

# Red Hat Ceph Storage for Content Distribution in the Modern Enterprise

Prepared for RedHat  
June 2016

## TABLE OF CONTENTS

TABLE OF CONTENTS.....	2
INTRODUCTION.....	3
THE PROBLEM OF CONTENT DISTRIBUTION.....	4
OBJECT STORES FOR CONTENT DISTRIBUTION.....	5
BOTTOM LINE.....	6
JUKU.....	7
WHY JUKU.....	7
AUTHOR.....	7

## INTRODUCTION

Enterprises are facing new challenges regarding storage infrastructures and data mobility. Basic characteristics of storage systems, like data protection, reliability and availability are common ground now, but the bar is raising again because users ask to access data from everywhere and any device. It's not easy to find a solution when scalability, performance and security are all involved at the same time. But this is what end users want and, to make it sustainable, require a low \$/GB.

The amount of data that must be stored safely and accessed from anywhere is growing exponentially. Now, with IT organizations implementing multiple cloud strategies, finding a future-proof platform capable of serving different workloads simultaneously, to local and remote locations, is even more challenging.

In this context, content distribution is no longer a function relegated to a few large streaming and telco operators. We are now entering a new phase where content distribution is becoming essential for any kind of organization that manages large file and media repositories. Industry-wide, the focus is quickly shifting from storing data safely and efficiently, which is now taken for granted, to distributing and sharing it quickly and safely.

*The focus is quickly shifting from storing data safely and efficiently, which is now taken for granted, to distributing and sharing it quickly and safely.*

## THE CHALLENGE OF CONTENT DISTRIBUTION

Even when archives are a reasonable size, on the order of a few Petabytes or 'just' hundreds of Terabytes, traditional file-based protocols, like SMB or NFS, are not adequate for long distance communication. They are notorious for being very 'chatty', the security layer is complex and they predate the internet and mobile device explosion of the last few years. This is why the wrong protocol can be a major problem when it comes to streaming data to remote devices using an unreliable connection or with unpredictable latency. For many practical reasons, the best protocol for content distribution continues to be HTTP/S and this is why the vast majority of audio and video streaming services are using it. Access to object storage is based on this

*Even when archives are not huge, in the order of few Petabytes or hundreds of Terabytes, traditional file-based protocols, like SMB or NFS, are not adequate for long distance communication.*

protocol (through HTTP-based RESTful APIs) and this, associated with the ability of object storage systems to maintain multiple copies of the same file in different locations, extremely simplifies the creation of Content Delivery Networks (CDN) of all sizes.

# OBJECT STORES FOR CONTENT DISTRIBUTION

Object storage is no longer relegated to the second or third tier in storage infrastructures, as it has been in the past. Some of its characteristics are perfectly suitable to be the backend of modern IT infrastructures which need scalability, performance and availability demanded by developers, end users and cloud applications.

*Object Storage is perfectly suitable to be the backend of modern IT infrastructures which need scalability, performance and availability demanded by developers, end users and cloud applications*

To retain its value in time, these data repositories must scale both in terms of capacity and bandwidth while maintaining a low \$/GB. Traditional approaches are not feasible, and only a modern, scale-out distributed and software-defined approach allows you to build a future-proof adaptable infrastructure.

It's also important to note that, because of the size of the storage infrastructure and the type of data stored within, there are some very significant characteristics to consider when implementing it:

- **Data protection:** in traditional storage systems, efficiency is always measured by the quality and quantity of data footprint reduction features. But this is not the case with large media files and, more generally, with content that is intended to be streamed out at some time. These objects are usually saved in already compressed formats and, in many cases now, they are also encrypted. This simply means that any further data reduction, like deduplication or compression, are quite ineffective. Alternatively, erasure coding techniques are preferred instead.
- **Automated replication and snapshots:** In this kind of high capacity storage infrastructure, backup is quite impossible and a policy based automated multi-site replication mechanism is necessary to prevent data loss in case of disaster while, to help prevent data losses at the file level, snapshots and file/object versioning are additional useful features.

Red Hat Ceph Storage has the right characteristics to be the storage layer of such an infrastructure. Thanks to its multi-tier architecture, it can scale to large capacity installations while maintaining a fast tier of cache for the most accessed files. It provides the right set of protocols and APIs to be accessed from any type of device while multi-site active/active configurations can help to sustain large amounts of traffic as well as to mitigate the risk of service disruption in case of a site failure. Recently introduced capabilities like the ability to maintain high throughput even on degraded objects, are clearly addressing content distribution challenges.

