): εργασίες υλοποιημένες σε άλλες πλατφόρμες δεν θα βαθμολογούνται.
7. Η παράδοση των παραδοτέων είναι υποχρεωτική: αν δεν παραδώσετε κάποιο παραδοτέο δεν θα βαθμολογηθείτε στην εργασία και δεν θα έχετε δικαίωμα συμμετοχής στις εξετάσεις. Σε κάθε επόμενο παραδοτέο μπορείτε να διορθώσετε, βελτιώσετε ή επεκτείνετε την εργασία σας (για παράδειγμα στο Παραδοτέο 2 μπορείτε να διορθώσετε ή επεκτείνετε κάτι που παραδώσατε στο Παραδοτέο 1). Τα παραδοτέα δεν είναι βαθμολογικά ισοδύναμα. Το Παραδοτέο 1 μετράει για περίπου 10-20% της συνολικής βαθμολογίας, το Παραδοτέο 2 για περίπου 30-40% και το Παραδοτέο 3 το υπόλοιπο ποσοστό.
8. Όλες οι προθεσμίες είναι αυστηρές. Δεν θα δοθεί παράταση.
9. Σε κάθε παραδοτέο πρέπει να παραδώσετε την αναφορά σας με συγκεκριμένο εξώφυλλο όπως αναλυτικά περιγράφεται παρακάτω. Αναφορές με ηλεκτρονικό ταχυδρομείο ή άλλο μέσο δεν γίνονται δεκτές.
10. Η εργασία ισχύει μόνο για το τρέχον ακαδημαϊκό έτος. Δεν μπορεί να μεταφερθεί σε επόμενο.
11. Συνοπτικές οδηγίες για τα κείμενά σας μπορείτε να βρείτε στο σύνδεσμο: <http://users.uop.gr/~spiros/courses/DB-writingnotes.html>. Πιο αναλυτικές μπορείτε να δείτε στο σύνδεσμο: <https://www.overleaf.com/read/frcdfqsdcrsx>.

## Περιγραφή εργασίας

Το θέμα της παρούσας εργασίας είναι ο σχεδιασμός, η ανάλυση και η υλοποίηση μιας βάσης δεδομένων, όπου καταχωρούνται πληροφορίες σχετικά με δήμους, δασικές πυρκαγιές και δεδομένα μετεωρολογικών σταθμών. Πιο συγκεκριμένα η βάση αποθηκεύει τις παρακάτω πληροφορίες:

**Δασικές πυρκαγιές.** Για κάθε δασική πυρκαγιά πρέπει να αποθηκεύονται: ένας αναγνωριστικός αριθμός, το όνομα του πυροσβεστικού τμήματος που επενέβη, η ώρα έναρξης, η ημερομηνία έναρξης, η ώρα κατάσβεσης, η ημερομηνία κατάσβεσης, η καμμένη έκταση (σε στρέμματα), το πλήθος του προσωπικού που επενέβη, το πλήθος των οχημάτων, και το πλήθος των εναέριων μέσων που χρησιμοποιήθηκαν.

**Δήμοι.** Για κάθε δήμο πρέπει να αποθηκεύονται: ένας αναγνωριστικός αριθμός, το όνομα της περιφέρειας και το όνομα του νομού στα οποία υπάγεται, το όνομα του δήμου, καθώς και το γεωγραφικό μήκος και το γεωγραφικό πλάτος του. Κάθε δήμος πρέπει να σχετίζεται με τις δασικές πυρκαγιές που εκδηλώθηκαν σε αυτόν, αλλά και με έναν μετεωρολογικό σταθμό αναφοράς.

**Μετεωρολογικοί σταθμοί.** Για κάθε μετεωρολογικό σταθμό πρέπει να αποθηκεύονται: ένας αναγνωριστικός αριθμός, το όνομα του μετεωρολογικού σταθμού, το γεωγραφικό μήκος και το γεωγραφικό πλάτος του, όπως και το υψόμετρο στο οποίο είναι εγκατεστημένος. Ένας μετεωρολογικός σταθμός μπορεί να αποτελεί σταθμό αναφοράς για παραπάνω από έναν δήμους, και πρέπει να σχετίζεται με τις καταγραφές των μετεωρολογικών του δεδομένων.

**Μετεωρολογικά δεδομένα.** Για κάθε ημέρα καταγραφής μετεωρολογικών δεδομένων των μετεωρολογικών σταθμών πρέπει να αποθηκεύονται: ένας αναγνωριστικός αριθμός, η ημερομηνία, η μέση, η μέγιστη και η ελάχιστη θερμοκρασία (σε βαθμούς Κελσίου), η μέση, η μέγιστη και η ελάχιστη υγρασία (σε ποσοστό επί τοις εκατό), η μέση, η μέγιστη και η ελάχιστη ατμοσφαιρική πίεση (σε εκατοπασκάλ), η ημερήσια βροχόπτωση (σε χιλιοστόμετρα), η μέση ταχύτητα του ανέμου (σε χιλιόμετρα ανά ώρα), η διεύθυνση του ανέμου, όπως και η μέγιστη ριπή του ανέμου (σε χιλιόμετρα ανά ώρα).

