

Object storage, cornerstone of a modern data management strategy

Prepared for: NetApp

Updated on September 2016

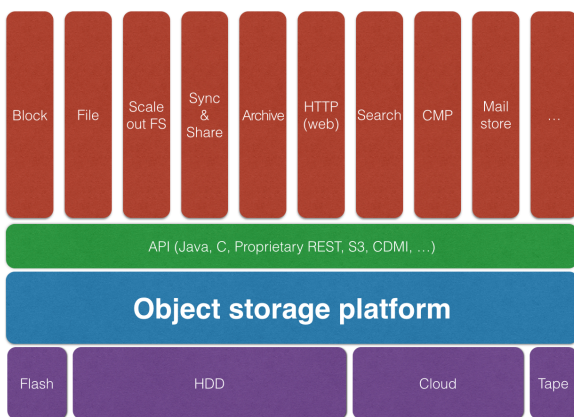
TABLE OF CONTENTS

EXECUTIVE SUMMARY	3
WHY OBJECT STORAGE TODAY	4
INFRASTRUCTURE SUSTAINABILITY	4
DISTRIBUTED PERFORMANCE	5
RISKLESS ADOPTION.....	5
A DATA PLATFORM FOR THE MODERN ENTERPRISE	7
APPLICATION RE-PATRIATION.....	7
MODERN DATA PROTECTION	8
PRIVATE CLOUD STORAGE	9
OBJECT STORAGE FOR EVERYONE.....	11
EASE OF USE	11
PERFORMANCE	12
MULTI-TENANCY	12
STRONG S3 COMPATIBILITY	12
BOTTOM LINE.....	14
JUKU	15
WHY JUKU	15
AUTHOR.....	15

EXECUTIVE SUMMARY

Business and end users expect much more from IT nowadays. Mobility of people and devices, new data access patterns, modern applications, and IT consumerization are creating new needs. Users want the best experience, unconstrained access to their data and flexibility while, at the same time, continuous data growth, security concerns and rigidity of traditional IT infrastructures make it very difficult to respond adequately.

Object storage is a cornerstone for a modern data storage strategy and can be used as a horizontal platform capable of serving many different needs including hybrid cloud storage deployments.



It is now evident that cloud computing, supported by a modern storage infrastructure, is fundamental for coping with this situation. Technology is finally maturing and many organizations are now leaning towards hybrid cloud infrastructures, where they find the agility, flexibility and savings they need to support users and business processes. Leveraging private and public cloud at the same time, allows to take full advantage of the reaction speed made possible by the public cloud and the savings consented by private infrastructures. Looking at this landscape, choosing the right hybrid multi-tier cloud

storage is essential for managing data locality, meeting regulations and company policies while providing tools to manage workload spikes and fast deployment of new projects.

Object storage is a cornerstone for modern data storage strategy and can be used as a horizontal platform capable of serving various needs including hybrid cloud storage deployments. Unmatched scalability, standard HTTP-based protocols plus the ability to geographically span across various locations, are its main differentiators and enablers. It can also be considered the perfect repository for any kind of data that has to be stored for a long time (like logs, for example) and computed at a later time if needed (Big Data analytics).

NetApp StorageGRID Webscale has all the characteristics to be leveraged as a crucial building block in these types of scenarios for both enterprises and xSPs. A rich feature set and flexible deployment options (i.e. software-only or appliance) make it suitable for many applications and use cases, including data archives, media repositories and web applications data. Last but not least, its potential is amplified by the ability to leverage different data protection schemes (i.e. local and global erasure coding as well as multiple object replicas), Amazon S3 compatibility and cloud tiering capabilities (useful for managing spikes at best while keeping costs under control).



WHY OBJECT STORAGE TODAY

Object storage can be considered a powerful backend platform upon which to build next generation applications. And it can also be leveraged to support the transformation of legacy infrastructures in modern data hubs capable of providing next generation services requested by end users and new business challenges.

