

2^ο Σύνολο Ασκήσεων

Καταληκτική Ημερομηνία Παράδοσης: Παρασκευή 11 Δεκεμβρίου 2020, 5μμ

Θεματική Ενότητα: Γλώσσες Ερωτήσεων: τυπικές γλώσσες (σχεσιακή άλγεβρα, σχεσιακός λογισμός), SQL.

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 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”)

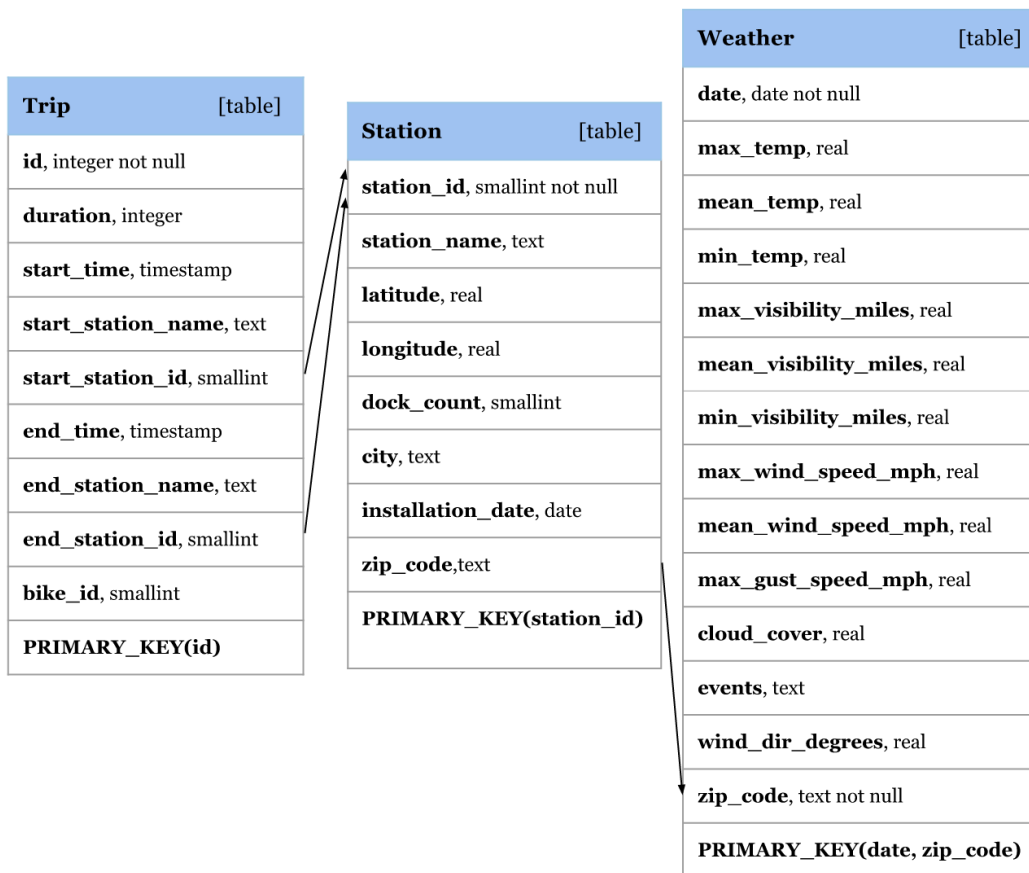
σημαίνει ότι ο σταθμός με **station_id** 2 και όνομα (**station_name**) “San Jose Diridon Caltrain Station” με γεωγραφικό πλάτος (**latitude**) 37.3297 και γεωμετρικό μήκος (**longitude**) -121.902 που βρίσκεται στην πόλη (**city**) “San Jose” έχει 27 θέσεις ποδηλάτων (**dock_count**) που εγκαταστάθηκαν (**installation_date**) στις “2013-08-06” και ο ταχυδρομικός του κώδικας (**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 γεγονότα (**events**) και η διεύθυνση του ανέμου (**wind_dir_degrees**) ήταν στους 286 βαθμούς.



Εικόνα 1. Το σχήμα της βάσης δεδομένων για τις Ασκήσεις 1, 2 και 3

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) Όλες τις κυκλικές διαδρομές, δηλαδή, τις διαδρομές που ξεκινούν και καταλήγουν στον ίδιο σταθμό. Συγκεκριμένα, ζεύγη (id, station-name) όπου το id το id της διαδρομής και station-name το όνομα του σταθμού.

(β) Τις πόλεις (το όνομα τους) στις οποίες εμφανίζονται τα μετεωρολογικά γεγονότα (events) ομίχλη (Fog) ή βροχή (Rain) αλλά όχι και τα δύο.

