

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

ΠΑΡΟΥΣΙΑΣΗ / ΕΞΕΤΑΣΗ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

**Καρνιαβούρα Φλώρα
Μεταπτυχιακή Φοιτήτρια**

**Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης
Επόπτης Μεταπτ. Εργασίας: Καθηγητής Δ. Πλεξουσάκης**

**Παρασκευή, 24/3/2017, 11:00
Αίθουσα B108, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης**

“ Πρόβλεψη της απόδοσης NoSQL συστημάτων με βάση συστηματικές μετρήσεις”

ΠΕΡΙΛΗΨΗ

Η δυνατότητα ακριβούς πρόβλεψης του τύπου και της ποσότητας των απαραίτητων πόρων για να διατηρηθεί ένα επιθυμητό επίπεδο απόδοσης αποτελεί σημαντικό στόχο κάθε συστήματος διαχείρισης υπηρεσιών λογισμικού. Η χρήση συστηματικών μετρήσεων με στόχο τη δημιουργία μοντέλων πρόβλεψης της απόδοσης λογισμικού έχει μελετηθεί ερευνητικά στο παρελθόν, ωστόσο τέτοια μοντέλα για εφαρμογές που χρησιμοποιούν NoSQL βάσεις δεδομένων δεν έχουν μελετηθεί αρκετά ως σήμερα. Οι τεχνικές πρόβλεψης απόδοσης που βασίζονται σε μετρήσεις συχνά περιορίζονται από το στενό εύρος χαρακτηριστικών διαθέσιμου υλικού στις ιδιωτικές υποδομές κάθε οργανισμού. Μια ευκαιρία για να βελτιωθεί αυτή η κατάσταση, είναι η εξάπλωση γεωγραφικά κατανεμημένων και υπηρεσιοστρεφών υποδομών έρευνας και εκπαίδευσης μεγάλης κλίμακας, οι οποίες προσφέρουν στην ερευνητική κοινότητα μια ποικιλία ετερογενούς υλικού συστημάτων. Η παρούσα διπλωματική εργασία παρουσιάζει μια μεθοδολογία βασισμένη σε μετρήσεις για την πρόβλεψη της απόδοσης του Yahoo Cloud Serving Benchmark (YCSB) πάνω από NoSQL συστήματα, σε τέτοιες

υποδομές. Λαμβάνουμε υπόψη τρεις τεχνικές regression: Multivariate adaptive regression splines (MARS), support vector regression (SVR), και artificial neural network (ANN) regression. Με βάση την ανάλυση πειραματικών αποτελεσμάτων καταλήγουμε στο συμπέρασμα πως και οι τρεις τεχνικές επιτυγχάνουν πρόβλεψη απόδοσης με μέσο όρο ακρίβειας πάνω από 90%, με την τεχνική MARS να επιτυγχάνει τα καλύτερα αποτελέσματα. Στη συνέχεια επεκτείνουμε τα αποτελέσματά μας εστιάζοντας στο MARS σε ένα εικονικό περιβάλλον νέφους, μελετώντας δυο NoSQL βάσεις δεδομένων, την MongoDB και την RethinkDB. Παρά τις διαφορές μεταξύ των δύο συστημάτων, επιβεβαιώνουμε πως το MARS επιτυγχάνει πρόβλεψη απόδοσης με ικανοποιητική ακρίβεια και στα δυο συστήματα, με μέσο όρο 95% στις περιπτώσεις που μελετήσαμε. Με βάση αυτά τα αποτελέσματα η παρούσα διπλωματική εργασία καταδεικνύει τη θετική προοπτική αυτής της μεθοδολογίας σε εφαρμογές που χρησιμοποιούν βάσεις δεδομένων τύπου NoSQL.

Karviavoura Flora

M.Sc. Thesis

Computer Science Department

University of Crete

Master's Thesis Supervisor: Professor Dimitris Pleksousakis

Friday, 24/3/2017, 11:00

Room B108, Computer Science Dept., University of Crete

“A measurement based approach to performance prediction in NoSQL systems”

ABSTRACT

The ability to accurately predict the amount and type of resources needed to sustain a desired level of service is an important enabler of goal-oriented application performance management. While the use of systematic measurements for building performance

prediction models is a well studied topic, little attention has been paid so far on the application space of data-intensive systems using NoSQL databases. In this thesis we introduce and evaluate a measurement-based approach to performance prediction of data-intensive applications over NoSQL systems. Measurement-based performance prediction approaches are often limited by a relatively narrow range of hardware characteristics available within each organization's private infrastructure. An opportunity to change this fact is the emergence of federated, large-scale, service-oriented research infrastructures, featuring a variety of heterogeneous hardware. This thesis demonstrates accurate measurement-based prediction of Yahoo Cloud Serving Benchmark (YCSB) performance over NoSQL systems in such infrastructures. We consider three regression techniques: Multivariate adaptive regression splines (MARS), support vector regression (SVR), and artificial neural network (ANN) regression. We find that all three techniques achieve performance prediction with average accuracy of over 90%, with MARS yielding the best results. We extend our results focusing on MARS and a virtualized private cloud environment with two NoSQL data stores, MongoDB and RethinkDB. Despite their differences, we find that MARS achieves accurate performance prediction on both data stores with an average accuracy of 95% across cases considered. This result points to the potential of applying our methodology to a broader set of NoSQL systems and deployment environments.