## 1<sup>ο</sup> ΠΑΡΑΔΟΤΕΟ

### Ερώτημα 1

1. Σχεδιάστε το μοντέλο οντοτήτων συσχετίσεων για την παραπάνω βάση δεδομένων. Για την σχεδίαση μπορείτε να χρησιμοποιήσετε την διαδικτυακή σχεδιαστική εφαρμογή στο <https://app.diagrams.net>. Παράδειγμα σχεδιασμού με την παραπάνω εφαρμογή μπορείτε να κατεβάσετε από τα αρχεία του μαθήματος στο [eclass](#).
2. Να αναφέρετε ξεκάθαρα τους περιορισμούς και τις υποθέσεις που κάνετε. Σημειώστε ότι κάποιες από τις παραπάνω πληροφορίες μπορούν να είναι γνωρίσματα: κάποιες όμως για να αναπαρασταθούν χρειάζονται οντότητες και συσχετίσεις. Η σχεδιάσή σας πρέπει να περιλαμβάνει τουλάχιστον 3 οντότητες και τουλάχιστον 2 συσχετίσεις.

### Ερώτημα 2

1. Μεταφράστε το μοντέλο οντοτήτων συσχετίσεων του Ερωτήματος 1 σε σχεσιακό μοντέλο.

2. Αποτυπώστε τις συναρτησιακές εξαρτήσεις κάθε σχέσης. Οι συναρτησιακές εξαρτήσεις προκύπτουν από το μικρόκοσμο που περιγράψαμε παραπάνω.
3. Σε ποια κανονική μορφή είναι οι σχέσεις που προκύπτουν με τις συναρτησιακές εξαρτήσεις τους;
4. Όσες σχέσεις δεν είναι σε Boyce-Codd κανονική μορφή, μετατρέψτε της, χρησιμοποιώντας την μέθοδο της ανάλυσης, σε Boyce-Codd κανονική μορφή ή αν δεν είναι αυτό εφικτό σε τρίτη κανονική μορφή.
5. Να διατυπώσετε τις κατάλληλες εντολές που κατασκευάζουν τις απαραίτητες σχέσεις σε ένα σχεσιακό σύστημα διαχείρισης βάσεων δεδομένων.

ΠΡΟΘΕΣΜΙΑ ΓΙΑ ΤΟ 1<sup>ο</sup> ΠΑΡΑΔΟΤΕΟ  
**Τετάρτη 14/12/2022 – 11:59μμ**

Πρέπει να παραδώσετε αναφορά, η οποία θα περιλαμβάνει τις απαντήσεις σας στα Ερωτήματα 1 και 2.

Για το παραδοτέο αυτό δεν θα χρειαστείτε να συνδεθείτε στο `hilon.dit.uop.gr`.

Η πρώτη σελίδα του παραδοτέου σας (εξώφυλλο) πρέπει υποχρεωτικά να παραχθεί από την σελίδα <http://users.uop.gr/~spiros/Cover/>.

## 2<sup>ο</sup> ΠΑΡΑΔΟΤΕΟ

### Πλαίσιο εργασίας

Τη βάση σας θα την υλοποιήσετε στο ΣΔΒΔ PostgreSQL, το οποίο βρίσκεται εγκαταστημένο στον υπολογιστή `hilon.dit.uop.gr`. Για τις λειτουργίες σας θα χρησιμοποιήσετε SQL, την οποία θα καλείτε μέσω PHP. Ο μεταγλωττιστής της PHP κι η απαραίτητη διασύνδεση με την PostgreSQL είναι εγκατεστημένη στο `hilon.dit.uop.gr`.

**Προσοχή:** Η εξέταση της εργασίας σας θα γίνει στον υπολογιστή `hilon.dit.uop.gr`. Εργασία που έχει υλοποιηθεί σε άλλο σύστημα δε θα βαθμολογηθεί!

### Ερώτημα 3 – απαιτείται πρόσβαση στο `hilon.dit.uop.gr`

Να σχεδιάσετε και να υλοποιήσετε τις παρακάτω λειτουργίες. Όλες οι λειτουργίες πρέπει να υποστηρίζουν ελληνικά.

