Think Global, Act Local!

A close look at application-aware object storage

Prepared for OpenIO
September 2016

Think Global, Act Local!

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Enterprise IT is facing an incredibly difficult challenge nowadays, especially when it comes to storage and data management. On one side there is the public cloud with its unmatched flexibility but with unpredictable and often, high costs. On the other hand, there are traditional infrastructures, which are more predictable, but seem unsustainable in the long term, if not sooner.

One of the biggest problems is unstructured data growth, and not only for rich media files but also big data sets, IoT sensor recordings and logs of all kinds are growing systematically. End users have been seeing an exponential trend for a while now. In fact, we are already in the range of 80 to 85% of unstructured data versus 15% of structured data sets (like in the case of relational databases), but the growing amount of never-delete retention policies and rich data types are quickly moving the needle towards an even more staggering unstructured/structured ratio, which has been experiencing great demand for highly scalable, automated and low TCO storage infrastructures.

To fully understand the scenario, other aspects are to be taken into consideration from the infrastructure standpoint. Some organizations are not large enough to build a private cloud infrastructure capable of solving all their needs; finding the right skills while making changes on how the infrastructure is managed takes a lot of time. This is also why most enterprises have already identified the hybrid approach as being the best compromise. In fact, it is not uncommon now to see the development and first deployment of new applications in the public cloud followed by a repatriation process when they get more stable in terms of resource consumption.

Consequently, enterprises are much more interested in solutions that can scale on-premises, as well in the cloud, capable of managing any boom in demand by responding immediately to all kinds of requests while allowing time to figure out the best data placement in terms of security, performance and cost. Having the ability to seamlessly leverage public cloud and on-premises infrastructure allows enterprise IT to provide a cloud-like and virtually-unlimited resource experience to end users.

IT organizations are aiming at next generation infrastructures capable of sustaining growth at a reasonable TCO but they also want to make them invisible to applications and end users. Object storage is a perfect component for this kind of infrastructure

Eventually, IT organizations will be aiming towards next-generation infrastructures capable of sustaining growth at a reasonable TCO (Total Cost of Ownership), but they also want to make it invisible to applications and end users, hence more flexible and agile. Object storage is a perfect component for this kind of infrastructure, but not all object storage systems have the same characteristics.

The ideal platform has some fundamental characteristics such as:

- being able to start very small (in the range of tens of Terabytes, or even less when intended for a development environment),

- being highly efficient and with a small footprint (a solution that could be delivered via containers if necessary).
- providing full S3 API compatibility (now the de facto standard in object storage, allowing developers to move applications without modifying them),
- granting multi-protocol access (to ease the transition from legacy file-based applications to object storage),
- being easy to deploy and use on any kind of hardware or virtual environment.

Having these characteristics and more, OpenIO is a good choice for users looking to build next-generation infrastructures in a cost-effective, yet flexible and scalable way. Some highlights are:

- It can scale from a few TBs to several PBs in a single domain namespace;
- Its modern design allows the system to adapt itself very quickly when the infrastructure underneath changes, taking advantage of new available resources immediately;
- It is set up to support different workloads and workflows thanks to native interfaces and connectors;
- It has an automated tiering capability which allows the end user to offload cold data from the primary on-premises system to public cloud services like BackBlaze B2.

Additionally, OpenIO is an open-source product, giving even more freedom to end users.

THE EVOLUTION IN OBJECT STORAGE DEMAND

We usually think about object storage as a huge backend infrastructure that starts becoming worthwhile when it is larger than 1 PB in capacity. This is in fact the most common scenario but, at the same time, there is a new trend which is favoring relatively smaller, faster and more agile object stores. And this is particularly true when these object stores serve a single vertical application or a specific use case.

Today with the right technology, object storage can be adopted by organizations of any size without undermining scalability over time. In other cases object storage must be faster than in traditional deployments.

