

On the Adoption of Erasure Code for Cloud Storage by Major Distributed Storage Systems

Aatish Chiniyah^{1*} and Avinash Mungur²

¹ Department of Digital Technologies,
Faculty of Information, Communication and Digital Technologies
University of Mauritius, a.chiniyah@uom.ac.mu

² Department of Information and Communication Technologies
Faculty of Information, Communication and Digital Technologies
University of Mauritius, a.mungur@uom.ac.mu

Abstract

INTRODUCTION: Traditional Cloud Systems are struggling to cope with the exponential growth of data in today's distributed application environment. The amount of data online has continuously increased since 2003. From an estimated 5 Exabyte in 2003 to 988 Exabyte in 2010. Presently it is estimated that 5 Exabyte of data are produced daily. To cope with such astronomical load of data, Distributed Storage Systems such as Amazon, Google and Microsoft Azure are becoming the de-facto method for the storage of data. Replication is the method used for providing redundancy. However Erasure Coding is a worthy alternative.

OBJECTIVES: The purpose of this paper is to assess the most used distributed storage systems using different evaluation criteria and identifying how erasure code can be integrated into them.

METHODS: This paper provides a survey of well-known Distributed Storage Systems by using the CAP (Consistency, Availability and Partition Tolerance) Theorem. We go by presenting the solution according to the objectives set and trade-off acknowledged by the designers.

RESULTS: A comprehensive survey is presented using five evaluation criteria (design principle, data model, failure detection and recovery, consistency and security). Adoption of erasure code in Distributed Storage Systems is discussed and its advantages are deliberated. Several open challenges are also put forward.

CONCLUSION: This paper provides researchers in the field with a comprehensive review of Distributed Storage Systems and how the adoption of erasure codes will enhance their capabilities.

Keywords: Distributed Storage Systems (DSS), Erasure Code, Cloud Storage, Redundancy, CAP Theorem.

Received on 07 June 2021, accepted on 31 August 2021, published on 14 September 2021

Copyright © 2021 Aatish Chiniyah *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [Creative Commons Attribution license](#), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/_____

1. Introduction

The world is experiencing an exponential growth of data [1] as everything is being digitalized: music, pictures,

videos, online gaming, and not to mention Internet related data such as Facebook's posts, Instagram's pictures, tweeter's tweets and so on. The new generation of web applications requires the processing of terabytes or even petabytes of data. Such processing power can only be achieved by using distributed/parallel processing [2]. In

*Corresponding author. Email: a.chiniyah@uom.ac.mu

