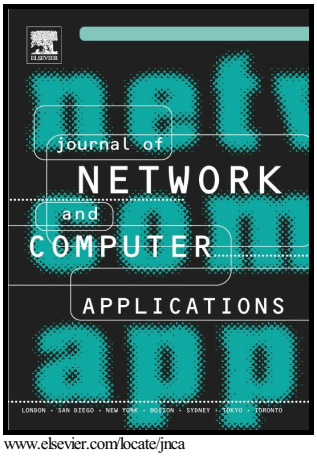
Cloud Storage Reliability for Big Data Applications: A State of the Art Survey

Rekha Nachiappan, Bahman Javadi, Rodrigo Calherios, Kenan Matawie



www.elsevier.com/locate/jnca

PII: S1084-8045(17)30273-4
DOI: http://dx.doi.org/10.1016/j.jnca.2017.08.011
Reference: YJNCA1957

To appear in: *Journal of Network and Computer Applications*

Received date: 12 May 2017
Revised date: 31 July 2017
Accepted date: 22 August 2017

Cite this article as: Rekha Nachiappan, Bahman Javadi, Rodrigo Calherios and Kenan Matawie, Cloud Storage Reliability for Big Data Applications: A State of the Art Survey, *Journal of Network and Computer Applications,* http://dx.doi.org/10.1016/j.jnca.2017.08.011

Cloud Storage Reliability for Big Data Applications: A State of the Art Survey

Rekha Nachiappan [a], Bahman Javadi [a], Rodrigo Calherios [a], Kenan Matawie [a]

[a] *School of Computing, Engineering and Mathematics, Western Sydney University, Australia*

Abstract

Cloud storage systems are now mature enough to handle a massive volume of heterogeneous and rapidly changing data, which is known as Big Data. However, failures are inevitable in cloud storage systems as they are composed of large scale hardware components. Improving fault tolerance in cloud storage systems for Big Data applications is a significant challenge. Replication and Erasure coding are the most important data reliability techniques employed in cloud storage systems. Both techniques have their own trade-off in various parameters such as durability, availability, storage overhead, network bandwidth and traffic, energy consumption and recovery performance. This survey explores the challenges involved in employing both techniques in cloud storage systems for Big Data applications with respect to the aforementioned parameters. In this paper, we also introduce a conceptual hybrid technique to further improve reliability, latency, bandwidth usage, and storage efficiency of Big Data applications on cloud computing.

Keywords: Fault tolerance; Big Data applications; Cloud storage; Replication; Erasure coding; Data Reliability.

1. Introduction

In the contemporary society of Big Data, the data volume is growing faster than storage capacity (Gantz and Reinsel, 2012). Each week, Facebook requires extra 60TB of storage just for new photos (Beaver et al., 2010). YouTube users upload over 400 hours of video every minute and it requires 1 Petabyte of new storage every day (Baesens, 2014). According to the Inter-