

ΔΕΝ ΕΠΙΤΡΕΠΕΤΑΙ Η ΧΡΗΣΗ ΚΙΝΗΤΩΝ ΤΗΛΕΦΩΝΩΝ - ΚΑΛΗ ΕΠΙΤΥΧΙΑ!

1. Κανονικοποίηση (20 μονάδες)

Σας δίνεται ότι για τη σχέση R (A, B, C, D, E, F) ισχύουν οι ακόλουθες συναρτησιακές εξαρτήσεις:

$A \rightarrow CD$

$F \rightarrow ABD$

$D \rightarrow E$

- 1) **(6 μονάδες)** Ποιες από τις 3 συναρτησιακές εξαρτήσεις παραβιάζουν (και γιατί;) τους κανόνες της BCNF;
- 2) **(14 μονάδες)** Εφαρμόστε τον αλγόριθμο αποσύνθεσης σε BCNF για να λάβετε μία χωρίς απώλειες αποσύνθεση της R σε ένα σύνολο από σχέσεις που είναι σε BCNF. Δείξτε αναλυτικά τη διαδικασία που ακολουθήσατε. Η απάντησή σας πρέπει να περιλαμβάνει (Α) τους πίνακες με τα χαρακτηριστικά τους και μία ένδειξη για τα κλειδιά κάθε πίνακα (με υπογράμμιση), καθώς και (Β) αιτιολόγηση για το πώς προέκυψε το κλειδί του κάθε πίνακα.

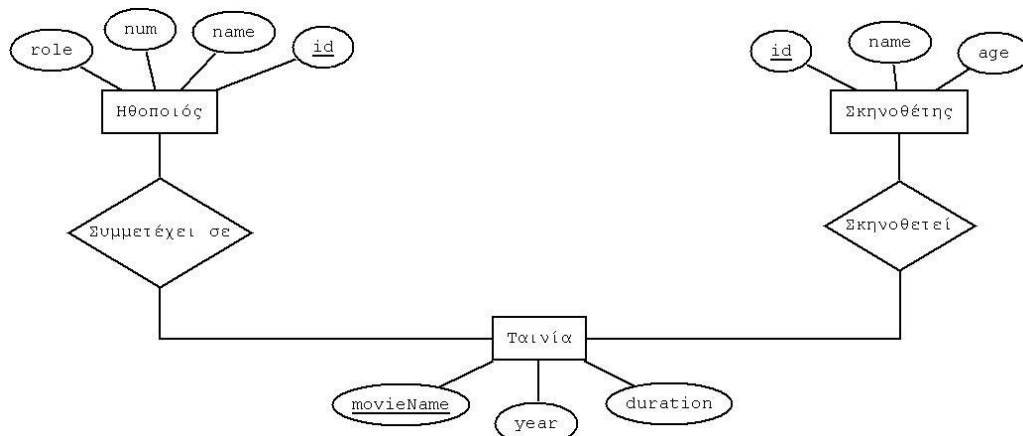
ΣΗΜΕΙΩΣΗ: Ζητείται συγκεκριμένος αλγόριθμος παραπάνω. Η εφαρμογή άλλου αλγόριθμου (πχ, σύνθεσης σε 3NF) οδηγεί αυτόματα σε μηδενισμό του συγκεκριμένου κομματιού.

2. ER διάγραμμα + Μετατροπή στο Σχεσιακό (28 μονάδες: 20 + 8)

Σας ζητείται να σχεδιάσετε το ER διάγραμμα για την ακόλουθη εφαρμογή καταγραφής στοιχείων για ταινίες του κινηματογράφου:

- **Σκηνοθέτης:** Κάθε σκηνοθέτης χαρακτηρίζεται μοναδικά από το αναγνωριστικό του (id). Για κάθε σκηνοθέτη κρατάμε επιπλέον πληροφορία για το ονοματεπώνυμό του (name) και την ηλικία του (age).
- **Ηθοποιός:** Κάθε ηθοποιός χαρακτηρίζεται μοναδικά από το αναγνωριστικό του (id). Για κάθε ηθοποιό κρατάμε επιπλέον πληροφορία για το ονοματεπώνυμό του (name) και το συνολικό αριθμό ταινιών στις οποίες έχει συμμετάσχει (num). Θεωρήστε ότι τα στοιχεία όλων αυτών των ταινιών είναι αποθηκευμένα στη βάση σας.
- Για κάθε ηθοποιό κρατάμε πληροφορία και για το ρόλο του (role) σε κάθε ταινία στην οποία έχει συμμετάσχει.
- Κάποιο άτομο (πχ, Κλιντ Ίστγουντ) μπορεί να είναι σκηνοθέτης σε κάποιες ταινίες και ηθοποιός σε άλλες. Αυτό κάπως πρέπει να αποτυπωθεί στο ER διάγραμμά σας, ώστε αν εμφανίζονται τέτοιες περιπτώσεις στα δεδομένα σας, να είναι φανερό ότι πρόκειται για το ίδιο άτομο που είναι στη μία περίπτωση ηθοποιός και στην άλλη σκηνοθέτης.
- **Ταινία:** Κάθε ταινία χαρακτηρίζεται μοναδικά από το συνδυασμό του ονόματός της (movieName) και του έτους παραγωγής της (year). Για κάθε ταινία κρατάμε επίσης πληροφορία για τη διάρκειά της (duration).
- Κάθε ηθοποιός συμμετέχει σε 0 ή περισσότερες ταινίες. Σε κάθε ταινία συμμετέχουν 1 ή περισσότεροι ηθοποιοί.
- Κάθε σκηνοθέτης σκηνοθετεί 1 ή περισσότερες ταινίες. Σε κάθε ταινία υπάρχει ακριβώς 1 σκηνοθέτης.