(γ) Μια δυάδα (duration1, duration2) όπου duration1 είναι η μικρότερη διάρκεια και duration2 η μεγαλύτερη διάρκεια των διαδρομών που εμφανίζονται στον πίνακα Trips (δηλαδή, τη διάρκεια της συντομότερης και της μακρότερης χρονικά διαδρομής).

(δ) Τα ποδήλατα που έχουν χρησιμοποιηθεί σε διαδρομές που έχουν είτε ως αφετηρία (start-station) είτε ως προορισμό (end-station) σταθμούς που βρίσκονται σε πόλεις που η ελάχιστη θερμοκρασία τους πέφτει κάτω από 50F (οποιαδήποτε μέρα, όχι απαραίτητα την ημέρα που έγινε η διαδρομή). Συγκεκριμένα, όλα τα ζεύγη (bike-id, city) όπου bike-id το id του ποδήλατου και city το όνομα της πόλης.

(ε) Τις πόλεις στους σταθμούς στους σταθμούς των οποίων έχει χρησιμοποιηθεί είτε ως αφετηρία (start-station) είτε ως προορισμό (end-station) το ποδήλατο με αριθμό 318.

Άσκηση 2 (σχεσιακός λογισμός)

Διατυπώστε σε σχεσιακό λογισμό:

(α) Την ερώτηση (β) της Άσκησης 1.

(β) Το ερώτημα (γ) της Άσκησης 2. Δώστε δύο διαφορετικές ερωτήσεις μία με χρήση του υπαρχαικού και μία με χρήση του καθολικού ποσοδείκτη.

Άσκηση 3 (SQL)

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

Διατυπώστε σε SQL ερωτήσεις που να έχουν ως αποτέλεσμα:

(i) Το ερώτημα (γ) της Άσκησης 1.

(ii) Το ερώτημα (ε) της Άσκησης 1.

Ερωτήματα που να έχουν ως αποτέλεσμα:

(iii) Το ποσοστό των κυκλικών διαδρομών επί όλων των διαδρομών (πχ 20% σημαίνει ότι το 20% των διαδρομών που εμφανίζονται στον πίνακα Trips είναι κυκλικές). Κυκλική είναι μια διαδρομή που ξεκινά και καταλήγει στον ίδιο σταθμό.

(iv) Τα ποδήλατα που έχουν χρησιμοποιήθηκαν σε διαδρομές που είχαν είτε ως αφετηρία (start-station) είτε ως προορισμό (end-station) σταθμούς που βρίσκονται σε πόλεις που εκείνη την ημέρα υπήρχε κάποιο έκτακτο μετεωρολογικό φαινόμενο (δηλαδή, το πεδίο event δεν ήταν κενό). Τυπώστε τον αριθμό (bike-id) του ποδήλατου και τον αριθμό των φορών που αυτό χρησιμοποιήθηκε.

(v) Τις 5 πιο δημοφιλείς πόλεις, δηλαδή τις πόλεις που οι σταθμοί τους εμφανίζονται είτε ως αφετηρία (start-station) είτε ως προορισμός στις περισσότερες διαδρομές. Συγκεκριμένα, ζεύγη (city, count) όπου city το όνομα της πόλης και count ο συνολικός αριθμός των διαδρομών.

(β) Δώστε μια SQL έκφραση που να προσθέτει στον πίνακα Trips για το ποδήλατο με αριθμό 318 διαδρομές που θα έχουν είτε ως αφετηρία (start-station) είτε ως προορισμό (end-station) όλους τους σταθμούς της πόλης (city) 'San Jose'. Στα υπόλοιπα πεδία δώστε όποιες τιμές θέλετε.

(γ) Στόχος της άσκησης είναι να χρησιμοποιήσετε SQL από Java ή Python. Για να συνδεθείτε (connect) με τη βάση δεδομένων, χρησιμοποιείστε τον κατάλληλο driver.

Γράψτε ένα πρόγραμμα που

(i) Ο χρήστης θα δίνει ως είσοδο το όνομα ενός πίνακα, το όνομα ενός πεδίου του και μια τιμή για αυτό το πεδίο και θα τυπώνει στην οθόνη τις πλειάδες του πίνακα που έχουν την συγκεκριμένη τιμή σε αυτό πεδίο.

(ii) Θα τυπώνει για κάθε πόλη τους σταθμούς της. Συγκεκριμένα, τα αποτελέσματα θα είναι μια ακολουθία από γραμμές. Κάθε γραμμή θα ξεκινά με το όνομα της πόλης και θα ακολουθούν τα ονόματα των σταθμών που υπάρχουν σε αυτήν.