

ΥΠΟΜΝΗΜΑ ΕΠΙΣΤΗΜΟΝΙΚΟΥ ΕΡΓΟΥ

ΙΩΑΝΝΗΣ Η. ΚΩΝΣΤΑΝΤΙΝΟΥ

Ηλεκτρολόγος Μηχανικός & Μηχανικός Υπολογιστών ΕΜΠ

PhD, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών ΕΜΠ

ΠΕΡΙΕΧΟΜΕΝΑ

Περιεχόμενα

1. Γενικά Στοιχεία.....	3
2. Σπουδές.....	3
3. Ξένες Γλώσσες.....	3
4. Διδακτική Εμπειρία.....	3
5. Επαγγελματική Εμπειρία	5
6. Ερευνητικά Ενδιαφέροντα	8
7. Ερευνητικά Έργα	9
8. Επιστημονική Δραστηριότητα σε Περιοδικά και Συνέδρια	16
9. Επιστημονικές Εργασίες.....	19
10. Βραβεία.....	48
11. Αναφορές στον τύπο	48
12. Ετερο-αναφορές.....	48

ΒΙΟΓΡΑΦΙΚΟ ΣΗΜΕΙΩΜΑ – ΥΠΟΜΝΗΜΑ ΣΤΑΔΙΟΔΡΟΜΙΑΣ

1. Γενικά Στοιχεία

Επώνυμο: Κωνσταντίνου
Όνομα: Ιωάννης
Πατρώνυμο: Ηλίας
Τόπος Γεννήσεως: Λαμία Φθιώτιδος
Ημ/νία Γεννήσεως: 3 Οκτωβρίου 1980
Κινητό: +30-694-5992906
e-mail: ikons@uth.gr

2. Σπουδές

- Απόφοιτος Γενικού Λυκείου Σπερχειιάδας (1998) με βαθμό Άριστα (19.1/20).
- Απόφοιτος της Σχολής Ηλεκτρολόγων Μηχ. και Μηχ. Υπολογιστών Ε.Μ.Π. (2004) με βαθμό Λίαν Καλώς. Τίτλος Διπλωματικής Εργασίας: «Δημιουργία – διόρθωση και παραγωγή εκτυπώσιμου οδηγού σπουδών με συνδυασμό στατικής και δυναμικής πληροφορίας μέσω διαδικτύου» (Επ. Καθ. Ν. Μήτρου).
- Κάτοχος Μεταπτυχιακού Διπλώματος Ειδίκευσης, Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών Τεχνοοικονομικά Συστήματα, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Η/Υ, ΕΜΠ 2007. Τίτλος Διπλωματικής Εργασίας: «Data Grid: Ενοποίηση ανομοιογενών και γεωγραφικά απομακρυσμένων αποθηκευτικών χώρων μέσω διαδικτύου» (Επ. Καθ. Ν. Κοζύρης)
- Διδάκτορας Μηχανικός, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Ηλεκτρονικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου (Ιούνιος 2011). Τίτλος Διδακτορικής Διατριβής: «Προσαρμοστικοί Αλγόριθμοι Εξισορρόπησης Φόρτου σε Κατανεμημένα Περιβάλλοντα (Δίκτυα Ομοτίμων και Υπολογιστικά Νέφη)» (Επιβλέπων. Καθ. Π. Τσανάκας).

3. Ξένες Γλώσσες

- *Αγγλικά*: Certificate of Proficiency in English (Cambridge University, 2002)
- *Γαλλικά*: DELF - Diplôme d'études en langue française (1995)

4. Διδακτική Εμπειρία

Αναπληρωτής Καθηγητής, Τμήμα Πληροφορικής και Τηλεπικοινωνιών, Πανεπιστήμιο Θεσσαλίας.

- ΦΕΚ διορισμού στην Βαθμίδα του Αναπληρωτή Καθηγητή 2835/τ.Γ'/13.09.2024
- ΦΕΚ Μονιμοποίησης στη βαθμίδα του Επίκουρου Καθηγητή 3100/τ.Γ'/27.11.2023
- ΦΕΚ Διορισμού στη βαθμίδα του Επίκουρου Καθηγητή επί θητεία 924/τ.Γ'/18.06.2020

Διδασκαλία μαθημάτων:

- Λειτουργικά Συστήματα Χειμερινό εξάμηνο 2020, 2021, 2022, 2023
- Προγραμματισμός ΙΙ, Χειμερινό εξάμηνο 2020, 2021, 2022, 2023
- Νεφρολογιστική, Εαρινό εξάμηνο 2021, 2022, 2023, 2024
- Συστήματα Διαχείρισης Μεγάλου Όγκου Δεδομένων, Εαρινό εξάμηνο 2021, 2022, 2023, 2024
- ΠΜΣ: Πληροφορική και Τηλεπικοινωνίες, Προγραμματισμός σε C, Εαρινό 2022
- ΠΜΣ: Πληροφορική και Τηλεπικοινωνίες, Εισαγωγή στις Βάσεις Δεδομένων, Χειμερινό εξάμηνο 2022

Επίκουρος Καθηγητής, Τμήμα Πληροφορικής, Οικονομικό Πανεπιστήμιο Αθηνών, ΔΠΜΣ: ΕΠΙΣΤΗΜΗ ΔΕΔΟΜΕΝΩΝ

Διδασκαλία Μαθημάτων

- Introduction to Data Management and Engineering Χειμερινό εξάμηνο 2020, 2021, Εαρινό εξάμηνο 2021, Χειμερινό εξάμηνο 2022
- Large Scale Data Management, Εαρινό εξάμηνο 2022, 2023

Λέκτορας (ΠΔ 407/80) Τμήμα Πληροφορικής, Οικονομικό Πανεπιστήμιο Αθηνών,

Διδασκαλία μαθημάτων

- Λειτουργικά Συστήματα Εαρινό εξάμηνο 2017
- Κατανεμημένα Συστήματα (Εργαστήριο), Εαρινό εξάμηνο 2020, 2021, 2022

Πανεπιστημιακός Υπότροφος, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Εθνικό Μετσόβιο Πολυτεχνείο Διδασκαλία μαθημάτων

- Προχωρημένα Θέματα Βάσεων Δεδομένων Χειμερινό εξάμηνο 2016, 2017, 2018, 2019
- Μεταπτυχιακό Πρόγραμμα Σπουδών Επιστήμη Δεδομένων και Μηχανική Μάθηση – Μάθημα: Διαχείριση Δεδομένων Μεγάλης Κλίμακας Εαρινό 2019, 2020, 2021, 2022, 2023, 2024

Πανεπιστημιακός Υπότροφος, Τμήμα Πληροφορικής και Τηλεπικοινωνιών, Πανεπιστήμιο Θεσσαλίας Διδασκαλία μαθημάτων

- Νεφρολογιστική, Εαρινό εξάμηνο 2020, 2021
- Συστήματα Διαχείρισης Μεγάλου Όγκου Δεδομένων, Εαρινό εξάμηνο 2020, 2021

Λέκτορας (ΠΔ 407/80), Τμήμα Πληροφορικής και Τηλεπικοινωνιών, Πανεπιστήμιο Θεσσαλίας. Διδασκαλία μαθημάτων:

- Λειτουργικά Συστήματα Χειμερινό εξάμηνο 2016, 2017, 2018, 2019
- Εισαγωγή στον Προγραμματισμό Χειμερινό εξάμηνο 2016, 2017, 2018
- Εξόρυξη Δεδομένων Χειμερινό εξάμηνο 2016
- Αντικειμενοστραφής Προγραμματισμός (Εργαστήριο) Εαρινό εξάμηνο 2017, 2018
- Διαδικτυακός και Ταυτόχρονος Προγραμματισμός (Εργαστήριο) Εαρινό εξάμηνο 2017, 2018, 2019
- Προγραμματισμός ΙΙ, Χειμερινό 2019

Επίκουρος Καθηγητής (ΠΔ 407/80), Τμήμα Μηχανικών Η/Υ & Πληροφορικής, Πανεπιστήμιο Πατρών. Διδασκαλία μαθημάτων:

- Λειτουργικά Συστήματα ΙΙ (Cloud Computing, Big Data, MapReduce) Εαρινό εξάμηνο 2012
- Λειτουργικά Συστήματα Ι Χειμερινό εξάμηνο 2012
- Λειτουργικά Συστήματα ΙΙ (Cloud Computing, Big Data, MapReduce) Εαρινό εξάμηνο 2013

Επισκέπτης Μεταδιδακτορικός Ερευνητής, Τμήμα Μηχανικών Η/Υ & Πληροφορικής, Πανεπιστήμιο Πατρών. Διδασκαλία μαθημάτων :

- Λειτουργικά Συστήματα Ι Χειμερινό εξάμηνο 2013

Μεταδιδακτορικός Ερευνητής Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου. Διδασκαλία μαθημάτων :

- Προχωρημένα Θέματα Βάσεων Δεδομένων Χειμερινό εξάμηνο 2013
- Προχωρημένα Θέματα Βάσεων Δεδομένων Χειμερινό εξάμηνο 2014
- Προχωρημένα Θέματα Βάσεων Δεδομένων Χειμερινό εξάμηνο 2015

Βοηθός εργαστηρίου, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου:

- Λειτουργικά Συστήματα (08-09)
- Λειτουργικά Συστήματα (09-10)
- Λειτουργικά Συστήματα (10-11)

5. Επαγγελματική Εμπειρία

Τμήμα Πληροφορικής και Τηλεπικοινωνιών, Πανεπιστήμιο Θεσσαλίας (6/2020-σήμερα)
ΑΝΑΠΛΗΡΩΤΗΣ ΚΑΘΗΓΗΤΗΣ

- ΦΕΚ διορισμού στην Βαθμίδα του Αναπληρωτή Καθηγητή 2835/τ.Γ'/13.09.2024
 - ΦΕΚ Μονιμοποίησης στη βαθμίδα του Επίκουρου Καθηγητή 3100/τ.Γ'/27.11.2023
 - ΦΕΚ Διορισμού στη βαθμίδα του Επίκουρου Καθηγητή επί θητεία 924/τ.Γ'/18.06.2020
- Γνωστικό αντικείμενο: «Καταναεμημένα Υπολογιστικά Συστήματα με έμφαση στα Υπολογιστικά Νέφη»

Διευθυντής εργαστηρίου «Εργαστήριο Τεχνολογίας Βίντεο και Νεφοϋπολογιστικής»

Συμμετοχή στις παρακάτω επιτροπές του Πανεπιστημίου

- Μέλος Επιτροπής Ψηφιακής Διακυβέρνησης του Πανεπιστημίου Θεσσαλίας

Συμμετοχή στις παρακάτω επιτροπές του Τμήματος

- Επιτροπή επιλογής διδασκόντων σύμφωνα με το Π.Δ.407/80 & ακαδημαϊκών υποτρόφων σύμφωνα με το άρθρο 29 του ν.4009/2011
- Επιτροπή αξιολόγησης υποψηφιοτήτων για την πρόσληψη διδασκόντων στο πλαίσιο του έργου «απόκτηση ακαδημαϊκής διδακτικής εμπειρίας σε νέους επιστήμονες κάτοχους διδακτορικού 2019-2022» για το ακαδ. έτος 2021-2022
- Εκπρόσωποι τμήματος στην επιτροπή κατάρτισης ωρολογίων προγραμμάτων μαθήματος και προγραμμάτων εξετάσεων για τις αίθουσες διδασκαλίας των κτιρίων Α & Β της ΣΘΕ (3ο χλμ. Π.Ε.Ο. Αθηνών- Λαμίας) (αναπληρωματικό μέλος)

- Επιτροπή αξιολόγησης ενστάσεων για την πρόσληψη διδασκόντων σύμφωνα με το Π.Δ.407/80 & ακαδημαϊκών υποτρόφων σύμφωνα με το άρθρο 29 του ν.4009/2011
- Επιτροπή επικαιροποίησης του εσωτερικού κανονισμού λειτουργίας του προπτυχιακού προγράμματος σπουδών του τμήματος για το ακαδημαϊκό έτος 2021-2022
- Επιτροπή επιλογής ακαδημαϊκών υποτρόφων σύμφωνα με το άρθρο 29 του ν.4009/2011 για το χειμ. εξάμηνο του ακαδ. έτους 2022-2023
- Επιτροπή κατανομής χωρών του τμήματος στο νότιο συγκρότημα των κτιριακών εγκαταστάσεων της ΣΘΕ

Εκπροσώπηση τμήματος

- Συντονιστής εξωστρέφειας τμήματος

Εργαστήριο Υπολογιστικών Συστημάτων, ΣΗΜΜΥ, ΕΜΠ

(07/2011-σήμερα)

ΜΕΤΑΔΙΔΑΚΤΟΡΙΚΟΣ ΕΡΕΥΝΗΤΗΣ.

Ερευνητικά ενδιαφέροντα: Κατανεμημένα Υπολογιστικά Συστήματα, Συστήματα Διαχείρισης Μεγάλου Όγκου Δεδομένων, Αρχιτεκτονικές λογισμικού Υπολογιστικών Νεφών, Δίκτυα Ομοτίμων και Υπολογιστικά Πλέγματα.

- Συγγραφή ερευνητικών προτάσεων για χρηματοδότηση από ΕΕ (FP7-Horizon 2020) και ΓΓΕΤ. Συνολικός προϋπολογισμός επιτυχημένων ερευνητικών προτάσεων στις οποίες συμμετείχε ή διεύθυνε την υποβολή: **+24M Euro**. (IW-NET, ZEROW, DataSource, BigOptiBase, E2Data, SELIS, CELAR, MoDiSENSE)

Χρηματοδότηση για το Εργαστήριο που προέκυψε από τις παραπάνω προτάσεις: **1,4M Euro**

- Επίβλεψη διπλωματικών εργασιών προπτυχιακού και μεταπτυχιακού επιπέδου
- Συμμετοχή σε τριμελείς επιτροπές εξέτασης

- 14 διπλωματικές εργασίες προπτυχιακού ΣΗΜΜΥ
- 9 διπλωματικές εργασίες μεταπτυχιακού ΕΔΕΜΜ
- 3 υποψήφιους διδάκτορες ΗΜΜΥ

- Επιστημονικός σύμβουλος, θέματα πληροφορικής για φορείς του δημοσίου τομέα όπως : Συμβούλιο της Επικρατείας, Υπουργείο Οικονομίας και Ανάπτυξης, Γενική Διεύθυνση Ιδιωτικών Επενδύσεων, ΟΑΣΑ κλπ.

Ερευνητικά Έργα:

- **OASA Analytics**: Εξαγωγή Δεικτών Απόδοσης του Συγκοινωνιακού Έργου της ΟΑΣΑ Α.Ε. και της Στάθμης Εξυπηρέτησης Επιβατών στις Αστικές Συγκοινωνίες Αθηνών με Χρήση Προηγμένων Τεχνικών Αναλυτικής Επεξεργασίας Δεδομένων (Προγραμματική Σύμβαση ΑΔΑ: ΨΨΡΧ46ΨΧΕ3-Ε2Η)

- **ZEROW**: Systemic Innovations Towards a Zero Food Waste Supply Chain. (GA 101036388, IA, H2020-LC-GD-2020-4)

- **PUZZLE**: Towards a Sophisticated SIEM Marketplace for Blockchain-based Threat Intelligence and Security-as-a-Service (GA 883540, IA, H2020-SU-DS-2019) WP4 Lead

- **DataSource**: Πλατφόρμα Εύρεσης, Απεικόνισης και Επεξεργασίας Μεγάλων Ανοιχτών Δεδομένων. T2EΔK-01231. Τεχνικός Υπεύθυνος Έργου

- **IW-NET**: Innovation driven Collaborative European Inland Waterways Transport Network. Υπεύθυνος για πλατφόρμα Big Data (H2020 MG-2-6-2019)

- **eThemisID**: Ενσωμάτωση των τεχνολογιών eIDAS και e-signature στο Ελληνικό Δικαστικό Σύστημα CEF/ICT/A2019/1926069 (CEF-Connecting Europe Facility) Τεχνικός Υπεύθυνος Έργου.
- **BigOptiBase**: Βελτιστοποίηση Κατανάλωσης Ενέργειας Σταθμών Βάσης με Τεχνολογίες Διαχείρισης Μεγάλου Όγκου Δεδομένων T1EΔΚ-04605. Τεχνικός Υπεύθυνος Έργου
- **E2Data** European Extreme Performing Big Data Stacks (Horizon 2020, H2020-ICT-2017-1 # 780245)
- **SELIS** Towards a Shared European Logistics Intelligent Information Space (Horizon 2020, MG-6.3-2015 GA No 690588) Τεχνικός υπεύθυνος για την περιοχή των Big Data Systems (Πακέτο εργασίας 4 – Υποδομή Υπολογιστικών Νεφών και Συστημάτων Μεγάλου Όγκου Δεδομένων).
- **MoDiSENSE**: Κατανεμημένη πλατφόρμα ανάπτυξης υπηρεσιών κοινωνικής δικτύωσης για κινητές συσκευές. (Πρόγραμμα Συνεργασία 2009, κωδικός έργου: 09ΣΥΝ-72_881) Τεχνικός υπεύθυνος έργου.
- **CELAR**: Automatic, multi-grained elasticity-provisioning for the Cloud, 7th Framework Programme, FP7-ICT-2011-8 #317790. Συμμετοχή στην τεχνική οργάνωση.
- **CLARIN-EL**: Υποστήριξη και ανάπτυξη ελληνικών εταιρών για συμμετοχή στις Κοινοπραξίες των Ευρωπαϊκών Ερευνητικών Υποδομών στα πλαίσια του ESFRI/2006. Υποδομή γλωσσικών πόρων, τεχνολογιών και υπηρεσιών ΟΠΣ 441451. Συμμετοχή στην τεχνική οργάνωση.
- **PredMine**: Πλατφόρμα Προβλεπτικής Ανάλυσης με Χρήση Δεδομένων Μεγάλου Όγκου σε Πραγματικό Χρόνο. Κωδικός έργου: ΕΣΠΑ-Ελλάδα-Ισραήλ 3487. Συμμετοχή στην τεχνική οργάνωση.
- **ARCOMEM**: ARchive COMmunities MEMories. European Commission, 7th Framework Programme, FP7 270239. Ερευνητής.
- Διδασκαλία μαθημάτων: Προχωρημένα Θέματα Βάσεων Δεδομένων, 9^ο Εξάμηνο (2013, 2014, 2015, 2016, 2017, 2018).

Κοινωνία της Πληροφορίας Α.Ε.

Μέλος Διοικητικού Συμβουλίου

Μέλος γνωμοδοτικής επιτροπής δράσης ICT4GROWTH

(05/2015-σήμερα)

(7/2019-Σήμερα)

(5/2015 – 5/2019)

Κέντρο Υδρολογίας και Υδροπληροφορικής, Πολιτικοί Μηχανικοί, ΕΜΠ (04/2014-2016)

ΜΕΤΑΔΙΔΑΚΤΟΡΙΚΟΣ ΕΡΕΥΝΗΤΗΣ.

- Συμβουλευτικές υπηρεσίες για την Ευρωπαϊκή Υπηρεσία Περιβάλλοντος (EEA) μέσω του European Topic Centre on Inland, Coastal and Marine Waters (ETC/ICM)
- Διαχείριση δεδομένων του of WISE Water Quantity dataflow
- Υποστήριξη IT για το Water Accounting
- Σχεδιασμός και υλοποίηση λογισμικού για την διαχείριση υδρολογικών δεδομένων
- Συγγραφή ερευνητικών προτάσεων για χρηματοδότηση μέσω της Ευρωπαϊκής Ένωσης (Horizon 2020)

Εργαστήριο Υπολογιστικών Συστημάτων, ΣΗΜΜΥ, ΕΜΠ

(11/2005-7/2011)

ΕΡΕΥΝΗΤΗΣ.

- Συμμετοχή στην τεχνική οργάνωση και ανάπτυξη των παρακάτω ερευνητικών έργων:
- **STRATUSLAB**: Enhancing Grid Infrastructures with Virtualization and Cloud Technologies (European Commission CP-CSA, RI-261552)

- **GRECIA:** Grid Enabled Access to Rich media Content). European Commission, 6th Framework Programme, FP6 34363
- **GRIDNEWS:** Κατανεμημένη πλατφόρμα πλέγματος (GRID) για τη διάθεση και προηγμένη αναζήτηση περιεχομένου σε ειδησεογραφικό οπτικοακουστικό υλικό. (ΚτΠ Μέτρο 3.3 κωδικός έργου 10)
- PRESEMT:** Pattern Recognition Based Statistically enhanced MT (European Commission, 6th Framework 63/711)
- E-VACATIONS:** Ανάπτυξη Πληροφοριακού Συστήματος σύνθετων και πολλαπλών κρατήσεων σε τουριστικές επιχειρήσεις των Ιονίων Νήσων με χρήση Grid/Webservices
- Greed:** Εξαγωγή, εξόρυξη και διαχείριση μονογλωσσικών και πολυγλωσσικών δεδομένων, καθώς και πολυμεσικών μεταδεδομένων για ποικιλία εφαρμογών
- Υπεύθυνος για τον σχεδιασμό, την υλοποίηση και την εγκατάσταση πρωτοτύπων κατανεμημένων συστημάτων διαχείρισης δεδομένων μεγάλης κλίμακας που βασίζονται σε τεχνικές δικτύων ομοτίμων, υπολογιστικών πλεγμάτων και υπολογιστικών νεφών.
- Συμμετοχή στην προετοιμασία και υποβολή ερευνητικών προτάσεων (proposals) για χρηματοδότηση από ΕΕ και ΓΓΕΤ.
- Καθοδήγηση σε προπτυχιακούς φοιτητές για την διεκπεραίωση των διπλωματικών εργασιών τους.
- Βοηθός Εργαστηρίου για το χειμερινό μάθημα Λειτουργικά Συστήματα (2008, 2009 και 2010)
 - Διαχειριστής του δικτυακού τόπου artemis.cslab.ntua.gr που δεικτοδοτεί και διαμοιράζει όλες τις διπλωματικές εργασίες της Σχολής ΗΜΜΥ του ΕΜΠ από το 2002 μέχρι σήμερα.

EXODUS SA, Αθήνα

(7/2002–12/2006)

ΜΗΧΑΝΙΚΟΣ ΛΟΓΙΣΜΙΚΟΥ

- Ανάπτυξη δικτυακών τόπων με το σύστημα διαχείρισης περιεχομένου eCentric (HTML, DHTML, JavaScript, ASP,ASP.NET SQL Server 2000)
- Υπεύθυνος συντήρησης και αποσφαλμάτωσης δικτυακών εφαρμογών του Ομίλου της Τράπεζας Πειραιώς
- Ανάπτυξη εφαρμογών Internet Banking (www.easypay.gr)
- Ανάπτυξη εφαρμογών με Visual Basic.
- Ανάπτυξη εφαρμογών Workflow Management με τεχνολογία JSP, JavaBeans (Jboss)
- Συμμετοχή στην ανάπτυξη εφαρμογών για το τμήμα R&D.

Wateron Multimedia, Αθήνα

(2000-2002)

ΜΗΧΑΝΙΚΟΣ ΛΟΓΙΣΜΙΚΟΥ

- Ανάπτυξη δικτυακού τόπου (www.thalassa.gr) για το Πρόγραμμα «Καθαρές Ακτές» του Υ.ΠΕ.ΧΩ.ΔΕ. (HTML, DHTML, JavaScript, ASP, SQL)
- Ανάπτυξη δικτυακού τόπου (www.thisavros.gr) για το Πρόγραμμα «Λίμνες και Ποτάμια» του Υ.ΠΕ.ΧΩ.ΔΕ. (HTML, DHTML, JavaScript, ASP, SQL)

6. Ερευνητικά Ενδιαφέροντα

Η έρευνα του εστιάζεται στην περιοχή των κατανεμημένων συστημάτων μεγάλης κλίμακας δεδομένων με έμφαση στα υπολογιστικά νέφη. Σε γενικές γραμμές ασχολείται με την σχεδίαση αποδοτικών κατανεμημένων συστημάτων δεδομένων, μεθόδων, αλγορίθμων και αρχιτεκτονικών οι οποίες μπορούν να κλιμακώνονται με ελαστικό και αυτόματο τρόπο προσαρμόζοντας το

μέγεθος των πόρων και τις ρυθμίσεις του λογισμικού. Οι μέθοδοι που σχεδιάζει καλύπτουν μια ευρεία περιοχή κατανεμημένων συστημάτων, όπως:

- Αναλυτική επεξεργασία δεδομένων μεγάλης κλίμακας με τεχνικές παράλληλης διαχείρισης δεδομένων και χρήση κατανεμημένων συστημάτων μεγάλης κλίμακας Big Data (NoSQL, MapReduce, κλπ).
- Ελαστικότητα πόρων με χρήση υπολογιστικών νεφών (Cloud Computing) μέσω της ρύθμισης των πόρων υλικού και λογισμικού με αυτόματο και προσαρμοστικό τρόπο χρησιμοποιώντας τεχνικές τεχνητής νοημοσύνης (ενισχυτική μάθηση, βαθιά ενισχυτική μάθηση, δέντρα αποφάσεων, κλπ.) σε ένα ευρύ φάσμα κατανεμημένων συστημάτων αναλυτικής επεξεργασίας δεδομένων και μηχανικής μάθησης (ML for systems).
- Βελτιστοποίηση κατανεμημένης εκπαίδευσης συστημάτων λογισμικού τεχνητής νοημοσύνης (Systems for ML).
- Σχεδιασμός, υλοποίηση και αποδοτική εκτέλεση κατανεμημένων συστημάτων λογισμικού με χρήση περιεκτών (containers) και μεθόδων άνευ διακομιστή (serverless) σε υπολογιστικά νέφη
- Εξισορρόπηση φόρτου εργασίας με εφαρμογές σε δίκτυα ομοτίμων (Peer-to-Peer Networks).

