

Project Report

Σπυριδούλα Κουτσούκου, 3170080

Φίλιππος Δουραχαλής, 3170045

Part A)

Για την δημιουργία της βάσης μας ακολουθήσαμε τα εξής βήματα:

1. Δημιουργήσαμε έναν πίνακα για κάθε αρχείο csv και κατόπιν κάναμε import τα αντίστοιχα δεδομένα
2. Αφαιρέσαμε τα διπλότυπα από τους πίνακες (πλην του Ratings και του Ratings_Small)
3. Στη συνέχεια καθορίσαμε διαισθητικά ποιες θα πρέπει να είναι οι συσχετίσεις μεταξύ των πινάκων και επαληθεύσαμε τις συσχετίσεις αυτές εκτελώντας κατάλληλα αιτήματα (για παράδειγμα inner και outer join για την εύρεση συσχετιζόμενων πεδίων) πάνω σε όλους τους πίνακες
4. Κατασκευάσαμε το διάγραμμα ER της βάσης για να οπτικοποιήσουμε τις σχέσεις μεταξύ των πινάκων.
5. Καθώς βρήκαμε τις συσχετίσεις μεταξύ των πινάκων, αφαιρέσαμε όλες τις πλειάδες που δεν υπήρχαν στον Movies_Metadata από όλους τους πίνακες,
6. Εφόσον είχαμε βρει τους περιορισμούς και τις εξαρτήσεις επί των πινάκων, ορίσαμε primary key στον πίνακα Movies_Metadata
7. Αφού πλέον είχαμε αφαιρέσει τις πλειάδες που θα μπορούσαν να δημιουργήσουν πρόβλημα στον ορισμό ξένων κλειδιών, προσθέσαμε τα foreign keys στους υπόλοιπους πίνακες, τα οποία αναφέρονται στο id του Movies_Metadata, ώστε να υπάρχει ακεραιότητα δεδομένων.
8. Τέλος αλλάξαμε τον τύπο του πεδίου genres του Movies_Metadata από varchar σε JSON, εκτελώντας τους κατάλληλους μετασχηματισμούς για να μπορούμε να εκτελέσουμε εύκολα ερωτήματα που αφορούν το συγκεκριμένο πεδίο και ταχύτερα απ' ότι θα γινόταν αν το πεδίο genres αποτελούσε ξεχωριστό πίνακα συσχετιζόμενο με τον Movies_Metadata μέσω ενός άλλου, ενδιάμεσου πίνακα.