JUKU

Think Global, Act Local!

Cloudian delivers object storage for next generation infrastructures

Prepared for: Cloudian

July 2014

JUKU
Think Global, Act Local!

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Traditional storage infrastructures are not designed to cope with emerging business needs originated by a surge in unstructured data, and they risk undermining the agility and the competitiveness of modern enterprise organizations.
It's not only about data growth. In fact, applications like Big Data analytics, mobile apps, active archiving, corporation "sync and share", and cloud-like services, just to name a few, work on wider and expanding data sets but also in a different manner when compared to legacy applications. Longer retention, better availability, low access rate, multi tenancy, geographical distributed access patterns, new resource provisioning/deprovisioning processes, as well as a general cost reduction should all be considered as key aspects of modern storage infrastructures.

At a higher level, cloud computing is emerging as the standard in realizing next generation infrastructures and, besides what is happening in the public cloud, object storage is the right companion for the private cloud too. Take Amazon AWS as an example, Simple Storage Service (S3) is one of its most successful products.

Modern object storage systems can be considered as an holistic platform capable of providing secure, available and reliable storage functions to a multitude of high end services. Services which range from infrastructure support (e.g. backup or archiving repositories) to backend for modern mobile services capable of improving end user experience and productivity (e.g. sync & share).

Cloudian Hyperstore® is enterprise ready public, private, and hybrid cloud storage that delivers a feature rich platform able to be deployed on commodity x86 hardware. Boasting limitless scalability, total S3 API compatibility, and auto-tiering capabilities, Cloudian Hyperstore provides the enterprise with total control while ensuring data can move seamlessly between private and public clouds.

INTRODUCTION

The challenges

We are getting used to what is becoming an exponential data growth. However, the problem is not about storing data but managing and taking advantage of them.

- Providing strong and secure distributed access interfaces (access from everywhere with any device),
- building next generation data services,
- building Big Data analytics repositories,
- managing lifetime long retentions policies,
- complying with laws and regulations,
- avoiding security breaches and data leaks,
- providing multiple-nines data availability and durability (which also means backup and disaster recovery).

These are only examples of what “managing data” means today, and facing these challenges with traditional storage systems, especially at scale, simply not possible at a reasonable cost and with the right amount of human resources.

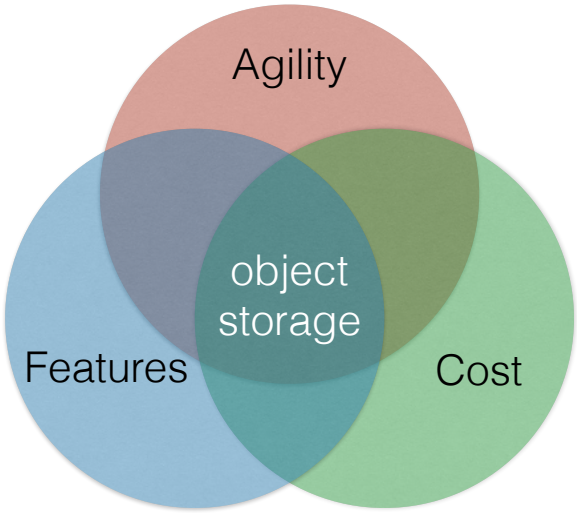
In fact, traditional storage architectures are limited in both scalability, automation capabilities and resiliency when it comes to large distributed infrastructures. On the opposite side, object-based cloud storage could be considered as an answer, while the choice between public and private implementations strictly depends on single organization needs.

Considering TCO and TCA

Total Cost of Ownership (TCO) and Total Cost of Acquisition (TCA) are two important key factors when we talk about storage infrastructures, especially when the numbers grow. Nowadays, the cost of most traditional storage systems is measured in $/GB. This type of cost is too high when you have to store hundreds of Terabytes or Petabytes of data, especially if they are not accessed very frequently. On the other hand, tape-based storage infrastructures promise costs in the range of ¢/GB but it has a very poor flexibility and speed when it's time to retrieve single items.

On one side we have DBs, ERPs, application servers and so on, on the other side there are all those kind of services and applications that rely on unstructured data: file servers, archiving, mail servers, web content, content management systems, backup.
Even some business applications like Big Data or traditional Data Warehouse can fit the model by storing data in the latter tier (i.e. huge data chunks, which are usually read sequentially, don't need many IOPS, and throughput is the most important metric) and running jobs in the high speed virtualized environment. In many cases the concept of unified storage, as we know it today, will no longer be applicable: in time you will need more specialized storage to solve both problems: one for speed, one for space.

