

Dailymotion selects OpenIO for easy scaling of its storage system and to optimize use of its machine pool

Stakes

- **100% hardware agnostic** with improved data reliability and security
- **50% reduction in TCO** and Opex model
- **Primary storage tier** thanks to **high performance**
- **Simplified expansion, reconfiguration and fault management**
- **Geographically distributed** infrastructure

Deployed data storage infrastructure

- **1 OpenIO cluster** consisting of **28 machines** • **1.4 Billion objects**
(800 million videos + associated thumbnails)
- **40 PB of usable storage**

dailymotion

ABOUT DAILYMOTION

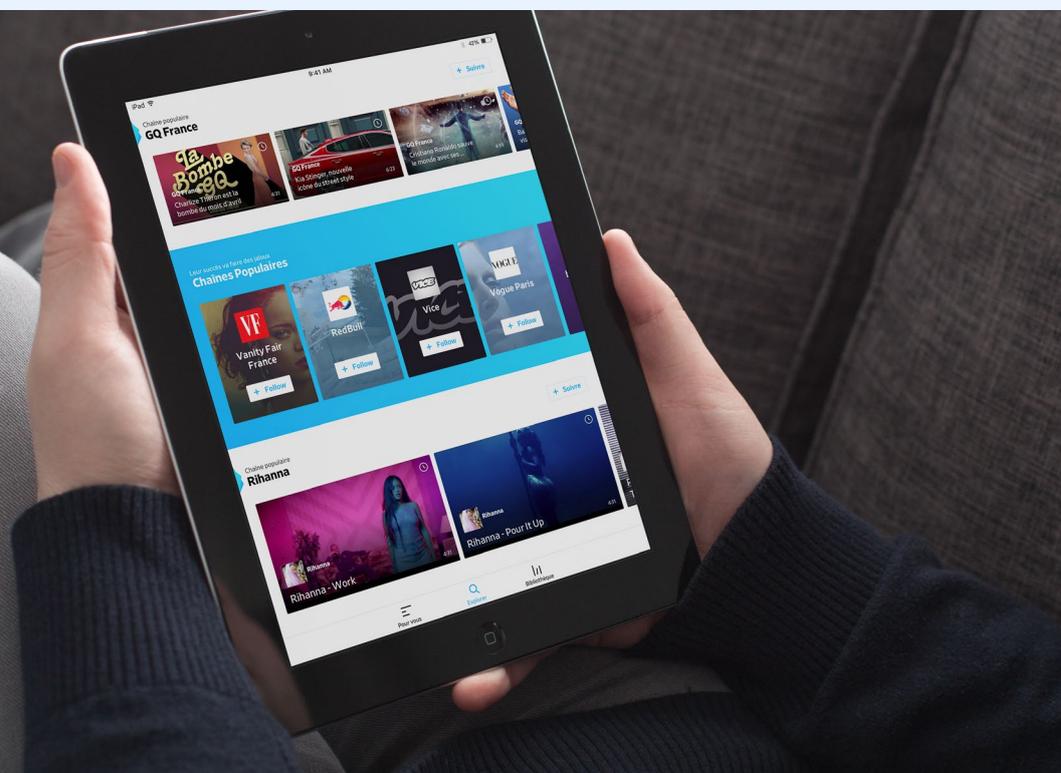
French company created in 2005.

Sector: Online platform for hosting, sharing and viewing video.

Locations: Head office: Paris.
Other locations: New York, Singapore, Marseille.

Success: Now part of the Vivendi group, Dailymotion is an international video streaming platform with more than 300 million users who can create their own personalized selection of videos related to their favorite subjects. Based on an intelligent player, an intuitive algorithm, and human recommendations, Dailymotion attracts more than 3.5 billion video views each month.

www.dailymotion.com





Q&A with

Alan Martins Head of Operations & Infrastructure at Dailymotion

and Thomas Gerbier System and Network Administrator at Dailymotion

You decided to make your application completely agnostic regarding back-end storage solutions. Why is that?