Today, with the right technology, object storage can be adopted by organizations of any size without undermining scalability over time. In other cases, object storage must be faster than in traditional deployments. Use cases like email storage, for example, or event driven workflows (like in the case of AWS Lambda), need a prompt response, throughput and higher performance. Particularly efficient and fast object stores are essential when the volumes are small because the limited amount of available resources, while internal and to-the-cloud automated tiering mechanisms contribute to improve overall efficiency and rapid expansion at a reasonable cost when necessary.

Furthermore, by analyzing adoption patterns of object storage, the vast majority of enterprises start small with a single application and then grow by adding more and more applications or by converting processes already in place (like for example when they migrate from traditional backup systems, like a VTL, to an object storage repository).

Small object stores are also in high demand by developers. The number of developers using S3 APIs and object storage as their primary storage system is dramatically growing, but only few of them need to access Petabytes of data at the same time. For them, even though object storage is the platform of choice, single application usually accesses just a few TBs of data at a time. Having a platform that allows both approaches (small and large) is the key to enable the developer and, consequently, the enterprise to choose the right configuration and deploy it without constraints or limitations.

Once again, having the ability to deploy an efficient object storage system enables the developers to build applications with embedded storage that can be easily installed on premises or in the public cloud, enabling an unmatched application portability, especially if you want data close to the application and complete control over it.

Having the ability to deploy an efficient object storage system enables the developers to build applications with embedded storage that can be easily installed on premises or on the public cloud.

BEYOND OBJECT STORAGE

Looking at the scenario depicted above, it is evident why for next generation, cloud-native solutions, developers are favoring object storage instead of other systems. It is much easier for them to deal with it while the rest of the IT organization has less to tend to when it comes to security and scalability.

However, in order to be really effective, object storage must be designed to be as easy as possible to deploy and manage. Smaller organizations and teams do not have the necessary skills to manage complex infrastructures and they want simple UIs as well as APIs while, for large enterprises, this brings a better TCO. Another aspect of manageability comes from the ability of the object store to be very granular in terms of cluster expansion and quick in taking advantage of new available resources, both when they are deployed on-premises in the form of additional nodes or through cloud tiering. In other words, the storage infrastructure must adapt immediately to infrastructure changes.

The next step for object storage is an evolution to become smarter and more application aware by running portions of code directly into the storage itself in a "server-less" fashion

Software-defined storage solutions fits perfectly in this model, allowing to build a storage infrastructure that can be separated from the hardware layer, becoming more transparent and closer to the applications. Even more so, the next step is an evolution to become smarter and more application aware by running portions of code directly into the storage itself. As with AWS Lambda, it is possible to enable the developer to offload event triggered and highly repetitive throughput-demanding tasks directly to the storage platform in a "server-less" fashion. There are several use cases, including face recognition, video transcoding, text/email object indexing and so on, spanning from a mere better usage of resources up to security and auditing needs.

With these kinds of features implemented at the storage level, development is simplified, performance can be more consistent and part of the processing is done closer to the data without moving it around, asynchronously while it is being stored. As a consequence, scalability of the application is simplified and becomes more aligned to the quantity of data under management and the size of the storage infrastructure.

WHY OPENIO

OpenIO is a young startup which is developing a next generation open source object storage system. Its characteristics are truly innovative, allowing to overcome some of the limits of traditional solutions.

Usually, object storage systems leverage consistent hashing algorithms, which enable great scalability and node balancing. Unfortunately, when the configuration of the cluster changes, a rebalance is necessary before taking advantage of new resources and, in the meantime, performance is penalized. OpenIO, thanks to a radically different design, avoids cluster rebalancing of this sort when nodes are added (or removed) and solves the problem. At the same time other automatic load balancing techniques are in place to ensure that the cluster always responds at its best.

Contrary to what usually happens to other object storage systems, OpenIO integrates an embedded data processing framework allowing to offload some operations directly to the storage platform.