Infrastructure sustainability

Storage capacity needs, imposed by the tremendous data growth we are experiencing these days, cannot find sustainable long term solutions in traditional storage approaches. There are several reasons why, but let me point out the most important ones:

Object storage can be leveraged to support the transformation of legacy infrastructures in modern data hubs.

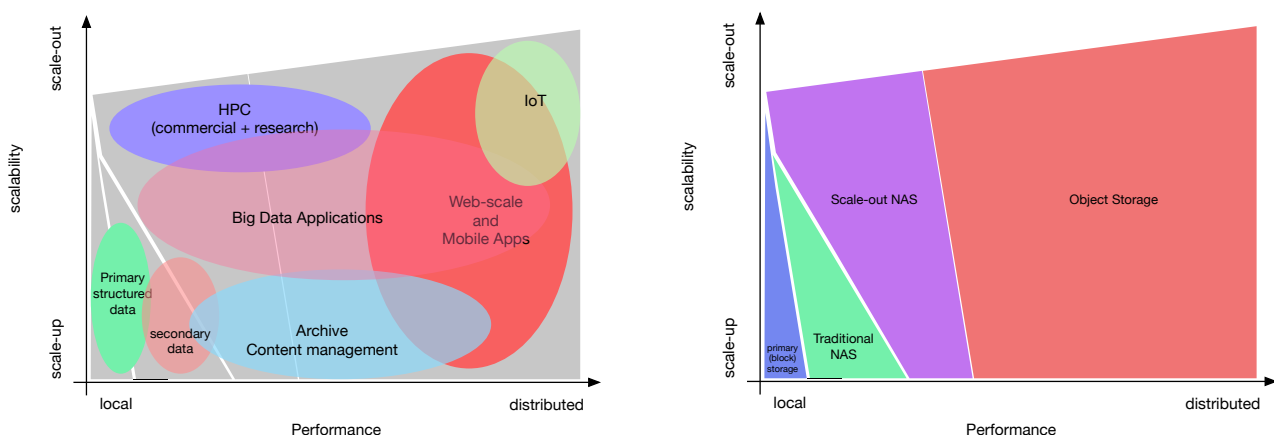
- **Cost:** TCA and TCO of storage infrastructures must decrease to become sustainable in the long term; Flash memory is of great help to drive down the costs of primary storage thanks to an undisputable \$/IOPS ratio but, when large capacity repositories are involved, \$/GB is the first parameter to consider and large SATA disks are still the preferred media. In this case it is not just the capacity of the single disk but how they are all organized and accessed that makes the difference.
- **Operations:** As a consequence of the previous point, operational models must change too. New technologies and tools are indispensable for SysAdmins to be able to manage several PBs per person instead of hundreds of TBs.
- **Scalability:** Traditional storage simply does not scale enough. It is not only about capacity or performance: If the object storage environment is correctly implemented, everything is designed to work at scale, availability and resiliency included. In addition to strong data protection schemes, like erasure coding, policy-based automation and distributed data placement ensure overall scalability of the system.
- **Data durability:** Large scale multi-Petabyte environments need the highest data availability. To assure data integrity and longevity, local and distributed erasure coding, alongside object replication, grant multiple nine availability with a much superior fault tolerance and overall resiliency. At the same time, at upper levels, automatic mechanisms to control data integrity are fundamental for data availability and longevity. Fingerprinting while data is ingested, continuous object checking and automatic rebuilds of suspected objects contribute significantly to increase overall data availability and resiliency.



Distributed performance

Another important aspect of object storage is performance. We usually think about storage performance in terms of IOPS, latency and throughput that are delivered within the data center. This concept is not applicable to distributed applications and to clients, like mobile devices which usually access data over long distance and from disparate locations. Distributed object storage architectures respond to the needs of distributed performance and throughput. In this case, performance is usually obtained through fast and reliable object streaming, load balancing and various caching mechanisms, which enable support for a multitude of concurrent clients simultaneously. This peculiarity, coupled with REST based protocols like Amazon S3, makes object storage particularly efficient for storing data coming from sensors and other sources, in any form and quantity; allowing enterprises to build an easy-to-manage data lake and to prepare for the Internet Of Things (IoT) era.

