

## 2<sup>ο</sup> Σύνολο Ασκήσεων

**Καταληκτική Ημερομηνία Παράδοσης:** Παρασκευή 16 Δεκεμβρίου 2022, 5μμ

**Θεματική Ενότητα:** Γλώσσες Ερωτήσεων: σχεσιακή άλγεβρα, SQL. Λογικός Σχεδιασμός.

Η παράδοση θα γίνει μέσω του ecourse  
Παραδώστε ένα pdf αρχείο με τις απαντήσεις σας σε όλες τις ασκήσεις.

Για γενικές οδηγίες σχετικά με τις ασκήσεις, συμβουλευτείτε την ιστοσελίδα του μαθήματος.

Οι Ασκήσεις 1 και 2 αφορούν μια βάση δεδομένων που διατηρεί πληροφορία για τη χρήση δημόσιων ποδηλάτων (bike sharing) σε μια περιοχή της Καλιφόρνιας (Bay Area). Το σχήμα της βάσης δεδομένων αποτελείται από τους τρεις πίνακες (σχήματα σχέσεων) της Εικόνας 1.

- Ο πίνακας **Trip** περιέχει πληροφορίες για τις διαδρομές που έγιναν με τα ποδήλατα.

Για παράδειγμα, η πλειάδα

(5088, 183, “2013-08-29 22:08:00”, “Market at 4th”, 76, “2013-08-29 22:12:00”, “Post at Kearney”, 47, 309)

σημαίνει ότι το ποδήλατο με **bike\_id** 309 έκανε τη διαδρομή με **id** 5088 και διάρκεια (**duration**) 183. Ξεκίνησε από τον σταθμό με **start\_station\_id** 76 και **start\_station\_name** “Market at 4th” στις (**start\_time**) “2013-08-29 22:08:00” και έφτασε στον σταθμό με **end\_station\_id** 47 και **end\_station\_name** “Post at Kenedy” στις (**end\_time**) 2013-08-29 22:12:00.

- Ο πίνακας **Station** περιέχει πληροφορίες για τους σταθμούς και τη διαθεσιμότητά τους σε ποδήλατα.

Για παράδειγμα, η πλειάδα

(2, “San Jose Diridon Caltrain Station”, 37.3297, -121.902, 27, “San Jose”, “2013-08-06”, “95113”)

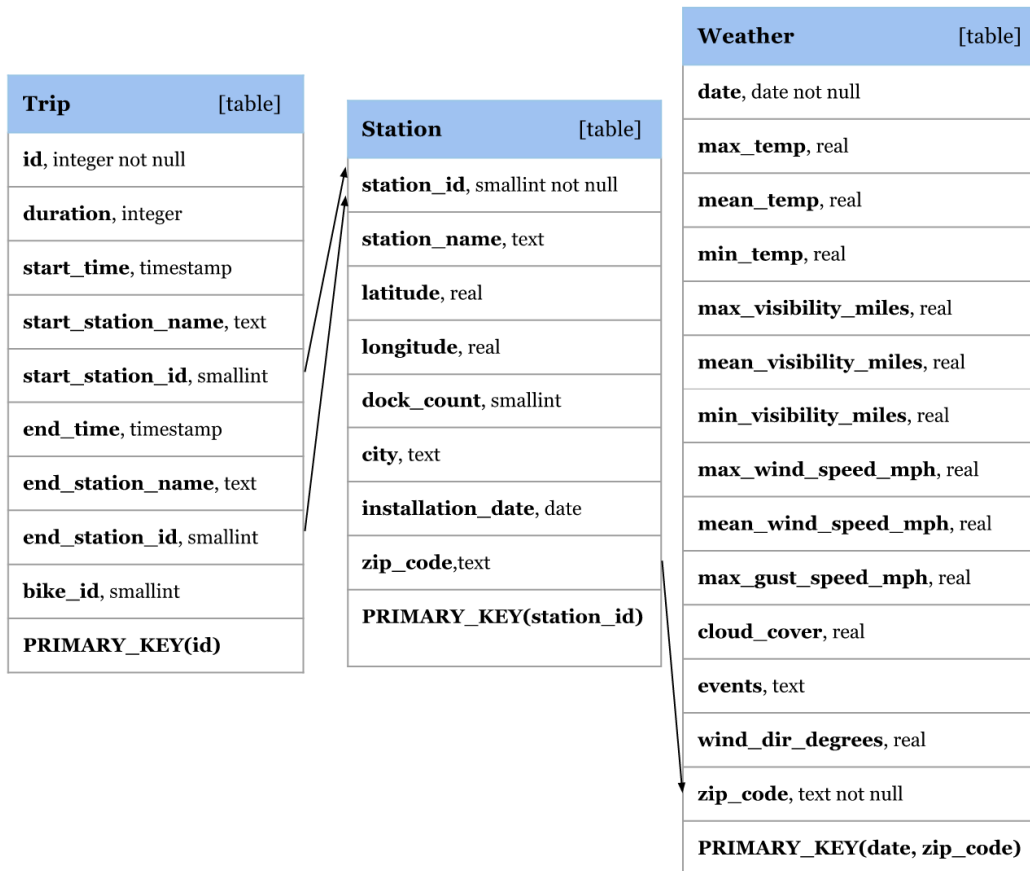
σημαίνει ότι ο σταθμός με **station\_id** 2 και όνομα (**station\_name**) “San Jose Diridon Caltrain Station” με γεωγραφικό πλάτος (**latitude**) 37.3297 και γεωμετρικό μήκος (**longitude**) -121.902 που βρίσκεται στην πόλη (**city**) “San Jose” έχει 27 θέσεις ποδηλάτων (**dock\_count**) που εγκαταστάθηκαν (**installation\_date**) στις “2013-08-06” και ο ταχυδρομικός κώδικας του (**zip code**) είναι 95113.

