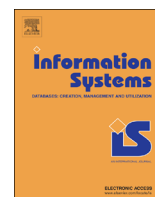




Contents lists available at ScienceDirect

Information Systems

journal homepage: www.elsevier.com/locate/infosys

The rise of “big data” on cloud computing: Review and open research issues



Ibrahim Abaker Targio Hashem^{a,*}, Ibrar Yaqoob^a, Nor Badrul Anuar^a,
Salimah Mokhtar^a, Abdullah Gani^a, Samee Ullah Khan^b

^a Faculty of Computer Science and information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia

^b NDSU-CIT Green Computing and Communications Laboratory, North Dakota State University, Fargo, ND 58108, USA

ARTICLE INFO

Article history:

Received 11 June 2014

Received in revised form

22 July 2014

Accepted 24 July 2014

Recommended by: Prof. D. Shasha

Available online 10 August 2014

Keywords:

Big data

Cloud computing

Hadoop

ABSTRACT

Cloud computing is a powerful technology to perform massive-scale and complex computing. It eliminates the need to maintain expensive computing hardware, dedicated space, and software. Massive growth in the scale of data or big data generated through cloud computing has been observed. Addressing big data is a challenging and time-demanding task that requires a large computational infrastructure to ensure successful data processing and analysis. The rise of big data in cloud computing is reviewed in this study. The definition, characteristics, and classification of big data along with some discussions on cloud computing are introduced. The relationship between big data and cloud computing, big data storage systems, and Hadoop technology are also discussed. Furthermore, research challenges are investigated, with focus on scalability, availability, data integrity, data transformation, data quality, data heterogeneity, privacy, legal and regulatory issues, and governance. Lastly, open research issues that require substantial research efforts are summarized.

© 2014 Elsevier Ltd. All rights reserved.

Contents

1. Introduction	99
2. Definition and characteristics of big data	99
2.1. Classification of big data	100
3. Cloud computing	101
4. Relationship between cloud computing and big data	102
5. Case studies	104
5.1. Organization case Studies from vendors	104
5.1.1. A. SwiftKey	104
5.1.2. B. 343 Industries	105
5.1.3. C. redBus	105
5.1.4. D. Nokia	105
5.1.5. E. Alacer	105

* Corresponding author. Tel.: +60 173946811.

E-mail addresses: targio@siswa.um.edu.my (I.A.T. Hashem), ibraryaqoob@siswa.um.edu.my (I. Yaqoob), badrul@um.edu.my (N.B. Anuar), salimah@um.edu.my (S. Mokhtar), abdullah@um.edu.my (A. Gani), samee.khan@ndsu.edu (S. Ullah Khan).

<http://dx.doi.org/10.1016/j.is.2014.07.006>

0306-4379/© 2014 Elsevier Ltd. All rights reserved.