Object storage is now mature to support many non primary storage needs, web applications and smart data services like sync&share.



Internal design of the object storage platform is indispensable in order to obtain a system capable of streaming large objects with good performance. When a fail occurs, the stream should not be impacted. This capability results in increased predictability of the systems and some applications, like in-place Big Data analytics, will experience a lot of benefits as a result. But, in more general terms, throughput is important as well. In fact, organizations and providers of all sizes are looking at object storage as a great opportunity to offer next generation services such as, flexible and efficient data protection through Cloud backup gateways.

Riskless adoption

Object storage, contrary to the past, is now mature enough to support many non-primary storage needs, web applications and smart data services like sync&share. Important roles to implement these types of services can be found in Amazon S3 API compatibility, which grants access to more than 4000 different solutions already available on the market, and the ability of the product to scale according to the needs of the customers. Some modern object storage platforms can start very small now (beginning with a few VMs and taking advantage of the storage infrastructure already in place) and then grow from there with dedicated resources or through cloud tiering mechanisms. Since services like Amazon S3 (or, even better, Glacier) are very cheap in terms of capacity, but very expensive when it comes to data movement, the best cloud tiering implementations use a policy-based engine capable of making decisions following rules imposed by the

system administrator. With the freedom enabled by a product that includes all these characteristics it is possible to implement a cost conscious object storage platform without risks.

Viable installations are now possible starting from 100TB of usable space and the expansion can be served by more nodes or leveraging a public S3 compatible service.

Development and testing purposes can be served by a small VM-based infrastructure, for example. By moving from virtual to physical nodes it is possible to take full advantage of the platform in terms of savings and build from there. Viable installations are now possible starting from capacities as small as 100TB of usable space and the expansion of the initial

capacity can be served by more nodes or, with a hybrid approach, leveraging a public S3 compatible service. In practice, any form of scalability is very easy and inherent by design, keeping costs low costs while supporting gradual adoption, one application at a time.

A DATA PLATFORM FOR THE MODERN ENTERPRISE

The number of use cases for object storage are endless. The potential for this kind of platform can easily cover 80% of data managed by traditional organizations ranging from deep archives and backup to storage for remote locations and private cloud deployments. Contrary to what commonly happens for primary storage, managing a few terabytes or petabytes of object storage results in almost the same effort and this is one of the major benefits of this type of storage system. Applications taking advantage of native or S3 APIs, hardware or software appliance as well as specific gateways can leverage object storage as a huge data hub which allows maximum availability and resiliency at a minimum cost.

Application re-patriation

Enterprises have been building mobile applications to support their workforce for some time now and, in many cases, these applications are used frequently even when users are in the office at their workstation. Developing these kinds of applications on the public cloud is just easier and faster but when they go into production, the amount of data grows rapidly and the costs associated with them can quickly become unsustainable.

In other cases, especially in European countries, regulations concerning about data locality are very strict and building a private infrastructure is often considered the best option. Once again, having the ability to play on both private and public cloud tables at the same time is the best option to have all the advantages and none of the risks.

Having the ability to play on both private and public cloud tables at the same time is the best option to have all the advantages and none of the risks.

Application re-patriation could also be very difficult and expensive without the right tools and platforms. Consequently, the risk of trading flexibility for savings is very high. Bringing this type of application back on premises means having to build an adequate private cloud infrastructure and an object storage system must be part of it. API compatibility plays an important role in this process, allowing the user to move applications without rewriting code and having to do extensive testing activity.

NetApp StorageGRID Webscale, thanks to its rich set of APIs, including Amazon S3, can be considered an interesting option to build an infrastructure ready to host applications previously developed for the public cloud. At the same time, for newly deployed large private cloud installations, StorageGRID Webscale has all the characteristics to reproduce complex multi-DC/zone environments at a fraction of the cost while maintaining the same flexibility as the public cloud.

