



Μελέτη περίπτωσης F1 Grand Prix

Το Grandprix είναι μια βάση δεδομένων που περιέχει δεδομένα σχετικά με τους αγώνες της Formula 1. Τέτοια δεδομένα μπορούν να βρεθούν ως ανοιχτά σύνολα δεδομένων διαθέσιμα μέσω του Διαδικτύου (π.χ., Kaggle). Θα υποθέσουμε ότι ένα τέτοιο σύνολο δεδομένων έχει ληφθεί και προετοιμαστεί για περαιτέρω ανάλυση. Συγκεκριμένα, το συνημμένο αρχείο .csv περιέχει ενδεικτικά δεδομένα σχετικά με τις θέσεις των 10 πρώτων οδηγών σε δύο GP - το Grand Prix του Μπουχρέιν και το Grand Prix της Σαουδικής Αραβίας.

Assignment 1 (Συναρτησιακές εξαρτήσεις και αποσύνθεση σχήματος)

Η πρώτη εργασία περιλαμβάνει δύο μέρη. Στο πρώτο μέρος, θα πρέπει να χρησιμοποιήσετε την PostgreSQL για να ορίσετε μια καθολική σχέση που θα περιέχει δεδομένα στο αρχείο F1.csv. Στη συνέχεια, χρησιμοποιώντας την εντολή COPY, ανεβάστε τα δεδομένα στο αρχείο csv. Επεκτείνετε τη βάση δεδομένων σας εισάγοντας παρόμοια δεδομένα για δύο διαφορετικά Grand Prix του 2022 της επιλογής σας. Μπορείτε να βρείτε ακατέργαστα δεδομένα στην επίσημη ιστοσελίδα της F1 επιλέγοντας ένα Grand Prix και εντοπίζοντας τα αποτελέσματα των Προκριματικών. Δεδομένου του συνόλου δεδομένων σας, διερευνήστε πιθανά υποψήφια κλειδιά για τον πίνακα που έχετε δημιουργήσει για να φιλοξενήσει τα δεδομένα. Απαιτείται συγκεκριμένα να διερευνήσετε εάν το ζεύγος (GRANDPRIXID, BASELINEPOSITION) μπορεί να είναι υποψήφιο κλειδί και να συνοψίσετε την απάντησή σας σε ένα σύντομο έγγραφο. Στο δεύτερο μέρος θα πρέπει να αξιολογήσετε πιθανές σχεσιακές αποσυνθέσεις. Συγκεκριμένα, στην προσπάθειά σας να αξιολογήσετε την πιθανότητα το ζεύγος (GRANDPRIXID, BASELINEPOSITION) να είναι υποψήφιο κλειδί για το σχήμα μίας σχέσης, μπορεί επίσης να έχετε εντοπίσει πρόσθετους συνδυασμούς χαρακτηριστικών ως υποψήφια κλειδιά. Σε αυτό το μέρος της εργασίας σας ζητείται να αξιολογήσετε κριτικά την επάρκεια του σχήματος μίας σχέσης και να σχεδιάσετε μια κατάλληλη αποσύνθεση στην οποία όλες οι σχέσεις βρίσκονται σε 3NF ή BCNF. Πρέπει να δικαιολογήσετε τη λύση σας αναλόγως. Μόλις είστε ικανοποιημένοι με την καθορισμένη αποσύνθεση, θα πρέπει να την εφαρμόσετε σε PostgreSQL χρησιμοποιώντας ένα κατάλληλο υποσύνολο συνόλου δεδομένων.

