Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií

Evidenčné číslo

Matúš Cimerman

Analýza prúdu údajov

Bakalárska práca

Vedúci práce: Ing. Jakub Ševcech
máj 2015

Abstract

Nowadays we can see Big Data processing and analysis in many domains. As a amounts of data growing, more people are focusing on this problem. The most affected domains are social media websites like Facebook or Twitter. A data from such a sources are streaming in huge amounts and changing in real-time, called data streams. We want to process and analyze data streams in real-time to provide users personalized and valuable outputs. The most common approach to handle data streams is map-reduce paradigm, e.g. batch data processing. Proposed methods are not meeting our requirement to process data streams in real-time. To achieve these requirements, we need use different approach called data stream processing which is built on Lambda Architecture.
Processing and analysis of big data streams is a complex task, because we need to provide low-latency, scalable and fault-tolerant solution. In our project, we analyze existing solutions and frameworks to analyze data streams. We provide verification of its characteristics in different kind of tasks. Accordingly to this, we propose a application for processing and analyzing big data streams (e.g. Twitter data stream), which allows users to get valuable outputs changing in real-time.

ToDo: upravit anglicku podla slovenskej. V slovenskej su male upravy...-

az po revizii slovenskej

Abstrakt

Poďakovanie

Thank all of you!

Matúš Cimerman

Contents

Abstract iii

Abstrakt v

Acknowledgement vii

1 Úvod 1

2 Metódy spracovania prúdu údajov 3

3 Existujúce riešenia pre spracovanie a anlýzu prúdu údajov 5

4 Metóda pre analýzu prúdu údajov v reálnom čase 7

5 Implementácia navrhnutej metódy 9

6 Vyhodnotenie a experimenty 11

7 Zhodnotenie a budúca práca 13

7.1 Zhodnotenie a záver 13

7.2 Budúca práca 13

A Plán pre letný semester **15**

B Druhá príloha **17**

Bibliography **19**

List of Figures

List of Tables

1. Úvod

ToDo: napisat uvod

2. Metódy spracovania prúdu údajov

3. Existujúce riešenia pre spracovanie a anlýzu prúdu údajov

4. Metóda pre analýzu prúdu údajov v reálnom čase

5. Implementácia navrhnutej metódy

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum. Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum. Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

6. Vyhodnotenie a experimenty

7. Zhodnotenie a budúca práca

7.1 Zhodnotenie a záver

7.2 Budúca práca

A. Plán pre letný semester

ToDo: Plan na druhy semester!

B. Druhá príloha

Bibliography