Modern data protection

Data grows and has to be protected, this is a fact. Traditional storage systems cannot cope with this kind of growth and the same is true for traditional backup. VTLs have only partially replaced tapes in large datacenters because the cost of maintaining long retention backups in them is too high. Furthermore, scale-up design of major VTL players create inefficiencies and scalability problems while costs more than double when it comes to adopting solutions like electronic vaulting for disaster recovery reasons.

Object storage, with the right cloud-enabled backup gateway, can easily improve cost and efficiency of backup procedures.

On the other hand, object storage, with the right cloud-enabled backup gateway, can easily improve cost and efficiency of backup procedures. This type of solution can be adopted by large organizations with their own infrastructure but also from smaller enterprises which subscribe to object storage capacity from

a service provider. By leveraging object storage at the backend most of the problems with traditional VTLs quickly disappear: for example, recent backups can be stored and restored locally, privileging speed, while long term backups can be seamlessly managed on cheaper and infinitely scalable cloud storage based on an object store.

Once the cloud-enabled backup gateway is adequately configured to manage the right throughput and some local storage for caching purposes, all the backend capacity lies on object storage. Electronic vaulting is no longer an issue because all data can be placed in a geographically distributed object storage repository. In the eventuality of a disaster, thanks to the fact that these appliances have all the configurations and metadata already stored in the object storage system, a new physical appliance or, even better, a virtual one, will be sufficient to recover all the data.

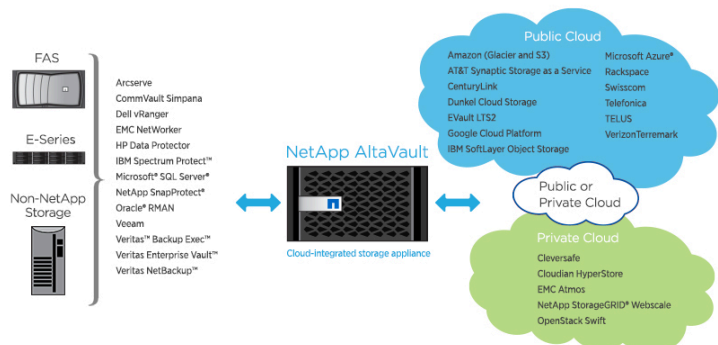
\$/GB of object storage can be very low and, even though it is not as low as tapes for the longest retentions, the overall TCO of the end-to-end solution based on a cloud-integrated VTL and object stores can be even better. In fact, an object-storage integrated VTL appliance can store short and long retention backup at the same time, with the same interface and without a practical limit in size. Savings occur at every level:

- simplification of the operations,
- better overall cost when compared to a traditional VTL and tapes solution,
- better flexibility with the adoption of virtual appliances that can cover remote and branch offices of any size.

This type of solution has the lowest impact on legacy backup procedures already in place (from the SysAdmin point of view it is just a VTL replacement) and finds its place in complex environments with multiple backup solutions. On the other hand, it is also true that modern backup software solutions now include direct support for saving data directly on object storage through S3 APIs, and in this case smaller organizations can directly take advantage of it without any additional layer.

AltaVault and StorageGRID with its tiering-to-the-cloud capability, is an end-to-end solution that can easily manage short and long retention backups within the same appliance.

NetApp, thanks to its AltaVault cloud-integrated storage appliance, has a complete end-to-end solution to address this need. Whether the end user chooses the public or private cloud, the appliance can leverage various object storage protocols and different back-end object stores. But even a greater value comes from the integration between AltaVault and StorageGRID Webscale with its tiering-to-the-cloud capability. This solution can easily manage short and long retention backups within the same appliance, bringing a tremendous simplification to the infrastructure and backup procedures.



