

# 5η Εργασία: Κανονικοποίηση και Ερωτήματα Ομαδοποίησης και Συνάθροισης

**Προθεσμία: Δευτέρα 08-04-2019**

## Σκοπός:

Συνεχίζουμε με μερικές ακόμα αλλαγές κανονικοποίησης και έπειτα τρέχουμε ερωτήματα ομαδοποίησης και συνάθροισης.

## Προαπαιτούμενα:

Θεωρούμε πως η βάση είναι στην κατάσταση που περιγράφεται στην 4η Εργασία.

## Ζητούμενα Εργασίας:

### Μέρος Α

- Κάντε τις απαιτούμενες μετατροπές στις τιμές του πεδίου `response_rate` του Πίνακα `Host` και μετατρέψτε το σε `numeric type`. Εάν η τιμή του είναι 'N/A' κάντε το `NULL`.
- Αντικαταστήστε το πεδίο `location` του `Host` με τα εξής 3 πεδία: `city`, `state`, `country`. Διαγράψτε το πεδίο `location`. Δεν χρειάζεται να μεριμνήσετε για τις λίγες εγγραφές που στο `location` δεν έχουν τα αναμενόμενα πεδία χωρισμένα με `' , '`.
- Παρατηρήστε ότι οι τιμές του πεδίου `amenities` στον Πίνακα `Listing` δεν είναι ατομικές. Προκειμένου να γίνει κανονικοποίηση, δημιουργήστε έναν νέο Πίνακα `Amenity` με πεδία `amenity_id` και `amenity_name`. Αποθηκεύστε στο `amenity_name` όλες τις μοναδικές ατομικές τιμές του πεδίου `amenities` αφού επεξεργαστούν κατάλληλα (Μπορείτε να χρησιμοποιήσετε τις `String functions` της `postgres` `regexp_split_to_table()`, `regexp_replace()` ή/και όποιες άλλες θεωρείτε ότι σας βολεύουν). Δηλώστε το `amenity_id` ως πεδίο αυτόματης προσαύξησης και κάντε το `Primary Key`.
- Κάντε τις ενέργειες που είναι απαραίτητες ώστε να υπάρχει σύνδεση του Πίνακα `Listing` με τον `Amenity`. Οι ενέργειες αυτές μπορεί να περιλαμβάνουν δημιουργία νέων πεδίων, ορισμό `Primary/Foreign Keys` μέχρι και δημιουργία νέου Πίνακα.
- Διαγράψτε το πεδίο `amenities` από τον `Listing`.
- Ενημερώστε το ER διάγραμμα της βάσης με τις αλλαγές.
- Ερώτηση: Βρίσκεται η βάση μας σε Κανονική Μορφή `BCNF`; Δικαιολογείστε για κάθε πίνακα. Εάν όχι, εντοπίστε τις "κακές" Συναρτησιακές Εξαρτήσεις.

Γράψτε όλες τις εντολές διαμόρφωσης της βάσης του Μέρους Α σε ένα αρχείο **part1.sql** και την τελευταία ερώτηση για `BCNF` σε ένα αρχείο **BCNF.docx** ή **BCNF.pdf**. Για το ER παραδώστε το τελικό αρχείο εικόνας.

## Μέρος B

Φτιάξτε 6 ερωτήματα συνάθροισης που να έχουν κάποιο νόημα.

- 3 από τα ερωτήματα θα πρέπει να έχουν τουλάχιστον ένα (inner ή outer) join.
- Τουλάχιστον 2 από τα ερωτήματα θα πρέπει να έχουν ένα WHERE.
- Τουλάχιστον 2 από τα ερωτήματα θα πρέπει να έχουν ένα ORDER BY.
- Τουλάχιστον 2 από τα ερωτήματα θα πρέπει να έχουν ένα HAVING.

Γράψτε όλα τα ερωτήματα σε ένα αρχείο **part2.sql**.

Προσθέστε τις σύντομες περιγραφές των ερωτημάτων και τα πλήθη των αποτελεσμάτων στο ίδιο αρχείο με τη μορφή σχολίων. Π.χ:

```
/* Find all rentals from all houses that took place in the 31st of  
December 2017
```

```
Output: 9663 rows
```

```
*/
```

```
SELECT listings.id, listings.listing_url, calendar.date_attr  
FROM listings  
INNER JOIN calendar  
ON listings.id = calendar.listing_id  
WHERE calendar.date_attr = '2017-12-31';
```

## Εργαλεία:

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

## Συμβουλές για την υλοποίηση:

- Προτού αρχίσετε να κάνετε αλλαγές σε κάθε πίνακα θα μπορούσατε να φτιάχνετε ένα αντίγραφο αυτού. Εάν 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.
- Τρέξτε και ελέγξτε κάθε ερώτημα στην Airbnb βάση σας.
- Επιβεβαιώστε ότι κανένα ερώτημα δεν είναι άνευ ουσίας όσον αφορά την εξερεύνηση των δεδομένων με την έννοια ότι δεν είναι απλή εμφάνιση κάποιου πίνακα. Το ζητούμενο είναι να υπάρχει συνδυασμός κριτηρίων ώστε να εξάγεται κάποια γνώση. αντι-π.χ: `select * from listings;` ή `select * from listings where id="123";`.
- Επιβεβαιώστε ότι ο χρήστης της βάσης του οποίου μας στέλνετε τα credentials όντως έχει πρόσβαση στη βάση σας.

## Χρήσιμα links:

<https://www.postgresql.org/docs/9.3/static/functions-matching.html>

<http://www.postgresqlonline.com/journal/archives/152-Regular-Expressions-in-PostgreSQL.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 part1.sql, BCNF.docx/pdf, part2.sql, το αρχείο εικόνας του ER και το αρχείο .txt σε ένα 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 / 5η Εργασία*.