Assignment 2 (Νέοι τύποι δεδομένων)

Η Εργασία 2 απαιτεί PostgreSQL, ώστε να μπορείτε να πειραματιστείτε με νέους τύπους δεδομένων και τις δυνατότητες που προσφέρονται. Για το σκοπό αυτό θα θεωρήσετε ότι η βάση δεδομένων GRANDPRIX πρέπει να επεκταθεί έτσι ώστε να περιλαμβάνει τη σύνθεση των οδηγών και των αυτοκινήτων τους σε κάθε Grand Prix. Ως σύνθεση θα λάβετε υπόψη τις θέσεις εκκίνησης των οδηγών πριν από τον πραγματικό αγώνα. Αυτό είναι προφανώς διαφορετικό από το τελικό αποτέλεσμα μετά τον αγώνα, το οποίο έχει κωδικοποιηθεί στα σύνολα δεδομένων csv σας μέχρι στιγμής. Στη συνέχεια, θα χρησιμοποιήσετε τον πίνακα Postgres για να δημιουργήσετε έναν κατάλληλο πίνακα που να απεικονίζει τη σύνθεση ως λίστες επιμονής (δηλαδή, ποιος ακολουθεί



/ ακολουθείται από έναν οδηγό στη σύνθεση). Μόλις δημιουργηθεί ο πίνακας, μπορείτε να εισάγετε ένα κατάλληλο σύνολο δεδομένων για ένα Grand Prix και να απαντήσετε σε ερωτήματα όπως τα ακόλουθα:

- Βρείτε τους οδηγούς που ξεκίνησαν τον αγώνα πίσω από τον οδηγό με έναν κωδικό της επιλογής σας στο grand prix της επιλογής σας
- Βρείτε τον οδηγό που ξεκίνησε τον αγώνα μία θέση πίσω από τον οδηγό με έναν κωδικό της επιλογής σας στο grand prix της επιλογής σας
- Βρείτε τον οδηγό που ξεκίνησε τον αγώνα δύο θέσεις πίσω από τον οδηγό με έναν κωδικό της επιλογής σας στο grand prix της επιλογής σας
- Βρείτε τον οδηγό που ξεκίνησε τον αγώνα στην πρώτη θέση στο grand prix της επιλογής σας

Assignment 3 (JSON/XML και ιεραρχίες εξειδίκευσης)

Η τρέχουσα άσκηση είναι η πρώτη σας προσπάθεια να εξετάσετε τη χρήση πολλαπλών διαφορετικών τύπων δεδομένων για την αντιμετώπιση των απαιτήσεων ενός προβλήματος. Καθώς τα ζητήματα που εγείρονται είναι ελαφρώς πιο απαιτητικά, η χρονική κλίμακα εντός της οποίας απαιτείται να ολοκληρώσετε την εργασία είναι της τάξης των μερικών εβδομάδων αντί για μερικών ημερών. Συγκεκριμένα, θα συνεχίσετε να εργάζεστε στη βάση δεδομένων GRANDPRIX, αλλά αυτή τη φορά θα εκτεθείτε σε πρόσθετους τύπους δεδομένων όπως σύνθετους τύπους δεδομένων που ορίζονται από τον χρήστη, JSON και κληρονομικότητα.

Η ομάδα ανάπτυξής σας έχει θέσει νέες απαιτήσεις που πρέπει να αντιμετωπιστούν στον σχεσιακό σας σχεδιασμό. Πρώτον, πρέπει να επεκτείνετε τη βάση δεδομένων σας ώστε να καταγράφει περιοδικές μετρήσεις (δηλαδή, κάθε 5 λεπτά ξεκινώντας από μια καθορισμένη ώρα) της εξωτερικής θερμοκρασίας στον τόπο όπου λαμβάνει χώρα ο αγώνας. Για τον σκοπό αυτό, οι διοργανωτές βασίζονται σε πολλά πανομοιότυπα κιτ μετρήσεων που είναι καταναμημένα σε τρεις διαφορετικές τοποθεσίες στην πίστα κάθε αγώνα. Κάθε μέτρηση που λαμβάνεται από αυτά τα κιτ καταγράφει τη χρονική σήμανση της μέτρησης, τη μετρούμενη μεταβλητή και την πραγματική τιμή. Όσον αφορά τις μεταβλητές, αυτές περιλαμβάνουν τη θερμοκρασία, τη βροχόπτωση, την υγρασία και τον άνεμο. Μια ενδεικτική μέτρηση θα μπορούσε επομένως να είναι η εξής:

Wednesday 10:00 AM, Mostly sunny Temperature: 13C, Precipitation: 0%, Humidity: 44%, Wind: 37 km/h

