Red Hat Ceph Storage, a next generation platform for web scale storage

Prepared for RedHat
June 2016

JUKU
Think Global, Act Local!

TABLE OF CONTENTS

EXECUTIVE SUMMARY

End user needs have been evolving quite rapidly in the last few years. Organizations of all sizes are experiencing a substantial diversification in the workloads served by their storage infrastructures. Consequently, requirements are no longer just about performance or capacity but about how storage resources are being consumed.

Organizations of all sizes are experiencing substantial diversification in the workloads served by their storage infrastructure. Consequently, requirements are no longer just about performance or capacity but they depend on how storage resources are being consumed.

Mobile devices, Big Data applications, IoT are the most visible examples, they are drastically changing access patterns to large scale storage systems. Security is another big issue - data is now moved over long distances and accessed from multiple devices and locations. With data being transferred over IP networks, and different protocols, server and perimeter-based security is becoming obsolete even for storage resources.

Clearly, most IT organizations are choosing a hybrid cloud strategy. The goal is to combine the flexibility of the public cloud with a lower, and much more predictable, cost of on premises infrastructures. It's not uncommon to find end users taking advantage of a multiple cloud approach, where they start the development of new applications in the public cloud only to repatriate them later when they become more stable in terms of capacity growth and compute resources. This means that public and private infrastructures should be similar, if not compatible, to ease the work of developers. In this scenario, from the storage point of view, S3 compatible object storage is fundamental for reproducing public cloud services on-premises.

Enterprises are hiring more cloud native developers which are used to thinking about ephemeral storage for their VM instances and S3-like object stores as their persistent and shared storage infrastructure.

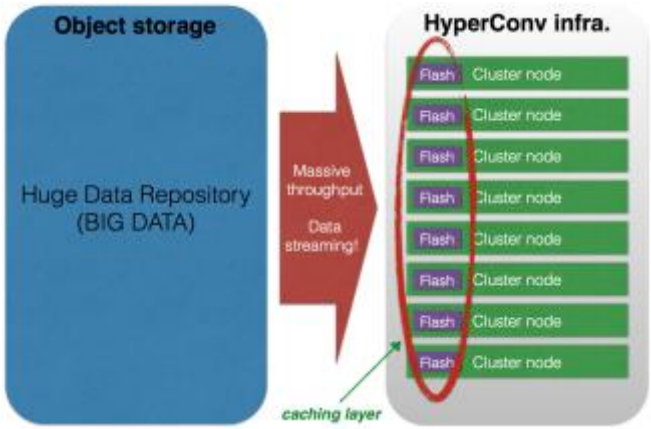
Enterprises are hiring increasingly more cloud native developers. These developers are used to thinking about ephemeral storage for their VM instances and S3-like object stores as their persistent and shared storage infrastructure, which means that they also expect more performance and predictability from the object store than in the past. This is one of the reasons behind the increasing demand for higher performance object storage systems, with solutions coming from traditional vendors starting to expose S3 APIs, as well as All-Flash based object storage infrastructures.

In a similar vein, we find Big Data analytics and, more precisely, in-memory applications, which pose another big challenge to traditional storage infrastructures where performance, throughput and scalability are involved. A tough problem to solve, where also HDFS, a file system specifically designed to massively scale, is not suitable for all kinds of workloads and file sizes.

OBJECT STORAGE, THE FOUNDATION OF MODERN DATA-DRIVEN INFRASTRUCTURES

New radical infrastructure designs are surfacing and becoming more common, with memory-based ephemeral storage or sophisticated caching mechanisms used as first tier and coupled with massive capacity-driven data stores at the back-end. In this scenario, object storage (which is historically considered suitable only for cheap, durable and cold data archives), is in the right position to become the perfect foundation for this new class of storage. Now, throughput with scalability should not be an issue - S3/Swift APIs are supported by an increasing number of software vendors and NFS/SMB gateways make it very easy to ingest data and remain compatible with legacy environments.

Object storage is no longer to be considered only a second or third tier storage solution. Some of its characteristics are perfectly suitable to be the backend of modern IT infrastructures which need scalability, performance and security.



For object storage systems, the process of streaming data to the compute cluster is seamless, while storing original data sets or results is less costly and more reliable than for any other kind of storage… and even more so now with the flexibility provided by cloud tiering mechanisms implemented by most modern platforms.

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.

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 FS and S3 gateway.

Its open source and software-defined nature is another key to its success. In fact, it's no coincidence that Ceph is already prevalent in OpenStack clusters and growing in popularity in a large number of object (and block) use cases.