Historically, our storage system was based on scale-out NAS technology, which means scalable, file-based network storage servers. About ten years ago this was the best solution available, but storage technologies have changed a lot since then. The growing popularity of object-based storage has been a real breakthrough, allowing the consolidation of “big” storage systems. The result is not just cost savings - in our case, up to a 50% reduction in TCO - but also improved data reliability and security.

Of course, when your platform hosts more than 800 million videos like ours at Dailymotion, the underlying technology is not something you can change overnight. Migration is a significant operation, storage is an essential component of our infrastructure and so the choice of technology has real strategic implications for our business. Dependence on a single technology appeared risky, so we wanted to make the Dailymotion application agnostic towards the storage technologies used, both from a hardware and software point of view.

With this kind of technological independence, it's possible to organize storage system redundancy across several different technologies. This helps us to control costs with competitive tendering, but also limits the risk of possible failures. It would be truly bad luck for a bug or human error to affect both technologies at the same time!

Finally, maintenance windows opened up significantly when you can switch from one system to the other in a way that is transparent to users.

Why was Dailymotion interested in OpenIO Object Storage software-defined solution?

The first step in our strategy of decentralizing our technology was to build a backup of the main storage system, in order to have a mirror of the platform in case of a major incident. To build this backup platform, we began deploying Object Storage, using a market-leading technology from France. This type of storage theoretically provides infinite scalability, unlike the NAS solution which forced us to create multiple storage silos.

Storage is at the heart of your business. What are you looking for when choosing a storage technology?

Cost is a key criterion given the sheer volume involved. We also want a good split between Capital Expenditure (Capex) and Operating expense (Opex). NAS-type technologies require large up-front investments in hardware, whereas with Object Storage you can start small and add additional hardware resources to scale-up as required.

Cost is also impacted by how well the technology allows us to optimize hardware resources. OpenIO is truly hardware agnostic – it is able to work with machines of different characteristics, capacities and generations. This allows us, for example, to plan for a gradual refresh of our server park, taking advantage of the latest generation of components without having to part with our older machines.

The second criterion is, as you suggested, scalability: the freedom to easily extend our platform capacity by adding new machines, and to see immediate benefits. Then comes the ease of

deployment, configuration, use and maintenance. And resilience: how the platform behaves in the event of a problem such as the loss of a disk, or a failure in one or more machines. How long does it take to rebuild? Of course, this all leads to a related question: is the support reactive, or even proactive as is the case with OpenIO?

The fact that a technology is open source is also an asset: it helps to guarantee the durability of the solution we're investing in. That's financial investment, but also the time and human resources we devote to ensuring our team learns about the new solution

What about performance?

We had initially considered Object Storage for our backup platform. However, due to its performance, Object Storage soon proved to be a satisfactory solution as primary storage, serving video directly to users of our web application behind a CDN.

With Object Storage, items can be accessed using HTTP protocols from any location or application. Object Storage brings the promise of linear growth in performance as the platform size increases, and unified access behind a de facto standard (S3 or Openstack Swift) that is simple to integrate into applications that handle only immutable data.

What's more, the data protection provided by this distributed storage approach is appropriate to both primary storage and as backup, provided the infrastructure is geographically distributed within several distinct incident zones.

Why were you surprised by OpenIO's performance?

We first tested OpenIO to back up our existing storage system and it allowed us to take control of the technology. Grid architecture offers many advantages. Cluster reconfigurations can be done in minutes with no impact on performance. Expansion, reconfiguration and fault management become simple operations that can be done during working hours, and the additional capacity of the new resources is available in just a few minutes, since there is no data rebalancing.

There are few constraints, the hardware is so totally agnostic that, as mentioned above, all different kinds of resources can be exploited to their maximum capacity. Storage policies are also highly customizable, which we demonstrated by adding a new requirement for the storage of thumbnail images in addition to videos.

Three years ago, we decided to make our OpenIO cluster the passive cluster of our main storage system. Several times we have switched the entire production to this cluster, for maintenance operations or due to an incident on the active cluster. The switchover has been completely transparent to users. This is why we are considering switching the system to active/active mode as part of our dual sourcing strategy, with a possible offline backup system. We always maintain the approach of not relying on a single technology, however reliable it may be!



Learn more:
www.openio.io