Η δεύτερη απαίτηση είναι ότι η βάση δεδομένων μας θα πρέπει επίσης να καταγράφει ζωντανά δεδομένα για κάθε αυτοκίνητο κατά τη διάρκεια του αγώνα. Τέτοια δεδομένα συλλέγονται από αισθητήρες που είναι εγκατεστημένοι σε όλο το αυτοκίνητο. Παραδείγματα τέτοιων αισθητήρων περιλαμβάνουν αισθητήρες θερμοκρασίας διαφόρων ειδών (π.χ. αισθητήρες θερμοκρασίας κινητήρα και κιβωτίου αέρα, αισθητήρες θερμοκρασίας χωρίς επαφή που μετρούν την τριβή μεταξύ των εξαρτημάτων με υπέρυθρη ενέργεια κ.λπ.), επιταχυνσιόμετρα που μετρούν τις



δυνάμεις g που δημιουργούνται κατά τη διάρκεια των στροφών ή του φρεναρίσματος, αισθητήρες πίεσης που μετρούν υδραυλικά συστήματα, αισθητήρες διπλού άξονα που μετρούν το φρενάρισμα και το τιμόνι κ.λπ. Θα πρέπει να επιλέξετε τουλάχιστον πέντε διαφορετικούς τύπους αισθητήρων (της επιλογής σας) και για κάθε αισθητήρα θα πρέπει να ορίσετε έναν κατάλληλο/κατάλληλο τύπο δεδομένων για τη διατήρηση μετρήσεων κατά τη διάρκεια ενός γύρου/αγώνα ενός grand prix. Λάβετε υπόψη ότι θα πρέπει να συμπεριλάβετε αισθητήρες των οποίων οι μετρήσεις καταγράφονται (i) συνεχώς (με χρονική σήμανση) κατά τη διάρκεια του γύρου (ii) περιοδικά, π.χ. κάθε 3 δευτερόλεπτα για συνολικά 5 φορές ανά γύρο και (iii) υπό όρους με την ανθρώπινη ενέργεια του οδηγού. Οι τελευταίες τρεις κατηγορίες τύπων αισθητήρων θα πρέπει να χρησιμοποιούνται για τη δημιουργία μιας ιεραρχίας κληρονομικότητας για τους δύο πέντε αισθητήρες που θα επιλεγούν.

Για να αντιμετωπίσετε τις παραπάνω απαιτήσεις, πρέπει να μελετήσετε το πρόβλημα που αντιμετωπίζετε, να διερευνήσετε τον/τους κατάλληλο/ους τύπο/ους δεδομένων που θα χρησιμοποιηθούν για κάθε πρόκληση και να συμπληρώσετε τη βάση δεδομένων σας με κατάλληλα δεδομένα. Λάβετε επίσης υπόψη ότι ορισμένα από τα δεδομένα μπορούν να κωδικοποιηθούν χειροκίνητα, ενώ άλλα μπορούν να ληφθούν ως δημόσια δεδομένα από ανοιχτά σύνολα δεδομένων. Η λύση σας θα πρέπει να περιλαμβάνει τύπους δεδομένων enum, πίνακες, τύπους δεδομένων που ορίζονται από τον χρήστη, τύπους δεδομένων JSON και XML, όπως κρίνετε κατάλληλους και κατάλληλους.

Assignment 4 (Γράφοι σε Postgres και ερωτήματα τροχιάς σε SQL)

Η άσκηση 4 έχει ως διττό στόχο αφενός να εξοικειώσει τους φοιτητές με την ανάπτυξη γράφων και αφετέρου να επιτρέψει την ανάλυση εναλλακτικών τρόπων αναπαράστασης γράφων με τη χρήση της σχεσιακής τεχνολογίας. Προοδευτικά η ανάλυση αυτή θα επιτρέψει την κατανόηση και εξάσκηση ορισμένων προχωρημένων τεχνικών της PostgreSQL όπως η αναδρομική λειτουργία για τη διάσχιση (traversal) και το μεταβατικό κλείσιμο (transitive closure) γράφων. Οι φοιτητές εξακολουθούν και εργάζονται στην τρέχουσα υλοποίηση της βάσης δεδομένων GRANDPRIX όπου θα χρειαστεί να υλοποιηθούν τα ακόλουθα:

