

ΑΚΑΔΗΜΑΪΚΟ ΕΤΟΣ 2019 – 2020

Project - Βάσεις Δεδομένων

Δανοπούλου Αιμιλία (3170033)

Μπαλή Νίκη (3170114)

Αρχικά δημιουργήσαμε ένα καινούριο schema στην βάση μας, το οποίο ονομάσαμε PROJECT.

Endpoint βάσης: database-1.cez6n49bmr6y.us-east-1.rds.amazonaws.com

Username: examiner

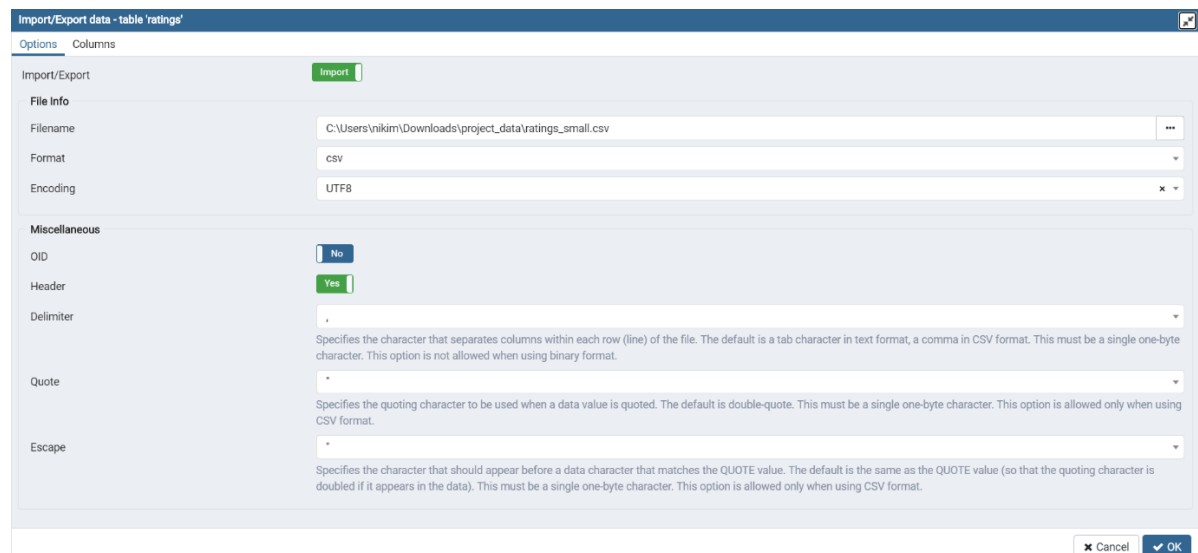
Password: examiner

Database: ActualDB

Schema: PROJECT

Αφού δημιουργήσαμε τους πίνακες, χρησιμοποιήσαμε το εργαλείο του pgadmin για να κάνουμε Import τα δεδομένα από τα αρχεία csv.

Ενδεικτικά:



Στη συνέχεια, διαγράψαμε διπλότυπα και θέσαμε primary key σε κάθε πίνακα, εκτός από τον Ratings.

Επιπρόσθετα, στον πίνακα Movies_Metadata αφαιρέσαμε τα γράμματα tt από όλα τα δεδομένα της στήλης imdb_id και αλλάξαμε τον τύπο σε int, και στη στήλη genres αντικαταστήσαμε τα μονά εισαγωγικά με διπλά ώστε τα δεδομένα να είναι σε μορφή json.

Μετά την επεξεργασία όλων των πινάκων διαγράψαμε δεδομένα ταινιών οι οποίες δεν υπάρχουν στον πίνακα Movies_Metadata αλλά υπάρχουν σε κάποιον από τους υπόλοιπους πίνακες και θέσαμε τα κατάλληλα ξένα κλειδιά.

«Τέλος δημιουργήστε ένα view table και αποθηκεύστε για κάθε χρήστη τον αριθμό των ratings που έχει κάνει καθώς και τη μέση βαθμολογία που έχει βάλει. Παίρνουμε κάποιο insight από αυτή τη σχέση?»

Το insight που παίρνουμε από το ViewTable είναι κατά πόσο επηρεάζεται η μέση βαθμολογία που δίνει κάθε κριτής ανάλογα με το πλήθος των κριτικών που κάνει, δηλαδή η αυστηρότητα του κριτή έχει σχέση με το πόσες κριτικές κάνει.