

# 4η Εργασία: Διαγράμματα Οντοτήτων-Συσχετίσεων και Όψεις

Προθεσμία: Δευτέρα 15 Μαΐου

## Σκοπός:

Κάνουμε αλλαγές στον σχεδιασμό της βάσης μας ώστε σε επόμενο βήμα να κανονικοποιηθεί και να αποφευχθούν ανωμαλίες. Στα πλαίσια αυτά θα πρέπει να:

- Καθαρίσουμε τα δεδομένα μας από θόρυβο και αχρείαστη πληροφορία.
- Απεικονίσουμε το σχήμα της βάσης σε ένα διάγραμμα οντοτήτων-συσχετίσεων (ER).
- Επεκτείνουμε το αρχικό μας σχήμα έτσι ώστε να περιέχει πρωτεύοντα κλειδιά για όλους τους πίνακες.
- Φτιάξουμε κάποιες όψεις και χρησιμοποιώντας τις όψεις αυτές να καθαρίσουμε τα δεδομένα της βάσης μας.

## Προαπαιτούμενα:

Θα πρέπει να έχετε δημιουργήσει τη βάση δεδομένων που περιγράφεται στην 2η εργασία και να έχετε εισάγει στους πίνακες τα δεδομένα ταινιών.

## Ζητούμενα Εργασίας:

Στην συγκεκριμένη Εργασία κάνουμε αλλαγές στην βάση μας ξεκινώντας από την κατάσταση που είχε όπως την ορίσαμε στην Εργασία 2. Στόχος μας είναι να δημιουργηθούν τα πρωτεύοντα κλειδιά για όλους τους πίνακες καθώς και το αντίστοιχο διάγραμμα οντοτήτων συσχετίσεων. Τα βήματα της εργασίας είναι τα ακόλουθα:

### Καθάρισμα των Πινάκων από Διπλότυπα

1. Να διαγραφούν οι διπλότυπες εγγραφές από τους πίνακες hasGenre, hasKeyword, belongsToCollection, hasProductioncompany. Για την διαγραφή να χρησιμοποιηθεί το script: removeDuplicates.sql που σας δίνεται στον φάκελο της άσκησης.

### Δημιουργία ER Διαγράμματος

2. Να δημιουργηθεί το διάγραμμα οντοτήτων-συσχετίσεων για την βάση δεδομένων του monielens. Στο διάγραμμα οντοτήτων-συσχετίσεων να εμφανίζονται τα αντίστοιχα γνωρίσματα κλειδιά υπογραμμισμένα, ενώ σε πίνακες με μεγάλο πλήθος γνωρισμάτων, να σχεδιαστούν ενδεικτικά κάποια από τα εναπομείναντα γνωρίσματα.

Θεωρείστε ότι:

- Το σύνολο οντοτήτων movie συνδέεται με το σύνολο οντοτήτων movie\_cast μέσω της συσχέτισης hasCast.  
Κάθε οντότητα του movie\_cast, σχετίζεται με **ακριβώς μία** οντότητα movie (να φαίνεται στο ER διάγραμμα).
- Το σύνολο οντοτήτων movie συνδέεται με το σύνολο οντοτήτων movie\_crew μέσω της συσχέτισης hasCrew.

Κάθε οντότητα του movie\_crew, σχετίζεται με **ακριβώς μία** οντότητα movie (να φαίνεται στο ER διάγραμμα).

3. Να απαντηθεί (το πολύ 5 γραμμές) η ακόλουθη ερώτηση:
  - ο Γιατί δεν υπάρχει ένας πίνακας στην βάση που να αντιστοιχεί στις συσχετίσεις hasCrew και hasCast;

### Δημιουργία Κλειδιών

4. Χρησιμοποιώντας την εντολή ALTER TABLE, βασισμένοι στην θεωρία των ER διαγραμμάτων, να δημιουργηθούν πρωτεύοντα κλειδιά για τους ακόλουθους πίνακες της βάσης: hasGenre, hasKeyword, belongsToCollection, hasProductioncompany, Ratings.

Σε περίπτωση που τα προαναφερθέντα κλειδιά είχαν δημιουργηθεί στην δεύτερη εργασία, να δοθεί η αντίστοιχη εντολή CREATE ή ALTER που χρησιμοποιήθηκε για την δημιουργία τους.

### Δημιουργία Όψεων (Views)

5. Να δημιουργήσετε τις ακόλουθες όψεις:
  - την **όψη Actor** από τα γνωρίσματα `distinct person_id, gender, name` του πίνακα Movie\_Cast.
  - την **όψη CrewMember** από τα γνωρίσματα `distinct person_id, gender, name` του πίνακα Movie\_Crew.
  - την **όψη Person** χρησιμοποιώντας την ένωση (union) των πινάκων Actor και CrewMember.

