

Τα βήματα που ακολουθήσαμε για την αποθήκευση και την επεξεργασία των δεδομένων είναι τα εξής:

1. Δημιουργήσαμε τους πίνακες με τα column τους και ύστερα κάναμε import τα δεδομένα του κάθε πίνακα από τα csv αρχεία . Μια αλλαγή που έγινε για να περαστούν τα data είναι η αλλαγή τύπου , από int σε varchar , του πεδίου revenue στον πίνακα Movies\_Metadata.
2. Για την διαγραφή διπλότυπων στους πίνακες Credits , Keywords ,Movies\_Metadata και Links δημιουργήσαμε καινούριους πίνακες με τα ίδια πεδία και κάναμε αρχικά insert το πεδίο που είναι primary key με insert distinct και ύστερα κάναμε update τον πίνακα με set τα υπόλοιπα πεδία .
3. Για την διαγραφή δεδομένων ταινιών που δεν υπάρχουν στον πίνακα Movies\_Metadata φτιάξαμε καινούριους πίνακες με join τον Movies\_Metadata ,ελέγχοντας και διαγράφοντας τα δεδομένα που δεν υπάρχουν σε αυτόν .