Next-Generation Object Storage and Serverless Computing Explained

June 2017

Table of Contents

Executive Summary

The world is entering what can be called the data-centric era. The quantity of machine-generated data is already overwhelming, and this is just the beginning. With big data, industrial internet of things, and machine learning initiatives becoming increasingly common among organizations of all sizes, we will need to rethink from the ground up the way we store and compute data, while keeping costs as low as possible. Traditional infrastructures are not designed to cope with these new challenges and, at the same time, scalability, flexibility, and efficiency are the foundation of overall infrastructure sustainability, and are fundamental to supporting future storage and data processing needs.

STORAGE TYPE	LOCATION	PROTOCOL	#DATA/ACCESS RATIO	PERFORMANCE	SCALABILITY
Block	Data center, close to servers	FC, iSCSI	N:1 - 1:1	High IOPS Low latency Consistency	Usually scale up (less than 1PB)
File (NAS)	Local network, close to clients	SMB, NFS	1:N	High latency Good IOPS Good throughput	Scale up + limited scale out (multi PB)
Object storage	Internet, accessible from everywhere and any device	HTTP, APIs	1:Millions	High throughput High parallelism Higher latency	Scale out (hundreds of PBs, Exabytes)

There are several types of storage systems, and each one is best for different workloads and applications. Object storage is ideal for large scale internet-based applications because data can be stored alongside metadata, and this data can be accessed from anywhere and on any device. It has always been very focused on low $/GB and high scalability but some of its characteristics make it rigid and not very user friendly.

OpenIO is changing this paradigm. Based on a modern and innovative lightweight design, it blends efficiency, flexibility, and ease of use with the benefits usually found on other platforms.

OpenIO SDS is open source software that can be installed on ARM and x86 servers, making it possible to build a hyper scalable storage and compute platform without the risk of lock-ins and with a very good TCO for the highest and fastest ROI.

OpenIO SDS is perfect for traditional use cases (such as active archiving, big data, private cloud) but at the same time, in conjunction with Grid for Apps, it opens the door to applications that need much more sophisticated back-ends to operate. These applications include industrial IoT, machine learning and artificial intelligence, as well as any other workflow or job that can benefit from automated tasks. The ability of the storage infrastructure to intercept events and trigger applications to offload specific data and metadata operations, seamlessly and transparently to the rest of the stack, is a game changer that simplifies software development and vastly improves overall infrastructure efficiency.

OpenIO SDS and Grid for Apps are the ideal choice for building flexible storage infrastructures with integrated compute capabilities that can handle several types of data and capacity-driven workloads. This software meets the needs of today and of the future.

Next-generation Object Storage and Serverless Computing; that's what OpenIO's software is about.

Next-Generation Object Storage

OpenIO has some unique characteristics that set our solution apart from others on the market.

If you look at the object storage landscape today, you'll find that most available solutions are able to do what you expect from an object store. Scalability, data protection, replication, S3 compatibility, and many other basic features are taken for granted now. While some systems implement a specific feature better than others, the differences are minimal, and various object storage systems are adopting the same feature set across the board.

The problem is exactly this: they all look alike. Based on similar concepts, and with a similar rigidity, they have the same limitations when it comes to deploying, managing, and scaling an infrastructure, and the results are nearly always comparable.

First, the overall TCO is higher than expected, and the constraints impose infrastructure design choices that are not always aligned with the evolution of a specific business. Second, in the real world, scalability is much more of an issue than you may think. It's not that these systems don't scale; but every time you introduce new resources in the system, everything has to be thoroughly planned in advance and has an impact on performance.

This is why we think OpenIO SDS can be considered Next Generation. SDS overcomes all the limitations of traditional object stores while giving the same, or better, functionalities.

The lightweight design of SDS, capable of running on a single CPU core and 512MB of RAM, coupled with the flexibility provided by Conscience Technology, make it easy to manage cluster resources very efficiently.

