



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ
ΕΡΓΑΣΤΗΡΙΟ ΥΠΟΛΟΓΙΣΤΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

Διπλωματική Εργασία 2011-2012

Αποθήκευση Χωροχρονικών Δεδομένων σε Περιβάλλον Cloud

Εισαγωγή

Η εξάπλωση των ασύρματων δικτύων σε συνδυασμό με την εξέλιξη των κινητών τηλεφώνων σε υπερσύγχρονες συσκευές με ιδιαίτερες δυνατότητες έχει προάγει την ευρεία χρήση κοινωνικών δικτύων και υπηρεσιών από κινητούς χρήστες. Τα δεδομένα που παράγουν ποικίλουν από ακατέργαστα δεδομένα που λαμβάνονται από τις κινητές συσκευές (π.χ. συντεταγμένες GPS) μέχρι μηνύματα κειμένου, φωτογραφίες, βίντεο με τα οποία οι χρήστες ανταλλάσσουν νέα, εκφράζουν ιδέες και γνώμες, μοιράζονται τις θέσεις τους, τη διάθεση ή τα συναισθήματά τους.

Επιπλέον, για κινητούς χρήστες υπάρχουν εξειδικευμένες υπηρεσίες, οι υπηρεσίες θέσης (Location Based Services or LBS) που προσπελάζονται μέσω κινητών συσκευών και κάνουν χρήση της γεωγραφικής τους θέσης. Για παράδειγμα το Foursquare [2] δίνει την δυνατότητα στους χρήστες να κάνουν check in στις διάφορες τοποθεσίες που επισκέπτονται (όπως εστιατόρια, καφετέριες, μουσεία κτλ), να μοιράζονται την πληροφορία αυτή με τους φίλους τους και να απολαμβάνουν προσωποποιημένες υπηρεσίες βάσει της θέσης τους.

Ο όγκος των δεδομένων που παράγουν οι κινητοί χρήστες είναι τεράστιος και αποτελεί μια πολύτιμη πηγή γνώσης και πληροφορίας που μπορεί να συσχετιστεί χωρικά, χρονικά και κοινωνικά για την εξαγωγή χρήσιμων συμπερασμάτων, όπως π.χ. την αναγνώριση αυθόρμητων γεγονότων που μπορεί να συμβούν σε κάποιο σημείο. Ο όγκος αυτός απαιτεί καταναμημένες λύσεις και αποτελεί μεγάλη πρόκληση για τον ερευνητικό κόσμο.

Υπάρχει πληθώρα κλασικών τεχνικών δεικτοδότησης χωροχρονικών δεδομένων στην περιοχή των βάσεων δεδομένων (Space Filling Curves, k-d trees, Quad trees, R-trees, κλπ.), που ωστόσο λειτουργούν μόνο κεντρικά, θέτοντας περιορισμούς στον όγκο των δεδομένων που μπορούν να χειριστούν. Από την άλλη πλευρά, υπάρχουν ευρέως διαδεδομένες καταναμημένες βάσεις, όπως η HBase [1], οι οποίες όμως έχουν περιορισμένη λειτουργικότητα (key-value primitives) και δε μπορούν να διαχειριστούν αποδοτικά ειδικούς τύπους δεδομένων όπως είναι τα χωροχρονικά.

Στην παρούσα διπλωματική θα ασχοληθούμε με την αποδοτική δεικτοδότηση και

αποθήκευση χωροχρονικών δεδομένων σε καταναμημένο περιβάλλον, ώστε να είναι δυνατή η μαζική επεξεργασία τους για την εξαγωγή πληροφορίας. Για το σκοπό αυτό θα ενσωματώσουμε τεχνικές από την περιοχή των βάσεων δεδομένων για τη δεικτοδότηση χωροχρονικών δεδομένων στην πλέον διαδεδομένη καταναμημένη βάση σε περιβάλλοντα Cloud, την HBase [3].

Επικοινωνία:

Νεκτάριος Κοζύρης, Καθηγητής nkoziris@cslab.ece.ntua.gr

Κατερίνα Δόκα, Μεταδ. Ερευνήτρια katerina@cslab.ece.ntua.gr

Βιβλιογραφία

[1] The Apache HBase Project. <http://hbase.apache.org/>. 1

[2] Foursquare. <https://foursquare.com/>. 1

[3] Shoji Nishimura, Sudipto Das, Divyakant Agrawal, and Amr El Abbadi. Md-hbase: a scalable multi-dimensional data infrastructure for location aware services. In *Mobile Data Management (MDM), 2011 12th IEEE International Conference on*, volume 1, pages 7–16. IEEE, 2011. 2