## BOTTOM LINE

For large storage repositories, like in the case of CDN, there are important characteristics to take into account and which make them sustainable in terms of both TCA (Total Cost of Acquisition) and TCO (Total Cost of Operation), especially in the long term:

- **Linear scalability:** the object storage backend is fundamental in sustaining any form of large scale capacity growth, especially when facing the challenges imposed by the cloud, Big Data and IoT applications.
- **High performance:** expressed more so in terms of high throughput rather than low latency. A common pattern for all unstructured data types and related workloads.
- **Low \$/GB:** indispensable for infrastructure sustainability over time.
- **Great flexibility:** freedom of choice for configurations to help build better optimized configurations to ease provisioning and operations

Software Defined Solutions like Red Hat Ceph Storage are clearly going toward this direction:

- More performance: by eliminating intermediate layers like local file systems and including optimizations for All-Flash configurations;
- More security: improved thanks to better encryption options;
- More interfaces: thanks to the maturity of scale-out file system and S3 gateway.

Red Hat Ceph Storage has all these characteristics and more. It's open source and software-defined nature eliminates any form of lock-in while providing a compelling roadmap for the future. The success of open source in the enterprise and the cloud is indisputable. And Software-Defined open source based storage is now getting a lot of traction because it provides better infrastructure longevity and overall cost while promising future-proof features and a cloud-like architecture for all kinds of organizations.

By adopting Red Hat Ceph Storage, enterprises can effectively face the challenges posed by data growth and diversity with a sustainable and cost effective infrastructure. It can be the ideal solution for building the core of data-driven infrastructures for today's and future needs.

*By adopting Red Hat Ceph Storage, enterprises can effectively face the challenges posed by data growth and diversity with a sustainable cost effective infrastructure.*

As is common for successful open source projects, the development is driven by a large and growing community led by end users who are implementing Red Hat Ceph Storage in their data-driven infrastructures. And this is also the reason why it is future proof and evolves fast: by following the latest trends in hyper-scale and modern infrastructures. These trends are now becoming of interest to smaller and traditional organizations as well, who are constantly seeking simplification and solutions which will lower the Total Cost of Ownership (TCO) of their infrastructures.

## JUKU

### WHY JUKU

Jukus are Japanese specialized cram schools and our philosophy is the same. Not to replace the traditional information channels, but to help decision makers in their IT environments, to inform and to discuss the technological side that we know better: IT infrastructure virtualization, cloud computing and storage.

Unlike the past, today those who live in the IT environment need to be aware of their surroundings: things are changing rapidly and there is a need to be constantly updated, to learn to adapt quickly and to support important decisions - but how? Through our support, our ideas, the result of our daily global interaction on the web and social networking with vendors, analysts, bloggers, journalists and consultants. But our work doesn't stop there - the comparison and the search is global, but the sharing and application of our ideas must be local and that is where our daily experience, with companies rooted in local areas, becomes essential in providing an honest and productive vision. That's why we have chosen: "think global, act local" as a payoff for Juku.

### AUTHOR



Enrico Signoretti is an analyst, trusted advisor and passionate blogger (not necessarily in that order). He has been immersed in IT environments for over 20 years. His career began with Assembler in the second half of the 80's before moving on to UNIX platforms until now when he joined the "Cloudland". During these years his job has changed from highly technical roles to management and customer relationship management. In 2012 he founded Juku consulting SRL, a new consultancy and advisory firm deeply focused on supporting end users, vendors and third parties in the development of their IT infrastructure strategies.

He keeps a vigil eye on how the market evolves and is constantly on the lookout for new ideas and innovative solutions. You can find Enrico's social profiles here: <http://about.me/esignoretti>

All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources Juku Consulting srl (Juku) considers to be reliable but is not warranted by Juku. This publication may contain opinions of Juku, which are subject to change from time to time. This publication is covered by [Creative Commons License \(CC BY 4.0\)](#): Licensees may cite, copy, distribute, display and perform the work and make derivative works based on this paper only if Enrico Signoretti and Juku consulting are credited. The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions and typographical errors. Juku consulting srl has a consulting relationship with Red Hat. This paper was commissioned by Red Hat. No employees at the firm hold any equity positions with Red Hat. Should you have any questions, please contact Juku consulting srl ([info@juku.it](mailto:info@juku.it) - <http://juku.it>).