7. Ερευνητικά Έργα

Έχει εργαστεί ως ερευνητής σε αρκετά ερευνητικά έργα. Παρακάτω παραθέτουμε ορισμένα από αυτά:

- **OASA Analytics:** Εξαγωγή Δεικτών Απόδοσης του Συγκοινωνιακού Έργου της ΟΑΣΑ Α.Ε. και της Στάθμης Εξυπηρέτησης Επιβατών στις Αστικές Συγκοινωνίες Αθηνών με Χρήση Προηγμένων Τεχνικών Αναλυτικής Επεξεργασίας Δεδομένων (Προγραμματική Σύμβαση ΑΔΑ: ΨΨΡΧ46ΨΧΕ3-Ε2Η)
- **ZEROW:** Systemic Innovations Towards a Zero Food Waste Supply Chain. (GA 101036388, IA, H2020-LC-GD-2020-4)
Η παγκόσμια πείνα αυξάνεται, ωστόσο το ένα τρίτο όλων των τροφίμων που παράγονται παγκοσμίως πηγαίνει στα σκουπίδια. Το χρηματοδοτούμενο από την ΕΕ έργο ZeroW στοχεύει να προσφέρει σημαντικές συνεισφορές στην πρόληψη και τη μείωση της σπατάλης τροφίμων και στην ενίσχυση της βιωσιμότητας των αλυσίδων εφοδιασμού τροφίμων. Συγκεκριμένα, το ZeroW θα πραγματοποιήσει την επίδειξη διαφορετικών καινοτομιών σε εννέα αλυσίδες φαγητού της πραγματικής ζωής, εφαρμόζοντας μια συστημική προσέγγιση καινοτομίας για την αποτελεσματική αντιμετώπιση του πολυδιάστατου ζητήματος της σπατάλης τροφίμων. Επιπλέον, το έργο θα δικαιολογήσει τον τρόπο με τον οποίο η κλιμάκωση των αποτελεσμάτων των καινοτομιών που επιδεικνύονται και ο συνδυασμός τους με κατάλληλες πολιτικές, μπορεί να μας φέρει πιο κοντά στους στόχους της Πράσινης Συμφωνίας για τη μείωση των απορριμμάτων τροφίμων για το 2030.
- **PUZZLE:** Towards a Sophisticated SIEM Marketplace for Blockchain-based Threat Intelligence and Security-as-a-Service (GA 883540, IA, H2020-SU-DS-2019) WP4 Lead
Για τις μικρές και μεσαίες επιχειρήσεις (ΜΜΕ) και τις πολύ μικρές επιχειρήσεις, η ψηφιακή μετάβαση είναι σημαντική για την ανάπτυξη των επιχειρήσεων. Ωστόσο, παρουσιάζει επίσης τον κίνδυνο επιθέσεων στον κυβερνοχώρο. Το χρηματοδοτούμενο από την ΕΕ έργο PUZZLE

αναπτύσσει εργαλεία κυβερνοασφάλειας που επιτρέπουν ακόμη και στις μικρότερες επιχειρήσεις να παρακολουθούν, να προβλέπουν, να αξιολογούν και να διαχειρίζονται τους κινδύνους στον κυβερνοχώρο. Συγκεκριμένα, το έργο θα παρακολουθεί τις σχέσεις μεταξύ των περιουσιακών στοιχείων στον κυβερνοχώρο κάθε μικρής και πολύ μικρής επιχείρησης. Θα εξετάσει επίσης το διαθέσιμο δίκτυο και τις υποδομές υπολογισμού και αποθήκευσης και θα τις χρησιμοποιήσει για τον αποτελεσματικό υπολογισμό μεμονωμένων, σωρευτικών και διαδεδομένων κινδύνων, καθώς και για να προτείνει και να εφαρμόσει μέτρα μετριασμού. Στόχος του έργου είναι να σχεδιάσει λύσεις που μπορούν εύκολα να ενσωματωθούν από εξωτερικούς παρόχους κυβερνοασφάλειας και να υιοθετηθούν απρόσκοπτα από τους τελικούς χρήστες.

- **DataSource:** Πλατφόρμα Εύρεσης, Απεικόνισης και Επεξεργασίας Μεγάλων Ανοιχτών Δεδομένων. T2EΔK-01231. Τεχνικός Υπεύθυνος Έργου

Ο στόχος του έργου DataSource είναι η δημιουργία μιας τεχνολογικής πλατφόρμας για την ανάπτυξη ενός οικοσυστήματος που θα παράγει υπεραξία από τη συνάθροιση και σύνθεση ανοιχτών δεδομένων που παρέχουν ανεξάρτητοι «μέτοχοι» της επιστημονικής αλυσίδας αξίας με σκοπό την διευκόλυνση της διεξαγωγής έρευνας που βασίζεται σε δεδομένα. Στα πλαίσια του έργου θα σχεδιαστεί και θα υλοποιηθεί ένα ολοκληρωμένο σύστημα βασισμένο σε τεχνολογίες αιχμής το οποίο θα μπορεί να συγκεντρώνει, οπτικοποιεί και επεξεργάζεται συνδυασμένα δεδομένα που προέρχονται από ανοιχτά σετ δεδομένων διαθέσιμα στο διαδίκτυο με την χρήση καταναμημένων συστημάτων επεξεργασίας μεγάλου όγκου δεδομένων πάνω από πλατφόρμες υπολογιστικών νεφών. Οι χρήστες θα έχουν την δυνατότητα να χρησιμοποιούν την πλατφόρμα για να εκτελούν ερωτήματα σε όλες τις διαθέσιμες διασυνδεδεμένες βάσεις δεδομένων με το DataSource χρησιμοποιώντας ένα εύχρηστο γραφικό περιβάλλον, θα μπορούν να έχουν μια σύντομη προ-επισκόπηση των διαθέσιμων αποτελεσμάτων, να εκτελούν κάποιους βασικούς μετασχηματισμούς στα συνδυασμένα δεδομένα, και να μπορούν να ανακτούν το τελικό προϊόν της επεξεργασίας μέσω των διαθέσιμων συνδέσμων. Θα αναπτυχθούν τα απαραίτητα υποσυστήματα οπτικοποίησης της πληροφορίας, συγκέντρωσης και αλγοριθμικής επεξεργασίας αυτής με την χρήση κλιμακώσιμων τεχνικών διαχείρισης δεδομένων. Το τελικό σύστημα θα μπορέσει να βοηθήσει χρήστες από διαφορετικούς τομείς όπως ερευνητές, υπάλληλους δημόσιων φορέων, κλπ. να διεξάγουν εύκολα τις έρευνές τους, ενώ ο φορέας που θα το αξιοποιήσει εμπορικά θα μπορεί μέσω της χρήσης υποδομών υπολογιστικών νεφών να δημιουργήσει ένα ευέλικτο οικονομικό μοντέλο χρέωσης ανάλογα με την χρήση της υπηρεσίας.

- **IW-NET:** Innovation driven Collaborative European Inland Waterways Transport Network. (H2020 MG-2-6-2019) Υπεύθυνος για πλατφόρμα Big Data

Το IW-NET θα παράσχει μια πολυτροπική διαδικασία βελτιστοποίησης στο σύνολο του Συστήματος Μεταφορών της ΕΕ, αυξάνοντας το μερίδιο των μέσων μεταφοράς της εσωτερικής ναυσιπλοΐας και υποστηρίζοντας τις φιλοδοξίες της ΕΚ να μειώσει τις εκπομπές αερίων θερμοκηπίου κατά δύο τρίτα έως το 2050. Βελτιώσεις στη βιώσιμη διαχείριση υποδομών και καινοτόμων σκαφών θα υποστηρίξουν τον ανταγωνιστικό τομέα IWT που αντιμετωπίζει σημεία συμφόρησης των υποδομών, ανεπαρκής ολοκλήρωση των τεχνολογιών πληροφορικής κατά μήκος της αλυσίδας και βραδεία υιοθέτηση τεχνολογιών, όπως νέοι τύποι πλοίων, εναλλακτικά καύσιμα, αυτοματισμοί, διαδίκτυο και μηχανική μάθηση. Το σενάριο χρήσης θα εφαρμόσει σενάρια εφαρμογών που βασίζονται σε χρήστες σε σημαντικούς διαδρόμους TEN-T, επιδεικνύοντας και αξιολογώντας τις επιπτώσεις σε προσομοιώσεις και δοκιμές που καλύπτουν τεχνολογικά, οργανωτικά, νομικά, οικονομικά,

οικολογικά και θέματα ασφάλειας / ασφάλειας: 1) Ψηφιοποίηση: πυκνές αστικές περιοχές με πρόβλεψη της δρομολόγησης της ζήτησης (Βρυξέλλες-Αμβέρσα-Courtrai-Lille-Valenciennes). βελτιστοποίηση πλοήγησης με χρήση δεδομένων σε αβέβαιες συνθήκες ύδατος (Δούναβης). 2) Βιώσιμη υποδομή και ευφυής διαχείριση της κυκλοφορίας: πρόβλεψη κλειδώματος μειώνοντας την αβεβαιότητα στον προγραμματισμό του ταξιδιού. σχεδιασμός κλειδαριών διαχείριση των τμημάτων των πλωτών οδών όπου απαγορεύονται οι συναντήσεις · προγραμματισμός αγκυροβόλησης με υποχρεωτική παροχή ηλεκτρικής ενέργειας στην ξηρά και άλλες υπηρεσίες (ενδοχώρα του Bremerhaven μέσω του καναλιού Weser / Mittelland). 3) Καινοτόμα σκάφη: νέα σχέδια φορτηγίδων που προσαρμόζουν τις συνθήκες του διαδρόμου και αγορές-στόχους: φορτηγίδες με υψηλό βαθμό αυτοματισμού για την αστική διανομή (Ανατολική Φλάνδρα-Γάνδη) · νέα φορτηγίδα για βυθοφόρα πλοία ικανά με χαμηλή / υψηλή στάθμη ύδατος για βελτιστοποίηση των δυνατοτήτων (Δούναβης από την Αυστρία στη Ρουμανία) · χρήση υπηρεσιών GALILEO για προηγμένη βοήθεια οδηγού, όπως καθοδήγηση, προειδοποίηση ύψους γέφυρας και είσοδος αυτόματης κλειδαριάς (πλωτή οδός Spree-Oder κοντά στο Βερολίνο). Οι συνοδευτικές δραστηριότητες είναι η εμπλοκή των ενδιαφερομένων μερών, η ανάπτυξη ικανοτήτων και η παράδοση ενός ευρωπαϊκού χάρτη πορείας για την ανάπτυξη της εσωτερικής ναυσιπλοΐας με συστάσεις πολιτικής για την αύξηση του μεριδίου της εσωτερικής ναυσιπλοΐας.

– **eThemisID: Ενσωμάτωση των τεχνολογιών eIDAS και e-signature στο Ελληνικό Δικαστικό Σύστημα /CEF/ICT/A2019/1926069 (CEF-Connecting Europe Facility)**

Η δράση στοχεύει στην ενσωμάτωση των ψηφιακών υπηρεσιών υποδομής (DSI) eID DSI και eSignature με τα υπάρχοντα συστήματα του ελληνικού συστήματος δικαιοσύνης προκειμένου να καταστεί δυνατή η διασυνοριακή πιστοποίηση ταυτότητας eIDAS και υπηρεσιών ηλεκτρονικής υπογραφής στους τελικούς χρήστες σε επίπεδο ΕΕ.

Ειδικότερα, η ελληνική πύλη του Συμβουλίου της Επικρατείας (ως Παροχέας Υπηρεσιών) θα συνδεθεί με την Ελληνική υποδομή eIDAS-Node, επιτρέποντας στους πολίτες της ΕΕ και τα νομικά πρόσωπα που έχουν διοικητικές / νομικές διαφορές με τον Ελληνικό Δημόσιο Τομέα να χρησιμοποιούν ηλεκτρονικές υπηρεσίες για την επικοινωνία με τις σχετικές ελληνικές αρχές. Η διασυνοριακή πιστοποίηση βάσει του eIDAS θα καλύπτει τις ακόλουθες ηλεκτρονικές υπηρεσίες του Ελληνικού Συμβουλίου Επικρατείας:

- Παρακολούθηση ηλεκτρονικών δικογράφων για τους πολίτες της ΕΕ:
- Ηλεκτρονική υποβολή δικογράφων για τους δικηγόρους της ΕΕ.
- Ηλεκτρονική παρακολούθηση και έλεγχο/έκδοση πιστοποιητικών για νομικές οντότητες.

Η υλοποίηση της διασύνδεσης της πύλης με το eIDAS θα προσφερθεί ως πακέτο κώδικα ανοιχτού λογισμικού (open-source) ώστε να μπορεί να χρησιμοποιηθεί από τα υπόλοιπα ελληνικά δικαστήρια για εύκολη ενσωμάτωση.

Επιπλέον, η Δράση θα επιτρέψει τις δυνατότητες eSeal (δυνατότητα ψηφιακής υπογραφής και επικύρωσης) για όλα τα ηλεκτρονικά έγγραφα που παράγονται μέσω αυτοματοποιημένων μηχανισμών παράδοσης στον ιστοτόπο του Συμβουλίου της Επικρατείας. Θα χρησιμοποιήσει ξανά τη μονάδα eID-DSS (eSignature DSI) για να προσθέσει τις δυνατότητες eSeal στην Ελληνική Πύλη Υπηρεσιών του Συμβουλίου της Επικρατείας. Με αυτόν τον τρόπο όλα τα έγγραφα που παράγονται ως απάντηση στο αίτημα του πολίτη / νομικής οντότητας θα υπογραφούν ψηφιακά πριν την προωθήσουν ως απάντηση στο αρχικό αίτημα. Οι δυνατότητες eSeal θα ενεργοποιηθούν για τα 41 διοικητικά δικαστήρια του Ελληνικού Συμβουλίου της Επικρατείας.

Το πακέτο εργαλείων ανοιχτού κώδικα θα προετοιμαστεί επίσης για άλλους οργανισμούς (όπως άλλα δικαστήρια) που επιθυμούν να χρησιμοποιήσουν την υπηρεσία αποστολής που είναι συμβατή με το eIDAS.

Μέχρι το τέλος της δράσης, η ελληνική πύλη του Συμβουλίου της Επικρατείας θα διατεθεί στο περιβάλλον παραγωγής για διασυνοριακή πιστοποίηση (μέσω eIDAS) και για eSeal δυνατότητες (με βάση το εργαλείο DSI DSS eSignature).

– **BigOptiBase: Βελτιστοποίηση Κατανάλωσης Ενέργειας Σταθμών Βάσης με Τεχνολογίες Διαχείρισης Μεγάλου Όγκου Δεδομένων T1EΔΚ-04605 (Ερευνώ - Δημιουργώ - Καινοτομώ)**

Ο στόχος του BigOptiBase είναι η υλοποίηση έξυπνης διαχείρισης της υποδομής για τη βελτιστοποίηση της ενεργειακής κατανάλωσης εξοπλισμού σταθμών βάσεων κινητής τηλεφωνίας με διατήρηση παράλληλα της ποιότητας υπηρεσιών που απολαμβάνουν οι χρήστες των δικτύων κινητών επικοινωνιών.

Στα πλαίσια του έργου θα σχεδιαστεί και θα υλοποιηθεί ένα ολοκληρωμένο σύστημα που θα συγκεντρώνει, αναλύει, επεξεργάζεται και οπτικοποιεί συνδυασμένα δεδομένα που προέρχονται από τηλεπικοινωνιακά στοιχεία χρήσης σταθμών βάσης, από δεδομένα τηλεμετρίας σταθμών βάσης καθώς και από διαθέσιμα ανοιχτά δεδομένα.

Θα αναπτυχθούν καινοτόμοι αλγόριθμοι και τεχνικές συλλογής, αποθήκευσης και διαχείρισης δεδομένων βασισμένες σε ανοιχτές τεχνολογίες λογισμικού και υλικού οι οποίες θα έχουν τη δυνατότητα να αναλύουν μεγάλο όγκο δεδομένων που παράγεται με μεγάλη ταχύτητα από το σύνολο των σταθμών βάσης και να το συνδυάζουν με δεδομένα από εξωτερικές πηγές. Παράλληλα, θα εφαρμοστούν τεχνικές καταναεμμένης μηχανικής μάθησης σε πραγματικό χρόνο για την εξαγωγή μοντέλων πρόγνωσης με σκοπό τη δυναμική, αυτόματη και βελτιστοποιημένη διαχείριση δικτυακών (πχ. κεραιές) και άλλων (πχ. κλιματιστικά) στοιχείων της δικτυακής υποδομής του ραδιοδικτύου. Επίσης, θα σχεδιαστεί και προγραμματιστεί απαραίτητο υλικό που θα εγκατασταθεί στους σταθμούς βάσης για την επικοινωνία, την αρχική συλλογή/επεξεργασία των δεδομένων και την απομακρυσμένη διαχείριση του εξοπλισμού. Η πειραματική αποτίμηση θα γίνει μέσα από τη δημιουργία ενός εκτεταμένου δικτυακού προσομοιωτή που θα αναπτυχθεί με στόχο να δίνει έμφαση σε θέματα ποιότητας υπηρεσιών χρησιμοποιώντας ιστορικά δεδομένα χρήσης από έναν ελληνικό πάροχο κινητής τηλεφωνίας. Τέλος, το σύστημα θα εγκατασταθεί για πιλοτική λειτουργία σε υπάρχοντες σταθμούς βάσης.

– **E2Data European Extreme Performing Big Data Stacks (Horizon 2020, H2020-ICT-2017-1 # 780245)**

Το E2Data θα παράσχει ένα νέο παράδειγμα λογισμικού Big Data για την επίτευξη της μέγιστης χρήσης πόρων για ετερογενή ανάπτυξη εφαρμογών υπολογιστικού νέφους που περιλαμβάνουν ποικιλία αρχιτεκτονικών όπως CPUs, GPUs, FPGAs και MICs, χωρίς να επηρεάζονται τα τρέχοντα πρότυπα προγραμματισμού. Η προτεινόμενη λύση υιοθετεί μια προσέγγιση σταυρωτής στρώσης επιτρέποντας την κάθετη επικοινωνία μεταξύ των τεσσάρων βασικών στρωμάτων των εφαρμογών Big Data (εφαρμογή, λογισμικό Big Data, scheduler / cloud provider και execution run-time) που θα επιτρέψουν σε μια E2Data-enabled εφαρμογή να απαντήσει την ακόλουθη ερώτηση:

"Πώς μπορεί ο χρήστης να δημιουργήσει για κάθε συγκεκριμένο επιχειρηματικό σενάριο την πιο αποδοτική και φθηνότερη διαμόρφωση υλικού;"

Το E2Data προτείνει μια λύση από άκρο σε άκρο για εφαρμογές Big Data, οι οποίες θα εκμεταλλευτούν πλήρως και θα προωθήσουν την προηγμένη τεχνολογία στις υποδομές, επιτυγχάνοντας αύξηση της απόδοσης έως και 10x, χρησιμοποιώντας έως και 50% λιγότερο υποδομές υπολογιστικών νεφών.

– **SELIS: Towards a Shared European Logistics Intelligent Information Space. (Horizon 2020, MG-6.3-2015 GA No 690588)**

Το έργο SELIS αποσκοπεί στη δημιουργία μιας «πλατφόρμας για πανευρωπαϊκές εφαρμογές logistics» μέσω των παρακάτω:

- Αγκαλιάζοντας ένα ευρύ φάσμα προοπτικών logistics και δημιουργώντας μια ενοποιημένη επιχειρησιακή και στρατηγική ατζέντα επιχειρηματικής καινοτομίας για πανευρωπαϊκά Green Logistics.

- Καθιέρωση μιας εξαιρετικά ισχυρής κοινοπραξίας των φορέων logistics και των παρόχων ΤΠΕ, που μπορούν να αξιοποιήσουν IP από πάνω από 40 έργα, για την δημιουργία μιας proof of concept κοινής πλατφόρμας επικοινωνίας και πλοήγησης για πανευρωπαϊκές εφαρμογές logistics.

- Δημιουργία ενός περιβάλλοντος της έρευνας και της καινοτομίας που θα δημιουργήσει δεδομένα τα οποία μπορούν να χρησιμοποιηθούν για την ανακάλυψη των νέων ιδεών που θα επιτρέψουν τη συνεχή δημιουργία αξίας υποστηρίζοντας την μεγάλης κλίμακας υιοθέτηση του Selis.

Το προτεινόμενο κοινό ευρωπαϊκό δίκτυο Logistics SELIS, είναι ένα δίκτυο ειδικών τοπικών κοινοτήτων logistics οι οποίες ονομάζονται κοινοτικοί κόμβοι SELIS (SCN). Οι SCNs είναι κατασκευασμένοι από μεμονωμένες κοινότητες logistics για να διευκολύνουν την επόμενη γενιά των συνεργατικών ευκίνητων και πράσινων αλυσίδων μεταφορών. Οι SCNs συνδέονται με τα υπάρχοντα συστήματα μέσω μιας ασφαλούς υποδομής και παρέχουν από κοινού και εργαλεία για την απόκτηση και τη χρήση των δεδομένων, σύμφωνα με μια «συμφωνία συνεργασίας». Οι διασυνδεδεμένοι κόμβοι παρέχουν μια κατανομημένη κοινή πλατφόρμα επικοινωνίας και πλοήγησης για Πανευρωπαϊκές εφαρμογές logistics. Κάθε κόμβος αποφασίζει ποιες πληροφορίες επιθυμεί να δημοσιεύσει και σε ποιες πληροφορίες θέλει να εγγραφεί.

Η βασική αρχή του SELIS είναι ότι παρέχει μια «ελαφριά δομή ΤΠΕ» για την ενεργοποίηση του διαμοιρασμού πληροφοριών για συνεργατικές βιώσιμες δομές logistics σε στρατηγικό και επιχειρησιακό επίπεδο.

– **MoDiSENSE: Κατανομημένη πλατφόρμα ανάπτυξης υπηρεσιών κοινωνικής δικτύωσης για κινητές συσκευές. (Πρόγραμμα Συνεργασία 2009, κωδικός έργου: 09ΣΥΝ-72_881)**

Το MoDisSENSE εμπλουτίζει τις υπηρεσίες κοινωνικής δικτύωσης με την αξιοποίηση των δεδομένων που προκύπτουν από την διαρκή, καθημερινή χρήση ενός τηλεφώνου, όπως τοποθεσιών που επισκέπτεται ο χρήστης, επαφών, κλήσεων, ημερολογίου (calendar) αλλά και των δεδομένων από την κοινωνική δικτύωση του χρήστη (φίλοι, προτιμήσεις, προφίλ χρηστών). Το έργο συνδυάζει τις ετερογενείς αυτές πηγές δεδομένων, γεωγραφικά, κοινωνικά logs, προφίλ και προτιμήσεις χρηστών (user preferences και context information) και προσφέρει υπηρεσίες που βασίζονται σε σύνθετες αναζητήσεις και συνδυαστικές ερωτήσεις πάνω σε όλες αυτές τις πηγές. Επίσης ασχολείται με την παραγωγή και διάθεση υπηρεσιών που αξιοποιούν τα χωροχρονικά δεδομένα που προκύπτουν από τις διαδρομές των χρηστών. Το MoDisSENSE θα αξιοποιήσει “state of the art” τεχνικές για το συνδυασμό και τη διαχείριση ετερογενών

δεδομένων που συλλέγονται από τις κινητές τερματικές συσκευές (ρεύματα δεδομένων συντεταγμένων, επαφές, προφίλ χρηστών, logs κλήσεων, προτιμήσεις χρηστών) και δεδομένων από την κοινωνική δικτύωση των χρηστών με αποτελεσματικό τρόπο αξιοποιώντας τεχνικές κατανεμημένης αποθήκευσης, επεξεργασίας, αναζήτησης και χωροχρονικής, πολυδιάστατης δεικτοδότησης. Η απόδοση της πλατφόρμας κρίνεται σημαντική για την ανάπτυξη των «κοινωνικά» επαυξημένων εφαρμογών που θα διατίθενται σε χρήστες μέσω έξυπνων τερματικών συσκευών. Το πρόβλημα που έχει να διαχειριστεί το MoDisSENSE είναι να παράγει με οικονομικά αποτελεσματικό και προσιτό τρόπο, ένα σύστημα ικανό να διαχειριστεί τα φορτία, τα ρεύματα και τα logs των πληροφοριών και τις σύνθετες ερωτήσεις και αναζητήσεις από τις «κοινωνικά» επαυξημένες εφαρμογές σε αποδεκτούς χρόνους, μέσα από ένα φιλικό περιβάλλον χρήσης.

– **CELAR: Automatic, multi-grained elasticity-provisioning for the Cloud, 7th Framework Programme, FP7-ICT-2011-8 #317790**

Η αυτό-ρύθμιση πόρων είναι ένα από τα κορυφαία εμπόδια και ευκαιρίες για το cloud computing: οι καταναλωτές μπορούν να ελαχιστοποιήσουν το χρόνο εκτέλεσης των εργασιών τους, χωρίς να υπερβαίνουν ένα συγκεκριμένο προϋπολογισμό. Οι πάροχοι Cloud μπορούν να μεγιστοποιήσουν το κέρδος τους, διατηρώντας παράλληλα τους πελάτες τους ικανοποιημένους ελαχιστοποιώντας το διοικητικό τους κόστος. Πολλά συστήματα ισχυρίζονται ότι προσφέρουν προσαρμοστική ελαστικότητα, αλλά η ρύθμιση γίνεται συνήθως χειροκίνητα, απαιτώντας από το χρήστη να υπολογίσει τις κατάλληλες συνθήκες κλιμάκωσης. Προκειμένου να αξιοποιήσουμε τα οφέλη της ελαστικής παροχής υπηρεσιών, είναι επιτακτικό να την εκτελούμε με ένα αυτοματοποιημένο, πλήρως προσαρμόσιμο τρόπο. Το CELAR προσφέρει ένα πλήρως αυτοματοποιημένο και ιδιαίτερα προσαρμόσιμο σύστημα για ελαστική παροχή πόρων σε πλατφόρμες cloud computing.