A SINGLE PLATFORM FOR SEVERAL USE CASES

It's not easy to find a solution for different use cases when scalability, performance and security are all involved at the same time. But this is what end users want for their storage infrastructures.

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 simultaneosuly, to local and remote locations, is even more challenging.

Red Hat Ceph Storage is one of the few solutions which stands out not only because of its software-defined approach and modern design, which leverages an object storage architecture at the backend, but also because it was concieved from the start to serve multiple protocols and applications. In fact, this is why use cases for Red Hat Ceph Storage range from block storage for OpenStack infrastructures down to large multi-Petabyte systems for archiving.

Ceph is one of the few solutions which stands out not only because of its software-defined approach and modern design, which leverages an object storage architecture at the backend, but also because it was conceived from the start to serve multiple protocols and applications.

As commonly occurs 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.

THE ROLE OF OBJECT STORAGE IN BIG DATA ANALYTICS

End users are promptly embracing Big Data analytics (sometimes synonymously referred to as Hadoop). There are some important aspects to consider here:

- Even though HDFS is designed to scale, over time, cluster resources must be scaled differently than originally assumed. Storage could grow much more than compute or vice versa. And data retention can quickly lead to large capacity needs while size of active data sets remain quite unchanged.
- Furthermore, with consolidation in Big Data lakes, organizations are starting to use different compute resources: cloud-based and on-premises cluster with different applications.

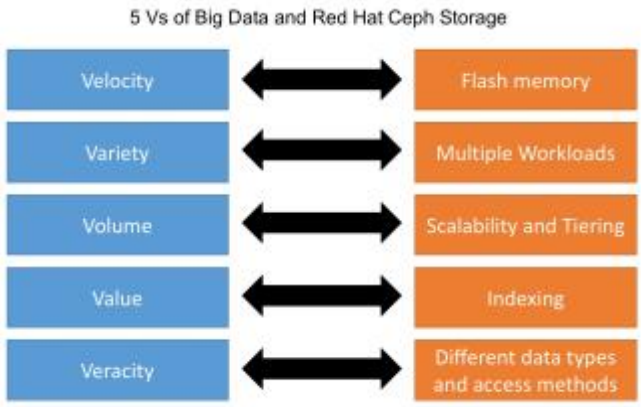
In this scenario, where flexibility and consolidation are key for top efficiency, two technologies are rapidly gaining attention for Big Data analytics: containers and in-memory databases.

- Big Data application containerization is happening: different application stacks use and access data differently and it is not unusual now to see activity running in real time alongside scheduled batch jobs. All of HDFS's limitations are exposed in these cases, especially when it comes to smaller files. An object storage system could be a good option here as it is much easier to access from remote compute resources, also enabling the expansion of the compute cluster using public cloud resources.
- Lately, in-memory databases have been chosen instead of Hadoop MapReduce-based application stacks for a greater number of use cases. This means the use of fewer large memory nodes. But the HDFS storage model, usually based on a large number of hard disks installed into the cluster nodes becomes unsustainable. These cluster nodes are alternatively filled with RAM and/or alternatively, NVMe devices acting as a ram extension, caching or first storage tier. As a direct consequence, data streaming capabilities provided by an external storage infrastructure become essential. Object storage is one of the best options in this case because of its ability to scale and sustain high throughputs at a relatively low cost while providing a highly reliable and durable persistent storage layer.

WHY USE OBJECT STORES IN BIG DATA INFRASTRUCTURES

Big Data Storage infrastructures must be scalable and accessible through various protocols. This is why S3 is the perfect match for HDFS. They are complementary to each other:

- S3-based object stores are typically less costly (especially when based on commodity servers), more scalable and durable;
- While HDFS can be faster, S3 could be more suitable for specific use cases, (i.e. IoT, log analysis or in any other case where a large number of small files is involved);
- HDFS can be implemented as front-end of an object store, allowing for more simple data movement and access to data through different protocols.



HDFS is typically considered as part of the first storage tier while S3 is preferred for near-line storage and specific use cases, (as in all those cases where a large number of small files is involved)

Many organizations are now taking great interest in new comprehensive data management strategies which involve building their own data lakes. These large-scale storage repositories are designed to save all kinds of

data and to be accessed by different compute resources, aiming towards the elimination of data silos and enabling the creation of a wider range of applications capable of better insights.

Object storage is the best option for building the foundation of a data lake.

