

ΚΑΤΑΝΕΜΗΜΕΝΗ ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ

Υποχρεωτική Ατομική Προγραμματιστική εργασία

Να δημιουργηθεί ένα πρόγραμμα (κατά προτίμηση σε Java ή Scala), το οποίο να έχει τις παρακάτω λειτουργίες:

1. Να γεμίζει με τυχαίες εγγραφές δύο πίνακες σε δύο διαφορετικές βάσεις δεδομένων τύπου Redis ή Ignite ή SQLite. Οι πίνακες να συνδέονται βάσει συνθήκης ισότητας και στα πεδία να υπάρχει πεδίο (τεχνητής) χρονοσφραγίδας.
2. Να υλοποιεί, υποβάλλοντας SQL ερωτήματα στις δύο βάσεις δεδομένων τους παρακάτω αλγόριθμους σύνδεσης **εξωτερικά** από τις βάσεις δεδομένων: **pipelined hash join**, **semi-join** και προαιρετικά (θα έχει bonus) **intersection bloom filter join**. Σε όλες τις συνδέσεις, εκτός από την ισότητα στο πεδίο σύνδεσης, να ελέγχεται οι χρονοσφραγίδες να μην απέχουν πάνω από ένα όριο.

Παραδοτέα:

- α) πηγαίος κώδικας με εκτενή σχόλια είτε στα ελληνικά είτε στα αγγλικά,
- β) εκτελέσιμο κατά προτίμηση χρησιμοποιώντας docker
- γ) σύντομες οδηγίες εκτέλεσης και
- δ) σύντομη τεχνική αναφορά το πολύ 4 σελίδων αναφορικά με τον τρόπο υλοποίησης, σύγκριση, σχεδιασμό και τις επισημάνσεις σας, δίνοντας έμφαση σε θέματα πως υλοποιήσατε αποδοτικά τον έλεγχο στη διαφορά των χρονοσφραγίδων και την τεχνική semi-join ώστε να έχει νόημα. Για τη σύγκριση, να συμπεριλαμβάνονται απλά πειράματα αποτίμησης της αποδοτικότητας.

ΠΡΟΘΕΣΜΙΑ: Παρασκευή 16/7/2023 11:59μμ