– **CLARIN-EL: Υποδομή γλωσσικών πόρων, τεχνολογιών και υπηρεσιών**

Το CLARIN-EL ασχολείται με την δημιουργία υποδομών για την αποθήκευση, ευρετηρίαση και διαμοιρασμό αποθετηρίων γλωσσικού περιεχομένου. Στα πλαίσια του έργου γίνεται ο σχεδιασμός μιας κατανεμημένης υποδομής πάνω από ένα περιβάλλον υπολογιστικού νέφους. Η υποδομή προσφέρει κατανεμημένη δεικτοδότηση, διαμοιρασμό, αποθήκευση και εκτέλεση αλγορίθμων με την χρήση καινοτόμων τεχνολογιών Big-Data.

– **PREDMINE: Πλατφόρμα Προβλεπτικής Ανάλυσης με Χρήση Δεδομένων Μεγάλου Όγκου σε Πραγματικό Χρόνο.**

Το PREDMINE ασχολείται με την εκτέλεση αλγορίθμων μηχανικής μάθησης (predictive analytics, machine learning, decision trees, κλπ.) πάνω από δεδομένα μεγάλου όγκου. Τα δεδομένα εμπλουτίζονται με επιπλέον πληροφορία που συλλέγεται από εξωτερικές πηγές (π.χ., Διαδίκτυο). Ο στόχος είναι η εξαγωγή πιο αναλυτικών μοντέλων πρόβλεψης και κατηγοριοποίησης με μεγαλύτερη ακρίβεια. Για τον σκοπό αυτό, κατανεμημένες τεχνικές αποθήκευσης, ευρετηρίασης και επεξεργασίας δεδομένων θα μελετηθούν. Καινοτόμες πλατφόρμες επεξεργασίας δεδομένων σε πραγματικό χρόνο θα αξιολογηθούν και θα ενσωματωθούν στην πλατφόρμα που θα αναπτυχθεί.

– **ARCOMEM: ARchive COmmunities MEMories. European Commission, 7th Framework Programme, FP7 270239**

Το ARCOMEM ασχολείται με οργανισμούς αρχειοθέτησης, μουσεία και βιβλιοθήκες στην εποχή του Κοινωνικού Ιστού. Σήμερα, τα ιδρύματα διατήρησης αρχείων είναι πιο σημαντικά από ποτέ: καθώς αντιμετωπίζουμε όλο και μεγαλύτερες οικονομικές και περιβαλλοντικές προκλήσεις χρειαζόμαστε την κατανόηση για το παρελθόν για να μας βοηθήσει να περιηγηθούμε σε ένα βιώσιμο μέλλον. Αυτή είναι μια βασική λειτουργία των δημοκρατιών, αλλά αυτή η λειτουργία αντιμετωπίζει σκληρές νέες προκλήσεις στο πρόσωπό του κοινωνικού ιστού, και των ριζικών αλλαγών στη δημιουργία πληροφοριών, την επικοινωνία και τη συμμετοχή των πολιτών που χαρακτηρίζει την σημερινή κοινωνία της πληροφορίας (π.χ., υπάρχουν πλέον περισσότερες επισκέψεις σε κοινωνικά δίκτυα από αναζητήσεις στο Google). Τα μέσα κοινωνικής δικτύωσης γίνονται όλο και πιο διεισδυτικά σε όλους τους τομείς της ζωής. Στο Ηνωμένο Βασίλειο, για παράδειγμα, είναι τώρα συνηθισμένο για έναν υπουργό της κυβέρνησης να απαντάει σε κοινοβουλευτικές ερωτήσεις μέσω του Twitter, και το υλικό αυτό είναι τόσο εφήμερο όσο και ιδιαίτερα παισιοστρεφές, κάτι που καθιστά όλο και πιο δύσκολο για έναν πολιτικό αρχειοθέτη να αποφασίσει τι πρέπει να διατηρηθεί. Στόχος του ARCOMEM είναι να βοηθήσει στην μετατροπή των αρχείων σε συλλογικές μνήμες που να είναι πιο στενά συνδεδεμένες με την κοινότητα των χρηστών. Επιπλέον, στοχεύει στην αξιοποίηση του κοινωνικού ιστού και την σοφία του πλήθους έτσι ώστε η αρχειοθέτηση του Web να γίνει μια πιο επιλεκτική κα με νόημα διαδικασία.

- **STRATUSLAB: Κατανεμημένη πλατφόρμα ανάπτυξης υπηρεσιών κοινωνικής δικτύωσης για κινητές συσκευές. (European Commission CP-CSA, RI-261552)**

Το StratusLab έχει ως στόχο την παροχή υπηρεσιών, τη δικτύωση και την έρευνα των υπολογιστικών νεφών και τεχνολογιών εικονικών μηχανών για την απλοποίηση και τη βελτιστοποίηση της χρήσης και λειτουργίας των υφιστάμενων κατανεμημένων υπολογιστικών υποδομών της Ευρωπαϊκής Υποδομής Πλέγματος (EGI). Το έργο αναπτύσσει την Εργαλειοθήκη StratusLab, μια διανομή ανοιχτού κώδικα για το cloud. Ενσωματώνει καινοτόμες τεχνολογίες υπολογιστικών νεφών και virtualization σε υπάρχουσες υποδομές πλέγματος, με την χρήση των τεχνολογιών και των υπηρεσιών cloud σε πλεγματοειδή sites. Επιπλέον, εμπλουτίζει τις υπάρχουσες υποδομές πληροφορικής με “Infrastructure as a Service” (IaaS) cloud υπηρεσίες.

- **GRIDIA: Πρόσβαση σε Πολυμεσικά Δεδομένα από Πλεγματοειδή Συστήματα (Grid Enabled Access to Rich media Content). European Commission, 6th Framework Programme, FP6 34363**

Αντικείμενο του έργου είναι η δημιουργία μιας αξιόπιστης πλατφόρμας βασισμένης στο πλεγματοειδές μοντέλο για το σχεδιασμό και ανάπτυξη ασφαλών και αποδοτικών επιχειρησιακών εφαρμογών. Βασικός στόχος είναι η ανάπτυξη ενός πλεγματοειδούς ενδιάμεσου επιπέδου (grid middleware) που βοηθά στην αποδοτική πρόσβαση κατανεμημένων πλουσίων πολυμεσικών δεδομένων (εικόνα, ήχος, video, άρθρα, κλπ). Επιπλέον, επιτρέπεται η απρόσκοπτη και ανάλογη με τις δυνατότητες του κάθε πελάτη πρόσβαση στα δεδομένα από κινητές συσκευές. Τα παραπάνω πιστοποιήθηκαν μέσα από δύο πραγματικές εφαρμογές από τον τραπεζικό χώρο και τον χώρο των Μ.Μ.Ε. Στο έργο συμμετέχει ένας συνδυασμός από κορυφαία ευρωπαϊκά πανεπιστήμια και εταιρίες με έναν 3-ετή ορίζοντα ολοκλήρωσης.

- **GRIDNEWS: Κατανεμημένη πλατφόρμα πλέγματος (GRID) για τη διάθεση και προηγμένη αναζήτηση περιεχομένου σε ειδησεογραφικό οπτικοακουστικό υλικό. (ΚτΠ Μέτρο 3.3 κωδικός έργου 10)**

Αντικείμενο της πρότασης είναι η ψηφιακή αποθήκευση με υψηλή πιστότητα οπτικοακουστικού υλικού που προέρχεται από τηλεοπτικό σήμα και στην συνέχεια η επεξεργασία αρχείου για τον αποδοτικό εντοπισμό των τμημάτων που περιέχουν κάποια προκαθορισμένη ή αυθαίρετη λέξη κλειδί. Στόχος του έργου είναι η δημιουργία ενός ολοκληρωμένου πληροφοριακού συστήματος μέσω του οποίου οι χρήστες θα έχουν τη δυνατότητα να εντοπίζουν και στη συνέχεια να αναπαράγουν τμήματα οπτικοακουστικού υλικού, πραγματοποιώντας αναζήτηση με λέξεις-κλειδιά. Το πληροφοριακό σύστημα εκμεταλλεύεται τους κατανεμημένους κόμβους μια υποδομής Grid, η οποία με τη χρήση κατάλληλου ενδιάμεσου λογισμικού (Grid middleware) μοιράζει τον υπολογιστικό φόρτο εκτέλεσης των αλγορίθμων επεξεργασίας του υλικού και τον αποθηκευτικό όγκο των ψηφιακών δεδομένων σε ένα πλήθος κόμβων, επιτυγχάνοντας υψηλή απόδοση και διαθεσιμότητα.

– **PRESEMT: (Pattern Recognition-Based Statistically Enhanced MT). (FP7 ICT-2009.2.2 248307)**

Το έργο PRESEMT προτείνει μια νέα προσέγγιση στο πρόβλημα της Μηχανικής Μετάφρασης με την εισαγωγή διεπιστημονικών τεχνικών από τις επιστημονικές περιοχές της μηχανικής μάθησης και υπολογιστικής νοημοσύνης. Για το σκοπό αυτό, ένα ευέλικτο σύστημα MT αναπτύχθηκε, το οποίο ενισχύθηκε με (α) προσεγγίσεις αναγνώρισης προτύπων (εκτεταμένη συσταδοποίηση ή νευρωνικά δίκτυα) για την ανάπτυξη μιας ανεξάρτητης από την γλώσσα ανάλυσης και (β) εξελικτικό υπολογισμό (όπως Γενετικοί Αλγόριθμοι ή Swarm Intelligence) για τη βελτιστοποίηση του συστήματος.

Τα χαρακτηριστικά του έργου είναι:

Η ανάπτυξη μιας νέας μεθόδου που βασίζεται σε γενικευμένες τεχνικές συσταδοποίησης, για τη δημιουργία ενός ανεξάρτητου από την γλώσσα *phrase aligner* που είναι προσαρμόσιμος στις αρχές διατύπωσης που ορίζονται από τους τελικούς χρήστες

Η χρήση της αναγνώρισης προτύπων για τον καθορισμό της συντακτικής δομής

Η χρήση τεχνικών εμπνευσμένων από τα λειτουργικά βιολογικά συστήματα μη αποσαφηνισμένες μεταφράσεις

Η εκτεταμένη χρήση αυτόματων τεχνικών βελτιστοποίησης για να καθοριστεί ένα ώριμο σύστημα για την μεθοδική τη βελτιστοποίηση των παραμέτρων του συστήματος

Η εφαρμογή των μεθόδων μηχανικής μάθησης με σκοπό την προσαρμογή του συστήματος

Η χρήση των παράλληλων υπολογιστικών αρχιτεκτονικών καθώς και ενσωμάτωση αρχιτεκτονικών πολλαπλών πυρήνων για την ουσιαστική πρόοδο στην ταχύτητα της μετάφρασης.

8. Επιστημονική Δραστηριότητα σε Περιοδικά και Συνέδρια

Κριτής σε Επιστημονικά Περιοδικά

1. **Guest Editor**, Journal of Grid Computing: From Grids to Cloud Federations. Special Issue: Orchestration of computing resources in the Cloud-to-Things continuum (Springer)
2. ACM Computing Surveys
3. The Computer Journal (Oxford University Press)
4. IEEE Transactions on Services Computing (IEEE)

5. IEEE Transactions on Knowledge and Data Engineering (IEEE)
6. IEEE Transactions on Parallel and Distributed Systems (IEEE)
7. IEEE Transactions on Cloud Computing (IEEE)
8. Data & Knowledge Engineering (Elsevier)
9. GeoInformatica (Springer)
10. Archives for Scientific Computing (Springer)
11. The Journal of Supercomputing (Springer)
12. Big Data Research (Elsevier)
13. Future Generation Computer Systems (Elsevier)
14. International Journal of Cooperative Information Systems (World Scientific)
15. Algorithms (MDPI)
16. Sustainability (MDPI)
17. Engineering Applications of Artificial Intelligence (Elsevier)
18. Computers and Electrical Engineering (Elsevier)
19. Simulation Modeling Practice and Theory (Elsevier)
20. Journal of Web Semantics (Elsevier)
21. Journal of Systems and Software (Elsevier)
22. Computer Networks (Elsevier)
23. International Journal on Artificial Intelligence Tools (World Scientific)
24. International Journal of Metadata, Semantics and Ontologies (Inderscience)

Μέλος σε Επιστημονικές Επιτροπές Συνεδρίων

1. PC Member, 11th IEEE International Conference on Data Science and Advance Analytics (IEEE DSAA 2024)
2. **PC Member, ACM SIGMOD** International Conference on Management of Data 2025
3. PC Member, 20th International Conference on Advanced Data Mining and Applications (ADMA 2024)
4. PC Member, 29th IEEE Symposium on Computers and Communications, 2024.
5. **Proceedings Chair**, Nineteenth ACM European Conference on Computer Systems (EuroSys 2024)
6. **Chair**, 4th Workshop on From Cloud to Things: harnessing pervasive data in the Computing Continuum - (Cloud2Things) 2024 - in conjunction with IEEE PERCOM 2024
7. **Chair**, 3rd Workshop on From Cloud to Things: harnessing pervasive data in the Computing Continuum - (Cloud2Things) 2023 - in conjunction with IEEE PERCOM 2023
8. PC Member, International Conference on Advanced Data Mining and Applications 2023
9. PC Member, 29th International European Conference on Parallel and Distributed Computing (EuroPar 2023)
10. PC Member, 10th IEEE International Conference on Data Science and Advance Analytics (IEEE DSAA 2023)
11. PC Member, IEEE International Conference on Metaverse Computing, Networking and Applications (IEEE MetaCom 2023)
12. PC Member, 28th International Conference on Database Systems for Advanced Applications (DASFAA 2023)
13. PC Member, DistInSys 2022 2st International Workshop on Distributed and Intelligent Systems in conjunction with IEEE ISCC 2022

14. **Chair**, 2nd Workshop on Cloud-to-Things continuum: towards the convergence of IoT, Edge and Cloud Computing - (Cloud2Things) 2022 - in conjunction with IEEE/ACM CCGRID 2022
15. PC Member, 27th International Conference on Database Systems for Advanced Applications (DASFAA 2022)
16. **PC Member, ACM SIGMOD** International Conference on Management of Data 2022
17. PC Member, International Workshop on Security and Privacy for SMEs (SME-SP 2021), in Conjunction with the 16th International Conference on Availability, Reliability and Security (ARES 2021)
18. **Chair**, 1st Workshop on Cloud-to-Things continuum: towards the convergence of IoT, Edge and Cloud Computing - (Cloud2Things) 2021 - in conjunction with IEEE/ACM CCGRID 2021
19. PC Member, 26th International Conference on Database Systems for Advanced Applications (DASFAA 2021)
20. PC Member, 10th International Conference on Cloud Computing and Services Science (CLOSER 2020)
21. PC Member, 1st Workshop on Secure IoT, Edge and Cloud systems - in conjunction with CCGRID 2020
22. PC Member, 20th Annual IEEE/ACM International Symposium in Cluster, Cloud, and Grid Computing (CCGrid 2020)
23. PC Member, 19th Annual IEEE/ACM International Symposium in Cluster, Cloud, and Grid Computing (CCGrid 2019)
24. PC Member, 9th International Conference on Cloud Computing and Services Science (CLOSER 2019)
25. **Chair**, Short Papers, Posters and Demos, IEEE CloudCom 2018
26. PC Member, 1st International Workshop on Next Generation Clouds for Extreme Data Analytics in conjunction with IEEE CloudCom 2018
27. PC member: International Workshop on Semantic Big Data (SBD 2018) in conjunction with SIGMOD 2018
28. PC member: 8th International Conference on Cloud Computing and Services Sciences (CLOSER 2018)
29. PC member: International Workshop on Semantic Big Data (SBD 2017) in conjunction with SIGMOD 2017
30. PC Member: International Workshop on Autonomic Systems for Big Data Analytics (ASBDA 2017) in conjunction with the 2017 IEEE International Conference on Cloud and Autonomic Computing (ICCAC)
31. Publicity co-chair, PC member: 1st International Workshop on Autonomic Management of Large Scale Container-Based Systems in conjunction with the 2017 IEEE International Conference on Cloud and Autonomic Computing (ICCAC)
32. PC member: International Workshop on Semantic Big Data (SBD 2016) in conjunction with SIGMOD 2016
33. PC member: IEEE International Symposium on Network Computing and Applications (NCA 2016)
34. PC member: 8th International Conference on Knowledge Discovery and Information Retrieval (KDIR 2016)

35. PC member: 2nd International Workshop on Algorithmic Aspects of Cloud Computing (ALGO CLOUD 2016)
36. PC member: Advances in High Dimensional Big Data, in Conjunction with IEEE BigData 2015
37. PC member: IEEE International Symposium on Network Computing and Applications (NCA 2015)
38. PC member: 1st International Workshop on Algorithmic Aspects of Cloud Computing (ALGO CLOUD 2015)
39. PC member: 24th ACM International Conference on Information and Knowledge Management (CIKM 2015)
40. PC member: 23rd ACM International Conference on Conference on Information and Knowledge Management (CIKM 2014)
41. PC member: Fourth International Workshop on Cloud Data Management (CloudDB 2012), in conjunction with CIKM 2012

9. Επιστημονικές Εργασίες

ΔΙΑΤΡΙΒΕΣ:

- Δ1. Ιωάννης Κωνσταντίνου, «**Προσαρμοστικοί Αλγόριθμοι Εξισορρόπησης Φόρτου σε Κατανεμημένα Περιβάλλοντα (Δίκτυα Ομοτίμων και Υπολογιστικά Νέφη)**» Διδακτορική διατριβή, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Η/Υ ΕΜΠ, Εργαστήριο Υπολογιστικών Συστημάτων, 2011.
- Δ2. Ιωάννης Κωνσταντίνου, «**Data Grid: Ενοποίηση ανομοιογενών και γεωγραφικά απομακρυσμένων αποθηκευτικών χώρων μέσω διαδικτύου**» Διπλωματική εργασία, Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών Τεχνο-Οικονομικά Συστήματα, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Η/Υ ΕΜΠ, 2007
- Δ3. Ιωάννης Κωνσταντίνου, «**Δημιουργία – διόρθωση και παραγωγή εκτυπώσιμου οδηγού σπουδών με συνδυασμό στατικής και δυναμικής πληροφορίας μέσω διαδικτύου**» Διπλωματική εργασία, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Η/Υ, ΕΜΠ, 2004.

ΠΡΑΚΤΙΚΑ ΔΙΕΘΝΩΝ ΣΥΝΕΔΡΙΩΝ:

- Σ1. A. Spyrou, I. Konstantinou, and N. Koziris. **A Decision Support System for Automated Configuration of Cloud Native ML Pipelines** In proceedings of the 4th International Workshop on From Cloud to Things: Harnessing pervasive data in the Computing Continuum - (Cloud2Things) in conjunction with IEEE Pervasive Computing 2024 PerCom, Biarritz, France, March 11-15 2024.

Τα συστήματα Big Data όπως το Apache Spark και το Hadoop είναι ακρογωνιαίοι λίθοι της μεγάλης κλίμακας επεξεργασίας δεδομένων. Ωστόσο, χρησιμοποιούνται μόνο σε ένα ή μερικά βήματα μιας μεγαλύτερης ροής επεξεργασίας δεδομένων. Σύνθετες ροές επεξεργασίας δεδομένων που περιλαμβάνουν διαφορετικά σύνολα δεδομένων, υποδομές και βιβλιοθήκες προγραμματισμού απλοποιούνται με τη χρήση εγγενών εργαλείων στο cloud, όπως το Kubernetes και το KubeFlow, που μοντελοποιούν εξαρτήσεις και διαχειρίζονται τον κύκλο

ζωής της εκτέλεσης. Ωστόσο, δεν υπάρχουν ολοκληρωμένες λύσεις που να είναι πλήρως διαλειτουργικές με το Apache Spark και το KubeFlow σε εγκατάσταση εσωτερικής εγκατάστασης. Σε αυτήν την εργασία, επεκτείνουμε το εργαλείο KubeFlow Pipelines για την υποστήριξη συστοιχιών Apache Spark εντός εγκατάστασης. Αξιολογούμε πειραματικά το εργαλείο μας σε ένα πρότυπο βιομηχανικής αξιολόγησης Big-Data με διάφορες διαμορφώσεις υποδομής και δεδομένων. Χρησιμοποιούμε τη συλλεγόμενη γνώση σχετικά με την απόδοση του Spark για να εκπαιδύσουμε ένα σύστημα ML που βασίζεται σε δέντρο αποφάσεων που μπορεί να ανιχνεύσει τη βέλτιστη διαμόρφωση συστοιχίας σύμφωνα με τους περιορισμούς του χρήστη και να προβλέψει τον χρόνο εκτέλεσης του ερωτήματος. Το σύστημα μπορεί να χρησιμοποιηθεί από μη ειδικούς μέσω ενός ολοκληρωμένου GUI. Τέλος, παρέχουμε υλοποιήσεις ανοιχτού κώδικα τόσο της ενοποίησης KubeFlow του Apache Spark όσο και του Συστήματος Υποστήριξης Αποφάσεων.

- Σ2. M. Chrysopoulos, I. Konstantinou, and N. Koziris: **Deep Reinforcement Learning in Cloud Elasticity through Offline Learning and Return Based Scaling** In proceedings of the 16th IEEE International Conference on Cloud Computing (CLOUD), Chicago, IL, USA, July 2-8 2023

Η ελαστική κατανομή πόρων είναι ένα επιθυμητό χαρακτηριστικό των περιβαλλόντων υπολογιστικών νεφών και ένας από τους κύριους λόγους για την ευρεία υιοθέτησή τους. Η ελαστικότητα πόρων επιτρέπει την προσαρμοστική και σε πραγματικό χρόνο κλιμάκωση της υποδομής που μπορεί να ακολουθεί τις διακυμάνσεις του φόρτου εργασίας με οικονομικά αποδοτικό τρόπο χωρίς να θυσιάζεται η απόδοση. Οι προσεγγίσεις της υποσχόμενης μηχανικής μάθησης (ML) χρησιμοποιούν την Ενισχυτική Μάθηση (RL) όπου ένας πράκτορας αλληλεπιδρά με το περιβάλλον του cloud εφαρμόζοντας ενέργειες, παρατηρώντας την ανταμοιβή του αποτελέσματός τους και τροποποιώντας ανάλογα τη στρατηγική του. Ωστόσο, ένα από τα κύρια προβλήματα του RL σε αυτή τη ρύθμιση είναι ότι για να αποκτήσει επαρκή αρχική γνώση του περιβάλλοντος, ο πράκτορας πρέπει να εκτελέσει πολλές χρονοβόρες και υποβαθμιστικές αλληλεπιδράσεις με το cloud. Σε αυτήν την εργασία σχεδιάζουμε και υλοποιούμε το RBS-CQL, ένα Deep-RL σύστημα πρακτόρων υλοποιημένο στην πλατφόρμα Kubernetes για την παρακολούθηση και την αυτόματη κλιμάκωση των περιεκτών μιας εφαρμογής NoSQL σύμφωνα με τον εισερχόμενο φόρτο εργασίας. Συνδυάζουμε τεχνικές βελτιστοποίησης εκπαίδευσης από τη σύγχρονη βιβλιογραφία καθώς και αλγόριθμους RL εκτός σύνδεσης για να μειώσουμε τον χρόνο εκπαίδευσης. Παρέχουμε εμπειρικά αποτελέσματα που δείχνουν ότι η RBS-CQL επιτυγχάνει συστηματική βελτίωση άνω του 10% σε σύγκριση με το επιγραμμικό της ισοδύναμο για δεδομένο αριθμό εμπειριών και ότι είναι σε θέση να εξάγει βελτιωμένες πολιτικές λήψης αποφάσεων ακόμη και από δεδομένα χαμηλότερης ποιότητας.

- Σ3. N. Nikitas, I. Konstantinou, V. Kalogeraki, and N. Koziris: **Cherry: A Distributed Task-Aware Shuffle Service for Serverless Analytics** In proceedings of the 2021 IEEE International Conference on Big Data (BigData 2021), December 15-18 2021

Ενώ έχει γίνει μεγάλη προσπάθεια τα τελευταία χρόνια για τη βελτιστοποίηση συστημάτων Big Data όπως το Apache Spark και το Hadoop, η μεταφορά δεδομένων από όλους μεταξύ ενός βήματος υπολογισμού MapReduce, δηλαδή ο μηχανισμός διαμοιρασμού δεδομένων

μεταξύ κόμβων της συστοιχίας παραμένει πάντα ένα σοβαρό εμπόδιο. Σε αυτήν την εργασία, παρουσιάζουμε το *Cherry*, μια υπηρεσία ανοιχτού κώδικα κατανεμημένης υπηρεσίας προσωρινής αποθήκευσης *sHuffle* για αναλύσεις χωρίς διακομιστή. Τα ενδεδειγμένα πειράματά μας σε ένα *cloud testbed* χρησιμοποιώντας ρεαλιστικούς και συνθετικούς φόρτους εργασίας δείχνουν ότι το *Cherry* μπορεί να επιτύχει σχεδόν 23% έως 39% μείωση στην ολοκλήρωση του σταδίου μείωσης με μικρά μεγέθη μπλοκ τυχαίας αναπαραγωγής, μείωση 10% στον χρόνο εκτέλεσης σε πραγματικούς φόρτους εργασίας, ενώ μπορεί να χειριστεί αποτελεσματικά τις αποτυχίες εκτέλεσης *Spark* με σταθερή επιβάρυνση επανυπολογισμού χρόνου εργασίας σε σύγκριση με τις υπάρχουσες προσεγγίσεις.