(Σημείωση: Οι όψεις Actor, CrewMember, Person να μην εμφανίζονται στο ER διάγραμμα. Μια σωστά σχεδιασμένη βάση θα είχε ξεχωριστούς πίνακες (οντοτήτων) Actor, CrewMember, Person. Θα δούμε σε επόμενη εργασία πώς μπορούμε να χρησιμοποιήσουμε την θεωρία σχεδιασμού για να διορθώσουμε τέτοια λάθη στην βάση μας.)

### Λάθος Εγγραφές

6. Για τον πίνακα Person:
  - Βρείτε τα person\_ids τα οποία εμφανίζονται πάνω από 1 φορά, είτε για διαφορετικό name, είτε για διαφορετικό gender (Σημείωση: χρησιμοποιείτε τους τελεστές group by και having με την συνάρτηση συνάθροισης count(\*))
  - Κάντε τα απαιτούμενα Updates έτσι ώστε το κάθε person\_id να αντιστοιχεί σε μία μοναδική εγγραφή στον Πίνακα Person. (Σημείωση: οι αλλαγές στην συγκεκριμένη περίπτωση μπορούν να γίνουν μόνο στους αρχικούς πίνακες movie\_cast και movie\_crew)

### Εργασία:

- Azure Data Studio.
- [Draw.io](https://draw.io)

## Συμβουλές για την υλοποίηση:

- Επιβεβαιώστε ότι ο χρήστης της βάσης του οποίου μας στέλνετε τα credentials όντως έχει πρόσβαση στη βάση σας και μπορεί να τρέχει SELECT queries.
- Χρησιμοποιείτε ALTER TABLE ... ADD PRIMARY/FOREIGN KEY ...
- Για τον σχεδιασμό του ER προτείνεται η χρήση [Draw.io](https://draw.io).

## Χρήσιμα links:

-ALTER TABLE:

<https://learn.microsoft.com/en-us/sql/t-sql/statements/alter-table-transact-sql?view=sql-server-ver16>

- Having:

<https://learn.microsoft.com/en-us/sql/t-sql/queries/select-having-transact-sql?view=sql-server-ver16>

- Create View:

<https://learn.microsoft.com/en-us/sql/t-sql/statements/create-view-transact-sql?view=sql-server-ver16>

## Παραδοτέα:

- A. Δημιουργήστε ένα αρχείο **a.info.txt** στο οποίο θα αναγράφονται το endpoint του Azure instance σας (Server name στο Overview tab του Azure), το όνομα της βάσης σας και το username και το password ενός χρήστη με read-only δικαιώματα, ώστε να μπορούμε να δούμε τους πίνακες της βάσης σας. Το .txt αρχείο θα πρέπει να έχει την παρακάτω μορφή:

Endpoint: <name\_of\_the\_endpoint>

Username: <username>

Password: <password>

Database: <name\_of\_the\_database>

- B. Να παραδοθεί το αρχείο **b.ER.pdf** το οποίο περιέχει το *ER διάγραμμα* της βάσης δεδομένων όπως περιγράφεται στο 2<sup>ο</sup> βήμα της άσκησης και να παραδοθεί σε μορφή pdf. Θα χρειαστεί να επεκτείνετε το ER διάγραμμα σε επόμενες ασκήσεις (2<sup>ο</sup> βήμα της άσκησης).
- C. Να παραδοθεί το αρχείο **c.answer.txt** στο οποίο απαντάτε στην ερώτηση για το 3<sup>ο</sup> βήμα της άσκησης (3<sup>ο</sup> βήμα της άσκησης).
- D. Να δημιουργήσετε το αρχείο **d.alterTables.sql**, το οποίο να περιέχει τις εντολές που άλλαξαν τους αντίστοιχους πίνακες προσθέτοντας πρωτεύοντα κλειδιά στο (4<sup>ο</sup> βήμα της άσκησης).
- E. Να δημιουργήσετε το αρχείο **e.createViews.sql**, το οποίο να περιέχει τα views τα οποία φτιάξαμε στο (5<sup>ο</sup> βήμα της άσκησης).
- F. Να δημιουργήσετε το αρχείο **f.inconsistencies.sql** το οποίο βρίσκει τα id των εγγράφων οι οποίες περιέχουν μη συνεπή πληροφορία καθώς και τα αντίστοιχα updates στους πίνακες movie\_cast και movie\_crew (6<sup>ο</sup> βήμα της άσκησης).
- Βάλτε όλα τα αρχεία σε έναν φάκελο. Το όνομα του φακέλου πρέπει να αποτελείται από τους αριθμούς μητρώου σας χωρισμένους με παύλα, δηλαδή *αριθμός\_μητρώου\_1-αριθμός\_μητρώου\_2*. Δημιουργήστε ένα **.zip** αρχείο αυτού του φακέλου, το οποίο θα έχει το ίδιο όνομα με τον φάκελο.
- Κάντε υποβολή το .zip αρχείο στο eclass στην ενότητα *Εργασίες / 4η Εργασία*.