Modern object storage systems show the right compromise: costs can be much lower than other types of storage but without waiving the features and functionalities. For example, most of the object storage systems can be heavily automated through management, retention or data protection policies, which free up system administrators' time from many repetitive management tasks. This makes them really agile and flexible, with the right characteristics for being the base of the next generation storage platform for any kind of organization.



WHAT OBJECT STORAGE CAN DO FOR YOU

Object Storage is an incredibly flexible technology and, when thought as a building block of a next generation infrastructure, it can be the solution for many kinds of different problems ranging from the consolidation of legacy services to the back-end for web and mobile applications.

When Object Storage makes sense

There are several use cases for Object Storage. In fact, object storage, usually accessed via RESTful API over HTTP as well as traditional file or block interfaces via protocol gateways, can be leveraged in many different scenarios:

You can think about object storage as a modern horizontal platform to manage many kinds of unstructured data without the limits and constraints of traditional storage.

- Next generation sync&share services for the distributed enterprise and mobile workforce.
- Large, secure and cost effective repositories for Big Data.
- Large, secure and cost effective repositories for Backups.
- Active archive and content management solutions.
- NAS consolidation and centralization of remote office (ROBO) data.
- Leverage Object Storage services to support Private and hybrid Cloud deployments.

These are only a few examples, but they should give a brief idea of the large range of possibilities of this technology. Most software vendors are

implementing object storage APIs in their products (Amazon S3, Swift CDMI) and the S3 API has become the standard option in most cases.

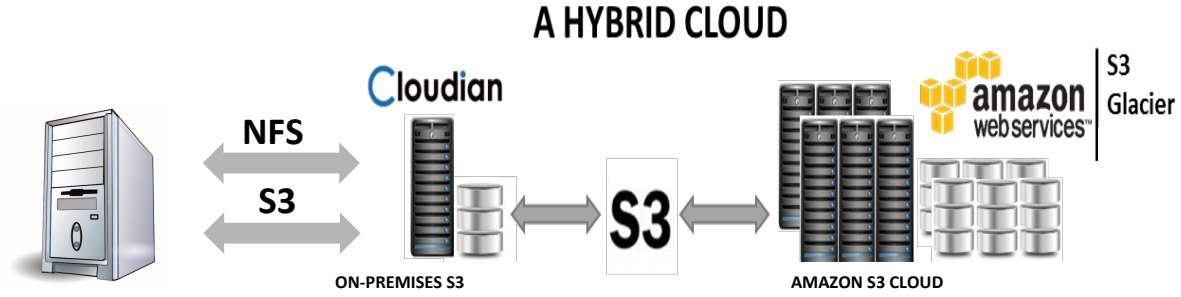
Through the adoption of Object Storage it is possible to cut costs while, at the same time, services are improved in terms of availability and resiliency.

Where Object Storage makes sense

From the infrastructure perspective Object Storage, public or private, shows its benefits when the organization has an important amount of data to manage and wants to target a lower TCO with an improved data availability and overall resiliency of the infrastructure.

Object Storage shows its value when scalability, “everywhere accessibility”, resiliency and availability are on top of the list of requisites. Cost is another key factor (as in every situation) but, in this case, for smaller quantities of data, public cloud could be much more affordable than private implementations.

With the latest generation of some object storage products it is possible to build hybrid infrastructures: for example, it could be possible to build a primary local site and use the public cloud for tiering or replication. This has a big impact on the architecture design and implementation costs, allowing medium sized organizations to take full advantage of object storage without the costs and the underlying infrastructure of a fully redundant architecture (i.e. multiple data centers).



AMAZON S3, THE DE-FACTO STANDARD

At the moment, Amazon, with its AWS (Amazon Web Services), is largely considered the most successful and largest public cloud service provider. Simple Storage Service, S3, is one of its most mature and adopted products, and it's widely considered the de-facto standard for modern public cloud storage implementations, especially for web applications.

What is S3

Amazon S3 is a highly distributed online storage service, primarily accessed via a REST interface.

S3 is a widely adopted de facto standard in the public cloud, but S3 compatibility is also a common request when the object storage is implemented on the premises.