- Σ4. N. Provas, I. Konstantinou, and N. Koziris: **Is Systematic Data Sharding able to Stabilize Asynchronous Parameter Server Training?** In proceedings of the 2021 IEEE International Conference on Big Data (BigData 2021), December 15-18 2021

Τα τελευταία χρόνια, η βαθιά εκμάθηση έχει αυξηθεί σε δημοτικότητα σε διάφορους τομείς, εισάγοντας πολύπλοκα μοντέλα για τη διαχείριση της έκρηξης δεδομένων. Ωστόσο, ενώ τέτοιες αρχιτεκτονικές μοντέλων μπορούν να υποστηρίξουν τον τεράστιο όγκο δεδομένων, ένας μόνο κόμβος υπολογιστών δεν μπορεί να εκπαιδεύσει το μοντέλο χρησιμοποιώντας ολόκληρο το σύνολο δεδομένων εγκαίρως. Έτσι, έχουν προταθεί εξειδικευμένες κατανεμημένες αρχιτεκτονικές, οι περισσότερες από τις οποίες ακολουθούν σχήματα παραλληλισμού δεδομένων, όπως για παράδειγμα η ευρέως χρησιμοποιούμενη προσέγγιση διακομιστή παραμέτρων. Σε αυτή τη περίπτωση, κάθε εργαζόμενος συμβάλλει στη διαδικασία εκπαίδευσης με ασύγχρονο τρόπο. Ενώ η ασύγχρονη εκπαίδευση δεν υποφέρει από γενικά έξοδα συγχρονισμού, εισάγει το πρόβλημα των παλιών κλίσεων που μπορεί να προκαλέσουν απόκλιση του μοντέλου κατά τη διάρκεια της εκπαιδευτικής διαδικασίας. Σε αυτό το άρθρο, εξετάζουμε διαφορετικά σχήματα ανάθεσης δεδομένων σε εργαζόμενους, τα οποία διευκολύνουν την προσέγγιση της ασύγχρονης μάθησης. Συγκεκριμένα, προτείνουμε δύο διαφορετικούς αλγόριθμους για την εκτέλεση της κοινής χρήσης δεδομένων. Η πειραματική μας αξιολόγηση έδειξε ότι, όταν λαμβάνεται υπόψη η διαστρωμάτωση, τα αποτελέσματα επικύρωσης παρουσιάζουν έως και 6 φορές λιγότερη διακύμανση σε σύγκριση με τη δημιουργία τυπικού διαμοιρασμού. Όταν εκτελείται περαιτέρω διερεύνηση δεδομένων για κρυφή διαστρωμάτωση, οι μετρήσεις επικύρωσης μπορούν να βελτιστοποιηθούν ελαφρώς. Αυτή η μέθοδος επιτυγχάνει επίσης τη μείωση της διακύμανσης των μετρήσεων εκπαίδευσης και επικύρωσης έως και 8X και 2X αντίστοιχα.

- Σ5. A. Krisilias, N. Provas, I. Konstantinou, and N. Koziris: **A Performance Evaluation of Distributed Deep Learning Frameworks on CPU Clusters Using Image Classification Workloads** In proceedings of the Fifth IEEE International Workshop on Benchmarking, Performance Tuning and Optimization for Big Data Applications (BPOD 2021 in conjunction with IEEE BigData 2021), virtual event, December 15-18 2021

Τα τελευταία χρόνια, η βαθιά μάθηση χρησιμοποιείται ευρέως σε μια ποικιλία διαφορετικών πεδίων και εφαρμογών. Η συνεχής αύξηση των δεδομένων που χρησιμοποιούνται για την εκπαίδευση σύνθετων μοντέλων, έχει ανοίξει την έρευνα στην κατανεμημένη μάθηση. Σε αυτόν τον τομέα, δύο κύριες αρχιτεκτονικές χρησιμοποιούνται για την εκπαίδευση μοντέλων με τρόπο διανομής, η *allreduce* και ο διακομιστής παραμέτρων. Και οι δύο υποστηρίζουν τη

σύγχρονη εκμάθηση, ενώ ο διακομιστής παραμέτρων υποστηρίζει επίσης ασύγχρονη μάθηση. Αυτές οι αρχιτεκτονικές υιοθετούνται από εταιρείες τεχνολογίας, οι οποίες έχουν αναπτύξει πολλαπλά συστήματα για το σκοπό αυτό. Μεταξύ των πιο δημοφιλών και ευρέως χρησιμοποιούμενων καταναμημένων συστημάτων βαθιάς μάθησης είναι το Google TensorFlow, το Facebook PyTorch και το Apache MXNet. Σε αυτό το άρθρο, ποσοτικοποιούμε το χάσμα απόδοσης μεταξύ αυτών των συστημάτων και παρουσιάζουμε μια λεπτομερή ανάλυση για να συζητήσουμε τις παραμέτρους που επηρεάζουν τον χρόνο εκτέλεσής τους. Συνολικά, στις ρυθμίσεις σύγχρονης εκμάθησης, το TensorFlow είναι πιο αργό σε σύγκριση με το PyTorch κατά μέσο όρο 2,65X, ενώ το τελευταίο υστερεί στο MXNet κατά μέσο όρο 1,38X. Όσον αφορά την ασύγχρονη μάθηση, το MXNet είναι ταχύτερο κατά μέσο όρο 3,22X σε σχέση με το TensorFlow.

- Σ6. A. Tsiourvas, C. Bitsakos, I. Konstantinou, D. Fotakis, and N. Koziris: **A Mechanism Design and Learning Approach for Revenue Maximization on Cloud Dynamic Spot Markets** In proceedings of the 14th IEEE International Conference on Cloud Computing (CLOUD), September 5-11 2021

Οι σύγχρονες εφαρμογές υπολογιστών μεγάλης κλίμακας αποτελούνται από πολύπλοκες ελαστικές εφαρμογές που εκτελούνται σε ομάδες μηχανών. Μια τρέχουσα τάση που υιοθετείται από τους παρόχους είναι να ορίζουν αχρησιμοποίητες εικονικές μηχανές, ή αλλιώς spot στιγμιότυπα, σε χαμηλές τιμές για να επωφεληθούν από την πλεονάζουσα χωρητικότητα. Σε αυτό το άρθρο παρουσιάζουμε μια ομάδα αποτελεσματικών πολιτικών κατανομής και τιμολόγησης που μπορούν να χρησιμοποιηθούν από τους πωλητές για τους μηχανισμούς άμεσης τιμής τους. Μοντελοποιούμε τη διαδικασία απόκτησης εικονικών μηχανών ως μια αληθινή δημοπρασία σακιδίου και εφαρμόζουμε κανόνες δυναμικής κατανομής και τιμολόγησης που επιτυγχάνουν σχεδόν βέλτιστα έσοδα και κοινωνική ευημερία. Καθώς το πρόβλημα είναι NP-hard, οι λύσεις μας βασίζονται σε κατά προσέγγιση αλγόριθμους. Πρώτον, προτείνουμε δύο λύσεις που δεν χρησιμοποιούν προηγούμενη γνώση. Στη συνέχεια, τα ενισχύουμε με τρεις αλγόριθμους εκμάθησης. Τις αξιολογούμε με προσομοιώσεις στο σύνολο δεδομένων Google Cluster και τις συγκρίνουμε με τους μηχανισμούς Ενιαία Τιμή, Βέλτιστη Ενιαία Τιμή και Ex-CORE. Ο προτεινόμενος δυναμικός μας μηχανισμός είναι στιβαρός, επιτυγχάνει έσοδα έως και 89% της δημοπρασίας Βέλτιστης Ενιαίας Τιμής και υπολογίζει την κατανομή σε πολυωνυμικό χρόνο, καθιστώντας τη συνεισφορά μας υπολογιστικά αποδοτική σε σενάρια πραγματικού χρόνου.

- Σ7. N. Provas, E. Kassela, N. Chalvantzis, A. Bakogiannis, I. Giannakopoulos, I. Konstantinou and N. Koziris: **SELIS BDA: Big Data Analytics for the Logistics Domain**. In proceedings of the 2020 Applications of Big Data Technology in the Transport Industry Workshop (in conjunction with IEEE BigData 2020), December 10-13 2020

Σε αυτό το άρθρο παρουσιάζουμε το SELIS Big Data Analytics and Machine Learning System (BDA), ένα ελαστικό σύστημα ανοιχτού κώδικα με υποστήριξη υπολογιστικών νεφών που έχει σχεδιαστεί και υλοποιηθεί προκειμένου να αντιμετωπίσει ζητήματα που σχετίζονται με δεδομένα από τον τομέα της εφοδιαστικής. Λαμβάνοντας υπόψη τις πραγματικές ανάγκες ανάλυσης δεδομένων από περισσότερους από 40 παρόχους εφοδιαστικής της ΕΕ,

παρουσιάζουμε τη λεπτομερή αρχιτεκτονική του SELIS BDA μαζί με τα γενικά μοντέλα δεδομένων και εκτέλεσης που σχεδιάστηκαν για να καλύψουν τις διαφορετικές ανάγκες τους. Περιγράφουμε τις κύριες τεχνολογίες που έχουμε χρησιμοποιήσει για να πραγματοποιήσουμε την αντίστοιχη προσφορά και να δικαιολογήσουμε τις επιλογές μας από την ευρύτερη κοινότητα των συστημάτων Μεγάλων Δεδομένων. Δοκιμάζουμε πειραματικά την προσφορά μας με διάφορους φόρτους εργασίας, όπου αποδεικνύουμε ότι μπορεί να κλιμακώσει την εξυπηρέτηση μεγάλου αριθμού ταυτόχρονων αιτημάτων, ενώ η αφαίρεση / ενορχήστρωση του δημιουργεί πολύ μικρό γενικό κόστος σε σύγκριση με τα επιμέρους συστήματα Μεγάλων Δεδομένων. Πιστεύουμε ότι το SELIS BDA μπορεί να είναι ένα εύκολο στη χρήση σημείο εισόδου για τον μεγάλο κόσμο των δεδομένων ανάλυσης για οποιαδήποτε εταιρεία logistics ειδικά από τον τομέα των MME.

Σ8. C. Kotselidis et al.: **Efficient Compilation and Execution of JVM-Based Data Processing Frameworks on Heterogeneous Co-Processors** In Proceedings of the Design, Automation and Test in Europe Conference (DATE 2020)

Αυτό το έγγραφο ασχολείται με το θεμελιώδες ζήτημα του τρόπου με τον οποίο τα σύγχρονα συστήματα Μεγάλου Όγκου Δεδομένων (Big Data) μπορούν να εκμεταλλεύονται δυναμικά και διαφανώς τους ετερογενείς επιταχυντές υλικού. Αφού παρουσιάσουμε τις μείζονες προκλήσεις που πρέπει να αντιμετωπιστούν για αυτόν τον στόχο, περιγράφουμε την προτεινόμενη αρχιτεκτονική μας για την αυτόματη και διαφανή επιτάχυνση υλικού των συστημάτων και των εφαρμογών Big Data. Το όραμά μας είναι να διατηρήσουμε το ομοιόμορφο μοντέλο προγραμματισμού των συστημάτων Big Data και να ενεργοποιήσουμε την αυτόματη, δυναμική σύνταξη Just-In-Time των υποψήφιων τμημάτων κώδικα που επωφελούνται από την επιτάχυνση υλικού στην αντίστοιχη μορφή. Σε συνδυασμό με την επιλογή συσκευής που βασίζεται σε μηχανική μάθηση, η οποία σέβεται τους περιορισμούς που ορίζονται από το χρήστη (π.χ. κόστος, χρόνος κλπ.), επιτρέπουμε τη δυναμική εκτέλεση κώδικα σε GPU και FPGA με διαφάνεια στον χρήστη. Επιπλέον, διορθώνουμε δυναμικά την εκτέλεση κατά το χρόνο εκτέλεσης με βάση τη διαθεσιμότητα πόρων. Τα προκαταρκτικά μας αποτελέσματα καταδεικνύουν ότι η προσέγγισή μας μπορεί να επιταχύνει μια υπάρχουσα εφαρμογή Apache Flink έως και 16,5x.

Σ9. E. Kassela, N. Provatias, I. Konstantinou, A. Floratou and N. Koziris: **General-Purpose vs. Specialized Data Analytics Systems: A Game of ML & SQL Thrones** In proceedings of the 2019 IEEE International Conference on Big Data (BigData 2019), Los Angeles, CA, USA December 9-12 2019

Κατά τη διάρκεια της τελευταίας δεκαετίας, έχουν προκύψει πολλά συστήματα για την υποστήριξη αναλύσεων δεδομένων σε διάφορους τομείς, όπως η SQL και η μηχανική μάθηση, μεταξύ άλλων. Σε κάθε τομέα ανάλυσης δεδομένων, υπάρχουν τώρα πολλά διαφορετικά εξειδικευμένα συστήματα που αξιοποιούν τις βελτιστοποιήσεις που σχετίζονται με το συγκεκριμένο τομέα για την αποτελεσματική εκτέλεση του φόρτου εργασίας τους. Μια εναλλακτική προσέγγιση είναι η δημιουργία ενός συστήματος γενικής ανάλυσης δεδομένων που χρησιμοποιεί μια κοινή μηχανή εκτέλεσης και ένα μοντέλο προγραμματισμού για την υποστήριξη φόρτων εργασίας σε διάφορους τομείς. Σε αυτό το έργο, επιλέγουμε αντιπροσωπευτικά συστήματα κάθε κατηγορίας (Spark, TensorFlow, Presto και Hive) και

συγκρίνουμε την απόδοσή τους σε μια μεγάλη ποικιλία φόρτων εργασίας SQL και μηχανικής μάθησης. Εκτελούμε μια εκτενή συγκριτική ανάλυση σχετικά με τα πλεονεκτήματα και τους περιορισμούς κάθε συστήματος και τονίζουμε τους βασικούς τομείς βελτίωσης για όλα τα συστήματα. Πιστεύουμε ότι οι βασικές γνώσεις που θα προκύψουν από αυτή τη μελέτη θα είναι χρήσιμες για τους προγραμματιστές να βελτιώσουν την απόδοση αυτών των συστημάτων.

- Σ10. E. Kassela, N. Provatias, A. Tsiourvas, I. Konstantinou, and N. Koziris: **BigOptiBase: Big Data Analytics for Base Station Energy Consumption Optimization** In proceedings of the 2019 IEEE International Conference on Big Data (BigData 2019), Los Angeles, CA, USA December 9-12 2019

Οι φορείς εκμετάλλευσης κινητών δικτύων αναπτύσσουν νέες τεχνολογίες, όπως το δίκτυο 5G, για να αντιμετωπίσουν τη διαρκώς αυξανόμενη κίνηση στο δίκτυο, ενώ προσπαθούν να βελτιστοποιήσουν τις λειτουργίες τους. Ωστόσο, οι ενεργειακά αποδοτικότερες προσεγγίσεις είναι απαραίτητες για τη μείωση του κόστους και τη συμμόρφωση με τις αρχές του ενεργειακού αποτυπώματος. Η χρήση του δικτύου και τα δεδομένα IoT που παράγονται στους σταθμούς βάσης μπορούν να χρησιμοποιηθούν για την ανάπτυξη τέτοιων προσεγγίσεων. Λαμβάνοντας υπόψη τα παραπάνω, προτείνουμε την πλατφόρμα BigOptiBase. Μέσω αυτής της πλατφόρμας θα προσφέρουμε ένα υποσύστημα αναλύσεων μεγάλων δεδομένων που αναπτύχθηκε για την παροχή ελαστικών ενεργειακά αποδοτικών λύσεων για τους σταθμούς βάσης χρησιμοποιώντας αναλύσεις δεδομένων και τεχνολογίες μηχανικής μάθησης.

- Σ11. N. Provatias, I. Konstantinou and N. Koziris: **Towards Faster Distributed Deep Learning Using Data Hashing Techniques** In proceedings of the 2019 IEEE International Conference on Big Data (BigData 2019), Los Angeles, CA, USA December 9-12 2019

Σήμερα, η βαθιά μηχανική μάθηση είναι ένα κρίσιμο κομμάτι μιας ποικιλίας εφαρμογών μεγάλων δεδομένων. Τόσο η τεράστια ποσότητα δεδομένων όσο και η μεγάλη πολυπλοκότητα των υπερσύγχρονων νευρωνικών δικτύων έχουν οδηγήσει στην εκμάθηση του δικτύου με καταναμημένο τρόπο σε ομάδες. Δεδομένου ότι τα γενικά έξοδα συγχρονισμού είναι συνήθως μη αποδεκτά για την απόδοση της κατάρτισης, συνήθως στις περιπτώσεις αυτές προτιμάται η ασύγχρονη εκπαίδευση. Ωστόσο, αυτός ο τρόπος εκπαίδευσης είναι ευαίσθητος σε αντικρουόμενες ενημερώσεις. Οι ενημερώσεις αυτές συμβαίνουν συχνότερα όταν οι εργαζόμενοι εκπαιδεύονται σε ένα εντελώς διαφορετικό μέρος των δεδομένων. Για να μειώσουμε αυτό το φαινόμενο, σε αυτή την εργασία προτείνουμε τη χρήση προγραμμάτων αντιστάθμισης κατά τη διανομή εκπαιδευτικών δεδομένων μεταξύ των εργαζομένων.

- Σ12. E. Kassela, I. Konstantinou and N. Koziris: **Towards a Multi-engine Query Optimizer for Complex SQL Queries on Big Data** In proceedings of the 2019 IEEE International Conference on Big Data (BigData 2019), Los Angeles, CA, USA December 9-12 2019

Σε μια εποχή όπου οι μεγάλες αναλύσεις δεδομένων έχουν γίνει μια απαίτηση πρώτης τάξης τόσο για τη βιομηχανική όσο και για την ακαδημαϊκή κοινότητα, κατασκευάζονται πολλά συστήματα για την εκτέλεση καταναμημένων ερωτημάτων αναλύσεων ειδικού σκοπού.

Ανάλυση δεδομένων με την χρήση SQL είναι ένας πολύ δημοφιλής αλλά και δύσκολος τομέας λόγω της πολυπλοκότητάς του που απαιτεί πολλαπλές βελτιστοποιήσεις ερωτημάτων κατά τον χρόνο εκτέλεσης. Τα δημοφιλή συστήματα, όπως το Presto και το SparkSQL, συνήθως ανακτούν δεδομένα από πολλαπλές πηγές και επεξεργάζονται τοπικά χρησιμοποιώντας βελτιστοποιητές συγκεκριμένου τομέα. Ωστόσο, πρόσφατες εργασίες δείχνουν ότι καμία μηχανή δεν προσφέρει τη βέλτιστη λύση all-in-one για όλους τους τύπους ερωτημάτων SQL. Λαμβάνοντας αυτό υπόψη, οραματιζόμαστε τη δημιουργία ενός βελτιστοποιητή για να διευκολύνουμε τα ταχύτερα καταναμημένα SQL analytics σε πολλαπλά συστήματα, ο οποίος θα πραγματοποιεί βελτιστοποίηση σε επίπεδο τελεστή χρησιμοποιώντας τεχνικές μηχανικής μάθησης και θα αξιοποιεί τις εξελιγμένες τοπικές βελτιστοποιήσεις μηχανών δεδομένων.

- Σ13. K. Bitsakos, I. Konstantinou and N. Koziris: **DERP: A Deep Reinforcement Learning Cloud System for Elastic Resource Provisioning** In proceedings of the 2018 IEEE International Conference on Cloud Computing Technology and Science (CloudCom), Nicosia, Cyprus December 10-13 2018

Οι σύγχρονες συστοιχίες υπολογιστών μεγάλης κλίμακας επωφελούνται σημαντικά από την ελαστικότητα. Η ελαστικότητα επιτρέπει σε μια συστοιχία να διαθέτει δυναμικά τους πόρους του υπολογιστή, με βάση τις διακυμάνσεις του φόρτου εργασίας του χρήστη. Πολλοί πάροχοι νέφους χρησιμοποιούν προσεγγίσεις που βασίζονται σε κατώτατα όρια (κατώφλια) όπου έχει αποδειχθεί ότι είναι δύσκολο να διαμορφωθούν και να βελτιστοποιηθούν, ενώ άλλοι χρησιμοποιούν μεθόδους ενισχυτικής μάθησης και προσεγγίσεις δέντρων αποφάσεων, οι οποίες αγωνίζονται όταν αντιμετωπίζουν μεγάλες πολυδιάστατες καταστάσεις μοντέλων. Σε αυτό το έργο χρησιμοποιούμε τεχνικές μάθησης βαθιάς ενίσχυσης για την επίτευξη αυτόματης ελαστικότητας. Χρησιμοποιούμε τρεις διαφορετικές προσεγγίσεις ενός πράκτορα εκμάθησης βαθιάς ενίσχυσης (DERP), που ονομάζεται DERP (Deep Elastic Resource Provisioning), που λαμβάνει ως εισροή την τρέχουσα πολυδιάστατη κατάσταση μιας συστοιχίας και καταφέρνει να εκπαιδεύσει και να συγκλίνει στη βέλτιστη συμπεριφορά ελαστικότητας μετά από μια πεπερασμένη ποσότητα εκπαιδευτικών βημάτων. Το σύστημα αποφασίζει αυτόματα και προχωρά στην αίτηση / απελευθέρωση πόρων VM από τον πάροχο και την ενορχήστρωση τους μέσα σε ένα cluster NoSQL σύμφωνα με τις πολιτικές / ανταμοιβές που ορίζονται από τον χρήστη. Συγκρίνουμε τον πράκτορά μας με τις πιο σύγχρονες μεθόδους μάθησης και ανάλυσης, σε απαιτητικά περιβάλλοντα προσομοίωσης και δείχνουμε ότι κερδίζει ανταμοιβές έως και 1,6 φορές καλύτερα κατά τη διάρκεια της ζωής του. Στη συνέχεια δοκιμάζουμε την προσέγγισή μας σε ένα πραγματικό περιβάλλον και δείχνουμε ότι το σύστημα μετασχηματίζει τις συστάδες σε πραγματικό χρόνο και προσαρμόζει την απόδοσή του μέσα από μια ποικιλία απαιτητικών στρατηγικών βελτιστοποίησης, φορτίων εισόδου και εκπαίδευσης.

- Σ14. Ioannis Mytilinis, Constantinos Bitsakos, Katerina Doka, Ioannis Konstantinou and Nectarios Koziris: **The Vision of a Heterogeneous Scheduler** In proceedings of the 1st International Workshop on Next Generation Clouds for Extreme Data Analytics (Xtreme-cloud - in conjunction with IEEE CloudCom 2018), Nicosia, Cyprus December 10-13 2018

Τα σύγχρονα συστήματα επεξεργασίας δεδομένων, οι πλατφόρμες προγραμματισμού και οι υποδομές cloud χρησιμοποιούν εξειδικευμένους επιταχυντές υλικού όπως GPUs, FPGAs,

TPUs, ASICs κ.λπ. για τη βελτιστοποίηση της εκτέλεσης φόρτων εργασίας με ένταση πόρων, όπως μηχανική μάθηση, τεχνητή νοημοσύνη ή γενικές εργασίες δεδομένων Analytics. Ωστόσο, αυτή η υποστήριξη είναι κυρίως μια χειρωνακτική διαδικασία που εξαρτάται από τους χρήστες και απαιτεί προσεκτικές και μορφομένες αποφάσεις τόσο για το ποσό όσο και για τον τύπο των απαιτούμενων πόρων για την εκμετάλλευση του υποκείμενου υλικού και για την επίτευξη πολιτικών υψηλού επιπέδου που ορίζονται από το χρήστη. Σε αυτό το έργο παρουσιάζουμε τον αρχικό σχεδιασμό του HeterogeneRous Scheduler (HRS), έναν έξυπνο διεκπαιρευτή που μπορεί να κάνει αυτοματοποιημένες αποφάσεις για το πώς και πού να χαρτογραφήσει αυθαίρετες εργασίες αναλύσεων δεδομένων στο υποκείμενο υλικό cloud που μπορεί να αποτελείται από ένα συνδυασμό επιταχυντών υλικού και συστοιχιών με CPU γενικού σκοπού. Προσπαθούμε να αξιολογήσουμε πειραματικά τους συνδυασμούς απόδοσης μεταξύ των επιταχυντών υλικού και των CPU, όπου δείχνουμε ότι υπάρχουν περιπτώσεις όπου η μία τεχνολογία ξεπερνάει την άλλη. Παρουσιάζουμε τελικά μια αρχική αρχιτεκτονική του HRS όπου παρουσιάζουμε τα διαφορετικά συστατικά στοιχεία και τις αλληλεπιδράσεις τους με το πλαίσιο Big Data και την υποδομή cloud.

- Σ15. N. Zacheilas, N. Chavantzis, I. Konstantinou, V. Kalogeraki and N. Koziris:
ORiON: Online ResOource Negotiator for multiple Big Data Analytics frameworks In proceedings of the 2018 IEEE International Conference on Autonomic Computing (ICAC 2018), Trento, Italy, September 3-7 2018.

Τα τελευταία χρόνια παρατηρούμε την ταχεία ανάπτυξη εφαρμογών μεγάλης κλίμακας σε ένα ευρύ φάσμα τομέων - από τις υποδομές υγειονομικής περίθαλψης έως τη διαχείριση της κυκλοφορίας. Ο μεγάλος όγκος δεδομένων που πρέπει να υποβληθούν σε επεξεργασία έχει προωθήσει την ανάπτυξη συστημάτων ειδικού σκοπού τα οποία χειρίζονται τον κατακλυσμό δεδομένων με παραλληλισμό της επεξεργασίας δεδομένων και ταυτόχρονα με τη χρήση πολλαπλών κόμβων υπολογιστών. Αυτά τα συστήματα διαφοροποιούνται σημαντικά από την άποψη των πολιτικών που ακολουθούν για να αποσυνθέσουν το φόρτο εργασίας τους σε πολλαπλές εργασίες και επίσης τον τρόπο με τον οποίο εκμεταλλεύονται τους διαθέσιμους υπολογιστικούς πόρους. Ως αποτέλεσμα, με βάση το σύστημα στο οποίο υλοποιήθηκαν οι εφαρμογές, παρατηρούμε σημαντικές διακυμάνσεις στη χρήση των πόρων και τον χρόνο εκτέλεσης. Επομένως, ο προσδιορισμός του κατάλληλου συστήματος για την εκτέλεση μιας μεγάλης εφαρμογής δεδομένων δεν είναι εύκολη υπόθεση. Σε αυτή την εργασία προτείνουμε τον Orion, έναν καινοτόμο διαπραγματευτή πόρων για υποδομές cloud που υποστηρίζουν πολλά μεγάλα συστήματα δεδομένων όπως το Apache Spark, το Apache Flink και το TensorFlow. Πιο συγκεκριμένα, σε μια αίτηση, το Orion καθορίζει το καταλληλότερο σύστημα για την ανάθεση της. Επιπλέον, ο Orion διατηρεί τους απαιτούμενους πόρους έτσι ώστε η εφαρμογή να είναι σε θέση να ανταποκριθεί στις απαιτήσεις επιδόσεων. Ο διαπραγματευτής μας εκμεταλλεύεται τις τελευταίες τεχνικές πρόβλεψης για τον υπολογισμό του χρόνου εκτέλεσης της εφαρμογής, όταν εκχωρείται σε ένα συγκεκριμένο πλαίσιο με διάφορες παραμέτρους διαμόρφωσης και πόρους επεξεργασίας. Τέλος, η λεπτομερής πειραματική αξιολόγηση, χρησιμοποιώντας πρακτικά μεγάλα δεδομένα για το τοπικό σύμπλεγμα, δείχνει ότι η προσέγγισή μας ξεπερνά τους ανταγωνιστές της.