Private cloud storage

Object storage has become a fundamental component of any successful private and/or public cloud deployment thanks to its scalability and multi-tenancy characteristics. Object-based private cloud storage is simple to implement and gives a fast ROI while improving end user experience.

Enterprises and Service Providers can leverage object storage to build modern data services without impacting the rest of the infrastructure and with a much lower TCO when compared to other solutions. Object stores can start very small and easily grow to petabytes just by adding new resources or, like in the case of StorageGRID Webscale, by taking advantage of cloud-tiering for inactive data.

Sync & share, a Dropbox-like experience but with the security and control only possible in a private environment, is always the first use case mentioned but the list is very long and includes:

- **Block/file storage gateways** for remote offices. Much easier to manage and maintain than local storage without the need of a local backup or remote replica for backups because data is actually stored in a remote data store.
- **Back-up gateways** for local and remote offices. As already discussed in the previous chapter, this solution brings higher efficiency and lower costs for data protection in those locations where local storage is still important.
- **Geographically distributed Name Space**. This kind of solution is very helpful for those users who concurrently work on large data repositories from various remote locations.
- **Content Management Platform repositories**. More than 20% of websites in the world run on Wordpress and 80% of Fortune 500 companies rely on Microsoft SharePoint (in most cases for



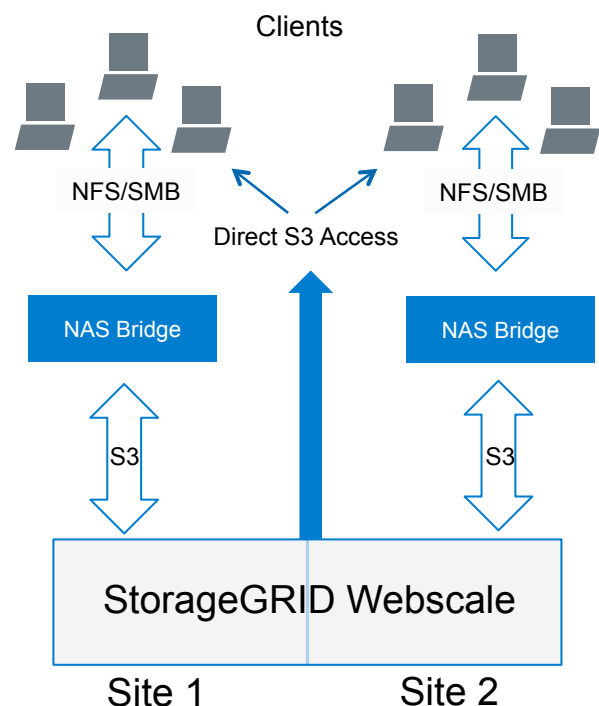
document management). Integration between CMPs and object stores can bring major savings and improve performance by off-loading bigger and/or inactive files from primary to secondary storage on the cloud.

- **Archiving applications.** This is a classic use case for object storage because it is much more reliable than any other form of file storage at a lower cost.

In adopting such a solution some principles can be assumed for both enterprises and service providers:

- Private and public object storage implementations should not be positioned in competition to Amazon S3 or any other big cloud provider. Price wars already in progress between Amazon, Google and Microsoft do not allow for other contenders. Both enterprises and service providers should always think about object storage as an enabler to build modern storage services that would be impossible to implement with other technologies. By adopting products that support cloud-tiering, services like S3 could become more of a partner than a competitor. With this same philosophy in mind, the value lies in the cloud storage services built upon it rather than the cost of the object store itself.
- Processes that are usually difficult with other forms of storage are automatic with object stores and can ease resource management as well as provisioning. Self-provisioning (a cloud pillar) is possible by design and control mechanisms such as chargeback, or Showback, and are much easier to implement thanks to how data and meta-data are organized and accessed. This approach to storage resources is not usually possible with traditional storage systems and resource provisioning can be a very long process in large organizations.