The services was launched in US in March 2006 and then in Europe the following year. Since then S3 has been adopted from many developers and now it's a fundamental component of many successful cloud services like, for example, Dropbox. Amazon S3 has reported to store more than 2 trillion objects as of April 2013.

While Amazon does not disclose the S3 architecture, it declares a 99.999999999% durability, with 99.99% availability and a design that grants very high scalability, relatively low latency, and low cost. Data is encrypted while in transit and at rest.

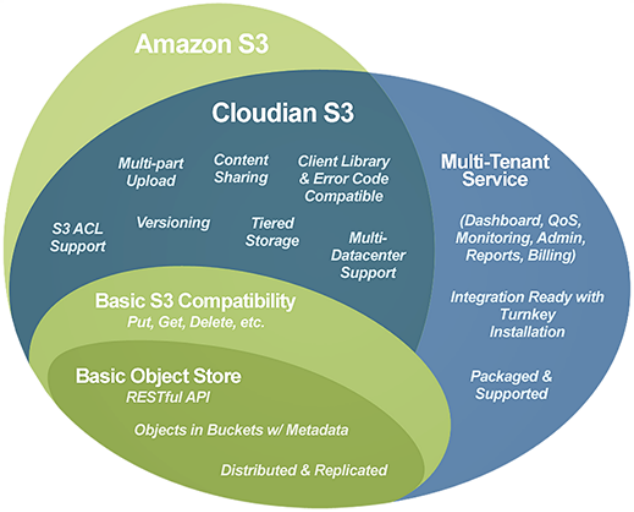
S3 is capable of storing files up to 5TB, alongside their metadata, in entities called buckets. Each bucket is owned by a single account and has its unique user-assigned key. A bucket can also be configured with ACLs (access control list) and objects are addressable through the REST interface as well as simple HTTP GET commands (or BitTorrent).

S3 APIs are not open and are controlled by Amazon, but they are very well documented. Even though SNIA (Storage Networking Industry Organization) has developed an open

standard (CDMI - Cloud Data Management Interface), the S3 PI is still, by a large margin, the most adopted API for object storage in both public and private implementations.
Today, there are thousands of applications that fully leverage the S3 API, from a variety of companies and organizations. At the same time, almost all Object Storage vendors have proprietary interfaces which are designed to leverage at best the characteristics of their product but, also in this case, users and third parties always ask for S3 compatibility to avoid lock-ins or the cost of developing specific implementations of their applications.

S3 for enterprise Object Storage

As already mentioned above, S3 is a widely adopted de facto standard in the public cloud, but S3 compatibility is also a common request when the object storage is implemented on the premises.



Being 100% S3 compatible has a tremendous advantage for the enterprise and in fact, many believe the S3 API should act as the sole standard. First of all there are no lock-ins to proprietary API and enterprises can use any of the hundreds of applications available through the S3 ecosystem without the painstaking process of rewriting or adding S3 support to existent applications. Secondly the end user has the freedom to choose what to deploy internally and what to deploy externally; and which services are better suited for the private or the public cloud. With S3 compatibility it is also easier to build true hybrid cloud architectures and, if provided by the object storage product of choice, implement automated tiering or other mechanisms aimed at saving money while providing a better service.

HOW CLOUDIAN DOES OBJECT STORAGE

Cloudian Inc. is a Silicon Valley based storage platform provider focused on delivering enterprise grade public, private, and hybrid cloud storage solutions.

A brief description of the company

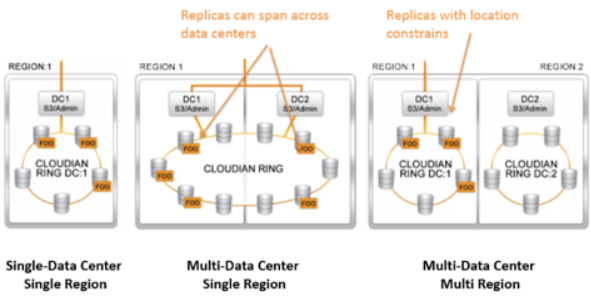
Cloudian, has offices in US, Japan and China with customers and operations in Europe too. Its product, Cloudian HyperStore®, is a S3 API compliant object store which can be used to build private and public cloud infrastructures by enterprises and internet service providers.

The company, born in 2001, was funded by strategic investors like Goldman Sachs and Intel Capital among others. It recently received a major round of funding from Intel Capital, Fidelity and INCJ.