- Σ16. A. Naskos, A. Gounaris and I. Konstantinou: **Elton: a Cloud Resource Scaling-out Manager for NoSQL Databases**. In proceedings of the 2018 IEEE International Conference of Data Engineering (ICDE 2018), Paris, France, April 16-19 2018

Παρουσιάζουμε το εργαλείο Elton, ένα σύστημα ανοιχτού κώδικα διαχείρισης της ελαστικότητας των πόρων του cloud προσαρμοσμένο στις βάσεις δεδομένων NoSQL. Ο Elton ενσωματώνεται στην πλατφόρμα web Ganetimg και προσφέρει μια εύχρηστη διεπαφή ιστού μέσω της οποίας μπορούν να εκτελεστούν παρακολούθηση και οριζόντια κλιμάκωση βάσεων δεδομένων NoSQL με ενεργοποιημένες τις ερωτήσεις ανάλυσης. Ο Elton χρησιμοποιεί τις διαδικασίες λήψης αποφάσεων Markov (MDPs) ως το υποκείμενο πλαίσιο μοντελοποίησης και ενσωματώνει πολιτικές οριζόντιας κλιμάκωσης που προσφέρουν διαφορετικές αντισταθμίσεις μεταξύ των επιδόσεων και του χρηματικού κόστους. Η κύρια καινοτομία του είναι ότι χρησιμοποιεί πιθανολογικούς ελέγχους μοντέλων για να επιτρέψει τόσο αποτελεσματικές αποφάσεις ελαστικότητας όσο και ανάλυση ενεργειών κλιμάκωσης και χρησιμεύει ως μελέτη περίπτωσης σχετικά με τα οφέλη του ελέγχου μοντέλου στην επιγραμμική (on-line) λήψη αποφάσεων και ανάλυση.

- Σ17. K. Lolos, I. Konstantinou, V. Kantere and N. Koziris: **Elastic Management of Cloud Applications using Adaptive Reinforcement Learning**. In proceedings of the 2017 IEEE International Conference on Big Data (BigData 2017), Boston, MA, USA December 11-14 2017

Οι σύγχρονες υπολογιστικές υποδομές μεγάλης κλίμακας αποτελούνται από σύνθετες εφαρμογές που εκτελούνται σε συστοιχίες μηχανών. Ένα σημαντικό ζήτημα σε αυτά είναι η προσφορά ελαστικότητας, δηλαδή η δυναμική κατανομή πόρων σε εφαρμογές για την κάλυψη των διακυμάνσεων των απαιτήσεων του φόρτου εργασίας. Οι προσεγγίσεις που βασίζονται στην χρήση κατωφλίων χρησιμοποιούνται συνήθως, ωστόσο είναι δύσκολο να βαθμονομηθούν και να βελτιστοποιηθούν. Έχουν προταθεί προσεγγίσεις βασισμένες στην ενισχυτική μάθηση (RL), αλλά απαιτούν μεγάλο αριθμό καταστάσεων προκειμένου να μοντελοποιήσουν μια σύνθετη συμπεριφορά μιας εφαρμογής. Έχουν προταθεί μέθοδοι που χωρίζουν προσαρμοστικά τον χώρο καταστάσεων, αλλά τα κριτήρια και οι στρατηγικές κατανομής τους είναι κατώτερα από τη βέλτιστη. Σε αυτή την εργασία παρουσιάζουμε τον MDP_DT, έναν νέο αλγόριθμο ενισχυτικής μάθησης ενίσχυσης πλήρους μοντέλου για τη διαχείριση ελαστικών πόρων που χρησιμοποιεί προσαρμοστικό διαχωρισμό χώρου καταστάσεων. Προτείνουμε δύο νέα στατιστικά κριτήρια και τρεις στρατηγικές και αποδεικνύουμε πειραματικά ότι αποφασίζουν σωστά τόσο πού και πότε πρέπει να χωριστούν, ξεπερνώντας τις υπάρχουσες προσεγγίσεις. Προσπαθούμε να αξιολογήσουμε πειραματικά το MDP_DT σε μια πραγματική συστοιχία μεγάλης κλίμακας πάνω από μεταβλητά πρωτοφανή φορτία εργασίας και αποδεικνύουμε ότι παίρνει πιο ενημερωμένες αποφάσεις σε σύγκριση με τις στατικές προσεγγίσεις χωρίς μοντέλα και βασισμένες σε κατώφλια, ενώ απαιτεί ένα ελάχιστο ποσό δεδομένων κατάρτισης. Δείχνουμε πειραματικά ότι αυτή η προσαρμογή επέτρεψε στο MDP_DT να βελτιστοποιήσει το επιτευχθέν κέρδος ενώ ήταν 40% φθηνότερο από τις βαθμονομημένες προσεγγίσεις RL και κατωφλίου.

- Σ18. G. Touloupas, I. Konstantinou and N. Koziris: **RASP: Real-time Network Analytics with Distributed NoSQL Stream Processing**. In proceedings of the 2nd

Workshop on Real-time and Stream Analytics in Big Data (in conjunction with IEEE BigData 2017), Boston, MA, USA December 11-14 2017

Σε αυτό το άρθρο παρουσιάζουμε το RASP, ένα σύστημα που συνδυάζει την επεξεργασία καταναλωμένων ροών και τα συστήματα NoSQL για να επιτρέψει σε πραγματικό χρόνο και με χαμηλή καθυστέρηση την αποθήκευση και τη σύνδεση των εισερχόμενων ροών δεδομένων με εξωτερικά σύνολα δεδομένων αυθαίρετων μεγεθών μέσω ενός επεκτάσιμου συμβατού με SQL τρόπου. Επιτυγχάνουμε εκτέλεση σε πραγματικό χρόνο με χαμηλή απόκριση, χρησιμοποιώντας τα συστήματα Kafka και Storm για να συνενώσουμε τις εισερχόμενες πλειάδες κατά την άφιξή τους, ενώ το αποκανονικοποιημένο αποτέλεσμα αποθηκεύεται στην HBase, ένα καταναλωμένο σύστημα NoSQL με τη χρήση του Phoenix, ένα σύστημα που υποστηρίζει πλήρως την SQL. Βελτιώσαμε την εκτέλεση της τοπολογίας για να επιτύχουμε μέγιστη απόδοση και εφαρμόζουμε επίσης ένα σύνολο βελτιστοποιήσεων τόσο στο σύστημα αποθήκευσης HBase όσο και στο πλαίσιο εκτέλεσης του SQL Phoenix. Χρησιμοποιούμε το RASP για να λύσουμε ένα πρόβλημα ανάλυσης δικτύου χρησιμοποιώντας πραγματικά δεδομένα. Το RASP εκτελεί τους υπολογισμούς του χρησιμοποιώντας έναν επεκτάσιμο αγωγό από Storm “bolts” που αυξάνουν σταδιακά τις εισερχόμενες πλειάδες με την εκτέλεση διαφορετικών αλγορίθμων. Εκτελούμε το σύστημά μας μέσω ενός cloud IaaS και αξιολογούμε την απόδοσή του για διάφορα φορτία εργασίας, μέγεθος συστοιχιών και διαμορφώσεις, όπου δείχνουμε ότι σε ορισμένες περιπτώσεις το RASP επιτυγχάνει αύξηση της απόδοσης κατά περισσότερο από 140% και μείωση της καθυστέρησης άνω του 65% σε σύγκριση με την απλή «εργοστασιακή» ρύθμιση.

- Σ19. T. Doudali, I. Konstantinou and N. Koziris: **Spaten: a Spatio-temporal and Textual Big Data Generator**. In proceedings of the 2017 Big Spatial Data Workshop (BSD 2017 in conjunction with IEEE BigData 2017), Boston, MA, USA December 11-14 2017

Οι χρήστες των δικτύων κοινωνικής δικτύωσης έχουν τη δυνατότητα να ελέγχουν τα σημεία ενδιαφέροντος (POI) και να συσχετίζουν την τοποθεσία με τις αναρτήσεις τους ή τα tweets τους, οδηγώντας στη δημιουργία Geo-Social Networks (GeoSNs). Υπάρχουν πολλά συστήματα που στοχεύουν στην αποτελεσματική αποθήκευση και ανάλυση καθαρών και κοινωνικά ενισχυμένων χωροχρονικών δεδομένων. Η σωστή αξιολόγηση αυτών των συστημάτων θα πρέπει να γίνει χρησιμοποιώντας πραγματικά δεδομένα από δημοφιλείς GeoSNs, όπως το Foursquare, το Facebook κ.λπ. Ωστόσο, οι περιορισμοί της ιδιωτικότητας απαγορεύουν την πρόσβαση σε τέτοια πραγματικά δεδομένα σε μεγάλη κλίμακα. Επομένως, οι αξιολογήσεις γίνονται με τη χρήση πραγματικών συνθετικών συνόλων δεδομένων που περιλαμβάνουν είτε χωρικά δεδομένα κειμένου (π.χ. tweets) είτε απλά χωρικά δεδομένα (π.χ. ίχνη GPS) τα οποία δεν είναι κοινωνικά ενισχυμένα. Σε αυτή την εργασία παρουσιάζουμε το Spaten, μία διαμορφώσιμη, ανοιχτού κώδικα γεννήτρια χωροχρονικών και κειμενικών δεδομένων, η οποία εξάγει ίχνη GPS από ρεαλιστικές διαδρομές χρησιμοποιώντας το Google Maps API, τις συνδυάζει με πραγματικά POI και σχετικά σχόλια χρηστών που ανιχνεύονται από το TripAdvisor και διαθέτει τα δεδομένα των αποτελεσμάτων για περαιτέρω ανάλυση. Η έγχυση κοινωνικών ιδιοτήτων που εξάγονται από υπάρχοντες γράφους Twitter στα παραγόμενα δεδομένα μαζί με περαιτέρω παραμετροποίηση οδηγεί σε ρεαλιστικά σύνολα δεδομένων GeoSN. Δημιουργούμε και προσφέρουμε δημόσια σύνολα δεδομένων μεγέθους GB με εκατομμύρια check-in και ίχνη GPS. Ως proof of concept, φορτώσαμε τα δεδομένα που δημιουργήσαμε σε ένα

σύστημα NoSQL και αξιολογήσαμε την δυνατότητα κλιμάκωσης του χρησιμοποιώντας ερωτήματα που απαντώνται συνήθως σε ιστότοπους κοινωνικής δικτύωσης. Ελπίζουμε ότι το Sraten μπορεί να προσφέρει στην ερευνητική κοινότητα τη δυνατότητα να παράγει ρεαλιστικά δεδομένα GeoSN σε μεγάλη κλίμακα, για να μπορεί να τα χρησιμοποιεί για πειραματική αποτίμηση.

- Σ20. K. Doka, I. Mytilinis, I. Giannakopoulos, I. Konstantinou, D. Tsitsigkos, M. Terrovitis and N. Koziris: **Exploiting Social Networking and Mobile Data for Crisis Detection and Management**. In Proceedings of the International Conference on Information Systems for Crisis Response and Management in Mediterranean Countries. Springer, Cham, 2017.

Κάθε μέρα, παράγονται και καταναλώνονται τεράστια ποσά δεδομένων κοινωνικής δικτύωσης με συνεχώς αυξανόμενο ρυθμό. Το ψηφιακό αποτύπωμα ενός χρήστη που προέρχεται από κοινωνικά δίκτυα ή κινητές συσκευές, όπως τα σχόλια, τα check-in και τα ίχνη GPS, περιέχει πολύτιμες πληροφορίες σχετικά με τη συμπεριφορά υπό κανονικές και επείγουσες συνθήκες. Η συλλογή και ανάλυση δεδομένων κινητής και κοινωνικής δικτύωσης πριν, κατά τη διάρκεια και μετά από μια καταστροφή ανοίγει νέες προοπτικές σε τομείς όπως ο εντοπισμός γεγονότων σε πραγματικό χρόνο, η διαχείριση κρίσεων και η εξατομίκευση και παρέχει πολύτιμες πληροφορίες για την έκταση της καταστροφής, του πληθυσμού και του ρυθμού ανάκαμψης. Τα παραδοσιακά συστήματα αποθήκευσης και επεξεργασίας δεν είναι σε θέση να αντεπεξέλθουν στο μέγεθος των συλλεγόμενων δεδομένων και στην πολυπλοκότητα της εφαρμοζόμενης ανάλυσης, έτσι χρησιμοποιούνται συνήθως καταναλωμένες προσεγγίσεις. Σε αυτή την εργασία προτείνουμε μια πλατφόρμα ανοικτού κώδικα που μπορεί να χρησιμεύσει ως βάση για εφαρμογές και υπηρεσίες που σχετίζονται με την ανίχνευση και τη διαχείριση κρίσεων, συνδυάζοντας δεδομένα χωρο-κειμένου (spatiotextual) που δημιουργούν οι χρήστες. Το σύστημα επικεντρώνεται στην επεκτασιμότητα και βασίζεται σε ένα συνδυασμό σύγχρονων συστημάτων Big Data. Υποστηρίζει τα πιο δημοφιλή κοινωνικά δίκτυα και είναι εύκολα επεκτάσιμο σε οποιαδήποτε κοινωνική πλατφόρμα. Η πειραματική αξιολόγηση του πρωτοτύπου μας επιβεβαιώνει την απόδοση και την επεκτασιμότητα του, ακόμα και κάτω από μεγάλο φορτίο, χρησιμοποιώντας διαφορετικούς τύπους ερωτήσεων σε διάφορα μεγέθη συστοιχιών.

- Σ21. K. Lolos, I. Konstantinou, V. Kantere and N. Koziris: **Rethinking Reinforcement Learning for Cloud Elasticity**. In Proceedings of ACM Symposium of Cloud Computing conference, Santa Clara, California USA, September 25-27 (SoCC'17)

Η ελαστικότητα του νέφους, δηλαδή η δυναμική κατανομή πόρων σε εφαρμογές που ανταποκρίνονται σε διακυμάνσεις των απαιτήσεων του φόρτου εργασίας, υπήρξε μία από τις μεγαλύτερες προκλήσεις στον τομέα των υπολογιστικών νεφών. Έχουν προταθεί προσεγγίσεις βασισμένες σε ενισχυτική μάθηση, αλλά απαιτούν μεγάλο αριθμό καταστάσεων προκειμένου να μοντελοποιήσουν μια σύνθετη συμπεριφορά μιας εφαρμογής. Σε αυτή την εργασία προτείνουμε μια νέα προσέγγιση ενισχυτικής μάθησης που χρησιμοποιεί προσαρμοστικό διαχωρισμό χώρων καταστάσεων. Η ιδέα είναι να ξεκινήσουμε από μια κατάσταση που αντιπροσωπεύει ολόκληρο το περιβάλλον και να το χωρίσουμε σε λεπτότερες καταστάσεις με προσαρμοστικότητα στο παρατηρούμενο φόρτο εργασίας και στη συμπεριφορά του συστήματος ακολουθώντας μια προσέγγιση δέντρων αποφάσεων. Εξετάζουμε νέα στατιστικά

κριτήρια και στρατηγικές που αποφασίζουν τόσο τις σωστές παραμέτρους όσο και τον κατάλληλο χρόνο εκτέλεσης της διαμέρισης.

- Σ22. K. Lolos, I. Konstantinou, V. Kantere and N. Koziris: **Adaptive State Space Partitioning of Markov Decision Processes for Elastic Resource Management**. In proceedings of the 2017 IEEE International Conference of Data Engineering, San Diego, CA, USA April 19-22 (ICDE 2017)

Τα σύγχρονα υπολογιστικά συστήματα μεγάλης κλίμακας αποτελούνται από πολύπλοκες εφαρμογές που τρέχουν πάνω από συστοιχίες υπολογιστών. Ένα σημαντικό ζήτημα σε αυτή την περίπτωση είναι η προσφορά της ελαστικότητας, δηλαδή, της δυναμικής κατανομής των πόρων σε εφαρμογές με σκοπό να ανταποκριθούν στις κυμαινόμενες απαιτήσεις φόρτου εργασίας. Προσεγγίσεις που βασίζονται σε κατώφλια τυπικά εφαρμόζονται, όμως η ρύθμιση των παραμέτρων είναι δύσκολο να γίνει και να βελτιστοποιηθεί. Προσεγγίσεις που βασίζονται στην τεχνική της ενισχυτικής μάθησης έχουν προταθεί, αλλά απαιτούν ένα μεγάλο αριθμό καταστάσεων προκειμένου να μοντελοποιήσουν μια σύνθετη συμπεριφορά μιας εφαρμογής. Μέθοδοι που προχωρούν σε προσαρμοστική κατάτμηση του χώρου καταστάσεων έχουν προταθεί, αλλά τα κριτήρια διαμέρισης και οι στρατηγικές είναι κατώτερες του βέλτιστου. Στην εργασία αυτή παρουσιάζουμε τον MDP_DT, έναν αλγόριθμο ενισχυτικής μάθησης πλήρους μοντέλου για την ελαστική διαχείριση πόρων που εφαρμόζει προσαρμοστική κατάτμηση του χώρου καταστάσεων. Προτείνουμε δύο καινούρια στατιστικά κριτήρια και τρεις στρατηγικές και αποδεικνύουμε πειραματικά ότι αποφασίζουν σωστά τόσο πού όσο και πότε να γίνει η κατάτμηση, ξεπερνώντας τις υπάρχουσες προσεγγίσεις. Αποτιμούμε πειραματικά τον MDP_DT σε μια πραγματική συστοιχία πάνω από μεταβλητά και όχι γνωστά εκ των προτέρων φορτία και δείχνουμε ότι παίρνει σωστές αποφάσεις σε σχέση με στατικές και model-free προσεγγίσεις, ενώ χρειάζεται ελάχιστο αριθμό δεδομένων εκπαίδευσης.

- Σ23. N. Chalcantzis, I. Konstantinou and N. Koziris : **BBQ: Elastic MapReduce over Cloud Platforms**. In proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, Madrid, Spain May 14-17 (CCGrid 2017).

Υπηρεσίες υπολογιστικών νεφών όπως η Amazon EMR επιτρέπουν στους χρήστες να έχουν πρόσβαση σε εξατομικευμένες συστοιχίες επεξεργασίας μεγάλου όγκου δεδομένων μέσα σε λίγα κλικ από τους φυλλομετρητές τους, χάρη στην ελαστική ιδιοκτησία των υπολογιστικών νεφών. Σε τέτοια περιβάλλοντα, η διαχείριση των πόρων είναι επιθυμητό να πραγματοποιηθεί κατά τρόπο που να βελτιστοποιεί την αξιοποίηση και κατά συνέπεια τη μεγιστοποίηση της αξίας των πόρων που αποκτήθηκαν. Καθώς οι υποδομές υπολογιστικών νεφών γίνονται όλο και πιο δημοφιλείς για την ανάλυση μεγάλου όγκου δεδομένων, η εκτέλεση των προγραμμάτων τηρώντας στόχους που έχει θέσει ο χρήστης, όπως προθεσμία ολοκλήρωσης εργασιών, εξακολουθεί να αποτελεί πρόκληση. Στην εργασία αυτή παρουσιάζουμε το BARBECUE (joB AwaRe Big-data Elasticity CloUd managEment System), ένα σύστημα που επιτρέπει σε μια εικονική συστοιχία Hadoop MapReduce να ρυθμίζεται αυτόματα ανάλογα με το μέγεθος του φόρτου εργασίας που απαιτείται για την εκτέλεση μεμονωμένων εργασιών τηρώντας τις προθεσμίες εκτέλεσης χωρίς να αποκτήσει περισσότερους πόρους από τα λιγότερο απαραίτητους. Η μονάδα αποφάσεων του BBQ

χρησιμοποιεί ένα μοντέλο απόδοσης για το *MapReduce*, το οποίο μπορεί να εκφράσει τους πόρους της συστοιχίας και τον χρόνο εκτέλεσης σε συνάρτηση με τον αριθμό των κόμβων. *BBQ* χρησιμοποιεί μια προσαρμογή του αλγόριθμου αναρρίχησης για τον υπολογισμό του βέλτιστου συνδυασμού αριθμού κόμβων και μεγέθους *reduce* κυμάτων δοθείσας μιας γνωστής εργασίας, του μεγέθους των δεδομένων εισόδου και μιας προθεσμίας εκτέλεσης. Οι συμμετέχοντες θα είναι σε θέση να παρακολουθήσουν το σύστημα εκτελεί αλλαγές του μεγέθους της συστοιχίας σε πραγματικό χρόνο προκειμένου να εκτελεστούν οι εργασίες που έχουν ανατεθεί.

- Σ24. Giannakopoulos, I. Konstantinou, D. Tsoumakos and N. Koziris : **AURA: Recovering from Transient Failures in Cloud Deployment**. In proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, Madrid, Spain May 14-17 (CCGrid 2017).

Σε αυτή την εργασία παρουσιάζουμε την *AURA*, ένα εργαλείο εγκατάστασης εφαρμογών για χρήση πάνω από παρόχους που τείνουν να παρουσιάζουν παροδικές αποτυχίες. Η πολυπλοκότητα των σύγχρονων υπολογιστικών νεφών προσδίδει μια επιρρεπή σε σφάλματα συμπεριφορά κατά τη διάρκεια της φάσης εγκατάστασης μιας εφαρμογής, κάτι που δυσχεραίνει τον αυτοματισμό και μεγεθύνει τα κόστη τόσο όσον αφορά το χρόνο και χρήματα. Για την αντιμετώπιση αυτής της πρόκλησης, προτείνουμε την *AURA*, ένα σύστημα που διαμορφώνει την εγκατάσταση μιας εφαρμογής σαν μια διάσχιση ενός ακυκλικού γράφου και επανεκτελεί τα τμήματα του γράφου που απέτυχαν. Η *AURA* επιτυγχάνει να εκτελέσει οποιοδήποτε *script* που ενημερώνει πόρους που αφορούν το σύστημα αρχείων με τρόπο *idempotent* μέσω της υιοθέτησης μιας πολυεπίπεδης τεχνικής συστήματος αρχείων. Στην παρακάτω επίδειξη, επιτρέπουμε στους χρήστες να περιγράψουν, να εγκαθιστούν και να παρακολουθούν τις εφαρμογές μέσα από ένα ολοκληρωμένο γραφικό περιβάλλον και επιδεικνύουμε την ικανότητα της *AURA* να ξεπερνά παροδικές αποτυχίες, ακόμη και στα πιο ασταθή περιβάλλοντα.

- Σ25. I. Giannakopoulos, I. Konstantinou, D. Tsoumakos and N. Koziris: **Recovering from Cloud Application Deployment Failures through Re-execution**. In proceedings of the 2nd International Workshop on Algorithmic Aspects of Cloud Computing (ALGO CLOUD 2016), in conjunction with the ALGO 2016 Conference, August 22, 2016, Aarhus, Denmark

Σε αυτή την εργασία μελετάμε το πρόβλημα της αυτοματοποιημένης ανάπτυξης και διαμόρφωσης εφαρμογών στο *cloud*. Οι παροδικές αποτυχίες που παρατηρούνται συνήθως σε υποδομές *cloud* αποδίδονται στην πολυπλοκότητα των στοιβών του λογισμικού και του υλικού που χρησιμοποιούνται. Αυτά τα σφάλματα επηρεάζουν την ανάπτυξη των *cloud* εφαρμογών, αναγκάζοντας τους χρήστες να ελέγχουν και να παρεμβαίνουν χειροκίνητα στην διαδικασία ανάπτυξης. Για την αντιμετώπιση αυτής της πρόκλησης, προτείνουμε μια απλή αλλά ισχυρή μεθοδολογία ανάπτυξης με την ικανότητα ανάκτησης σφάλματος που βασίζει τη λειτουργία της στην αναγνώριση των εξαρτήσεων μεταξύ των *scripts* και την εκ νέου εκτέλεση των απαραίτητων ενεργειών. Για να εξασφαλιστεί η έννοια του *idempotence*, έχουμε υιοθετήσει ένα μηχανισμό παραγωγής στιγμιότυπων συστήματος αρχείων που επιτρέπει στη μεθοδό μας να επανέλθει σε μια υγιή κατάσταση σε περίπτωση αποτυχίας εκτέλεσης του

σεναρίου. Η πειραματική ανάλυση μας δείχνει ότι η προσέγγισή μας μπορεί να επιλύσει οποιαδήποτε παροδική αποτυχία εγκατάστασης που εμφανίζεται κατά τη διάρκεια της φάσης ανάπτυξης, ακόμη και σε εξαιρετικά απρόβλεπτα περιβάλλοντα cloud.