- Ο πίνακας **Weather** περιέχει πληροφορίες για τις καιρικές συνθήκες.

Για παράδειγμα, η πλειάδα

(“2013-08-29”, 74, 68, 61, 10, 10, 10, 23, 11, 28, 4, “”, 286, “94107”)

σημαίνει ότι στις (**date**) 2013-08-29, η πόλη με ταχυδρομικό κώδικα (**zip code**) 94107 είχε μέγιστη θερμοκρασία (**max\_temp**) 74 F, μέση θερμοκρασία (**mean\_temp**) 68 F, ελάχιστη θερμοκρασία (**min\_temp**) 61 F, μέγιστη ορατότητα (**max\_visibility\_miles**) 10 μίλια, μέση ορατότητα (**mean\_visibility\_miles**) 10 μίλια, ελάχιστη ορατότητα (**min\_visibility\_miles**) 10 μίλια, μέγιστη ταχύτητα ανέμου (**max\_wind\_speed\_mph**) 23 μίλια την ώρα, ελάχιστη ταχύτητα ανέμου (**min\_wind\_speed\_mph**) 11 μίλια την ώρα, μέγιστη ριπή ανέμων (**max\_gust\_speed\_mph**) 28 μίλια την ώρα, νεφοκάλυψη (**cloud\_cover**) 4, δεν είχε ειδικά μετεωρολογικά φαινόμενα (**events**) - όπως ομίχλη (fog), η βροχή (rain)-, και η διεύθυνση του ανέμου (**wind\_dir\_degrees**) ήταν στους 286 βαθμούς.



Εικόνα 1. Το σχήμα της βάσης δεδομένων για τις Ασκήσεις 1 και 2.

### Άσκηση 1 (σχεσιακή άλγεβρα)

Διατυπώστε σε σχεσιακή άλγεβρα (χωρίς χρήση συναθροιστικών συναρτήσεων τις οποίες δεν καλύψαμε στο μάθημα) ερωτήσεις που να δίνουν ως αποτέλεσμα:

- (α) Τις κυκλικές διαδρομές που αφορούν σταθμούς της πόλης San Jose. Συγκεκριμένα, ζεύγη (id, station-name) όπου το id είναι το id της κυκλικής διαδρομής και station-name το όνομα του σταθμού αφετηρίας και προορισμού. Μια διαδρομή ορίζεται ως κυκλική αν έχει τον ίδιο σταθμό ως αφετηρία και ως προορισμό.
- (β) Τις πόλεις που δεν έχουν ποτέ ούτε ομίχλη ούτε βροχή, δηλαδή, δεν έχει εμφανιστεί ποτέ το μετεωρολογικό φαινόμενο Fog ή το μετεωρολογικό φαινόμενο Rain ή και τα δύο.
- (γ) Τις διαδρομές (id) που τουλάχιστον ένας από τους σταθμούς τους είναι στο San Jose.
- (δ) Μια δυάδα (min\_max\_temp, max\_max\_temp) όπου min\_max\_temp είναι η μικρότερη μέγιστη θερμοκρασία και max\_max\_temp είναι η μεγαλύτερη μέγιστη θερμοκρασία που έχουν καταγραφεί στον πίνακα Weather.
- (ε) Την ημερομηνία που σημειώθηκε η μεγαλύτερη μέγιστη θερμοκρασία.