1. **Εισαγωγή.** Σχεδιάστε και υλοποιήστε τις λειτουργίες για την εισαγωγή/διαγραφή:

1. Δασικής πυρκαγιάς
2. Δήμου
3. Μετεωρολογικού σταθμού
4. Εγγγραφής ημερησίων μετεωρολογικών δεδομένων

Για την εισαγωγή, ο χρήστης θα πληκτρολογεί τα απαραίτητα στοιχεία για κάθε περίπτωση και η λειτουργία θα εισάγει τα δεδομένα στη βάση. Πρέπει να υλοποιήσετε όλους τους απαραίτητους ελέγχους για την εισαγωγή των δεδομένων (π.χ., να ελέγχετε αν μια εγγραφή υπάρχει ήδη) αλλά και τους περιορισμούς συμμετοχής και πληθικότητας.

Για την διαγραφή, ο χρήστης θα πληκτρολογεί τα απαραίτητα στοιχεία και η λειτουργία θα διαγράφει τα αντίστοιχα δεδομένα από τη βάση.

- 2. Εισαγωγή από αρχείο κειμένου.** Ως είσοδος αυτής της λειτουργίας θα είναι τα αρχεία κειμένου που μπορείτε να βρείτε στη διαδρομή `/home/Data/2022-23` στο `hilon.dit.uop.gr`. Συγκεκριμένα υπάρχουν τα:

**fire\_data.csv:** Περιέχει πληροφορία σχετική με τις δασικές πυρκαγιές. Σε κάθε σειρά αναγράφονται: το όνομα του αρμόδιου πυροσβεστικού τμήματος που ανέλαβε την κατάσβεση, το όνομα του Νομού και του Δήμου στον οποίο εκδηλώθηκε η δασική πυρκαγιά, η ημερομηνία και η ώρα έναρξης της πυρκαγιάς, η ημερομηνία και η ώρα κατάσβεσης της πυρκαγιάς, το μέγεθος της καμμένης έκτασης (σε στρέμματα), το πλήθος του προσωπικού, των οχημάτων, και των εναέριων μέσων που πήραν μέρος στην κατάσβεση της πυρκαγιάς.

**locations\_data.csv:** Περιέχει πληροφορία σχετική με τους Δήμους. Σε κάθε σειρά αναγράφονται: το όνομα της Περιφέρειας, του Νομού και του Δήμου, το γεωγραφικό μήκος και το γεωγραφικό πλάτος του, καθώς και το όνομα του μετεωρολογικού σταθμού αναφοράς για τον Δήμο.

**stations\_list.csv:** Περιέχει πληροφορία σχετικά με τους μετεωρολογικούς σταθμούς. Σε κάθε σειρά αναγράφονται: το όνομα του μετεωρολογικού σταθμού, το γεωγραφικό μήκος και το γεωγραφικό πλάτος του, καθώς και το υψόμετρο στο οποίο βρίσκεται.

**meteo\_data.csv:** Περιέχει πληροφορία σχετική με τα καταγεγραμμένα μετεωρολογικά δεδομένα των μετεωρολογικών σταθμών. Σε κάθε σειρά αναγράφονται: το όνομα του μετεωρολογικού σταθμού που κατέγραψε τα δεδομένα, η ημερομηνία καταγραφής, η μέση, η μέγιστη και η ελάχιστη θερμοκρασία (σε βαθμούς Κελσίου), η μέση, η μέγιστη και η ελάχιστη υγρασία (σε ποσοστό επί τοις εκατό), η μέση, η μέγιστη και η ελάχιστη ατμοσφαιρική πίεση (σε εκατοπασκάλ), η ημερήσια βροχόπτωση (σε χιλιοστόμετρα), η μέση ταχύτητα του ανέμου (σε χιλιόμετρα ανά ώρα), η διεύθυνση του ανέμου, καθώς και η μέγιστη ριπή του ανέμου (σε χιλιόμετρα ανά ώρα).

Οι πληροφορίες σε κάθε γραμμή των αρχείων είναι διαχωρισμένες με το Ελληνικό ερωτηματικό ';'.

### **Δώστε προσοχή στα παρακάτω:**

Όπως θα παρατηρήσετε, στα αρχεία δεν περιλαμβάνονται μοναδικοί κωδικοί, παρά μόνο τα ονόματα των πυροσβεστικών τμημάτων, των Δήμων, και των μετεωρολογικών σταθμών. Όπως έχουμε πει στην θεωρία, αλλά και στο πρώτο παραδοτέο, το πρωτεύον κλειδί κάθε σχέσης θα πρέπει να είναι ένας ακέραιος αριθμός. Με άλλα λόγια δηλαδή *δεν μπορείτε να χρησιμοποιήσετε ως πρωτεύοντα κλειδιά τα ονόματα που υπάρχουν στα αρχεία*. Αυτά παρέχονται με τρόπο μοναδικό, ώστε να μπορέσετε να συνδέσετε όμοια πληροφορία μεταξύ διαφορετικών αρχείων, για να βάλετε δεδομένα στις τελικές σας σχέσεις, και όχι για να τα χρησιμοποιήσετε ως πρωτεύοντα κλειδιά.

