

Accepted Manuscript

An Efficient Time Optimized Scheme for Progressive Analytics in Big Data

Kostas Kolomvatsos, Christos Anagnostopoulos,
Stathes Hadjiefthymiades

PII: S2214-5796(15)00010-6
DOI: [10.1016/j.bdr.2015.02.001](http://dx.doi.org/10.1016/j.bdr.2015.02.001)
Reference: BDR 16

To appear in: *Big Data Research*

Received date: 11 June 2014
Revised date: 3 December 2014
Accepted date: 3 February 2015

Please cite this article in press as: K. Kolomvatsos et al., An Efficient Time Optimized Scheme for Progressive Analytics in Big Data, *Big Data Research* (2015), <http://dx.doi.org/10.1016/j.bdr.2015.02.001>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



An Efficient Time Optimized Scheme for Progressive Analytics in Big Data

Kostas Kolomvatsos

Department of Computer Science, University of Thessaly, 35100, Greece

Christos Anagnostopoulos

School of Computing Science, University of Glasgow, G12 8QQ, UK

Stathes Hadjiefthymiades

Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, 15784, Greece

Abstract

Big data analytics is the key research subject for future data driven decision making applications. Due to the large amount of data, progressive analytics could provide an efficient way for querying big data clusters. Each cluster contains only a piece of the examined data. Continuous queries over these data sources require intelligent mechanism to result the final outcome (query response) in the minimum time with the maximum performance. A *Query Controller (QC)* is responsible to manage continuous / sequential queries and return the final outcome to users or applications. In this paper, we propose a mechanism that can be adopted by the *QC* capable of managing partial results retrieved by a number of processors each one responsible for each cluster. Each processor executes a query over a specific cluster of data. The proposed mechanism adopts two sequential decision making models for handling the incoming partial results. The first model is based on a finite horizon time-optimized model and the second one is based on an infinite horizon optimally scheduled model. We provide mathematical formulations for solving the discussed problem and present simulation results. Through a large number of experiments, we reveal the advantages of the proposed models and give numerical results comparing them with a deterministic model. These results indicate that the proposed models can efficiently reduce the required time for returning the final outcome to the user / application while

Download English Version:

<https://daneshyari.com/en/article/10327328>

Download Persian Version:

<https://daneshyari.com/article/10327328>

[Daneshyari.com](https://daneshyari.com)