Object storage is the best option for building the foundation of a data lake. For example, data could come from the most disparate sources and can be easily tagged and indexed during the ingestion process. Once data is safely stored as objects, it's available for metadata searching to all application stacks. Through specific API and commands, S3 based storage can also be leveraged as a primary store or as a secondary, but persistent, storage system. And if the object storage has tiering capabilities, it is possible to implement tiering (and retention) policies to move data across flash, disk and cloud or tape to improve the overall $/GB of the entire storage infrastructure.

Red Hat Ceph Storage, thanks to its flexibility and scalability, could be the best suited for building data lakes of any size. It's modern architecture, the improved performance of the latest version (version 2.0 released in June 2016) and its security features allow you to start off small and grow when needed. It's new policy engine associated to S3 protocol support will enable the construction of a multi-tier infrastructure that can span from low latency flash memory to the cloud to cover a very wide spectrum of big data applications, workloads and data types. It's flexibility allows enterprises to build storage infrastructures suitable for primary or secondary workloads depending on other Big Data technology already in use or planned.

BUILDING SUSTAINABLE LOG ARCHIVES

Even the smallest of data centers can generate Gigabytes of logs per day. In fact, many organizations are beginning to collect logs from remote offices and all kinds of devices under their control for the most varied reasons like auditing, security, monitoring, billing and so on. This trend will soon be characterized by an even greater flow of logs and sensors, due to the massive number of machines and devices connected to the internet, which continuously send home information.

Even the smallest of datacenters can generate Gigabytes of logs per day, and many organizations are beginning to collect logs from remote offices and all kinds of devices for the most varied reasons.

Logs are collected for many different reasons and they are processed accordingly. They are utilized in real time for monitoring or through batch jobs to build reports, trend analysis and so on. Sometimes they are archived as is, because of specific regulations or compliance standards. And in the latter case, it's highly likely that data must reside in the country of origin and cannot be uploaded to a public cloud storage (many European based firms have strict requirements for keeping their data 'in country' for example).

Finding the appropriate storage infrastructure to store log is very challenging, since its characteristics are very demanding and costs need to be kept low as well.

Finding the appropriate storage infrastructure for this task is very challenging, since its characteristics are very demanding and so is the cost, which needs to be relatively low: reliability, durability and scalability are on top of the list, but $/GB is crucial as well. The Total Cost of Acquisition (TCA) has to be very low and features which drive down the Total Cost of Ownership (TCO) are needed to make the infrastructure sustainable in the long term.

In particular, looking at TCO, taking data protection and other basic capabilities for granted, the most significant features should be related to how the log lifecycle is managed:

- **Multi-protocol access:** Logs should be able to be imported by different methods depending on the source. Only the most modern applications and devices can use HTTP-based protocols and in many cases, for example, file access is still the only way to move logs from the original application/source.
- **Performance:** the number of logs collected simultaneously could be massive, and this can generate a very large amount of network and storage traffic. Many sequential streams can be joined by uploads of single small and large objects depending on how logs are generated. In some cases, log analysis is performed in real time while data is saved on the storage system, or immediately after. In any case, the entire infrastructure has to be designed to avoid bottlenecks.
- **Caching or tiering capabilities:** It's highly likely that most recent logs are being accessed frequently for analytics and reports but will become inactive quite quickly. Mechanisms to grant

faster access to most active data while managing a low $/GB for capacity growth are at the base of infrastructure sustainability.

- **Automation:** by automating retentions management it is possible to alleviate sysadmins from tedious tasks, improve efficiency and avoid mistakes. This is especially true when logs must be stored safely for a long time for compliance with laws and regulations.

WHY USE OBJECT STORAGE FOR LOG ARCHIVES

In other words, a very flexible and cost effective storage system is necessary for an effective solution to the problem of log archiving.

Generally speaking, object storage can be a very effective platform for log archives, especially at scale. But Red Hat Ceph Storage has the right characteristics to be an ideal choice for this task. In fact, it is not only an object store - it can offer more when it comes to flexibility and efficiency. With the particular nature of logs, where a long term archive could be associated to a very active short term repository accessed by applications that need high throughput for analytics and reports, its architecture can make the difference.

Object storage can be a very effective platform for log archives, especially at scale. Red Hat Ceph Storage has the right characteristics to be an ideal choice as an object store for log archives.

