

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

ΠΑΡΟΥΣΙΑΣΗ / ΕΞΕΤΑΣΗ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

**Μποφίλη- Αρβανίτη Ιωάννα- Μαρία
Μεταπτυχιακή Φοιτήτρια**

Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης

Επόπτης Μεταπτυχιακής Εργασίας: Αναπλ. Καθηγητής, Κ. Μαγκούτης

Τετάρτη, 24 Φεβρουαρίου 2021 , ώρα 10:00 π.μ.

Join Zoom Meeting

<https://zoom.us/j/4837766995>

**“ Διερεύνηση τεχνικών ανάλυσης δεδομένων σε πραγματικό χρόνο με χρήση
κατανεμημένων συστημάτων επεξεργασίας ροών”**

Περίληψη

Η κλιμακώσιμη επεξεργασία ροών δεδομένων είναι πολύ σημαντική τεχνολογία στους τομείς των υπηρεσιών του Διαδικτύου των Πραγμάτων (IoT), της κοινωνικής δικτύωσης, και άλλων τομέων των ηλεκτρονικών υπηρεσιών, έχοντας πρόσφατα οδηγήσει την ανάπτυξη μιας νέας γενιάς τέτοιων συστημάτων. Σε αυτή την εργασία εφαρμόζουμε μοντέρνες τεχνολογίες επεξεργασίας ροών προκειμένου να βελτιώσουμε την τεχνολογική στάθμιση σε δύο εφαρμογές που βασίζονται στον εντοπισμό στο φυσικό χώρο σε πραγματικό χρόνο: την δημιουργία του προφίλ κίνησης των επισκεπτών σε εκθεσιακούς χώρους και την πρόβλεψη παραβίασης συμφωνιών επιπέδου υπηρεσίας στον τομέα των μέσων μαζικής μεταφοράς. Προκειμένου να εξασφαλίσουμε πως τα κλιμακώσιμα συστήματα επεξεργασίας ροών μπορούν να προσαρμοστούν σε αλλαγές του φόρτου με αυτόματη προσαρμογή των πόρων τους, σε αυτή την εργασία υλοποιούμε ένα μηχανισμό με τον οποίο τα συστήματα επεξεργασίας ροών μπορούν να κλιμακωθούν δυναμικά ακόμα κι όταν αυτή η δυνατότητα δεν υποστηρίζεται από τα ίδια τα συστήματα και την υποκείμενη πλατφόρμα διαχείρισης πόρων.

Η πρώτη εφαρμογή των τεχνολογιών επεξεργασίας ροών δεδομένων σε συστήματα ανάλυσης δεδομένων σε πραγματικό χρόνο είναι στην ανάπτυξη δυναμικών προφίλ της συμπεριφοράς των επισκεπτών εκθεσιακών χώρων με βάση την κίνηση τους στο φυσικό χώρο. Η παρούσα εργασία εφαρμόζει τεχνολογίες ροών δεδομένων για την υποστήριξη θεωριών συμπεριφοράς από τις κοινωνικές επιστήμες και για την συλλογή πληροφορίας για πιθανά ενδιαφέροντα των χρηστών. Τέτοια προφίλ μπορούν να χρησιμοποιηθούν για την παραγωγή προτάσεων για τους επισκέπτες σχετικά με το ποια εκθέματα να επισκεφτούν, ποιο είναι το περιεχόμενο που θα ενδιέφερε περισσότερο το χρήστη ή για τη σχεδίαση προσωποποιημένων ερωτηματολογίων. Η δεύτερη εφαρμογή των τεχνολογιών επεξεργασίας ροών σε συστήματα ανάλυσης δεδομένων σε πραγματικό χρόνο είναι στην εκπαίδευση κατάλληλων μοντέλων για την πρόβλεψη οχημάτων που είναι πιθανό να παραβιάσουν τους στόχους που ορίζει η συμφωνία επιπέδου υπηρεσίας τους κατά τη διάρκεια δρομολογίου τους. Σε αυτή την εργασία επεκτείνουμε υπάρχουσες τεχνικές πρόβλεψης καθυστέρησης με τη δυνατότητα να εφαρμοστούν προβλέψεις σε πραγματικό χρόνο.

Σε αυτή την εργασία εξετάζουμε τις κατάλληλες ρυθμίσεις συστήματος για την επίτευξη κλιμακώσιμης και προσαρμόσιμης ανάλυσης δεδομένων ελέγχοντας την επίδραση διαφόρων παραμέτρων στο συνολικό σύστημα επεξεργασίας ροών δεδομένων. Ακόμα και σε περιπτώσεις που η δυναμική κλιμακωσιμότητα δεν υποστηρίζεται από την πλατφόρμα επεξεργασίας ροών, παρουσιάζουμε μια τεχνική που μπορεί να επιτύχει οριζόντια κλιμακωσιμότητα με μικρό χρόνο μη διαθεσιμότητας κατά την προσαρμογή.

University of Crete

Computer Science Department

M.Sc. Thesis presentation / examination

Bofili- Arvaniti Ioanna-Maria

Master's Thesis Supervisor: Associate Professor, K. Magoutis

Wednesday, 24 February 2021, 10:00 a.m

Join Zoom Meeting

<https://zoom.us/j/4837766995>

“Exploring real-time data analytics using distributed stream-processing systems”

Abstract

Processing high-volume streaming data is an important enabling technology in IoT-driven, social networking, and other e-services, having given rise to a new generation of stream-processing systems (SPS). In this thesis, we apply modern SPS technologies to improve the state of the art in two application areas involving real-time position tracking in physical space: maintaining profiles of visitor movement in exhibit spaces and predicting service-level objective violations in mass transit systems. To ensure that scalable SPSs are able to seamlessly adapt to varying levels of load by adjusting their resources, in this thesis we implement a mechanism by which SPSs can scale dynamically even when such capability is not natively supported by the SPS and the underlying resource management platform.

Our first application of SPS technologies to real-time data analytics is on the development of dynamic behavioral profiles of visitors in exhibit spaces based on their movements in physical space. While related approaches have been explored in the past, this thesis applies for the first time stream-processing technologies to materialize behavioral theories developed in social sciences and to collect richer information about visitors' interests. Such profiles can be used to produce recommendations for the visitors about exhibits they should visit, to decide the best content to present to them, or to design personalized questionnaires. Our second application of SPS technologies to real-time data analytics is on training appropriate models for predicting mass-transit vehicles that are likely to violate service-level objectives in their route duration. In this thesis we extend previous delay prediction techniques with the ability to apply predictions in a real-time fashion.

In this thesis we also address the necessary tuning and support for scalable, adaptive data analytics by examining various parameters of the SPS and its ingest engine. Even in cases where dynamic scaling is not explicitly supported by the SPS platform, we demonstrate a technique that achieves scale-out with low downtime.