

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
PROGRAMŲ SISTEMŲ KATEDRA

Didelių duomenų srautų analizė, anomalijų aptikimas

Large data flow analysis, detection of anomalies

Kursinis darbas

Atliko: 3 kurso 3 grupės studentas
Jokūbas Rusakevičius (parašas)

Darbo vadovas: dr. Vytautas Valaitis (parašas)

Vilnius – 2018

TURINYS

ĮVADAS	2
1. MACROBASE DUOMENŲ ANALIZĖS SISTEMA	3
REZULTATAI IR IŠVADOS	4
LITERATŪRA	5
SANTRUMPOS	6
PRIEDAI	6

Įvadas

Surenkamų duomenų kiekiai nuolatos didėja ir gerokai lenkia žmonių sugebėjimą juos apdoroti, peržiūrėti ar analizuoti. Didžiosios socialinių tinklų kompanijos Twitter, Facebook ir LinkedIn praneša kiekviena atskirai fiksuojanti iki 12 milijonų įvykių per sekundę [Ast16; PFT⁺15; Woo15]. Taip pat, vis daug duomenų yra surenkama iš automatizuotų duomenų šaltinių (pvz.: „Dalykų Internetui“ (angl. „Internet of Things“)). Kylantis automatizuotų duomenų šaltinių populiarumas, pinganti techninė įranga, išvystyti komunikaciniai tinklai bei mažėjančios duomenų saugojimo kainos paskatino dešimčių milijardų dolerių naujų komercinių investicijų šių technologijų vystimui [MCB⁺15]. Numatoma, kad kiekvienais metais bendras duomenų kiekis išaugs po 40% [EMC14].

1. MacroBase duomenų analizės sistema

MacroBase yra didelių ir/ar greitų duomenų ar duomenų srautų analitinė stebėjimo sistema.

Rezultatai ir išvados

Rezultatų ir išvadų dalyje turi būti aiškiai išdėstomi pagrindiniai darbo rezultatai (kažkas išanalizuota, kažkas sukurta, kažkas įdiegta) ir pateikiamos išvados (daromi nagrinėtų problemų sprendimo metodų palyginimai, teikiamos rekomendacijos, akcentuojamos naujovės).

Literatūra

- [Ast16] Anthony Asta. Observability at twitter: technical overview, part i. 2016. URL: https://blog.twitter.com/engineering/en_us/a/2016/observability-at-twitter-technical-overview-part-i.html.
- [EMC14] Dell EMC, red. The digital universe of opportunities: rich data and the increasing value of the internet of things. 2014. URL: <http://www.emc.com/leadership/digital-universe/>.
- [MCB⁺15] James Manyika, Michael Chui, Peter Bisson, Jonathan Woetzel, Richard Dobbs ir Jacques Bughin Dan Aharon. *McKinsey Global Institute: The internet of things: mapping the value beyond the hype*. 2015.
- [PFT⁺15] Tuomas Pelkonen, Scott Franklin, Justin Teller, Paul Cavallaro, Qi Huang, Justin Meza ir Kaushik Veeraraghavan. Gorilla: a fast, scalable, in-memory time series database. *VLDB Endowment - Proceedings of the 41st International Conference on Very Large Data Bases*, p. 1816–1827, Kohala Coast, Hawaii. Facebook Inc., 2015.
- [Woo15] Alex Woodie. Kafka tops 1 trillion messages per day at linkedin. 2015. URL: <https://www.datanami.com/2015/09/02/kafka-tops-1-trillion-messages-per-day-at-linkedin/>.

Santrumpos

Sąvokų apibrėžimai ir santrumpų sąrašas sudaromas tada, kai darbo tekste vartojami specialūs paaiškinimo reikalaujantys terminai ir rečiau sutinkamos santrumpos.