- Σ26. Copil, G., Moldovan, D., Le, H. D., Truong, H., Dustdar, S., Sofokleous, C., Loulloudes, N., Trihinas, D., Pallis, G., Dikaiakos, M. D., Giannakopoulos, I., Papailiou, N., Konstantinou, I., Sheridan, C., Loverdos, C. K., Floros, E., Star, K., & Xing, W. **On Controlling Elasticity of Cloud Applications in CELAR**. In S. Bagchi (Ed.), *Emerging Research in Cloud Distributed Computing Systems* (pp. 222-252). Hershey, PA.

Οι σημερινές πολύπλοκες cloud εφαρμογές αποτελούνται από πολλά συστατικά που εκτελούνται σε περιβάλλοντα multi-cloud. Η δυνατότητα να διαχειρίζονται και να ελέγχουν το κόστος, την ποιότητά τους, και την ελαστικότητα των πόρων είναι υψίστης σημασίας για τέτοιες εφαρμογές. Ωστόσο, δεδομένου ότι το κόστος των διαφόρων υπηρεσιών που προσφέρονται από τους παρόχους cloud μπορεί να διαφέρει πολύ με την ποιότητα / απόδοση τους, οι ελεγκτές της ελαστικότητας πρέπει να λαμβάνουν υπόψη όχι μόνο σύνθετες, πολυδιάστατες προτιμήσεις και ικανότητες από τα ενδιαφερόμενα μέρη, αλλά και διάφορες πληροφορίες χρόνου εκτέλεσης σχετικά με τις εφαρμογές cloud και τα περιβάλλοντα εκτέλεσης. Σε αυτό το κεφάλαιο, οι συγγραφείς παρουσιάζουν την προσέγγιση ελέγχου ελαστικότητας του έργου CELAR της ΕΕ, η οποία ασχολείται με πολυδιάστατες απαιτήσεις ελαστικότητας και εξασφαλίζει τον έλεγχο της πολύ-επίπεδης ελαστικότητας για την εκπλήρωση των απαιτήσεων των χρηστών. Δείχνουν τους μηχανισμούς ελέγχου και ελαστικότητας του έργου CELAR, από την περιγραφή της εφαρμογής μέχρι τον έλεγχο της ελαστικότητας πολλαπλών επιπέδων. Οι συγγραφείς τονίζουν τη χρησιμότητα των μηχανισμών του CELAR για τους χρήστες, οι οποίοι μπορούν να χρησιμοποιήσουν μια διαισθητική, φιλική προς το χρήστη διεπαφή για να περιγράψουν και στη συνέχεια να ελέγξουν την συμπεριφορά της ελαστικότητας της εφαρμογής τους.

- Σ27. E. Kassela, I. Konstantinou and N. Koziris: **A Generic Architecture for Scalable and Highly Available Content Serving Applications in the Cloud** In proceedings of the 4th IEEE Symposium on Network Cloud Computing and Applications (NCCA 2015), Munich, Germany, June 11-12 2015.

Το cloud computing μοντέλο επιτρέπει στους παρόχους υπηρεσιών να προσφέρουν κλιμακωτές και υψηλής διαθεσιμότητας εφαρμογές στους τελικούς χρήστες τους. Τυπικές περιπτώσεις όπου αυτό απαιτείται είναι εφαρμογές που σερβίρουν περιεχόμενο, όπου ένας μεγάλος αριθμός των συνδεδεμένων χρηστών διαχειρίζεται μεγάλες ποσότητες δεδομένων. Στην εποχή του Big Data, όπου η ποσότητα των πληροφοριών που παράγονται και καταναλώνονται αυξάνεται εκθετικά, οι κεντροποιημένες εφαρμογές είναι αναποτελεσματικές, καθώς δεν μπορούν να αναβαθμιστούν επαρκώς, ανάλογα με τον αριθμό των συνδεδεμένων χρηστών ή με τα μεγέθη του σύνολου των δεδομένων. Σε αυτές τις περιπτώσεις, μια αποτελεσματική μεταφορά στο cloud των εφαρμογών απαιτείται, προκειμένου να επωφεληθούν από τις προσφορές του υπολογιστικού νέφους. Σε αυτή την εργασία, παρουσιάζουμε μια γενική αρχιτεκτονική που μπορεί να χρησιμοποιηθεί από σχεδόν οποιαδήποτε εφαρμογή που εξυπηρετεί περιεχόμενο προκειμένου να προσφέρει επεκτάσιμες διαθέσιμες λειτουργίες διαχείρισης δεδομένων για τους χρήστες της, με τη χρήση τεχνικών

διαχείρισης υπολογιστικών νεφών. Περιγράφουμε τα αρχιτεκτονικά στοιχεία της προσέγγισής μας, μαζί με το πώς μπορούν να χρησιμοποιηθούν αποτελεσματικά σε ένα περιβάλλον cloud. Τεκμηριώνουμε τις εμπειρίες μας με μια πραγματική ανάπτυξη μιας τυπικής εφαρμογής περιεχομένου πάνω από τον οkeanos, μια δημόσια υπηρεσία cloud. Περιγράφουμε τις βιβλιοθήκες ανοικτού κώδικα που έχουμε επιλέξει από μια πληθώρα των υφιστάμενων εργαλείων, δικαιολογούμε τις επιλογές μας και περιγράφουμε τις αρχικές μας παρατηρήσεις κατά τη διάρκεια της λειτουργίας τους. Τέλος, παρουσιάζουμε την αρχική αξιολόγηση των επιδόσεων μας, όπου επιδεικνύουμε την ικανότητα του συστήματος να χειριστεί τον αυξανόμενο φόρτο εργασίας με ελαστική κλιμάκωση των πόρων.

- Σ28. I. Mytilinis, I. Giannakopoulos, I. Konstantinou, K. Doka, D. Tsitsigkos, M. Terrovitis, L. Giampouras and N. Koziris: **MoDisSENSE: A Distributed Spatio-Temporal and Textual Processing Platform for Social Networking Services**. In Proceedings of the 2015 ACM SIGMOD/PODS International Conference on Management of Data (SIGMOD'15), Demo track, Melbourne, Victoria, Australia, 2015.

Η ποσότητα των δεδομένων κοινωνικής δικτύωσης που παράγεται και καταναλώνεται καθημερινά είναι τεράστια και αυξάνεται συνεχώς. Το ψηφιακό αποτύπωμα ενός χρήστη που προέρχονται από κοινωνικά δίκτυα ή κινητές συσκευές, όπως σχόλια και check-ins περιέχει πολύτιμες πληροφορίες σχετικά με τις προτιμήσεις του. Η συλλογή και η ανάλυση αυτών των αποτυπωμάτων χρησιμοποιώντας επίσης πληροφορίες για τους φίλους των χρηστών και τα ίχνη τους προσφέρουν πολλές ευκαιρίες σε τομείς όπως η εξατομικευμένη αναζήτηση, συστάσεις, κλπ. Όταν το μέγεθος των συλλεγομένων δεδομένων ή η πολυπλοκότητα των εφαρμοζόμενων μεθόδων αυξάνει, τα παραδοσιακά συστήματα αποθήκευσης και τα συστήματα επεξεργασίας δεν είναι αρκετά και τότε κατανομημένες προσεγγίσεις χρησιμοποιούνται. Σε αυτή την εργασία, παρουσιάζουμε το MoDisSENSE, μια κατανομημένη πλατφόρμα ανοικτού κώδικα που παρέχει εξατομικευμένη αναζήτηση για σημεία ενδιαφέροντος και trending γεγονότα με βάση το κοινωνικό γράφημα του χρήστη και με το συνδυασμό χωρο-κειμενικών δεδομένων που δημιουργούνται από το χρήστη. Το σύστημα έχει σχεδιαστεί με δυνατότητα κλιμάκωσης, είναι χτισμένο χρησιμοποιώντας ένα συνδυασμό από τα τελευταία εργαλεία διαχείρισης μεγάλων δεδομένων και η λειτουργικότητά του προσφέρεται μέσω εύκολων στη χρήση κινητών και web πελατών που υποστηρίζουν τα πιο δημοφιλή κοινωνικά δίκτυα. Δίνουμε μια επισκόπηση των αρχιτεκτονικών και τεχνολογιών και αξιολογούμε τις επιδόσεις και την επεκτασιμότητα του με τη χρήση διαφορετικών τύπων ερωτημάτων πάνω σε διάφορα μεγέθη συστοιχιών. Χρησιμοποιώντας το διαδίκτυο ή το κινητό τους, οι χρήστες έχουν τη δυνατότητα να εγγραφούν οι ίδιοι με τα δικά τους διαπιστευτήρια κοινωνικών δικτύων και να εκτελέσουν κοινωνικά ενισχυμένα ερωτήματα για σημεία ενδιαφέροντος, να περιηγηθούν στα αποτελέσματα και να εξερευνήσουν την αυτόματη λειτουργία δημιουργίας blog που εξάγεται από την ανάλυση των ιχνών GPS που έχουν ήδη συλλεχθεί.

- Σ29. A. Naskos, E. Stachtari, A. Gounaris, P. Katsaros, D. Tsoumakos, I. Konstantinou and S. Sioutas **Dependable Horizontal Scaling Based On Probabilistic Model Checking** In proceedings of the 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), Shenzhen, Guangdong, China, May 4-7, 2015 (**best paper award nomination**)

Το επίκεντρο αυτής της εργασίας είναι η on-demand παροχή πόρων στο cloud computing, η οποία συνήθως αναφέρεται ως η ελαστικότητα. Αν και πολλή προσπάθεια έχει επενδυθεί στην ανάπτυξη συστημάτων και μηχανισμών που επιτρέπουν την ελαστικότητα, οι πολιτικές απόφασης τείνουν να είναι σχεδιασμένες χωρίς ποσοτική εγγύηση της ποιότητας της λειτουργίας τους. Παρουσιάζουμε μια προσέγγιση για την ανάπτυξη πιο επίσημων και αξιόπιστων πολιτικών ελαστικότητας. Κάνουμε δύο διαφορετικές εισφορές. Πρώτον, προτείνουμε μια επεκτάσιμη προσέγγιση για την επιβολή της ελαστικότητας μέσω της δυναμικής υλοποίησης και απευθείας ποσοτικής επαλήθευσης της απόφασης Markov Διεργασιών (MDP) χρησιμοποιώντας πιθανολογικό έλεγχο μοντέλου (probabilistic model checking). Δεύτερον, μελετιούνται διάφορα συγκεκριμένα μοντέλα ελαστικότητας και πολιτικές ελαστικότητας. Αξιολογούμε τις πολιτικές αποφάσεων χρησιμοποιώντας ίχνη από μια πραγματική βάση δεδομένων NoSQL κάτω συνεχώς εξελισσόμενο εξωτερικό φορτίο. Σχολιάζουμε την συμπεριφορά των διαφόρων επιλογών πολιτικής μοντελοποίησης και δείχνουμε ότι η πρότασή μας μπορεί να βελτιώσει το state-of-the-art μειώνοντας σημαντικά το underprovisioning αποφεύγοντας overprovisioning.

- Σ30. E. Kassela, C. Boumpouka, I. Konstantinou and N. Koziris: **Automated Workload-aware Elasticity of NoSQL Clusters in the Cloud**. In proceedings of the 2014 IEEE International Conference on Big Data (BigData 2014), Washington DC, USA, October 27-30 2014.

Η χρήση των υπηρεσιών υπολογιστικών νεφών έχει αποκτήσει ακραία δημοτικότητα. Μέσα από τις πλατφόρμες νέφους που παρέχουν υπηρεσίες (IaaS), οι χρήστες μπορούν να διαχειριστούν ελαστικά την παροχή πόρων επιτρέποντας την αυτοματοποιημένη επιτάχυνση εφαρμογών. Συνήθως, η κλιμάκωση πραγματοποιείται είτε χειροκίνητα είτε μέσω μιας υπηρεσίας που ενοποιεί δυναμικά πόρους που βασίζεται σε μια προκαθορισμένη πολιτική. Ωστόσο, αυτές οι πολιτικές είναι απλοϊκές, είναι βασισμένες στην χρήση απλών καταφυγίων και μπορεί να μην είναι σε θέση να συλλάβουν συγκεκριμένες συμπεριφορές εφαρμογών σύμφωνα με τις παραμέτρους διαμόρφωσης και τον φόρτο εργασίας. Στην εργασία αυτή, επεκτείνουμε τον TIRAMOLA, ένα σύστημα που επιτρέπει την αυτοματοποιημένη αλλαγή μεγέθους NoSQL συστοιχιών, προκειμένου να αναγνωρίσει διαφορετικούς τύπους φόρτου εργασίας και να εφαρμόσει την πιο ενεργητική δράση κλιμάκωσης ανάλογα με τις πολιτικές που ορίζει ο χρήστης. Εκτελούμε μια ενδελεχή ανάλυση του πώς οι διαφορετικοί τύποι ερωτημάτων επιλύονται από σύγχρονα συστήματα NoSQL και αξιολογούμε την απόδοση μιας συστοιχίας NoSQL ποικίλου μεγέθους, χρησιμοποιώντας φόρτο εργασίας μικτού τύπου και μεγέθους. Χρησιμοποιούμε αυτή τη γνώση για να τελειοποιήσουμε τις πολιτικές του εκτεταμένου TIRAMOLA, προκειμένου να λάβει ακριβείς αποφάσεις κλιμάκωσης. Πραγματοποιούμε μια εκτενή πειραματική αξιολόγηση του εκτεταμένου TIRAMOLA σε σχέση με την αρχική του έκδοση σε μια συστοιχία HBase και η ανάλυσή μας επιβεβαιώνει ότι ο πώτος μπορεί να λειτουργήσει με επιτυχία σε οποιοδήποτε περιβάλλον, με την ανάλογη συμπεριφορά σε οποιοδήποτε φορτίο εισόδου.

- Σ31. I. Mytilinis, I. Giannakopoulos, I. Konstantinou, K. Doka and N. Koziris: **MoDisSENSE: A Distributed Platform for Social Networking Services over Mobile Devices**. In proceedings of the 2014 IEEE International Conference on Big Data (BigData 2014), Washington DC, USA, October 27-30 2014.

Σε αυτή την εργασία παρουσιάζουμε το MoDisSENSE, μια κατακεντρωμένη πλατφόρμα αναλυτικής επεξεργασίας δεδομένων για υπηρεσίες κοινωνικής δικτύωσης μέσω κινητών συσκευών. Το MoDisSENSE συλλέγει και αποθηκεύει δεδομένα από ετερογενείς πηγές, όπως GPS ίχνη από κινητά τηλέφωνα, πληροφορίες για το προφίλ του χρήστη και σχόλια από κοινωνικά δίκτυα που συνδέονται με την πλατφόρμα. Αυτά συνδυάζονται με ανάλυση σε επίπεδο χωροχρονικών και δεδομένων κειμένου, που διεξάγεται με κατακεντρωμένο τρόπο, προκειμένου να αντληθεί επιπλέον γνώση και να γίνονται έξυπνες προτάσεις με σκοπό την βελτίωση της εμπειρίας του χρήστη. Το σύστημα αποθήκευσης δεδομένων ακολουθεί μια υβριδική προσέγγιση για να χειριστεί τόσο ακατέργαστο και επεξεργασμένο δεδομένα, καλύπτοντας ταυτόχρονα την ανάγκη για επεκτασιμότητα και γρήγορη επεξεργασία επερωτήσεων. Έτσι, η πλατφόρμα είναι σε θέση να επιλύσει πολύπλοκα, πολύ-παραμετρικά, κοινωνικά φορτισμένα ερωτήματα της τάξης των χιλιοστών του δευτερολέπτου πάνω από σημεία ενδιαφέροντος, ακόμη και κάτω από βαρύ φορτίο.

- Σ32. I. Giannakopoulos, N. Papailiou, C. Mantas, I. Konstantinou, D. Tsoumakos and N. Koziris: **CELAR: Automated Application Elasticity Platform**. In proceedings of the 2014 IEEE International Conference on Big Data (BigData 2014), Washington DC, USA, October 27-30 2014.

Μια από τις κύριες υποσχέσεις του μοντέλου των υπολογιστικών νεφών είναι η δυνατότητα της αυξομείωσης των πόρων σύμφωνα με την χρήση. Αυτή η ιδιότητα χαρακτηρίζει την εποχή των υπολογιστικών νεφών, όπου η επιβάρυνση των πρώτων δαπανών για υποδομές αποβάλλεται. Καινοτόμες υπηρεσίες είναι έτσι σε θέση να εισέλθουν στην αγορά πιο γρήγορα και να προσαρμοστούν πιο γρήγορα στις νέες προκλήσεις και τη ζήτηση των χρηστών. Μία από τις κύριες όψεις αυτού του είδους της προσαρμοστικότητας είναι η έννοια της «ελαστικότητας», δηλαδή, η ικανότητα της αυτόνομης προσαφαίρεσης πόρων αντιδρώντας στις αλλαγές του εισερχόμενου φορτίου. Μια ελαστική υπηρεσία είναι σε θέση να λειτουργεί με βέλτιστο κόστος με την επέκταση και την συμπίεση των πόρων της κατά το χρόνο εκτέλεσης και ανάλογα με τη ζήτηση. Με αυτό τον τρόπο δεν ελαχιστοποιείται μόνο το κόστος λειτουργίας, αλλά αποφεύγονται επίσης διακοπές λόγω περιόδων αιχμών στην χρήση των υπηρεσιών. Παρόλο που οι επιμέρους στρώσεις που αποτελούν μια υπηρεσία υπολογιστικού νέφους μπορούν να κλιμακωθούν αυτόνομα, αυτό δεν συμβαίνει με ένα ενιαίο τρόπο. Το όραμα του CELAR είναι να παρέχει μια πλήρως ολοκληρωμένη στοίβα λογισμικού που διαχειρίζεται την κατανομή των πόρων για τις εφαρμογές cloud με έναν αυτόνομο, αποτελεσματικό και γενικό τρόπο. Προκειμένου να επιτευχθεί αυτό, το CELAR ενσωματώνει νέες μεθοδολογίες για την περιγραφή των εφαρμογών υπολογιστικού νέφους, την παρακολούθηση της χρήσης των διαφόρων πόρων, την αξιολόγηση του κόστους, τη λήψη ενημερωμένων αποφάσεων και την αλληλεπίδραση με την υποκείμενη υποδομή υπολογιστικού νέφους. Ο στόχος μας είναι διττός: Από τη μία πλευρά η ανάπτυξη μεθοδολογιών για την επίτευξη πολλαπλού επιπέδου, αυτόματο έλεγχο ελαστικότητας τόσο σε επίπεδο εφαρμογής όσο και υποδομών. Από την άλλη πλευρά, είναι η ανάπτυξη των εργαλείων ανοιχτού κώδικα που εφαρμόζουν αυτές τις μεθόδους με ολοκληρωμένο τρόπο. Σας παραθέτουμε μια επισκόπηση της πλατφόρμας CELAR, εξηγώντας τα αρχιτεκτονικά στοιχεία και κάποιες βασικές ροές εργασίας που δείχνουν πώς αλληλεπιδρούν για την επίτευξη των βασικών λειτουργιών.

- Σ33. N. Papailiou, D. Tsoumakos, I. Konstantinou, P.Karras and N. Koziris: **H2RDF+: An Efficient Data Management System for Big RDF Graphs**. In Proceedings of the 2014 ACM SIGMOD/PODS International Conference on Management of Data (Demo Track), Snowbird, Utah, USA.

Ο πολλαπλασιασμός των δεδομένων σε μορφή RDF έχει ως αποτέλεσμα την εμφάνιση μιας πληθώρας εξειδικευμένων συστημάτων διαχείρισης. Ενώ η ικανότητα να προσαρμόζονται στην περιπλοκότητα ενός ερωτήματος SPARQL - δεδομένου των εγγενών ποικιλομορφιών τους - είναι ζωτικής σημασίας, οι τρέχουσες προσεγγίσεις δεν κλιμακώνονται καλά όταν έρχονται αντιμέτωπες με σημαντικά πολύπλοκα, μη-επιλεκτικά ερωτήματα συνένωσης, με αποτέλεσμα την εκθετική αύξηση των χρόνων εκτέλεσης. Σε αυτή την επίδειξη παρουσιάζουμε το H2RDF+, ένα σύστημα RDF που εκτελεί αποτελεσματικά καταναμημένα ερωτήματα συνένωση Merge και Sort-Merge χρησιμοποιώντας πολλαπλούς δείκτες πάνω από HBase ευρετήρια. Μέσω ενός «άπληστου» σχεδιασμού που ενσωματώνει το μοντέλο υπολογισμού του κόστους, επιλέγει προσαρμοστικά είτε για απλή ή για καταναμημένη εκτέλεση επερωτήσεων με βάση την πολυπλοκότητά τους. Σε αυτή την εργασία, παρουσιάζουμε τις βασικές επιστημονικές συνεισφορές μας και επιτρέπουμε στους συμμετέχοντες να αλληλεπιδράσουν με το H2RDF+ πάνω από μια υποδομή Cloud. Χρησιμοποιώντας ένα web-based GUI, επιτρέπεται στους χρήστες να φορτώσουν διαφορετικά σύνολα δεδομένων (πραγματικά και συνθετικά), και να εφαρμόσουν ερωτήματα (προσαρμοσμένο ή προαποφασισμένα) και να παρακολουθήσουν την εκτέλεσή τους. Επιτρέποντας τον έλεγχο σε πραγματικό χρόνο της κατάστασης της συστοιχίας, τους χρόνους απόκρισης και πόρους που χρησιμοποιήθηκαν, το κοινό θα αξιολογήσει την εγκυρότητα των ισχυρισμών H2RDF+ 's και θα εκτελέσει απευθείας συγκρίσεις με δύο άλλες υλοποιήσεις RDF συστημάτων.

- Σ34. N. Papailiou, D. Tsoumakos, I. Konstantinou, P.Karras and N. Koziris: **Scalable Indexing and Adaptive Querying of RDF Data in the cloud**. In Proceedings of the 6th International Workshop on Semantic Web Information Management - SWIM 2014 (in conjunction with the 2014 ACM International Conference on Management of Data), Snowbird, Utah, USA.

Αποτελεσματικά συστήματα διαχείρισης δεδομένων RDF είναι κεντρικής σημασίας για την πραγματοποίηση του οράματος του Σημασιολογικού Ιστού. Η τεράστια αύξηση τόσο των δεδομένων των χρηστών όσο και των δεδομένων που παράγονται από μηχανές υπαγορεύει για κλιμακούμενες λύσεις σε συστήματα triple store. Τα σύγχρονα συστήματα διαχειρίζονται την αποκέντρωση ορισμένων ή όλων των σταδίων της διαχείρισης των δεδομένων RDF, κλιμακώνοντας σε μεγάλο αριθμό δεδομένων. Ωστόσο, τα συστήματα αυτά έχουν πολύ μικρή ευελιξία στην προσαρμογή της συμπεριφοράς τους σε σχέση με την απάντηση των ερωτημάτων. Ερωτήματα πάνω από triple data περιλαμβάνουν πολλαπλές συνενώσεις με διάφορους βαθμούς επιλεκτικότητας και του κόστους. Σε πολλές περιπτώσεις, η επίλυση ενός ερωτήματος συνένωσης σε ένα κεντρικό κόμβο είναι πιο συμφέρουσα. Επομένως, σωστή σχεδίαση του πλάνου ενός ερωτήματος και επιλεκτική εκτέλεση αυτού είναι απαραίτητη τη βέλτιστη απόδοση σε τόσο επιλεκτικά και μη-επιλεκτικά ερωτήματα. Προς αυτή την κατεύθυνση, θα περιγράψουμε το H2RDF+, ένα σύστημα RDF που εκτελεί αποτελεσματικά ερωτήματα συνένωσης πάνω από ένα πολλαπλό σύστημα δεικτοδότησης. Το H2RDF+

υλοποιεί 6 δείκτες RDF και λεπτομερή στατιστικά στοιχεία με τη χρήση της HBase. Στην εργασία αυτή, δίνουμε έμφαση στην πρωτότυπη, επεκτάσιμη και αποδοτική διαδικασία ευρετηρίασης MapReduce που επιτρέπει το H2RDF+ για να χειριστεί αυθαίρετα μεγάλα σύνολα δεδομένων RDF. Μια επιθετική συμπίεση σε επίπεδο byte χρησιμοποιείται για να μειώσει τις απαιτήσεις σε χώρο αποθήκευσης του συστήματος. Το H2RDF+ μπορεί επίσης να επεξεργάζεται προσαρμοστικά τόσο πολύπλοκα όσο και επιλεκτικά ερωτήματα προσαρμοστικά επιλέγοντας το ποσό των πόρων που διατίθενται για κάθε ερώτημα, με βάση την εκτίμηση της πολυπλοκότητας μέσω στατιστικών δεικτών.

- Σ35. N. Papailiou, I. Konstantinou, D. Tsoumakos, P.Karras and N. Koziris: **H2RDF+: High-performance Distributed Joins over Large-scale RDF Graphs**. In proceedings of the 2013 IEEE International Conference on Big Data (BigData 2013), Santa Clara, CA, USA, October 6-9 2013

Ο πολλαπλασιασμός των δεδομένων σε μορφή RDF απαιτεί αποτελεσματικές και κλιμακούμενες λύσεις για τη διαχείρισή τους. Ενώ η κλιμακωσιμότητα στην εποχή των μεγάλων δεδομένων είναι μια σκληρή απαίτηση, τα σύγχρονα συστήματα αποτυγχάνουν να προσαρμοστούν με βάση την πολυπλοκότητα των ερωτημάτων. Οι τρέχουσες προσεγγίσεις δεν κλιμακώνονται καλά όταν αντιμετωπίζουν ουσιαστικά περίπλοκα, μη-επιλεκτικά ερωτήματα συνένωσης, με αποτέλεσμα την εκθετική αύξηση των χρόνων εκτέλεσης. Σε αυτή την εργασία παρουσιάζουμε το H2RDF+, ένα σύστημα RDF που εκτελεί αποτελεσματικά καταναμημένα ερωτήματα συνένωσης Merge και Sort-Merge πάνω από ένα πολλαπλό σύστημα δεικτοδότησης. Το H2RDF+ είναι εξαιρετικά κλιμακώσιμο, χρησιμοποιώντας καταναμημένη επεξεργασία βασισμένη σε MapReduce και σε ευρετήρια HBase. Αξιοποιώντας επιθετική συμπίεση σε επίπεδο ψηφίου και ομαδοποίηση αποτελεσμάτων με γρήγορες σαρώσεις, μπορεί να επεξεργαστεί τόσο πολύπλοκα και επιλεκτικά ερωτήματα συνένωσης με ένα άκρως αποτελεσματικό τρόπο. Επιπλέον, επιλέγει προσαρμοστικά είτε κεντροποιημένη είτε καταναμημένη εκτέλεση με βάση την πολυπλοκότητα του ερωτήματος μέσω των στατιστικών δεικτών. Μια εκτεταμένη αξιολόγηση μας δείχνει ότι το H2RDF+ απαντά αποτελεσματικά μη-επιλεκτικά ερωτήματα συνένωσης τάξη μεγέθους ταχύτερα σε σχέση με state-of-the-art υπάρχοντα συστήματα, ενώ είναι μόλις δέκατα του δευτερολέπτου πιο αργό σε απλά ερωτήματα, κλιμακώνοντας γραμμικά με το ποσό των διαθέσιμων πόρων.