The scalability of this type of platform is another key point. In fact, end users usually start collecting logs from few sources at the beginning with more and more consolidation happening over time. In this particular case, a scale-out architecture, like Red Hat Ceph Storage, is critical when aiming for a low TCO. The ability to start very small, in the order of a few tens of Terabytes, and grow to the Petabyte scale without changing the way the storage infrastructure is operated, added to the flexibility brought by the freedom of choice in its configuration, is a great advantage when compared to traditional solutions.

Furthermore, Red Hat Ceph Storage is a software-defined solution which runs on Linux and, for larger environments, it makes it possible to build a specialized hyperconverged-like infrastructure for log analytics and archive. Third party log analytics solutions as well as NoSQL databases can easily be installed alongside Red Hat Ceph Storage, in the same cluster, simplifying the infrastructure while optimizing the process.

CONTENT DISTRIBUTION FOR THE MODERN ENTERPRISE

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.

Even when archives are a reasonable size, on the order of a few Petabytes or hundreds of Terabytes, traditional file-based protocols, like SMB or NFS, are not adequate for long distance communication. They are 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 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.

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.

And now, with the rise of commercial HPC (High Performance Computing) and Big Data applications, the shortcomings of traditional file based protocols are becoming ever more evident at today's scale. HPC and Big Data applications are not only about business analytics and they do not produce only reports but rather, large and complex data sets and files. Examples are everywhere: for instance, it is now possible to compute your full genome profile for less than $1,000 (and quickly dropping) and this process produces around 1TB of data for each single person. It can be data that has to be stored forever and that can be potentially shared for clinical research or, in the future, streamed out to a research center to produce personalized medical treatment . Thanks to the Internet of Things (IoT) and many research projects, made possible by incredibly affordable and distributed compute resources, the number of open data repositories is rapidly growing both in the academic and enterprise fields.

WHY OBJECT STORAGE FOR CONTENT DISTRIBUTION

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 in general, 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 cache for 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

The different use cases for object storage, and Red Hat Ceph Storage described in this paper have more in common than one might think.

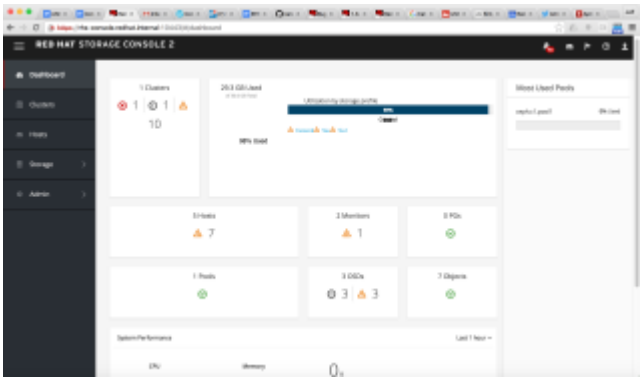
In all of these use cases it's plain to see that there are common characteristics which make these types of storage systems sustainable in terms of both TCA and TCO, especially in the long term:

Thanks to object storage technology, it is possible to build a data lake capable of concurrently streaming data, being the back-end for Big Data analytics and storing any other form of unstructured and semi-structured data, including logs.

- **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

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.

Red Hat Ceph Storage is ready to support multiple applications and workloads at the same time. In fact, all the applications described in this paper can fit in the Big Data category and they can all contribute to building a single, large and accessible enterprise data lake.



Red Hat Ceph Storage enables a smart data infrastructure that can start very small, in the order of tens of Terabytes, and grow to any size (into the multi Petabytes range) while being easy to use and manage, thanks to tools like the Red Hat Storage Console (see side graphic). New in Red Hat Ceph Storage 2.0, the Red Hat Storage Console simplifies both the deployment and operational management of Ceph, making scale-out storage accessible to a wider range of users while reducing deployment time.

Thanks to object storage technology it is possible to build a data lake capable of concurrently streaming data, being the back-end for Big Data analytics and storing any other form of unstructured and semi-structured data, including logs.

WHY RED HAT CEPH STORAGE?

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 sustainable 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.

The multi-protocol front-end, the object storage architecture back-end, it's modern distributed design, hardware independence and open, community based development model, all contribute to making Red Hat Ceph Storage a unique solution in the market, covering a growing number of use cases.

In a modern, two tier storage strategy which sees latency-sensitive workloads served by All-flash arrays and capacity-driven applications covered by high throughput scale-out storage, Red Hat Ceph Storage can play a perfect role in the latter category and, thanks to the latest improvements with Red Hat Ceph Storage 2.0, it can most definitely aspire to something more in the future.

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