Flexibility, a synonym of freedom of choice in this case, allows our customers to build all-flash as well as all-disk and hybrid configurations, starting with 3 nodes and growing up to hundreds, while still having the ability to add a single hard disk to a node if necessary. Users can mix any combination of ARM and X86 nodes in the same cluster, and performance is usually better than for the competition. This is because a small footprint means that the system is optimized and efficient. The flexibility that comes from supporting heterogeneous hardware is made possible by Conscience technology and the dynamic load balancing mechanism it provides.

Next Generation comes from all of this. It's a totally new way of thinking about an object storage cluster and how it works. Services and features are quite similar to what you can expect from the competition but, again, it's the simplicity, efficiency, performance, ease of use, and overall flexibility that sets us apart.

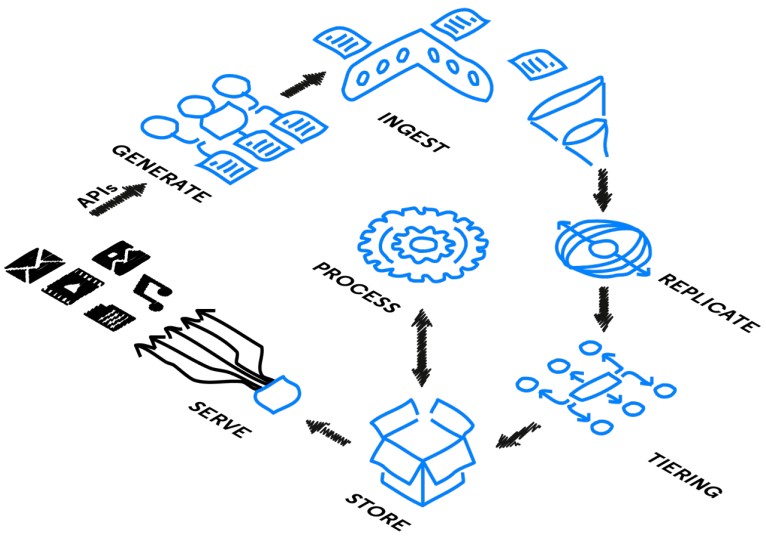
Serverless Computing

OpenIO SDS is an open source software solution that runs on commodity hardware. If you have a modern datacenter server, and your software only needs only one core and 512MB of RAM to run, what do you do with all the rest of the available resources? Why not use them to offload compute tasks from the rest of the infrastructure? This is exactly why we decided to use the word Serverless.

Grid for Apps is an event-driven compute framework that works on top of OpenIO SDS. It intercepts all the events that happen at the storage layer, and it is able to trigger specific applications or scripts to act on the data (and metadata) stored in the object store. By consolidating data and applications on the storage infrastructure, you save on external servers and have fewer other components to manage.

There were a number of words that could describe this, but Serverless was the most appropriate. Grid for Apps is also very similar to what you get from AWS S3+Lambda in the public cloud.

Our customers love Grid for Apps, and they are already using it for applications such as metadata enrichment, data indexing and search, pattern recognition, machine learning, data filtering, video transcoding, and so on. Grid for Apps simplifies many operations and workflows, and is very easy to adopt. So Serverless could mean that you need fewer servers to do the same job; or, to put it another way, you don't need external servers to run a lot of external tasks.



Takeways

The first goal of any IT professional is to avoid complexity. OpenIO SDS simplifies the life of developers and system administrators while contributing to lower TCO. By adapting OpenIO SDS to our customers' requirements, we provide flexibility and efficiency. Next-Generation Object Storage: because OpenIO SDS is different at the core.

The same goes for Serverless computing. The small resource footprint needed to run SDS allowed OpenIO to build more around the core and take advantage of all the resources available in the cluster. This improves efficiency (less data moving back and forth) and saves money (no external resources needed to run compute tasks that can be offloaded to the storage infrastructure). Once again, Grid for Apps reduces overall infrastructure complexity while doing more with less!.

Next-Generation Object Storage and Serverless Computing

openio.io

FR
3 avenue Antoine Pinay
Parc d'Activité des 4 Vents
59510 Hem

US
180 Sansome Street, FI4
San Francisco, 94104 CA

JP
1-35-2 Grains Bldg. #61
Nihonbashi-Kakigara-cho,
Chuo-ku, Tokyo , Japan 103-0014