Accepted Manuscript

A multi-layered performance analysis for cloud-based topic detection and tracking in Big Data applications

Meisong Wang, Prem Prakash Jayaraman, Ellis Solaiman, Lydia Y. Chen, Zheng Li, Song Jun, Dimitrios Georgakopoulos, Rajiv Ranjan

PII: S0167-739X(17)31593-5
DOI: https://doi.org/10.1016/j.future.2018.01.047
Reference: FUTURE 3954

To appear in: *Future Generation Computer Systems*

Received date : 19 July 2017
Revised date : 21 November 2017
Accepted date : 22 January 2018

Please cite this article as: M. Wang, P.P. Jayaraman, E. Solaiman, L.Y. Chen, Z. Li, S. Jun, D. Georgakopoulos, R. Ranjan, A multi-layered performance analysis for cloud-based topic detection and tracking in Big Data applications, *Future Generation Computer Systems* (2018), https://doi.org/10.1016/j.future.2018.01.047

A Multi-layered Performance Analysis for Cloud-based Topic Detection and Tracking in Big Data Applications

Meisong Wang[a], Prem Prakash Jayaraman[b], Ellis Solaiman[c], Lydia Y. Chen[d], Zheng Li[e], Song Jun[f], Dimitrios Georgakopoulos[b], Rajiv Ranjan[c]

[a]*School of Computer Science, Australian National University, ACT, Australia*
[b]*Faculty of Science, Engineering and Technology, Swinburne University of Technology, Melbourne, Australia*
[c]*School of Computer Science, Newcastle University, Newcastle Upon Tyne, UK*
[d]*Zurich Research Laboratory, IBM, Zurich, Switzerland*
[e]*Department of Electrical and Information Technology, Lund University, Sweden*
[f]*Department of Computer Science, Chinese University of Geosciences, Wuhan*

Abstract

In the era of the Internet of Things and social media; communities, governments, and corporations are increasingly eager to exploit new technological innovations in order to track and keep up to date with important new events. Examples of such events include the news, health related incidents, and other major occurrences such as earthquakes and landslides. This area of research commonly referred to as Topic Detection and Tracking (TDT) is proving to be an important component of the current generation of Internet-based applications, where it is of critical importance to have early detection and timely response to important incidents such as those mentioned above. The advent of Big data though beneficial to TDT applications also brings about the enormous challenge of dealing with data variety, velocity and volume (3Vs). A promising solution is to employ Cloud Computing, which enables users to access powerful and scalable computational and storage resources in a "pay-as-you-go" fashion. However, the efficient use of Cloud resources to boost the performance of mission critical applications employing TDT is still an open topic that has not

Email addresses: deanmeisong@gmail.com (Meisong Wang), pjayaraman@swin.edu.au (Prem Prakash Jayaraman), ellis.solaiman@ncl.ac.uk (Ellis Solaiman), yic@zurich.ibm.com (Lydia Y. Chen), zheng.li@eit.lth.se (Zheng Li), songjun@cug.edu.cn (Song Jun), dgeorgakopoulos@swin.edu.au (Dimitrios Georgakopoulos), rranjans@gmail.com, Corresponding Author (Rajiv Ranjan)