### Άσκηση 2 (SQL)

Τα παρακάτω πρέπει να υλοποιηθούν στη MySQL. Δημιουργείστε τους πίνακες. Φορτώστε τα δεδομένα που υπάρχουν στη σελίδα του μαθήματος.

- (α) Διατυπώστε σε SQL ερωτήσεις που να δίνουν ως αποτέλεσμα:
  - (i) Ό,τι η ερώτηση 1(β).
  - (ii) Ό,τι η ερώτηση 1(ε).

(iii) Για κάθε zip\_code, το ποσοστό των ημερών στις οποίες δεν υπήρχε κάποιο μετεωρολογικό φαινόμενο, δηλαδή, το πεδίο event ήταν κενό, συγκεκριμένα ζεύγη της μορφής (zip\_code, percentage), όπου για παράδειγμα ένα ζεύγος (95110, 30), σημαίνει ότι το 30% των ημερών στο zip\_code 95110 δεν υπάρχουν μετεωρολογικά φαινόμενα.

(iv) Τους 5 πιο δημοφιλείς σταθμούς για κυκλικές διαδρομές, δηλαδή τους σταθμούς που εμφανίζονται στις περισσότερες κυκλικές διαδρομές.

(v) Θεωρείστε ότι η μέση διάρκεια διαδρομών για μια πόλη ορίζεται ως ο μέσος όρος της διάρκειας όλων των διαδρομών που έχουν την πόλη αυτή ως προορισμό. Δώστε την πόλη με τη μεγαλύτερη μέση διάρκεια διαδρομών.

(β) Δώστε μια SQL έκφραση που να προσθέτει στον πίνακα Weather μια επιπρόσθετη εγγραφή για κάθε zip\_code που υπάρχει στη βάση δεδομένων. Για την ημερομηνία θεωρείστε μια οποιαδήποτε ημερομηνία του Νοεμβρίου του 2022 και για τα υπόλοιπα γνωρίσματα όποιες τιμές θέλετε.

(γ) [προαιρετική +1 μονάδα] Στόχος της άσκησης είναι να χρησιμοποιήσετε SQL από Java ή Python. Για να συνδεθείτε (connect) με τη βάση δεδομένων, χρησιμοποιήστε τον κατάλληλο driver.

Γράψτε ένα πρόγραμμα που

(i) Ο χρήστης θα δίνει ως είσοδο το όνομα μια πόλης και θα τυπώνονται στην οθόνη τα id των σταθμών αυτής της πόλης.

(ii) Θα τυπώνει για κάθε ποδήλατο τις διαδρομές του. Συγκεκριμένα, τα αποτελέσματα θα είναι μια ακολουθία από γραμμές. Κάθε γραμμή θα ξεκινά με το id του ποδήλατου και θα ακολουθούν τα id των διαδρομών στις οποίες χρησιμοποιήθηκε αυτό το ποδήλατο χωρισμένες με κόμμα.

### Άσκηση 3 (λογικός σχεδιασμός)

Θεωρείστε το σχεσιακό σχήμα  $R = (A, B, C, D, E, F)$  και το σύνολο συναρτησιακών εξαρτήσεων  $F = \{AB \rightarrow C, BC \rightarrow A, BD \rightarrow E, A \rightarrow D, AC \rightarrow B, E \rightarrow F\}$ .

(α) Η F είναι ή όχι ελάχιστο κάλυμμα. Εξηγήστε την απάντησή σας.

(β) Έστω η διάσπαση της R σε  $R1 = \{A, B\}$ ,  $R2 = \{A, C\}$ ,  $R3 = \{A, B, D, E\}$ ,  $R4 = \{E, F\}$ . Εξετάστε αν η διάσπαση είναι χωρίς απώλειες και αν διατηρεί ή όχι τις εξαρτήσεις.

(γ) Εξετάστε αν η  $R1 = \{A, B, C\}$  είναι σε BCNF ή 3NF.

(δ) Εξετάστε αν η  $R2 = \{A, B, C, E, F\}$  σε BCNF ή 3NF.