Contrary to what usually happens to other object storage systems, OpenIO integrates an embedded data processing framework, called Grid for Apps, allowing to offload some operations directly to the storage platform. This functionality is very helpful when developers need a scalable backend for intensive data processing applications but it is also useful to improve overall infrastructure efficiency. In fact, thanks to Conscience technology, a continuous metric data collection from the cluster nodes associated to algorithms to calculate the best match for any I/O request in real time, an OpenIO cluster is always dynamically balanced while releasing unutilized compute resources to applications.

The use cases for Grid for Apps technology are many (including image recognition, video transcoding, log analysis, email filtering and so on), but considering a hybrid infrastructure, with compute nodes running on private as well as public clouds and data stored on a local cluster, overall efficiency and cost savings can be very relevant, allowing to deploy smaller and cheaper cloud instances while the heavy lifting is offloaded to storage. This kind of storage system can lead to event triggered data processing systems similar to what has already been seen with Amazon AWS Lambda, but on a private infrastructure.

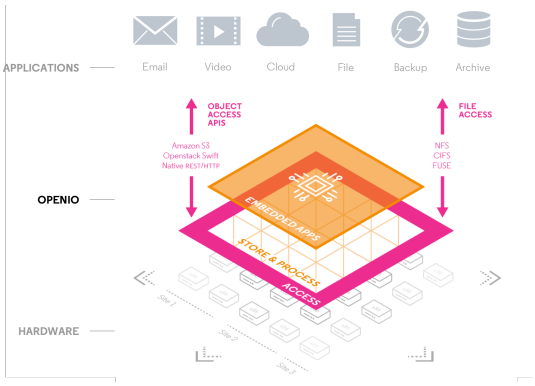
Thanks to OpenIO it is possible to build a modern smart storage infrastructure which can offer advanced data services capable of offloading specific application tasks and freeing compute resources.

The OpenIO small footprint allows to start very small: from 3 nodes and a few Terabytes, up to hundreds of PBs and trillions of objects. The nodes can be physical, virtual or even containers while the installation process takes no more than a couple of minutes and a few of clicks per node. It is incredibly simple to use and its modern UI contributes to lowering the bar for its adoption in smaller teams, and from developer.

Thanks to OpenIO it is possible to build a modern smart storage infrastructure which can offer advanced data services capable of offloading specific application tasks and freeing compute resources. This capability

is not entirely unique in the storage landscape, but this is the first time it has been implemented in a product so accessible from the TCA (Total Cost of Acquisition) point of view and which has an open source license.

Grid for Apps and Conscience technology are key features which allow OpenIO to redefine the concept of storage scalability and make it more versatile.



BOTTOM LINE

Object storage has been growing in popularity since 2006 with the introduction of Amazon AWS S3. Now standard capabilities of object storage platforms include S3 and Swift APIs compatibility, as well as scalability, resiliency and availability, but the most interesting products are taking a step forward.

Startups like Open IO are embracing a new concept of smarter storage, which enables a new class of data services very similar to what you can find on public cloud services like Amazon AWS but on premises. These products are leaner, more flexible and faster, with a new architecture design that adapts more rapidly to infrastructure changes, allowing the storage layer to become more invisible and easier to deploy.

OpenIO is embracing a new concept of smarter storage which promises an unmatched efficiency while offloading some data-driven operations directly to the storage infrastructure.

OpenIO (http://openio.io) is clearly part of this movement. Some technical choices make this product interesting for small installations or vertical applications without limiting future growth. The well designed UI makes it easy to use, but it is the back-end framework for application that makes the difference. It promises an unmatched efficiency while offloading some data-driven operations directly to the storage infrastructure. At the same time the product has all the features you usually expect from a traditional object storage platform, like erasure coding, multi-site and multi-geo replication, Automated and manual tiering as well as file gateways, like NFS, which allows to access objects as files.

OpenIO is clearly aiming at building more than a traditional object storage system, but a platform that integrates storage and advanced data services for next generation applications.

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 are 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