Επίσης για να συσχετίσετε κάποιες σχέσεις μεταξύ τους θα χρειαστεί να χρησιμοποιήσετε γνωρίσματα που παρέχονται στα αρχεία. Για παράδειγμα, η οντότητα **δήμοι** μπορεί να

συσχετιστεί με την οντότητα **δασικές\_πυρκαγιές** μέσω των γνωρισμάτων **νομός** και **δήμος**, τα οποία σας παρέχονται και στα δύο σχετικά αρχεία.

Για να μεταφορτώσετε τα δεδομένα πρέπει **υποχρεωτικά** να ακολουθήσετε την παρακάτω διαδικασία:

**Κατασκευή προσωρινών σχέσεων:** Αρχικά θα μεταφορτώσετε τα δεδομένα από τα αρχεία `fire_data.csv`, `locations_data.csv`, `meteo_data.csv` και `stations_list.csv`, σε 4 προσωρινές σχέσεις (έστω `fire_tmp`, `locations_tmp`, `meteo_tmp` και `stations_tmp`). Για το βήμα αυτό μπορείτε να χρησιμοποιήσετε την εντολή `\copy` της `postgres`, η οποία δεν χρειάζεται δικαιώματα υπερχρήστη.

**Κατασκευή τελικών σχέσεων:** Για κάθε σχέση της σχεδίασής σας πρέπει να φτιάξετε μια εντολή εισαγωγής δεδομένων αντλώντας δεδομένα από τις παραπάνω προσωρινές σχέσεις.

Για παράδειγμα, αν στην σχεδίασή σας έχετε μια σχέση `meteo_data`, για τις καταγραφές των μετεωρολογικών δεδομένων, πρέπει να φτιάξετε μια εντολή της μορφής:

```
INSERT INTO meteo_data(attribute1, attribute2, ... )
SELECT values for attribute1, attribute2, ...
FROM meteo_tmp or/and any other needed temp relation
WHERE ...
```

Θα πρέπει να παρουσιάσετε αντίστοιχες εκφράσεις για όλες τις σχέσεις που υπάρχουν στην σχεδίασή σας.

**Σημείωση 1:** Δεν πρέπει να αλλάξετε την σχεδίαση σας για να μεταφορτώσετε πιο εύκολα τα δεδομένα. Μπορείτε όμως να διορθώσετε λάθη και παραλήψεις. Δηλαδή μπορείτε να ενημερώσετε ότι παραδώσατε στο πρώτο παραδοτέο.

**Σημείωση 2:** Τα πρωτεύοντα κλειδιά σας πρέπει να είναι **ακέραιοι**. Επίσης οι τελικές σας σχέσεις πρέπει να είναι περισσότερες από 4 που είναι ο αριθμός των αρχείων που σας έχουμε δώσει.

**Σημείωση 3:** Όλη διαδικασία μεταφόρτωσης των δεδομένων πρέπει να γίνει **μόνο** με **SQL**. Μάλιστα για κάθε σχέση που θα μεταφορτώσετε δεδομένα θα γράψετε ένα ερώτημα χωρίς εμφώλευση και χωρίς χρήση τελεστών που αποκλείσαμε (**JOIN**, **IN**, **EXISTS** κλπ. Δείτε και την τελευταία διαφάνεια των σημειώσεων).

## ΠΡΟΘΕΣΜΙΑ ΓΙΑ ΤΟ 2<sup>ο</sup> ΠΑΡΑΔΟΤΕΟ

### Τετάρτη 11/1/2023 – 11:59μμ

Πρέπει να παραδώσετε *αναφορά*, η οποία θα περιλαμβάνει τις απαντήσεις στο Ερώτημα 1, στο Ερώτημα 2 (δηλαδή ότι παραδώσατε στο 1ο παραδοτέο κατάλληλα ενημερωμένο) και στο Ερώτημα 3. Δεν θα παραδώσετε κώδικα!

Στο παραδοτέο απλά περιγράψτε πώς σχεδιάσατε και υλοποιήσατε τις διαδικασίες. Ο κώδικάς σας θα πρέπει να βρίσκεται στο `hilon.dit.uop.gr` κι η βάση πρέπει να διατηρεί τα δεδομένα που περιγράφονται στα αρχεία που σας έχουν δοθεί.

Η πρώτη σελίδα του παραδοτέου σας (εξώφυλλο) πρέπει υποχρεωτικά να παραχθεί από την σελίδα <http://users.uop.gr/~spiros/Cover/>