NetApp, thanks to the recent introduction of its NAS bridge has added a powerful end-to-end solution to its ecosystem which leverages NetApp's leadership in file services and brings it to the private cloud. These NAS gateways, usually deployed as VMs, vastly improve file services in remote locations and branch offices with integrated data protection and Disaster Recovery capabilities. Enterprises can leverage the same back-end for both backup and remote site data management and protection to give end users more reliable services while improving the \$/GB for the whole storage infrastructure without additional costs, since this feature is included with StorageGRID Webscale.



OBJECT STORAGE FOR EVERYONE

There are several types of object storage platforms now but some of them have specific features that are much more aligned to enterprise needs than others. In fact, looking the enterprise field for example, we see object storage growing very quickly in the number of installations but they are relatively small.

Generally, the majority of end users begin adopting object storage for one use case like, for example, backup, remote NAS or Sync & Share and then they expand over time adding more and more features. It is not unusual now to find installations well under 200TB, sometimes in the range of 50-100TB which are also growing in capacity and number of stored objects quite quickly. This could be considered a sort of entry level for object storage now.

There are several types of object storage platforms now but some of them have specific features that are much more aligned to enterprise needs than others.

In fact, the reality is that enterprises have always looked at the application side of all solutions and not just to the technical or infrastructure aspects. With object storage quickly becoming the back-end of many solutions and the first choice for a large number of developers, no matter the quantity or quality of stored data, it is consequently gaining a lot of traction and more enterprises are adopting it.

Despite the overall scalability of object storage being taken for granted, achieving this goal is not that simple and the term scalability can have different meanings for enterprises or ISPs. Other characteristics, not always found in traditional object storage products, can be very important for a successful deployment, especially when the trend described above is taken into consideration. Even small ISPs, which usually start to deploy object storage only for a specific solution, or a limited set of end users, are looking at these characteristics favorably:

Ease of use

TCO is usually the first reason for an enterprise to look at object storage. \$/GB is crucial to have a sustainable storage infrastructure over time, and while the first look usually goes to TCO and features which are about the overall system efficiency, one of the most important characteristics, common to any other type of storage system, is ease of use. In fact, even though systems could start small, the idea is to have the ability to manage Petabytes of storage per SysAdmin instead of Tens of Terabytes.

In small IT teams, where any SysAdmin has multiple roles, a good GUI, a simple management API and, more in general, ease of use have a strong impact on day to day activities and maintenance operations.

A good user interface, is fundamental to achieve this goal and move the needle from TB/SysAdmin to PB/SysAdmin. The life of an object storage cluster is not only about basic maintenance operations, like disk

failures for example, but over time the system grows and needs to be continuously expanded with new hardware nodes, as well as updated, including the decommissioning of older storage servers.

In small IT teams, where any SysAdmin has multiple roles (ranging from server to storage and network admin), a good GUI, a simple management API and, more in general, ease of use have a strong impact on day to day activities and maintenance operations.

Some Big Data infrastructures are now based on in-memory applications which can leverage object storage as a persistent storage layer asking more performance than usual from it.

Performance

Object storage is not associated with high performance, but even though it will never be as fast as a primary storage system, because of its nature, its performance is becoming more and more relevant.

There are two aspects to consider regarding object storage performance. First of all, with the deployment of a larger number of small object stores (as previously mentioned, 200TB or less in capacity) performance becomes more important than ever, the ability of the storage cluster to save and serve objects relatively quickly with a limited set of resources is crucial and enables smaller installations. Secondly, and more in general, the number of applications leveraging object storage as a back-end is growing exponentially. Some applications use it as a media or file repository but, with the increase of IoT and Big Data analytics applications, the size and quantity of objects stored is changing dramatically. Even more so, some Big Data infrastructures are now based on in-memory applications which can leverage object storage as a persistent storage layer, expecting higher performance than usual from it.

Multi-tenancy

End users even the smaller ones, who usually start with only one application, quickly realize the power of object storage and migrate more and more data to it (e.g. Sync & Share, Backup, remote NAS gateways and so on), and they keep adding new applications over time.