- Σ36.I. Konstantinou, D. Tsoumakos, I. Mytilinis and N. Koziris: **DBalancer: Distributed Load Balancing for NoSQL Data-stores**. In proceedings of the 2013 ACM SIGMOD/PODS International Conference on Management of Data, New York, USA, June 22-27 2013.

Απρόβλεπτες αιχμές φορτίου ή μη ομοιόμορφη πρόσβαση στα δεδομένα μπορεί να οδηγήσει σε σοβαρή υποβάθμιση των επιδόσεων σε εφαρμογές δεδομένων, ένα τυπικό πρόβλημα που παρουσιάζεται στις καταναμημένες βάσεις NoSQL. Σε αυτές τις περιπτώσεις, η εξισορρόπηση φορτίου είναι μία αναγκαία ενέργεια. Σε αυτήν την επίδειξη, παρουσιάζουμε τον DBalancer, ένα καταναμημένο πρόγραμμα που μπορεί να εγκατασταθεί πάνω από μια τυπική NoSQL βάση και να προσφέρει αποτελεσματική και διαμορφώσιμη εξισορρόπηση φορτίου. Η εξισορρόπηση γίνεται με απλή ανταλλαγή μηνυμάτων και με τις τυπικές λειτουργίες μετακίνησης δεδομένων που υποστηρίζονται από τα περισσότερα σύγχρονα δεδομένα NoSQL

συστήματα. Παρουσιάζουμε την αρχιτεκτονική του συστήματος, περιγράφουμε λεπτομερώς τις ενότητες του και την αλληλεπίδρασή τους και υλοποιούμε μια σειρά από διαφορετικούς αλγόριθμους πάνω από αυτό. Μέσω ενός διαδικτυακού διαδραστικού GUI επιτρέπουμε στους χρήστες να ξεκινήσουν NoSQL βάσεις διαφόρων μεγεθών, να εφαρμόσουν πολλές διαφορετικά είδη φορτίων και να συγκρίνουν τους υλοποιημένους αλγόριθμους εξισορρόπησης φορτίου. Βίντεο και γραφήματα που καταδεικνύουν την επίδραση κάθε αλγόριθμου σε μια σειρά από ενδεικτικές μετρήσεις απόδοσης και κόστους θα δημιουργηθεί για κάθε εγκατάσταση. Παρατηρώντας τα αποτελέσματα της κάθε εκτέλεσης οι χρήστες θα είναι σε θέση να κατανοήσουν τους μηχανισμούς εξισορρόπησης του κάθε αλγόριθμου και την επίδραση στην απόδοση.

- Σ37. I. Konstantinou, V. Kantere, D. Tsoumakos, and N. Koziris: **COCCUS: Self-Configured Cost-Based Query Services in the Cloud**. In proceedings of the 2013 ACM SIGMOD/PODS International Conference on Management of Data, New York, USA, June 22-27 2013.

Πρόσφατα, ένας μεγάλος αριθμός pay-as-you-go υπηρεσιών δεδομένων παρέχονται μέσω υποδομών cloud. Οι πάροχοι υπηρεσιών δεδομένων χρειάζονται κατάλληλους και ευέλικτους μηχανισμούς χρέωσης και βελτιστοποίησης ερωτημάτων που να λαμβάνουν υπόψη το κόστος λειτουργίας στο cloud, στρατηγικές τιμολόγησης και προτιμήσεις των χρηστών. Ωστόσο, οι υπάρχουσες λύσεις είναι στατικές και μη ρυθμιζόμενες. Σε αυτή την δουλειά παρουσιάζουμε το COCCUS, ένα σπονδυλωτό σύστημα για εκτέλεση ερωτημάτων λαμβάνοντας υπόψη το κόστος, με προσαρμοστική χρέωση των ερωτημάτων και βελτιστοποίηση των υπηρεσιών δεδομένων cloud. Το κοινό μπορεί να ρυθμίσει τα ερωτήματά του, μαζί με τις προτιμήσεις της εκτέλεσης και τους περιορισμούς του προϋπολογισμού του, και το COCCUS προσαρμοστικά καθορίζει την χρέωση ερωτημάτων και διαχειρίζεται δευτερογενείς δομές δεδομένων, σύμφωνα με διάφορες οικονομικές πολιτικές. Παρουσιάζουμε την λειτουργία του COCCUS πάνω από κεντρικές και shared nothing βάσεις δεδομένων πάνω από δημόσια και ιδιωτικά cloud. Το κοινό έχει τη δυνατότητα να καθορίσει τις οικονομικές πολιτικές και να εκτελέσει διάφορα φορτία μέσα από ένα ολοκληρωμένο γραφικό περιβάλλον. Η προσαρμοστικότητα του COCCUS παρουσιάζεται χρησιμοποιώντας σε πραγματικό χρόνο γραφικές παραστάσεις που απεικονίζουν μια σειρά από βασικές μετρήσεις απόδοσης.

- Σ38.D. Tsoumakos, I. Konstantinou, C. Boumpouka, S. Sioutas and N. Koziris: **Automated, Elastic Resource Provisioning for NoSQL Clusters Using TIRAMOLA**. In proceedings of the 13th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), Delft, The Netherlands, May 13-16, 2013 (**best paper award**).

Η εργασία αυτή παρουσιάζει τον TIRAMOLA, ένα cloud-enabled σύστημα ανοιχτού λογισμικού για την εκτέλεση αυτόματων αλλαγών του μεγέθους NoSQL συστοιχιών σύμφωνα με πολιτικές οριζόμενες από τον χρήστη. Οι αποφάσεις σχετικά με την προσθήκη ή την αφαίρεση εικονικών μηχανών από μια συστοιχία μοντελοποιούνται ως μια διαδικασία λήψης αποφάσεων Markov και λαμβάνονται σε πραγματικό χρόνο. Το σύστημα αποφασίζει αυτόματα την συμφέρον μέγεθος συστοιχίας, σύμφωνα με καθορισμένες από το χρήστη πολιτικές. Στην συνέχεια δεσμεύει ή αποδεσμεύει εικονικές μηχανές από τον πάροχο του cloud και αναλαμβάνει τον συντονισμό τους σε μια NoSQL συστοιχία. Η αρθρωτή

αρχιτεκτονική του TIRAMOLA σε συνδυασμό με την υποστήριξη API επιτρέπει την αλληλεπίδραση με τις πιο πρόσφατες πλατφόρμες IaaS και την αύξηση της προσαρμογής. Μια εκτενής πειραματική αξιολόγηση σε μια συστοιχία HBase επιβεβαιώνει τους ισχυρισμούς μας: Το σύστημα αλλάζει το μέγεθος του cluster σε πραγματικό χρόνο και προσαρμόζει την απόδοσή του μέσα από διαφορετικές στρατηγικές βελτιστοποίησης, διαφορετικές επιτρεπτές ενέργειες, διαφορετικά φορτία εισόδου και εκπαίδευσης. Εκτός από την αυτοματοποίηση της διαδικασίας, παρουσιάζει χαρακτηριστικά αυτόματης μάθησης τα οποία επιτρέπουν να λαμβάνει πολύ κοντά σε βέλτιστες αποφάσεις ακόμα και με φορτία εισόδου 130% μεγαλύτερα ή εναλλασσόμενα 10 φορές γρηγορότερα σε σύγκριση με την αποθηκευμένη γνώση.

Σ39.I. Konstantinou and E. Floros and N. Koziris **Public vs Private Cloud Usage Costs: The StratusLab Case** In proceedings of the 2nd International Workshop on Cloud Computing Platforms (CloudCP 2012 - In conjunction with EuroSys), Bern, Switzerland, 2012

Τα υπολογιστικά νέφη ισχυρίζονται ότι προσφέρουν πολλαπλά πλεονεκτήματα σε σύγκριση με παραδοσιακές υποδομές πληροφορικής. Αυτά περιλαμβάνουν μεταξύ άλλων: καλύτερη ενεργειακή απόδοση, μείωση των συνολικών δαπανών διαχείρισης, καλύτερη αξιοποίηση των υλικών πόρων με την φιλοξενία πολλαπλών υπηρεσιών, κλπ. Η παρούσα εργασία αναφέρει τις εμπειρίες που αποκτήθηκαν κατά την παροχή μιας IaaS υπηρεσίας νέφους στα πλαίσια του ευρωπαϊκού έργου StratusLab και παρέχει ποσοτική και οικονομική ανάλυση του συνολικού κόστους της ιδιοκτησίας της υποδομής υπολογίζοντας το κόστος υποδομής, συντήρησης και λειτουργίας. Η ανάλυση συνεχίζεται με τη σύγκριση του κόστους της ιδιωτικής υπηρεσίας νέφους σε σχέση με την δημόσια δημοφιλή υπηρεσία νέφους EC2 της Amazon με την αξιοποίηση των συλλεγόμενων ιχνών χρήσης του StratusLab για μια περίοδο ενός έτους. Δείξαμε ότι με μέσο όρο χρήσης 70%, μια μικρή συστοιχία 20 μηχανών θα μπορούσε να αποσβέσει το συνολικό κόστος απόκτησης της υποδομής σε μια περίοδο 2-3 ετών, ενώ θα μπορούσε να προσφέρει τις ίδιες τιμές σε σύγκριση με το νέφος της Amazon εάν προσφέρονταν σε βάση pay-as-you-go.

Σ40.E. Angelou and N. Papailiou and I. Konstantinou and D. Tsoumakos and N. Koziris **Automatic Scaling of Selective SPARQL Joins Using the TIRAMOLA System** In proceedings of the 4th International Workshop on Semantic Web Information Management (SWIM 2012 - In conjunction with SIGMOD), Scottsdale, Arizona, USA, 20 May 2012

Οι σύγχρονες υποδομές υπολογιστικών νεφών που βασίζονται σε εικονικές μηχανές προσφέρουν νέες ευκαιρίες και προκλήσεις για τους προγραμματιστές και διαχειριστές συστημάτων. Από τις πιο αξιοσημείωτες ευκαιρίες είναι η υπόσχεση της ελαστικότητας των πόρων, κατά την οποία η υποδομή μπορεί να αυξομειώσει το μέγεθός της ανάλογα με τη ζήτηση. Αξιοποιώντας ελαστικούς πόρους, οι εφαρμογές μπορούν να παρέχουν καλύτερη ποιότητα των υπηρεσιών και να μειώνουν το κόστος τους πληρώνοντας μόνο για το απαιτούμενο ποσό των πόρων. Σε αυτή την εργασία, μελετάμε εκτενώς την απόδοση ορισμένων δημοφιλών NoSQL βάσεων δεδομένων πάνω από μια ελαστική υποδομή cloud.

Οι NoSQL βάσεις δεδομένων επικεντρώνονται σε αναλυτική επεξεργασία μεγάλων συνόλων δεδομένων κλίμακας, προσφέροντας αυξημένη επεκτασιμότητα πάνω από απλές υπολογιστικές μονάδες. Στη συνέχεια προχωρούμε στην περιγραφή του TIRAMOLA, ενός σπονδυλωτού πλαισίου βασισμένου σε τεχνολογίες υπολογιστικών νεφών, για την παρακολούθηση και την προσαρμοστική αλλαγή του μεγέθους οποιονδήποτε NoSQL συστοιχιών. Το σύστημά μας διαχειρίζεται τους πόρους συστοιχίας εικονικών μηχανών σύμφωνα με την εφαρμογή συγκεκριμένων περιορισμών που θέτει ο χρήστης ή ο διαχειριστής, μέσω μιας επεκτάσιμης υπομονάδας παρακολούθησης και εκτέλεσης εντολών. Οι χρήστες μπορούν εύκολα να τροποποιούν τις πολιτικές αλλαγής μεγέθους με βάση συγκεκριμένα μετρικά και έτσι να εκμεταλλευτούν πλήρως την ελαστικότητα της υποκείμενης υποδομής. Σαν παράδειγμα ρεαλιστικής χρήσης, εφαρμόζουμε το πλαίσιο πάνω από ένα πλήρως καταναμημένο κατάστημα RDF που υποστηρίζεται από μια ελαστική NoSQL βάση δεδομένων. Αφήνοντας τον TIRAMOLA να διαχειριστεί τον αριθμό των δεσμευμένων πόρων έχει σαν αποτέλεσμα την αυτοματοποιημένη εκτέλεση ενεργειών αλλαγής μεγέθους της συστοιχίας και τη μεγιστοποίηση της απόδοσης, ενώ οι διαχειριστές χρειάζεται να παρέχουν απλούς κανόνες για την ελαστικότητα.

Σ41.Ι. Konstantinou and E. Angelou and C. Boumpouka and D. Tsoumakos and N. Koziris and S. Sioutas **TIRAMOLA: Elastic NoSQL Provisioning through A Cloud Management Platform** In proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD demo track), Scottsdale, Arizona, USA, 2012

Οι βάσεις δεδομένων NoSQL επικεντρώνονται στην αναλυτική επεξεργασία μεγάλης κλίμακας δεδομένων, προσφέροντας αυξημένη δυνατότητα κλιμάκωσης πάνω από συστοιχίες απλών υπολογιστών. Ένα από τα ισχυρότερα χαρακτηριστικά τους είναι η ελαστικότητα, η οποία επιτρέπει υψηλή ποιότητα απόδοσης. Ωστόσο, η διαδικασία της προσαρμοστικής διαστολής και συστολής των πόρων περιλαμβάνει συνήθως πολλή χειρωνακτική προσπάθεια και συχνά απαιτεί τον καθορισμό των προϋποθέσεων για αυτό να παρέχεται από τους χρήστες. Μέχρι σήμερα, δεν υπάρχει σύστημα ανοιχτού κώδικα για την αυτόματη μεταβολή του μεγέθους των συστοιχιών NoSQL. Σε αυτή την επίδειξη, παρουσιάζουμε τον TIRAMOLA, ένα σπονδυλωτό πλαίσιο βασισμένο σε τεχνολογίες υπολογιστικών νεφών, για την παρακολούθηση και την προσαρμοστική αλλαγή του μεγέθους NoSQL συστοιχιών. Το σύστημα περιλαμβάνει μια μονάδα λήψης αποφάσεων που υπολογίζει τις βέλτιστες δράσεις αλλαγής μεγέθους της συστοιχίας, ώστε να μεγιστοποιείται οποιαδήποτε μετρήσιμη συνάρτηση ανταμοιβής με στόχο την δια βίου προσαρμογή σε αλλαγές του φόρτου εργασίας ή της υποδομής. Το κοινό θα είναι σε θέση να ξεκινήσει συστοιχίες HBase διαφόρων μεγεθών και εφαρμόσει μεταβαλλόμενο φόρτο εργασίας μέσω πολλαπλών πελατών YCSB. Οι συμμετέχοντες θα είναι σε θέση να παρακολουθήσουν, σε πραγματικό χρόνο, το σύστημα να εκτελεί αυτόματα προσθήκες και αφαιρέσεις εικονικών μηχανών καθώς και την μεταβολή μετρικών επίδοσης της συστοιχίας σε σχέση με την ρύθμιση παραμέτρων βελτιστοποίησης της επιλογής τους.

Σ42.Ν. Papailiou and I. Konstantinou and D. Tsoumakos and N. Koziris **H2RDF: Adaptive Query Processing on RDF Data in the Cloud** In Proceedings of the 21th International Conference on World Wide Web (WWW demo track), Lyon, France, 2012

Στην εργασία αυτή παρουσιάζουμε το σύστημα H2RDF, ένα πλήρως καταναμημένο σύστημα αποθήκευσης RDF που συνδυάζει το πλαίσιο επεξεργασίας MapReduce μαζί με καταναμημένες βάσεις δεδομένων NoSQL. Το σύστημα διαθέτει δύο μοναδικά χαρακτηριστικά που επιτρέπουν την αποτελεσματική επεξεργασία τόσο των απλών όσο και των multi-join SPARQL ερωτημάτων σε απεριόριστο αριθμό τριπλέτων RDF: αλγόριθμοι εκτέλεσης ερωτημάτων σύζευξης που λαμβάνουν υπόψη την επιλεκτικότητα των ερωτημάτων με σκοπό την μείωση του χρόνου επεξεργασίας και αλγόριθμους προσαρμοστικής επιλογής μεταξύ κεντροποιημένης και καταναμημένης εκτέλεσης (με βάση το MapReduce) ερωτημάτων ανάλογα με το μέγεθος των προς επεξεργασία δεδομένων. Το σύστημα απαντά αποτελεσματικά τόσο απλά όσο και σύνθετα ερωτήματα σύζευξης πολλών μεταβλητών και εύκολα κλιμακώνεται σε 3 δισεκατομμύρια τριπλέτες χρησιμοποιώντας μια μικρή συστοιχία από 9 υπολογιστικούς κόμβους. Το H2RDF ξεπερνά state-of-the-art καταναμημένες λύσεις σε μη-επιλεκτικά και πολλαπλής σύζευξης ερωτήματα επιτυγχάνοντας παράλληλα απόδοση συγκρίσιμη με κεντροποιημένες λύσεις σε επιλεκτικά ερωτήματα. Σε αυτή την επίδειξη παρουσιάζουμε τη λειτουργικότητα του συστήματος, μέσω ενός διαδραστικού GUI. Οι χρήστες μπορούν να εκτελέσουν προκαθορισμένα ή όχι SPARQL ερωτήματα σε δεδομένα διαφόρων μεγεθών, χρησιμοποιώντας διαφορετικούς αλγόριθμους σύζευξης. Επιπλέον, μπορούν να επαναλάβουν όλα τα ερωτήματα σε συστοιχίες διαφορετικών μεγεθών. Χρησιμοποιώντας real-time συστήματα παρακολούθησης επίδοσης συστοιχιών με λεπτομερή στατιστικά στοιχεία, οι συμμετέχοντες θα είναι σε θέση να κατανοήσουν τα πλεονεκτήματα των διαφορετικών αλγορίθμων εκτέλεσης σε σχέση με τα δεδομένα εισόδου καθώς και την δυνατότητα κλιμάκωσης του H2RDF τόσο σε σχέση με το μέγεθος των δεδομένων όσο και με τους διαθέσιμους πόρους της συστοιχίας.

Σ43.P. Antonopoulos and I. Konstantinou and D. Tsoumakos and N. Koziris **Efficient Updates for Web-scale Indexes over the Cloud** In Proceedings of the International Workshop of Data Management in The Cloud (DMC2012 - In conjunction with ICDE 2012)

Στην εργασία αυτή, παρουσιάζουμε ένα καταναμημένο σύστημα που επιτρέπει γρήγορες και συχνές ενημερώσεις σε ανεστραμμένα ευρετήρια μεγάλου όγκου. Η προτεινόμενη τεχνική ενημέρωσης επιτρέπει την σταδιακή επεξεργασία των νέων ή τροποποιημένων δεδομένων και ελαχιστοποιεί τις απαιτούμενες αλλαγές του ευρετηρίου, μειώνοντας σημαντικά το χρόνο ενημέρωσης και καθιστώντας τον ανεξάρτητο από το μέγεθος του υφιστάμενου ευρετηρίου. Χρησιμοποιώντας το Hadoop MapReduce για την παραλληλοποίηση των πράξεων ενημέρωσης, και την HBase για τη κατανομή του ανεστραμμένου ευρετηρίου δημιουργήσαμε ένα πλήρως καταναμημένο υψηλής απόδοσης σύστημα για την δημιουργία και την ενημέρωση του ευρετηρίου. Πρόκειται για το πρώτο σύστημα ανοιχτού κώδικα που δημιουργεί, ενημερώνει και εξυπηρετεί μεγάλης κλίμακας ευρετήρια με πλήρως καταναμημένο τρόπο. Πειράματα με πάνω από 23 εκατομμύρια έγγραφα της Wikipedia αποδεικνύουν την ταχύτητα και την ευρωστία της εφαρμογής μας: κλιμακώνεται γραμμικά με το μέγεθος των ενημερώσεων και το βαθμό της αλλαγής στα έγγραφα και δείχνει έναν σταθερό χρόνο ενημέρωσης ανεξάρτητα από το μέγεθος του υποκείμενου ευρετηρίου. Επιπλέον, η προσέγγισή μας αυξάνει σημαντικά τις επιδόσεις της, με την απόκτηση περισσότερων υπολογιστικών πόρων: Μια παρτίδα ενημέρωσης μεγέθους 15.4GB σε ένα σύνολο

δεδομένων 64.2GB γίνεται σε 21 λεπτά με μόλις 12 απλούς υπολογιστικούς κόμβους, 3,3 φορές πιο γρήγορα σε σύγκριση με τη χρήση δύο κόμβων

- Σ44. Preuß S. and Keffer, H. and Schmidt, P. and Goumas, G. and Asiki, A., and Konstantinou, I. **User Adaptation in a Hybrid MT System.** In International Conference on Text, Speech and Dialogue (pp. 362-369). Springer Berlin Heidelberg. (2012, September).

Σε αυτή την εργασία παρουσιάζουμε τη μονάδα προσαρμογής χρήστη (UA) που υλοποιείται ως μέρος ενός νέου υβριδικού συστήματος MT μετάφρασης. Η προτεινόμενη μονάδα UA επιτρέπει στο χρήστη να ενισχύσει συστατικά του πυρήνα του συστήματος, όπως σύγχρονες γραμματικές και λεξικά συστήματος κατά το χρόνο εκτέλεσης. Είναι γνωστό ότι επιτρέποντας στους χρήστες να τροποποιήσουν τη συμπεριφορά του συστήματος αυξάνει την προθυμία να συνεργαστούν με συστήματα MT. Ωστόσο, σε στατιστικά συστήματα MT τα σχόλια των χρηστών είναι μόνο «μια σταγόνα στον ωκεανό» των στατιστικών πόρων. Το υβριδικό σύστημα MT που προτείνεται εδώ χρησιμοποιεί τον σύγχρονες γραμματικές με βάση κανόνες που εξάγονται αυτόματα από μικρά παράλληλα σχολιασμένα δίγλωσσα σώματα. Αντιπροσωπεύουν τις διαρθρωτικές αντιστοιχίσεις από τη γλώσσα πηγή στην γλώσσα στόχο. Μεταγενέστερα μονόγλωσσα στατιστικά συστατικά αποσαφηνίζουν περαιτέρω τη δομή της γλώσσας-στόχου. Αυτή η προσέγγιση παρέχει ένα κατάλληλο υπόστρωμα για να ενσωματώσει ένα ελαφρύ και αποτελεσματικό στοιχείο UA. Οι διορθώσεις του χρήστη συλλέγονται από μια μηχανή μετα-επεξεργασίας και προστίθεται στο δίγλωσσο corpus, ενώ οι προκύπτουσες πρόσθετες διαρθρωτικές αντιστοιχίσεις παρέχονται στο σύστημα κατά το χρόνο εκτέλεσης. Οι χρήστες μπορούν επίσης να ενισχύσουν το λεξικό του συστήματος. Η προσαρμογή των χρηστών είναι οργανωμένη σε ένα συγκεκριμένο κύκλο commit and review που επιτρέπει στο χρήστη να αναθεωρήσει την είσοδο προσαρμογής χρήστη. Η προκαταρκτική πειραματική αξιολόγηση δείχνει ενθαρρυντικά αποτελέσματα για την ικανότητα του συστήματος να προσαρμοστεί στις διαρθρωτικές προτιμήσεις του χρήστη.

- Σ45. I. Konstantinou and E. Angelou and C. Boumpouka and D. Tsoumakos and N. Koziris **On the Elasticity of NoSQL Databases over Cloud Management Platforms** In Proceedings of the 20th International Conference on Information and Knowledge Management (CIKM), Glaskow, Scotland, UK, 2011, pp. 2385-2388

Οι βάσεις NoSQL εστιάζουν στην αναλυτική επεξεργασία μεγάλης κλίμακας δεδομένων, προσφέροντας αυξημένη επεκτασιμότητα πάνω από απλές υπολογιστικές μονάδες. Ένα από τα κύρια χαρακτηριστικά τους είναι η ελαστικότητα, η οποία επιτρέπει δίκαιο διαμοιρασμό πόρων και αποδόσεις υψηλής ποιότητας. Ωστόσο, η διαδικασία της προσαρμοστικής αύξησης και μείωσης των πόρων περιλαμβάνει συνήθως πολλή χειροκίνητη προσπάθεια κατά τη διάρκεια της ρύθμισης της συστοιχίας των υπολογιστών. Μέχρι σήμερα, δεν υπάρχει κάποια συγκριτική μελέτη για την ποσοτικοποίηση του κόστους αυτού και την μέτρηση της αποτελεσματικότητας των NoSQL συστημάτων που προσφέρουν αυτή την δυνατότητα. Στην παρούσα εργασία, παρουσιάζουμε ένα σύστημα βασισμένο σε τεχνολογίες υπολογιστικών νεφών, για προσαρμοστική παρακολούθηση NoSQL συστημάτων. Μελετούμε διεξοδικά τις δυνατότητες ελαστικότητας των πιο δημοφιλών βάσεων NoSQL πάνω από μια ανοιχτού κώδικα υποδομή υπολογιστικού νέφους. Με βάση αυτές τις μετρήσεις, παρουσιάζουμε μια πρωτότυπη υλοποίηση ενός συστήματος λήψης αποφάσεων το οποίο εφαρμόζει ελαστικές

λειτουργίες με αυτόματο τρόπο σε οποιοδήποτε σύστημα NoSQL σύμφωνα με περιορισμούς που ορίζονται από τον διαχειριστή.