Cloudian has compiled a long list of technology partners and integrations with prominent cloud platforms aimed at building a complete ecosystem of solutions for managing next generation content, file and infrastructure services.

A look at the architecture

Cloudian HyperStore, thanks to its flexible design, can be deployed in different ways ranging from a single instance into a Virtual Machine node (for test and development purposes) up to multiple Data Center topologies which are capable of granting the highest availability and resiliency.



HyperStore is based on distributed scale-out architecture and, as such, each node of the

cluster contributes with its resources to the total amount of space and performance.

Cluster nodes

HyperStore Nodes are usually based on standard x86 Linux servers and local disk space, while VMs can be considered as an option for smaller installations and development tasks. Although it is not considered a best practice in terms of load balancing. And different nodes can be mixed together in the same cluster.

Cluster nodes, which are organized in a ring topology, run a set of services to manage data, metadata and administration tasks. The ring can be distributed on multiple data centers for high availability and resiliency. Objects are transparently and homogeneously replicated on different nodes as a function of the chosen protection policy. They can also be logically replicated to remote sites providing an even better protection. If a logical or physical failure occurs, the system immediately checks which objects are involved and starts the necessary procedures to restore the optimal situation.

Cloudian Hyperstore® technology

Cloudian HyperStore® software, based on File System enhanced Cassandra NoSQL DB engine, allows to automatically manage small and large files with different protection schemes at the same time. The end user has the freedom to decide whether the system should maintain multiple object copies or leverage erasure coding to protect them. Almost all the configurations are very granular and can be applied system-wide, to a single tenant or down to the bucket level.

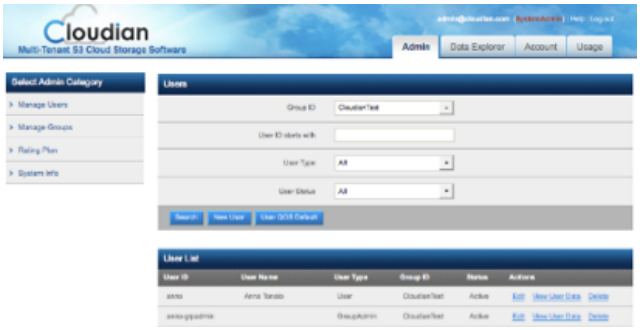
Multi-tenancy and QoS controls

QoS (Quality of Service) is becoming one of the most important features of modern storage infrastructures. Now, with many different applications and users accessing simultaneously at the same underlaying platform, it is necessary to provide QoS mechanisms aimed at avoiding resource contentions and "noisy neighbor" effects. Cloudian Hyperstore® implements granular QoS policies which enable to set various limits for different types of resources (region, group, user, system) in terms of requests/second, Bytes and number of objects transferred per unit of time, and so on.

Another important aspect regarding the multi-tenancy capabilities of this product is the useful set of features aimed at providing chargeback and showback functionalities.

Integrated management console

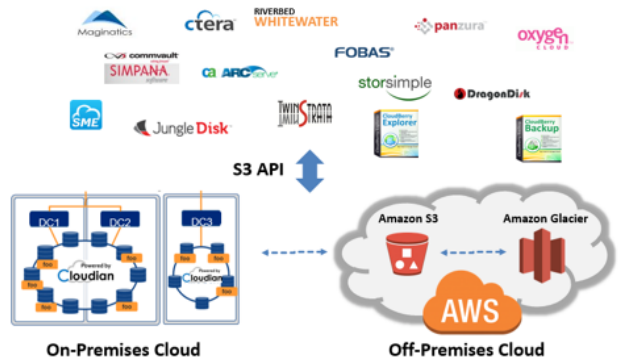
An easy to use web-based user interface is available for all the roles in the organization and helps to simplify configuration, monitoring and management operations. This complete management interface allows the product to be immediately deployed and used, since there is no need for additional integrations.



According to the scope of the product, all administration tasks are simplified to grant ease of use at scale. Automation tasks like node lifecycle, as well as its maintenance, are managed through a complete set of tools which come together with the product.