Strong platform multi-tenancy is now more important for enterprises and ISPs than for the largest private object storage implementations, which usually serve less, but huge, application data repositories. Another benefit of multi-tenancy is the simplifications of storage charge-back or show-back processes, simplifying them while giving more tools to the enterprise and the ISP.

Strong S3 compatibility

Nowadays S3 is the most important protocol to access an object store and S3 compatibility is critical for successful object storage deployments. No one wants to check API compatibility with its own applications, S3 is the standard and end users need compatibility because it allows seamless application migration (like in the case of application repatriation discussed earlier). At the same time, a good S3 compatibility allows for an unmatched freedom of choice when it comes to frontend applications and gateways, which are usually designed to support, at least, S3 APIs.

NetApp has recently released a new version of StorageGRID Webscale which is clearly embracing all these enterprise and small ISP needs allowing a larger number of end users to adopt object storage and take full advantage of it. The product which had already shown its scalability, data protection features and a built-in solution for NAS environments, as well as backup, in previous releases have now improved usability, performance and S3 compatibility widening the scope of the product and its use cases.

Nowadays S3 compatibility is the most important protocol to access an object store. And no one wants to check API compatibility with applications, S3 is the standard and end users pretend it.

BOTTOM LINE

Object storage enables Enterprises and Service Providers to manage multi petabyte secondary storage environments with ease, and now also smaller end users can take advantage of it. Products like NetApp StorageGRID Webscale are clearly looking to improve efficiency (also in terms of performance and with a limited amount of resources) and ease of use, allowing any kind of storage admin to deploy the solution.

This is not in direct competition with traditional file and block storage, which is usually intended to serve frequently accessed data and transactional workloads. The goal with Object Storage adoption should rely on the implementation of next generation data services and on renewing part of the legacy IT infrastructures by using it as a horizontal platform capable of scaling independently from the rest of the stack and with the multi-tenancy capabilities needed to support many different applications and users at the same time.

Implementing object storage today can result in an immediate improvement of the competitiveness of the IT infrastructure and, consequently, business functions. Its characteristics are aligned with the most recent IT paradigms of cloud computing bringing more flexibility, hybrid cloud functionalities, unmatched scalability, lower costs and simplification to the entire infrastructure.

In this scenario NetApp StorageGRID Webscale can suit for both Enterprises and Service Providers. Its rich feature set, strong S3 compatibility and multiple data protection mechanisms can respond to a wide range of needs from small development projects to multi-petabyte distributed environments. And now, thanks to the improved UI and higher performance, it can address a larger number of use cases. The risk in its adoption is even more limited thanks to the availability of the product as a software, that can be also installed on VMware or OpenStack VMs, up to high-density appliances.

The product has seen a lot of improvements in the last 24 months and has a solid roadmap ahead. Recently improved features like stronger S3 compatibility and a modern GUI, are just a follow up of the work seen in the previous releases, focusing on the goal of widening the number of possible use cases.

One last note goes to the ecosystem. NetApp, thanks to the AltaVault appliances, and its VM-based remote NAS gateway is able to propose a powerful and brilliant end-to-end solution for the distributed enterprise and the ISP that is interested in building a strong set of ready-to-market solutions based on end-to-end enterprise grade storage. At the moment, NetApp is also the only company which can leverage an object storage repository, like StorageGRID Webscale, as a back-end infrastructure for snapshots offloading from primary storage thanks to Solidfire, a kind of feature which is promising to become common to all its products thanks to the Data Fabric vision the company is pursuing.

Other solutions, which address vertical markets, are available thanks to partners, while the number of compatible third party solutions is increasing as well thanks to S3 compatibility.

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 NetApp. This paper was commissioned by NetApp. No employees at the firm hold any equity positions with NetApp. Should you have any questions, please contact Juku consulting srl (info@juku.it - <http://juku.it>).