- Σ46.Ι. Konstantinou and E. Angelou and D. Tsoumakos and N. Koziris: **Distributed Indexing of Web Scale Datasets for the Cloud**. In Proceedings of the International Workshop on Massive Data Analytics on the Cloud (**MDAC2010** - in conjunction with **WWW 2010**), Raleigh, NC, USA, 26 April 2010.

Σε αυτή την εργασία παρουσιάζουμε ένα καταναμημένο σύστημα κατάλληλο για δεικτοδότηση, αποθήκευση και διαμοιρασμό μεγάλου όγκου δεδομένων (σε μεγέθη TB και περισσότερο) κάτω από υψηλό φόρτο εργασίας. Οι χρήστες τροφοδοτούν το σύστημα με τα πρωτογενή δεδομένα καθώς και με κανόνες δεικτοδότησης που έχουν να κάνουν με τον τύπο των δεδομένων αυτών, και το σύστημα τα επεξεργάζεται ανάλογα. Η επαυξημένη πληροφορία που προέρχεται από την επεξεργασία των δεδομένων εξάγεται με τη μορφή ενός καταναμημένου ευρετηρίου και διαμοιράζεται σε ένα μεγάλο αριθμό ταυτόχρονων χρηστών. Για την επιτάχυνση της διαδικασίας, η υπολογιστικά και αποθηκευτικά απαιτητική δεικτοδότηση χρησιμοποιεί το καινοτόμο σύστημα MapReduce. Για την επίτευξη χαμηλών χρόνων απόκρισης κάτω από υψηλό φορτίο ταυτόχρονων ερωτημάτων, τα αιτήματα των χρηστών εξυπηρετούνται μέσω της HBase, μια υλοποίηση ανοιχτού κώδικα του BigTable της Google.

- Σ47.Ι. Konstantinou and D. Tsoumakos and N. Koziris: **Measuring the Cost of Online Load-Balancing in Distributed Range-Queryable Systems**. In Proceedings of the 9th International IEEE Conference on Peer-to-Peer Computing (**P2P**), Seattle, WA, USA, 8-11 September 2009.

Καταναμημένα συστήματα όπως τα δίκτυα ομοτίμων (peer to peer) υποστηρίζουν αποδοτικά την επεξεργασία ερωτημάτων εύρους τιμών σε μεγάλο αριθμό κόμβων. Σε τέτοια συστήματα, ανομοιόμορφες κατανομές φόρτου πρέπει να αντιμετωπίζονται για την ελαχιστοποίηση υπερφορτωμένων κόμβων. Σε αυτή την εργασία εντοπίζουμε και αναλύουμε τις δύο βασικές μεθοδολογίες για την εξισορρόπηση φόρτου: Ανταλλαγές Αντικειμένων μεταξύ Γειτόνων και Μεταναστεύσεις κόμβων. Με βάση αυτή την ανάλυση προτείνουμε έναν υβριδικό αλγόριθμο που χρησιμοποιεί τις δυο τεχνικές για να επιτύχει γρήγορη και αποδοτική εξισορρόπηση φόρτου. Σαν παράδειγμα, παρουσιάζουμε μια υλοποίηση σε έναν Skip Γράφο, μια καταναμημένη δομή που επιτρέπει την εκτέλεση ερωτημάτων εύρους τιμών, την οποία χρησιμοποιούμε για να ελέγξουμε την απόδοση του αλγορίθμου κάτω από διαφορετικές συνθήκες κίνησης και φορτίου. Η πειραματική μας ανάλυση δείχνει ότι ο υβριδικός αλγόριθμος συγκλίνει πιο γρήγορα και καταναλώνει πολύ λιγότερο εύρος ζώνης σε σχέση με τους απλούς αλγορίθμους.

- Σ48.Δ. Dimitriadis, A. Metallinou, I. Konstantinou, G. Goumas, P. Maragos and N. Koziris: **GRIDNEWS: A Distributed Automatic Greek Broadcast Transcriptions System**. Proc. of IEEE Intern. Conf. on Acoustics, Speech and Signal Processing ICASSP-09, Taipei, Taiwan, March 2009.

Στην παρούσα εργασία, παρουσιάζεται ένα καταναμημένο σύστημα αποθήκευσης και ανάκτησης δεδομένων ειδησεογραφικών δελτίων από την ελληνική τηλεόραση. Τα πολυτροπικά αυτά δεδομένα υποβάλλονται σε επεξεργασία από ένα υπολογιστικό πλέγμα που διασυνδέει καταναμημένα συστήματα αποθήκευσης και επεξεργασίας δεδομένων. Το καινοτόμο στοιχείο του συστήματος είναι η εφαρμογή αλγορίθμων επεξεργασίας σήματος σε υπολογιστικά πλέγματα, προσφέροντας επιπλέον ευελιξία και υπολογιστική ισχύ. Ανάμεσα στα ανεπτυγμένα υποσυστήματα επεξεργασίας σήματος είναι: ο *Segmentor*, ο οποίος κόβει το αρχικό βίντεο σε κομμάτια μικρότερου μεγέθους, ο *Classifier*, που αναγνωρίζει κατά πόσον αυτά τα σύντομα βίντεο περιέχουν ομιλία ή όχι, το ελληνικό σύστημα αναγνώρισης ομιλίας που μετατρέπει τον λόγο σε γραπτό κείμενο, και, τέλος, η μηχανή αναζήτησης κειμένου και ανάκτησης *video*. Όλα τα επεξεργασμένα δεδομένα αποθηκεύονται και ανακτώνται σε γεωγραφικά καταναμημένα μέσα αποθήκευσης. Ένα φιλικό προς το χρήστη *web-based interface*, έχει αναπτυχθεί, να διευκολυνθεί η εισαγωγή και η αποθήκευση των νέων πολυτροπικών δεδομένων, η *off-line* επεξεργασία τους και τέλος, η αναζήτηση και ανάκτησή τους.

Σ49.Ι. Konstantinou and D. Tsoumakos and N. Koziris: **PASS It ON (PASSION): An Adaptive Online Load-Balancing Algorithm for Distributed Range-Query specialized Systems.** In Proceedings of the 16th International Conference on Cooperative Information Systems (**COOPIS**), Monterrey, Mexico, 12-14 November 2008

Σε αυτή την εργασία παρουσιάζουμε μια αρχική προσέγγιση στο πρόβλημα της ισοκατανομής φορτίου στους κόμβους ενός δικτύου ομοτίμων ειδικά σχεδιασμένου να απαντά ερωτήσεις συνόλου. Καθώς στην πλειοψηφία των εφαρμογών το φορτίο παρατηρείται ότι είναι αρκετά ανισοκαταναμημένο, παρουσιάζουμε μια προσαρμοστική μέθοδο κατά την οποία οι κόμβοι, ανταλλάσσοντας διαστήματα αντικειμένων μεταξύ τους, καταφέρνουν σταδιακά να ισοκατανεύουν το φορτίο.

Σ50.Α. Asiki, K. Doka, I. Konstantinou, A. Zissimos, and N. Koziris: **A Distributed Architecture for Multi-Dimensional Indexing and Data Retrieval in Grid Environments.** In Proceedings of the Cracow 2007 Grid Workshop (CGW'07), Krakow, Poland, October 16-17, 2007.

Σε αυτή την εργασία, περιγράφουμε μια *service-oriented* αρχιτεκτονική μιας γενικής πλατφόρμας *middleware*, η οποία παρέχει τις απαιτούμενες υπηρεσίες για τη διαχείριση δεδομένων σε ένα καταναμημένο περιβάλλον. Ο σχεδιασμός μας εισάγει έννοιες από δίκτυα *Peer-to-Peer*, προκειμένου να παρέχει μια επεκτάσιμη και αξιόπιστη υποδομή για την αποθήκευση, αναζήτηση και ανάκτηση του εμπλουτισμένου περιεχομένου. Για να εξασφαλιστεί η γρήγορη αναζήτηση στα καταναμημένα αποθετήρια ενός εικονικού Οργανισμού, το σύστημά μας ενσωματώνει ένα πολυδιάστατο σύστημα ευρετηρίασης, το οποίο εξυπηρετεί την ανάγκη για την υποστήριξη τόσο ερωτημάτων σημείου όσο και ερωτημάτων εύρους πάνω από ένα σύνολο μεταδεδομένων. Τέλος, οι μεταφορές αρχείων διεξάγονται με τη χρήση του *GridTorrent*, ενός πλεγματικού, *Peer-to-Peer* μηχανισμού μεταφοράς αρχείων που επιτρέπει τη συνολική απόδοση κατά την μεταφορά δεδομένων να κλιμακώνεται αποτελεσματικά και να τα καταφέρνει σε περιπτώσεις "flash crowd".

Σ51.Ι. Konstantinou, K. Doka, A. Asiki, A. Zissimos, and N. Koziris: **Gredia Middleware Architecture**. In Proceedings of the Cracow 2007 Grid Workshop (CGW'07), Krakow, Poland, October 16-17, 2007.

Στην παρούσα εργασία, παρουσιάζουμε το RDLS (Rich Data Location Services), ένα πλεγματο middleware διαχείρισης δεδομένων για να ανταποκρίνεται στις απαιτήσεις που τίθενται από το σχέδιο GREDIA. Ο κύριος σκοπός του συστήματος είναι η παροχή υπηρεσιών για την αποθήκευση, αναζήτηση και ανάκτηση σχολιασμένης, πολυμεσικής πληροφορίας σε ένα μεγάλης κλίμακας κατανεμημένο περιβάλλον. Οι υπηρεσίες αυτές μπορούν να ενσωματωθούν αποτελεσματικά σε υπάρχοντα Grid middleware, όπως το Globus Toolkit 4. Οι χρήστες της πλατφόρμας GREDIA είναι σε θέση να ανεβάζουν τα αρχεία τους στο κατανεμημένο αποθετήριο του GREDIA και να εκτελούν σύνθετες αναζητήσεις στα μεταδεδομένα του αποθηκευμένου περιεχομένου. Ένα σύστημα Peer-to-Peer χρησιμοποιείται για την αποθήκευση των μεταδεδομένων. Κάθε κόμβος θα είναι σε θέση να ενεργεί όχι μόνο ως καταναλωτής αποθηκευτικών πόρων, μέσω της αναζήτησης αρχείων, αλλά και ως φορέας παροχής των αρχείων μέσω των υπηρεσιών δεδομένων

ΠΕΡΙΟΔΙΚΑ:

Π1.Ν. Chalvantzis, A. Vontzalidis, E. Kassela, A. Spyrou, N. Nikitas, N. Provatias, I. Konstantinou and N. Koziris: **IW-NET BDA: A Big Data Infrastructure for Predictive and Geotemporal Analytics of Inland Waterways**, in IEEE Access, vol. 12, pp. 52503-52523, 2024, doi: 10.1109/ACCESS.2024.3387315.

Η πρόσφατη στροφή προς την ψηφιοποίηση σε παραδοσιακούς τομείς όπως η εφοδιαστική αλυσίδα και οι μεταφορές έχει ξεκλειδώσει νέους δρόμους για την απόκτηση πολύτιμων γνώσεων και τον εξορθολογισμό των λειτουργιών. Αυτός ο μετασχηματισμός διευκολύνεται από την αφθονία και την ιδιαιτερότητα των δεδομένων που είναι πλέον διαθέσιμα, συμπεριλαμβανομένων των δεδομένων IoT, των επιχειρησιακών συναλλαγών και των ειδοποιήσεων συμβάντων. Αυτές οι επιχειρήσεις αφήνουν ένα σημαντικό ψηφιακό αποτύπωμα, ώριμο για ανάλυση όταν συνδυάζονται με εξωτερικές πηγές δεδομένων. Ωστόσο, η αξιοποίηση αυτών των πληροφοριών απαιτεί ισχυρή υπολογιστική υποδομή και προσαρμόσιμο λογισμικό ικανό να χειρίζεται τεράστιες ποσότητες δεδομένων. Σε αυτό το άρθρο, παρουσιάζουμε το IW-NET BDA, ένα σύστημα ανάλυσης μεγάλων δεδομένων που βασίζεται σε τεχνολογίες ανοιχτού κώδικα για την αντιμετώπιση των απαιτήσεων αποθήκευσης και επεξεργασίας μαζικών συνόλων δεδομένων από διάφορες προελεύσεις. Το σύστημα αναπτύχθηκε στο πλαίσιο του χρηματοδοτούμενου από την ΕΕ έργου έρευνας και καινοτομίας IW-NET (Innovation driven Collaborative European Inland Waterways Transport Network). Το σύστημα απευθύνεται στον τομέα της εφοδιαστικής αλλά προσφέρει μια ευέλικτη βάση υπηρεσιών πληροφορικής λόγω του αγνωστικιστικού του σχεδιασμού, εστιάζοντας στην υποδομή -παροχή ως υπηρεσία. Επιπλέον, επιτρέπει την ανάπτυξη και την ανάπτυξη εφαρμογών που ενσωματώνουν την επιχειρηματική λογική, προσαρμοσμένες έτσι στις συγκεκριμένες επιχειρηματικές ανάγκες. Στις επόμενες ενότητες, εμβαθύνουμε στις αρχές σχεδιασμού, τα αρχιτεκτονικά στοιχεία και τις δυνατότητες ανάπτυξης του IW-NET BDA. Επιπλέον, παρουσιάζουμε δύο ενδεικτικές περιπτώσεις χρήσης: πρώτον, η αυτοματοποιημένη ανίχνευση περιοχών ενδιαφέροντος και η παρακολούθηση της δραστηριότητας του σκάφους

για αναλύσεις γεω-χρονικών δεδομένων κατά μήκος του διαδρόμου του ποταμού *Weser*. Δεύτερον, η χρήση επαναλαμβανόμενων νευρωνικών δικτύων για την πρόβλεψη της στάθμης του νερού στον διάδρομο του ποταμού Δούναβη. Αυτά τα παραδείγματα υπογραμμίζουν την προσαρμοστικότητα και την αποτελεσματικότητα του *IW-NET BDA* στην αντιμετώπιση διαφορετικών προκλήσεων σε διαφορετικά περιβάλλοντα, υπογραμμίζοντας την ευελιξία και τη χρησιμότητά του.

Π2.Ι. Giannakopoulos, I. Konstantinou, D. Tsoumakos and N. Koziris: **Cloud Application Deployment with Transient Failure Recovery**. *Journal of Cloud Computing*, Springer, 2018 7:11

Η εκτέλεση εφαρμογών είναι μια σημαντική λειτουργία για τους σύγχρονους παρόχους *cloud*. Η δυνατότητα δυναμικής κατανομής πόρων και εκτέλεσης μιας νέας εφαρμογής, βασισμένης σε μια περιγραφή που παρέχεται από τον χρήστη με έναν πλήρως αυτοματοποιημένο τρόπο, έχει μεγάλη σημασία για τους χρήστες του *cloud* καθώς διευκολύνει την παραγωγή πλήρως αναπαραγωγίμων περιβαλλόντων εφαρμογών με ελάχιστη προσπάθεια. Ωστόσο, οι περισσότερες σύγχρονες λύσεις ανάπτυξης δεν λαμβάνουν υπόψη την επιρρεπή σε σφάλματα φύση του *cloud*: Συχνά συναντώνται δυσλειτουργίες δικτύου, κακός συγχρονισμός μεταξύ διαφορετικών υπηρεσιών και άλλου λογισμικού ή δυσλειτουργίες σχετιζόμενες με την υποδομή με μεταβατικά χαρακτηριστικά. Ακόμη και αν αυτές οι αποτυχίες μπορεί να είναι ανεκτές κατά τη διάρκεια της ζωής μιας εφαρμογής, κατά τη διάρκεια της φάσης εγκατάστασης μπορούν να προκαλέσουν σοβαρά σφάλματα και να την οδηγήσουν σε αποτυχία. Προκειμένου να αντιμετωπιστεί αυτή η πρόκληση, σε αυτό το άρθρο προτείνουμε το *AURA*, ένα σύστημα ανοιχτού κώδικα που επιτρέπει την ανάπτυξη εφαρμογών *cloud* με δυνατότητες αποκατάστασης παροδικής αποτυχίας. Η *AURA* διατυπώνει την ανάπτυξη της εφαρμογής ως ένα κατευθυνόμενο ακυκλικό γράφημα. Κάθε φορά που παρουσιάζεται μια παροδική αποτυχία, διασχίζει το γράφημα, εντοπίζει τα τμήματα του που αποτυγχάνουν και εκτελούν εκ νέου τα αντίστοιχα σενάρια, με βάση το γεγονός ότι όταν η προσωρινή αποτυχία εξαφανιστεί, η εκτέλεση του σεναρίου θα επιτύχει. Επιπλέον, προκειμένου να διασφαλιστεί ότι κάθε εκτέλεση σεναρίου είναι *idempotent*, η *AURA* υιοθετεί ένα ελαφρύ μηχανισμό βασισμένο σε στιγμιότυπα συστήματος αρχείων που στοχεύει στην ακύρωση των παρενεργειών των αποτυχημένων *scripts*. Η εμπειριστατωμένη αξιολόγηση μας έδειξε ότι το *AURA* είναι ικανό να εκτελέσει ποικίλες εφαρμογές πραγματικού κόσμου σε περιβάλλοντα που παρουσιάζουν υψηλές πιθανότητες σφάλματος, εισάγοντας ένα ελάχιστο χρονικό διάστημα, ανάλογα με την πιθανότητα αποτυχίας των σεναρίων ανάπτυξης.

Π3.Δ. Sarlis, N. Papailiou, I. Konstantinou, G. Smaragdakis and N. Koziris: **Datix: A System for Scalable Network Analytics** *ACM SIGCOMM Computer Communication Review*, 45(5), October 2015.

Η συνεχώς αυξανόμενη κίνηση στο Διαδίκτυο δημιουργεί προκλήσεις για τους φορείς εκμετάλλευσης δικτύων και τους διαχειριστές που πρέπει να αναλύσουν μεγάλα σύνολα δεδομένων δίκτυου σε εύθετο χρόνο για τη λήψη αποφάσεων σχετικά με τη δρομολόγηση του δικτύου, τη διαστασιολόγηση και την ασφάλεια. Σύνολα δεδομένων δικτύου που συλλέγονται

σε μεγάλα δίκτυα όπως οι πάροχοι υπηρεσιών Διαδικτύου (ISP) ή σημεία ανταλλαγής Διαδικτύου (IXPs) μπορεί να είναι της τάξης των Terabytes ανά ώρα. Δυστυχώς, οι περισσότερες από τις τρέχουσες προσεγγίσεις ανάλυσης δικτύου είναι *ad-hoc* και συγκεντρωτικές, και ως εκ τούτου δεν είναι επεκτάσιμες. Σε αυτή την εργασία, παρουσιάζουμε το *Datix*, ένα πλήρως αποκεντρωμένο, *open-source* σύστημα ανάλυσης πληροφορίας για δεδομένα κίνησης δικτύου που βασίζεται σε έξυπνα συστήματα αποθήκευσης που υποστηρίζουν την εκτέλεση γρήγορων αλγορίθμων συνένωσης και την αποτελεσματική εκτέλεση ερωτημάτων φιλτραρίσματος. Περιγράφουμε την αρχιτεκτονική και το σχεδιασμό του *Datix* και σας παρουσιάζουμε την αξιολόγηση του *Datix* χρησιμοποιώντας πραγματικά ίχνη από ένα επιχειρησιακό IXP. Το *Datix* είναι ένα σύστημα που ασχολείται με ένα σημαντικό πρόβλημα στη διασταύρωση της διαχείρισης των δεδομένων και την παρακολούθηση του δικτύου, ενώ αξιοποιεί *state-of-the-art* καταναεμημένα εργαλεία επεξεργασίας. Εν συντομία, το *Datix* καταφέρνει να απαντήσει αποτελεσματικά ερωτήματα μέσα σε λίγα λεπτά σε σύγκριση με πάνω από 24 ώρες που χρειάζονται κατά την εκτέλεση υπάρχοντα κώδικα Python που βασίζεται σε κεντροποιημένες αρχιτεκτονικές. *Datix* επιτυγχάνει επίσης σχεδόν το 70% επιτάχυνση σε σχέση με βασικές υλοποιήσεις ερωτημάτων των δημοφιλών μηχανών ανάλυσης μεγάλου όγκου δεδομένων, όπως το *Hive* ή το *Shark*

Π4.Ι. Konstantinou and D. Tsoumakos and N. Koziris: **Fast and Cost-Effective Online Load-Balancing in Distributed Range-Queryable Systems**. *IEEE Transactions on Parallel and Distributed Systems (TPDS)* vol. 22, August 2011, pp. 1350-1364.

Σε αυτή την εργασία παρουσιάζεται ο *NIXMIG*, ένας προσαρμοστικός επιγραμματικός (*online*) αλγόριθμος με σκοπό την εξισορρόπηση φόρτου εργασίας σε καταναεμημένες δομές δεδομένων που υποστηρίζουν την δρομολόγηση ερωτημάτων εύρους τιμών (*range queries*). Ο αλγόριθμος αποτελεί επέκταση των εργασιών Σ5 και Σ3. Ο προτεινόμενος αλγόριθμος αντιμετωπίζει προβλήματα άνισων κατανομών φόρτου εργασίας που προκύπτουν όταν οι εξυπηρετητές διαμοιράζουν αντικείμενα διαφορετικής δημοτικότητας. Ο αλγόριθμος υλοποιήθηκε και εφαρμόστηκε σε έναν *skip-γράφο*, ένα δομημένο δίκτυο ομότιμων κόμβων ικανό να δρομολογεί ερωτήματα εύρους τιμών. Ο αλγόριθμος συγκρίθηκε πειραματικά και θεωρητικά κάτω από διαφορετικές συνθήκες κίνησης με άλλους παρόμοιους αλγόριθμους. Η ανάλυση αποδεικνύει ότι ο προτεινόμενος αλγόριθμος είναι πιο γρήγορος και καταναλώνει λιγότερους δικτυακούς πόρους κατά την διαδικασία της εξισορρόπησης.

Π5.Α. Asiki and A. Doka and I. Konstantinou and A. Zissimos and D. Tsoumakos and N. Koziris and P. Tsanakas: **A Grid middleware for data management exploiting Peer-to-Peer techniques**. *Future Generation Computer Systems*, vol. 25, Apr. 2009, pp. 426-435.

Σε αυτή την εργασία περιγράφουμε μια αρχιτεκτονική προσανατολισμένη σε υπηρεσίες για πλεγματικά περιβάλλοντα. Συγκεκριμένα, περιγράφουμε την αρχιτεκτονική που σχεδιάστηκε και υλοποιήθηκε στα πλαίσια του ευρωπαϊκού έργου *GRECIA*, με σκοπό να παρέχει αξιόπιστες, και αποδοτικές υπηρεσίες για την αποθήκευση, αναζήτηση και ανάκτηση πολυμεσικών δεδομένων. Για να εξασφαλίσουμε γρήγορη αναζήτηση εφαρμόζουμε ένα πολυδιάστατο σχήμα δεικτοδότησης το οποίο μπορεί να απαντήσει και ακριβείς και ερωτήσεις συνόλου. Επιπλέον, οι μετακινήσεις δεδομένων εκτελούνται με το *GridTorrent*, έναν πλεγματικό, συμβατό με τη φιλοσοφία των δικτύων ομοτίμων μηχανισμό που επιτελεί

αποδοτική μεταφορά μέσω της συνεργασίας μεταξύ των παρόχων. Η προτεινόμενη αρχιτεκτονική χρησιμοποιείται ως middleware component στην υλοποίηση του συγκεκριμένου έργου.

10. Βραβεία

Από τις παραπάνω ερευνητικές εργασίες έχουν προκύψει τα παρακάτω βραβεία:

- 1) Best Paper Award, 13th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid) για την εργασία Σ38
- 2) Best Paper Award Nomination, 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid) για την εργασία Σ29
- 3) 2009, 2011 ΕΜΠ, Θωμάϊδιο βραβείο για δημοσιευμένη εργασία σε έγκριτο επιστημονικό περιοδικό ή σε πρακτικά συνεδρίου για τις εργασίες Σ47 και Π4

11. Αναφορές στον τύπο

- 1) Αλλάζουν οι μεταφορές με ελληνική «σφραγίδα» (Καθημερινή)

<https://www.kathimerini.gr/society/1042967/allazoy-n-oi-metafores-me-elliniki-sfragida/>

- 2) European Commission Intelligence Body Completes Global Supply Chain Mission (Forbes)

<https://www.forbes.com/sites/adrianbridgwater/2019/09/03/european-commission-intelligence-body-completes-global-supply-chain-mission/?sh=7b95dccb16df>

- 3) Διεθνές βραβείο σε εργασία Ελλήνων για το i-cloud (Καθημερινή)

<https://www.kathimerini.gr/society/496105/diethnes-vraveio-se-ergasia-ellinon-gia-to-i-cloud/>

- 4) Ο... Ττραμόλα που κέρδισε το Διεθνές Συμπόσιο Υπολογιστών (Πατρίς)

<https://archive.patris.gr/articles/247457#.UiSppx0pOfk>

12. Ετερο-αναφορές

Οι παραπάνω εργασίες έχουν λάβει πάνω από 1350 ετερο-αναφορές, και η τιμή του δείκτη h-index είναι 20, σύμφωνα με το Google Scholar:
<http://scholar.google.gr/citations?hl=el&user=4LnVNmIAAAAJ>