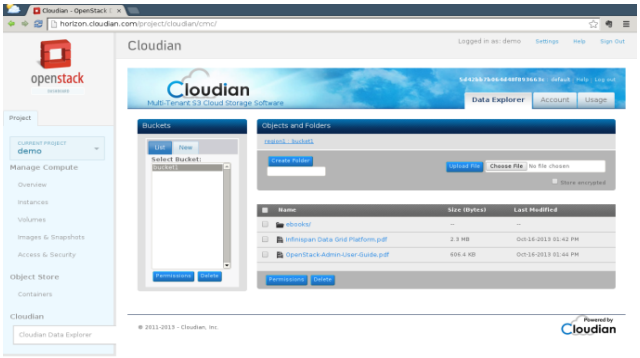
Cloud and third party integrations

Cloudian HyperStore, thanks to its strong S3 compatibility, can be implemented in hybrid cloud deployments where a private Object Store can be coupled with other S3 compatible public services. HyperStore's tiering functionalities allow the choice of maintaining critical



(and most used) data locally on premises, while the rest can be moved away to a remote cloud storage (wither a remote private cloud storage, or a remote public S3 cloud storage). This particular type of configuration can be of use in many cases, especially for small and medium sized organizations which could trade performance for great scalability at a lower cost.

Furthermore, integrations with most common cloud management platforms, like, Citrix Cloud Platform (based on Apache CloudStack),Citrix Cloud Portal, and OpenStack, are available and the end user can easily configure the product to be used in conjunction with them and implement STaaS (Storage as a Service) and all the other Object Storage services.



At the same time, Cloudian HyperStore is certified to work with many third party vendors to provide end-to-end solutions for sync&share, cloud NAS and other next generation services.

Why it is important for you

Cloudian HyperStore is a highly scalable S3-compatible object storage solution which is capable of covering needs of small and large organizations. The flexibility of the product and the licensing mechanism (pricing is on a per-TB basis) allow to start with a small infrastructure and grow, in function of future needs, on-premises or on the public cloud.

The characteristics of usability and resiliency are fundamental for building a private cloud storage environment. Its compatibility with the most common cloud management platforms out there enables the realization of an end-to-end private or hybrid cloud infrastructure.

BOTTOM LINE

Object Storage is attracting a lot of attention from service providers and enterprises. If the success of this technology is already consolidated for the ISPs (thanks, for example, to services like Amazon S3), for the latter it's a new emerging trend.

The first reason why enterprises usually look at object storage is because it's cheaper than traditional storage, but later they discover its key characteristics: the fact that it can be accessed from everywhere with any device; the fact that it can be reside everywhere; the fact that it can seamless scale; that makes it a very attractive solution for a much greater number of potential applications and opportunities. On the same wavelength is the growing landscape of solutions capable of taking full advantage from Object Storage: it is growing day by day, and it has to be considered a synonymous of market acceptance and success.

We, at Juku, strongly believe in object storage and we think that enterprises should take strong consideration this type of platforms when it's time to build a next generation infrastructure. In fact, object storage, can be considered a solution in all those cases where data doesn't have to be very tightly coupled to computing resources (like, for example, DBs or VMs).

Organizations of all sizes rely on Cloudian Hyperstore to deliver secure and total control over their data while dramatically reducing IT overhead and TCO. Hyperstore technology balances capacity with latency, giving users the power to set multiple data protection schemes (replication and erasure coding) within the same cluster. Combined with total S3 API compatibility, limitless scalability, and powerful features such as QoS controls and hybrid auto-tiering, Cloudian Hyperstore is a really interesting cloud storage platform for the enterprise.

Think Global, Act Local!

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 those who make decisions for their IT environments, to inform and discuss the technological side that we know better: IT infrastructure virtualization, cloud computing and storage.

Unlike the past, today those who live in IT should look around themselves: things are changing rapidly and there is the need to stay informed, learn quickly and to support important decisions, but how? Through our support, our ideas, the result of our daily interaction that we have globally 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 to provide a sincere and helpful vision. That's why we have chosen: "think global, act local" as a payoff for Juku.

Author

Enrico Signoretti, consultant, trusted advisor and passionate blogger (not necessarily in that order). Having immersed into 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 (but always with the Mac at heart) until now when he joined the "Cloudland". During these years his job has changed from deep technical roles to management and customer relationship management. In 2012 he founded Juku consulting SRL, a new consultancy and advisory firm highly focused on supporting end users, vendors and third parties in the development of their IT infrastructure strategies. He is constantly keeping an eye on how market evolves and continuously looking for new ideas and innovative solutions. You can find Enrico's social profiles here: http://about.me/esignoretti