Α. Διορθώστε το παρακάτω ER διάγραμμα, το οποίο ίσως περιέχει κάποια λάθη.



B. Δώστε την εντολή CREATE TABLE για τον πίνακα (και μόνο για αυτόν τον πίνακα) που θα δημιουργήσετε κατά τη μετατροπή του Τύπου Οντοτήτων «**Ηθοποιός**» στο σχεσιακό σχήμα. Συμπεριλάβετε όλους τους απαιτούμενους περιορισμούς (constraints).

3. SQL (32 μονάδες: 10 + 10 + 12)

Για μία εφαρμογή παρόμοια (αλλά όχι εντελώς ίδια) με το Ερώτημα 2, θεωρήστε το ακόλουθο σχεσιακό σχήμα που περιγράφει Ταινίες, Ηθοποιούς, καθώς και ταινίες στις οποίες έχουν συμμετοχή ηθοποιοί:

Movie (<u>mid</u> , name, year, minutes)	// Αναγνωριστικό, Τίτλος, Έτος Παραγωγής, Διάρκεια σε Λεπτά
Appears (name, <u>lastName</u> , <u>movieID</u>)	// Όνομα Ηθοποιού, Επώνυμο Ηθοποιού, Αναγνωριστικό Ταινίας
Actor (name, <u>lastName</u> , age)	// Όνομα Ηθοποιού, Επώνυμο Ηθοποιού, Ηλικία

Προσέξτε ότι (σε αυτό το ερώτημα) ένας ηθοποιός μπορεί να μην έχει συμμετοχή σε καμία ταινία, αλλά σε κάθε ταινία έχει συμμετάσχει τουλάχιστον ένας ηθοποιός.

Δώστε τα ακόλουθα ερωτήματα σε SQL:

- Τυπώστε για κάθε ηθοποιό το όνομά του, το επώνυμό του, καθώς και το συνολικό αριθμό των ταινιών στις οποίες έχει συμμετάσχει. Για ηθοποιούς που δεν έχουν συμμετοχή σε καμία ταινία, ο αριθμός αυτός θα πρέπει να είναι ίσος με 0.
- Τυπώστε τα διακριτά ζευγάρια αναγνωριστικών από ταινίες που δεν έχουν μεταξύ τους κάποιο κοινό ηθοποιό. Πχ, αν οι ταινίες με αναγνωριστικά 1, 2 δεν έχουν μεταξύ τους κάποιον κοινό ηθοποιό, θα πρέπει να τυπωθεί είτε το ζευγάρι
1 2
ή το ζευγάρι
2 1
αλλά όχι και τα 2 ζευγάρια.
Για ταινίες με 0 ηθοποιούς, θα θεωρήσω σωστό είτε να τις εμφανίσετε (σε συνδυασμό με όλες τις άλλες ταινίες), είτε όχι.
- Τυπώστε όλα τα στοιχεία των ταινιών στις οποίες έχουν συμμετάσχει ηθοποιοί (ένας ή περισσότεροι) με τη μεγαλύτερη ηλικία (από όλους τους ηθοποιούς, γενικώς).

4. Βελτιστοποίηση Ερωτήσεων (20 μονάδες – 8 + 4 + 8)

Θα υπολογίσουμε το κόστος του πλάνου στα δεξιά, πάνω στους πίνακες:

R(A, B), S(Γ, Δ)

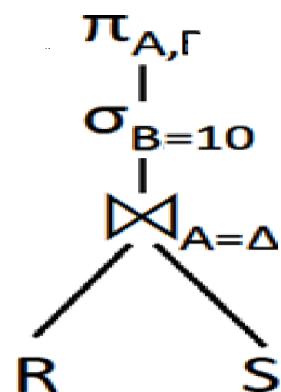
Θεωρήστε ότι έχετε τα ακόλουθα στατιστικά:

T(R) = 10000, B(R) = 500

T(S) = 5000, B(S) = 250

και γνωρίζετε ότι:

- το γνώρισμα R.B λαμβάνει 50 ακέραιες τιμές,
- το γνώρισμα S.Δ λαμβάνει 100 ακέραιες τιμές,
- το γνώρισμα S.Δ είναι **ξένο κλειδί** στο γνώρισμα R.A ,
- όλα τα γνώρισμα των σχέσεων R και S έχουν το ίδιο μέγεθος.



Θεωρήστε ότι υπάρχει **μη-συσταδοποιημένο** hash ευρετήριο στο S.Δ.

Θεωρήστε ότι υπάρχει **συσταδοποιημένο** B+-tree ευρετήριο στο R.B.

A. Υπολογίστε το κόστος όλου του πλάνου αν για το Join χρησιμοποιήσετε **index nested loop join**.

B. Εφαρμόστε τους **Ευριστικούς Κανόνες Βελτιστοποίησης** για να μετατρέψετε το πλάνο εκτέλεσης σε ένα πιο αποδοτικό πλάνο.

Γ. Υπολογίστε το κόστος όλου του πλάνου που δημιουργήσατε στο B υποερώτημα αν για το Join χρησιμοποιήσετε **block nested loop join** και γνωρίζετε ότι έχετε διαθέσιμη μνήμη ίση με 11 σελίδες για το join.