4A (Δημιουργία ενός *property graph* και υλοποίηση σε PostgreSQL): Θα ξεκινήσετε αναπτύσσοντας (αρχικά σε χαρτί ή χρησιμοποιώντας ένα εργαλείο) ένα γράφημα ιδιοτήτων (βλ. θεωρία) που θα αποτυπώνει: (α) τις σχέσεις μεταξύ οδηγών και κατασκευαστών (δηλαδή, ο Luis Hamilton είναι μέλος της ομάδας Ferrari) και (β) τον οδηγό που κάνει τον ταχύτερο γύρο σε ένα grand prix (δηλαδή, ο Luis Hamilton έκανε τον ταχύτερο γύρο στο grand prix του Μπαχρέιν το 2025). Θα πρέπει να στοχεύσετε στη συμπερίληψη επαρκών δεδομένων για να έχετε νόημα στην υποβολή ερωτημάτων στο γράφημα. Αφού σχεδιάσετε το γράφημα ιδιοτήτων, θα πρέπει στη συνέχεια να αναπαραστήσετε το γράφημα ιδιοτήτων χρησιμοποιώντας σχεσιακή τεχνολογία. Είναι στη διακριτική σας ευχέρεια να επιλέξετε μια τεχνική από αυτές που έχετε μάθει μέχρι



στιγμής, δηλαδή τη χρήση κανονικοποιημένων σχέσεων, μη κανονικοποιημένων σχέσεων, λιστών γειτνίασης κ.λπ., ενώ θα πρέπει να μελετήσετε τον τρόπο με τον οποίο θα υποστηρίζονται οι εναλλακτικοί κόμβοι και οι ακμές που θα υπάρχουν στο γράφημά σας (βλ. μηχανισμό κληρονομικότητας PostgreSQL).

4B (Ερωτήματα τροχιάς με αναδρομή): Έχοντας αναπαραστήσει το property graph με σχέσεις θα πρέπει να εξοικειωθείτε με σύγχρονες τεχνικές διάσχισης γράφων χρησιμοποιώντας αναδρομικές κλήσεις της SQL. Συγκεκριμένα, στην τρέχουσα έκδοση της βάσης δεδομένων Grandprix που έχετε υλοποιήσει θα πρέπει να απαντήσετε στα παρακάτω ερωτήματα:

1. Για κάθε / ένα κόμβο υπολογίστε τους επόμενους (σε μια κατεύθυνση της επιλογής σας)
2. Για κάθε / ένα κόμβο υπολογίστε τους προηγούμενους (σε μια κατεύθυνση επιλογής σας)
3. Για κάθε / ένα κόμβο υπολογίστε τους επόμενους ανεξαρτήτως κατεύθυνσης
4. Για κάθε / ένα κόμβο υπολογίστε τους προηγούμενους ανεξαρτήτως κατεύθυνσης
5. Υπολογίστε όλες τις πιθανές διαδρομές που συνδέουν τους καταγεγραμμένους developers στο property graph που υλοποιήσατε

Assignment 5 (NoSQL)

Η τελική εργασία είναι να επιλέξετε ένα σύστημα NoSQL της επιλογής σας για να επανεξετάσετε (ένα σχετικό υποσύνολο) της βάσης δεδομένων GRANDPRIX. Η ιδέα είναι να αποκτήσετε πρακτική εμπειρία (μόνο) με ένα σύστημα NoSQL και όχι να δημιουργήσετε μια ακόμη (ολοκαίνουργια ή συγκρίσιμη) υλοποίηση με την έκδοση PostgreSQL. Για το σκοπό αυτό, παρακαλώ κάθε φοιτητή να με ειδοποιήσει (μέσω email στη διεύθυνση da@hmu.gr) σχετικά με το σύστημα NoSQL που θα χρησιμοποιηθεί και μια ενδεικτική προθεσμία για την παρουσίαση της εργασίας στην τάξη.