

4η Εργασία: Επανασχεδιασμός της Βάσης μας

Προθεσμία: Κυριακή 31-03-2019

Σκοπός:

Κάνουμε αλλαγές στον σχεδιασμό της βάσης μας ώστε να κανονικοποιηθεί και να αποφευχθούν ανωμαλίες.

Προαπαιτούμενα:

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

Ζητούμενα Εργασίας:

Στην συγκεκριμένη Εργασία κάνουμε αλλαγές στην βάση μας ξεκινώντας από την κατάσταση που είχε όπως την ορίσαμε στην Εργασία 2. Για όλες αυτές τις αλλαγές, παραδοτέα θα είναι 6 sql αρχεία και ένα αρχείο εικόνας.

1. Πίνακας Host - αρχείο **host_table.sql**

- Δημιουργήστε έναν νέο πίνακα Host ο οποίος θα έχει τα εξής πεδία του Listings: table Host: host_id, host_url, host_name, host_since, host_location, host_about, host_response_time, host_response_rate, host_acceptance_rate, host_is_superhost, host_thumbnail_url, host_picture_url, host_neighbourhood, host_listings_count, host_total_listings_count, host_verifications, host_has_profile_pic, host_identity_verified, calculated_host_listings_count. Δηλαδή τα πεδία αυτά να πάρουν τις τιμές του Listings.
- Φτιάξτε τα πεδία του Host να έχουν ονόματα χωρίς το 'host_' πρόθεμα.
- Διαγράψτε τα αντιγραφέντα host πεδία από τον Listings εκτός από το host_id.
- Δηλώστε κατάλληλο Primary Key στον Host.
- Δηλώστε κατάλληλο Foreign Key στον Listings με αναφορά στον Host.

2. Πίνακας Calendar_Summary - αρχείο **calendar_summary_table.sql**

- Δημιουργήστε έναν νέο πίνακα Calendar_Summary ο οποίος θα έχει τα εξής πεδία του Listings: table Calendar_Summary: id, calendar_updated, calendar_last_scraped, availability_30, availability_60, availability_90. Δηλαδή τα πεδία αυτά να πάρουν τιμές του Listings.
- Μετονομάστε το πεδίο id σε listing_id.
- Μετονομάστε το πεδίο calendar_last_scraped σε from_date.
- Δηλώστε κατάλληλο σύνθετο Primary Key στον Calendar_Summary.
- Δηλώστε κατάλληλο Foreign Key στον Calendar_Summary με αναφορά στον Listings.

- Διαγράψτε τα αντιγραφέντα πεδία από τον Listings **εκτός από το id**.

3. Πίνακας Neighborhood - αρχείο **neighborhood_table.sql**

- Δημιουργήστε έναν νέο πίνακα Neighborhood που θα έχει distinct τιμές για ζευγάρια πεδίων neighborhood_name και zip_code προερχόμενα από τον πίνακα Listings. Για παράδειγμα: Rosedale, 78756.
- Στον Neighborhood μετονομάστε το μεταφερθέν neighbourhood σε neighborhood_name.
- Διαγράψτε τις εγγραφές με NULL τιμή στο neighborhood_name από τον πίνακα Neighborhood (εάν υπάρχουν τέτοιες).
- Στον Neighborhood μετονομάστε το μεταφερθέν neighbourhood_cleansed σε zip_code.
- Διαγράψτε τις εγγραφές με NULL τιμή στο zip_code από τον πίνακα Neighborhood (εάν υπάρχουν τέτοιες).
- Διαγράψτε το πεδίο neighbourhood_group_cleansed από τον Listings.
- Διαγράψτε τον παλιό πίνακα Neighbourhoods.
- Στον Listings μετονομάστε το neighbourhood σε neighborhood.
- Δηλώστε κατάλληλο σύνθετο Primary Key στον πίνακα Neighborhood.
- Δηλώστε Foreign Keys στον Listings για τα neighborhood και neighbourhood_cleansed στα αντίστοιχα πεδία του Neighborhood.
- Μετονομάστε το neighbourhood του Summary_Listings σε zip_code.
- Διαγράψτε το πεδίο neighbourhood_group του Summary_Listings.

4. Πίνακας Listings - αρχείο **listings.sql**

- Αλλάξτε τις τιμές του πεδίου street ώστε να περιέχουν μόνο την οδό χωρίς την πολιτεία, τον TK και την χώρα. Για παράδειγμα, η τιμή "Marathon Boulevard, Austin, TX 78756, United States" θέλουμε να γίνει UPDATE σε "Marathon Boulevard". Χρησιμοποιήστε την String function `split_part()` της Postgres.
- Αλλάξτε τις τιμές των πεδίων price, weekly_price, monthly_price, security_deposit, cleaning_fee αφαιρώντας όπου υπάρχουν τα σύμβολα του δολαρίου '\$' και τα κόμματα ','. Χρησιμοποιήστε την String function `replace()` της Postgres. Κατόπιν αυτού, κάντε ALTER τον τύπο αυτών των πεδίων από varchar σε numeric.

