OpenIO

Kayrros uses Big Data to make the energy market more transparent

Stakes

- **The business is data centric**: data is the most valuable asset

- **Control the storage and operating costs of a growing volume of data**

- **Have a well performing storage system, so that datasets can be used by the machine learning algorithms** (Hadoop cluster + Spark; results collected in Elasticsearch and presented to clients through an online interface)

- **Be able to develop the storage system over time** to meet the needs of the company, whether in rapid development or a diversification of its activity to other energy sectors.

KAYRROS

ABOUT KAYRROS

French start-up created in 2016. Already raised 32 million euros.

Sector: Economic intelligence/strategic analysis of data of physical assets.

Locations: Paris (head office), New York, Houston, London and Singapore.

**Clients**: Industrial companies, banks and financial institutions, governments, NGOs.

Activity: Kayrros leverages cutting-edge technologies, particularly satellite imagery and artificial intelligence, to provide an innovative technological platform for asset observation. First the company applied this vision to the energy market, and they are now expanding to observe any large physical asset across the globe.

Success: Technology and industry expertise are key to Kayrros' success. It's a combination of the tech side—artificial intelligence, big data, satellite imagery and machine-learning algorithms—with a deep understanding and expertise in the assets that they cover, such as ships, oilfields, dams and power plants, that has propelled Kayrros to where they are today.

www.kayrros.com

Deployed data storage infrastructure

- **1 OpenIO cluster** made up of **20 machines** (server hosting provided by Iguane Solutions)

- **2PB of usable storage** (of which 80% has been used)

- **Steady growth of + 2–3TB of data each month**, sometimes with much **larger data spills of + 100 TB** (for example a study into a new region)

- **1GB average size of a file** (satellite image)
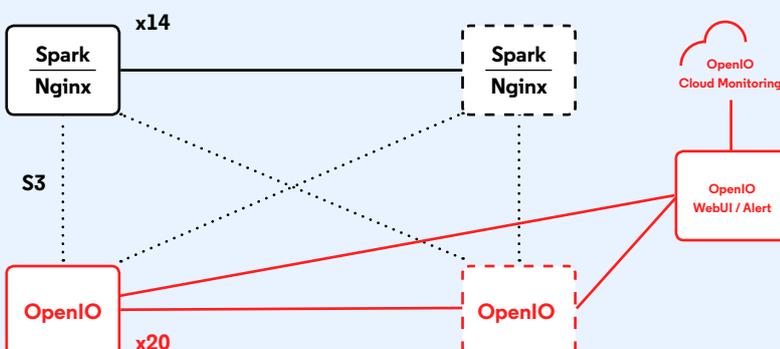
1x SERVER SPECS

- **Intel® Xeon® Silver 4114**
- **128 GB RAM**
- **2x 10 GbE**
- **16x 10Tb HDD**
- **2x 380Gb SSD**

Q&A with Sébastien Blot

VP for Infrastructure & Security at Kayrros

In 2018, Kayrros chose OpenIO to establish your own storage cluster. Where did you store your data before?

Kayrros used a leading Public Cloud provider of object storage, which seemed like an ideal solution to quickly launch a project as it offers great agility: no machines to be deployed, no minimum requirements, pay-per-use.

However, with the growing volume of data that we accumulated in the framework of our activity, costs started to soar and there was no possibility to reduce them with a pay as you go method. In reality, it's the exact opposite: the Hadoop cluster that we used to analyse data wasn't hosted on this service, so data had to be transferred from their Cloud for each job, which led to an "overbilling" (depending on the number of requests and bandwidth consumed). In other words, using Public Cloud Object Storage was not financially sustainable for Kayrros. And we had to act quick: the more data we accumulated, the higher the bandwidth costs to retrieve it.

Which reasons led you to choose the software-defined storage technology developed by OpenIO?

Kayrros chose OpenIO after researching several object storage solutions. OpenIO's performance meets our needs, and the core of the project is open source. This guarantees the continuity of the solution, and it allows our team to understand the main operating principles, something we couldn't do with proprietary solutions whose design and mechanisms are often opaque.

To build our OpenIO cluster, Kayrros asked OpenIO to recommend the best servers suited for our needs, and that would work with the web host Iguane Solutions, who provides outsourcing.

In the end, we recreated the services that the Public Cloud provider offered us, using our own services and by building our own private cloud, which is not accessible from the public network. In doing so, we have improved our performance, and we have reduced our costs by 40%. And we're still agile: simply by anticipating needs, the web host adds nodes when necessary to increase the capabilities of the platform.

Plus, OpenIO supports hardware heterogeneity: today we have an OpenIO cluster made up of 20 identical machines. In the future, we will be able to take advantage of the latest hardware developments (servers as well as media storage) without having to renew our park of servers. It's an undeniable advantage!

Compared to the Public Cloud Storage that you used previously, the TCO of data storage has already decreased by 40%. Is it still possible to further optimise your storage infrastructure?

Today, Kayrros has not yet ventured into data-tiering; all data that we collect is considered hot. Data-tiering, which is part of our project, could further reduce the TCO, for example by placing data that's older than 3 years in cheaper storage media, or by decreasing its protection level.

THE BENEFITS OF OPENIO FOR KAYRROS, ACCORDING TO SÉBASTIEN

- **Ideal performance for Big Data**
- **TCO reduction of 40 %**
- **Pay less, do more**: given that there is no billing for data usage, more computing campaigns can be launched without being constrained by financial issues
- **Control and prediction of costs**
- **Longevity, open source** is at the core of the solution
- **Possibility to carry out data-tiering** within the same cluster to optimise costs and performance