

Why the future of cloud storage is globally
distributed

| Featuring
Forrester

Watch the on-demand webinar
→

**STORJ**[Developers](#)[Product](#)[Learn](#)[Partners](#)[Pricing](#)[Company](#)[Sign in](#)[Start for free](#)

Using Storj DCS to Share Large Datasets with DeSci

Dr. Antonin Portelli

April 4, 2022



**Subscribe to our
newsletter**

Email*

Submit

Our work as theoretical particle physicists requires computer simulations on large scale supercomputer systems, with a workflow that has to handle enormous datasets. Projects often involve international collaboration and several supercomputers scattered around the planet. This creates some unique challenges, especially in terms of data storage.

Not only is the scientific data that we generate extremely large, it's extremely expensive to generate. Some of the datasets take years of computing to generate, which in turn significantly increases the value of the data. Additionally, as scientists we have the duty to ensure the results of our work are reproducible in the long-term future, which requires the preservation and curation of the data that led to scientific

Share this blog post



Related blog posts



**Test Storj for Video
Sharing**

publications. For these reasons, we have a critical requirement of being able to store these high-value datasets on a long-term basis in a highly resilient manner.

Another aspect of the data challenge that we face is that we need to be able to make these large datasets accessible to other groups of scientists scattered across multiple continents around the globe—scientific groups that often work independently of each other. The fact that scientific collaborations are often naturally decentralized, combined with the other challenges we face, prompted me to explore how Storj Decentralized Cloud Storage (DCS) might be able to help us.



Customers and
Featured Forrester
Analyst Advise
Enterprises on
Decentralized
Storage



Test Storj for Video
Streaming

Decentralization in the Scientific World: DeSci

The truth is that the usage of decentralized technology isn't new to the world of science or particle physics. A great example of this is the work being done with CERN's Large Hadron Collider (LHC). Even though the LHC is located in Geneva, Switzerland, the scientists doing the data analysis that produce the associated physics are scattered around the world. In other words, the European Organization for Nuclear Research (CERN) has been deploying a large distributed storage and computing system for years, the [LHC Grid](#).

My goal with Storj DCS was to assess its performance and resilience capabilities for our large datasets. This included not only a large number of files, but also very large files in regard to size. And we were particularly interested in its performance stability in relation to accessibility from diverse geographic locations. As a result, we conducted a series of tests based on synthetic data from January to February 2021, and published our results in a detailed report on October 13, 2021. In short, our ultimate findings were favorable in terms of how the multilayered parallelism of Storj DCS optimized edge-based performance for data transfer and geographic accessibility. We are now continuing this collaboration with Storj and aiming at reassessing this outcome through real scientific use cases.

Study Process and Results

The main objective of the study was to qualitatively determine the efficiency of Storj DCS for synthetic data by simulating typical large datasets used in high performance computing. To test the performance of the Storj DCS network our study employed three different approaches:

- Transferring 4GB files in parallel using the Rclone client
- Attempting to increase parallelism by manually splitting the files into smaller 64MB chunks
- Transferring 128GB files in parallel using Storj DCS native parallelism

We used the approach of manually splitting files into smaller chunks to give us a baseline of what would be the highest expected performance. However, in reality manual chunking wouldn't be the most practical approach, and our tests didn't take into account the overhead and time of manually splitting and reconstituting files. That said, the Rclone utility and DCS' native parallelism still achieved impressive results in comparison.

For example, Rclone transfers of 4GB files uploaded at 1GBs compared to 5.2GBs of a manually chunked file. Even more impressive was that an upload of a 128GB file with native DCS parallelism achieved 4 GBs compared to 4.8 GBs with manual chunking.* Native parallelism downloads performed at about 2.7GBs compared to 5.7GBs of a manually chunked file, which as mentioned above does not account for the chunked file reconstruction time.

One important result that our tests showed is that there is a minimal difference in download performance based on location with Storj DCS. We generally had excellent download rates whether they occurred in Edinburgh or various locations in the U.S.

Overall, we were impressed with Storj DCS' out-of-the-box performance in moving large datasets, which we feel is a direct

benefit of the way its decentralized network is structured. The excellent download speeds that we observed from locations that were far away from the upload location is something that would also be valuable to us.

Even though higher performance and throughput aren't usually associated with decentralized storage solutions, we found Storj DCS has the potential to enable multi-GB speed with increased parallelism, redundancy, and resiliency. We are looking forward to extending our collaboration with Storj, and extending our study on how their decentralized technology can provide solutions to the ever-growing challenges of data sharing in high-performance scientific computing.

For more details on the study, [download the full report](#) or view a [webinar of Dr. Portelli discussing the report's findings](#).

Download the Report

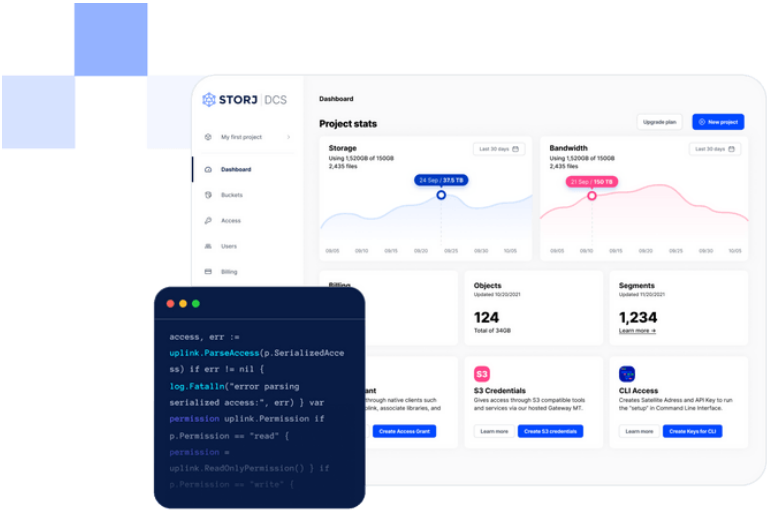
Watch the Webinar

* Note: the Rclone utility will have access to Storj DCS native parallelism in the future, which should enable it to achieve native parallelism's higher results.

Build on the decentralized cloud.

Get S3-compatible object storage with better security, performance and cost.

Start for free →



Switch to Storj and get ridiculously resilient cloud object storage with blazing performance and zero-trust security.

Subscribe to our newsletter

Email*

Submit

PRODUCT	RESOURCES	COMPANY	LEGAL	CONTACT US
Home	Docs	About	Privacy policy	✉ info@storj.io
How it works	Blog	Careers	Terms of service	📍 1450 W.
Pricing	Forum	DEI	Terms of use	Peachtree St.
Integrations	Resources	Contact	Node T&C	NW #200
				PMB 75268