5. Πίνακας Calendar - αρχείο **calendar.sql**

- Ομοίως με τα αντίστοιχα πεδία του Listings, αφαιρέστε τα σύμβολα που χρειάζεται ώστε να επιτρέπεται το ALTER σε τύπο numeric για το πεδίο price. Μετατρέψτε τον τύπο σε numeric.

Για όλους τους Πίνακες

- Μετονομάστε όλους τους πίνακες από πληθυντικό σε ενικό αριθμό. Δηλαδή Listing, Review, κλπ. Τις εντολές αυτές παραδώστε τες σε αρχείο **all_tables.sql**
- Αλλάξτε το ER διάγραμμα ώστε να είναι σύμφωνο με τις αλλαγές που κάνατε. Φυσικά δεν θα απεικονίζονται όλες οι αλλαγές στο ER και επίσης υπενθυμίζουμε σε πίνακες με πολύ μεγάλο πλήθος πεδίων απλώς σχεδιάστε μερικά ενδεικτικά και όχι όλα. Παραδώστε το ER σαν αρχείο εικόνας.

Εργαλεία:

- AWS RDS Postgres instance
- Postgres psql ή/και PgAdmin
- [Draw.io](https://draw.io)

Συμβουλές για την υλοποίηση:

- Προτού αρχίσετε να κάνετε αλλαγές σε κάθε πίνακα θα μπορούσατε να φτιάχνετε ένα αντίγραφο αυτού. Σε περίπτωση που γίνει κάποιο λάθος και σας είναι δύσκολο να επιστρέψετε στην προηγούμενη κατάσταση που ήταν ο πίνακάς σας απλώς ξεκινάτε να δουλεύετε με το αντίγραφό του. Σε περίπτωση που φτιάξετε τέτοια αντίγραφα πινάκων, όμως, όταν ολοκληρώσετε την Εργασία παρακαλούμε σβήστε τα αντίγραφα για να διευκολυνθούμε στην διόρθωση.
- Επιβεβαιώστε ότι ο χρήστης της βάσης του οποίου μας στέλνετε τα credentials όντως έχει πρόσβαση στη βάση σας και μπορεί να τρέχει `SELECT` queries.
- Χρησιμοποιήστε την εντολή `CREATE TABLE AS SELECT` για δημιουργία πίνακα από τα αποτελέσματα ενός `SELECT` ερωτήματος.
- Χρησιμοποιήστε το `SELECT DISTINCT` για την απαλοιφή διπλοτύπων.
- Χρησιμοποιείτε `ALTER TABLE ... ADD PRIMARY/FOREIGN KEY ...`
- Χρησιμοποιείτε `ALTER TABLE ADD/DROP/RENAME COLUMN ...`
- Χρησιμοποιείτε `ALTER TABLE RENAME TO ...`
- Χρησιμοποιείτε `ALTER TABLE ALTER COLUMN TYPE ...` και εάν χρειαστεί και το `USING ...` <http://www.postgresqltutorial.com/postgresql-change-column-type/>
- Χρησιμοποιήστε την `split_part()` συνάρτηση για να πάρετε ένα τμήμα String από varchar πεδίο το οποίο χωρίζεται με delimiter `'.'`
- Χρησιμοποιήστε την `replace()` συνάρτηση για να αφαιρέσετε κάποιον χαρακτήρα από varchar πεδίο.
- Για τον σχεδιασμό του ER, όπως και στην Εργασία 2 προτείνεται η χρήση Draw.io

Χρήσιμα links:

-CREATE TABLE AS: <https://www.postgresql.org/docs/9.6/static/sql-createtablesas.html>
-DISTINCT: <https://www.postgresql.org/docs/9.6/static/sql-select.html#SQL-DISTINCT>
-ALTER TABLE: <https://www.postgresql.org/docs/9.6/static/sql-altertable.html>
-String functions: <https://www.postgresql.org/docs/9.6/static/functions-string.html>

Παραδοτέα:

- Δημιουργήστε ένα .txt αρχείο στο οποίο θα αναγράφονται το endpoint του AWS instance σας (μπορείτε να το δείτε στο AWS console, *RDS > Databases > db_identifier > Connectivity section*), 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>

- Βάλτε τα αρχεία sql, το αρχείο εικόνας και το αρχείο .txt 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*.