flexFS Cryo-EM Analysis Benchmark

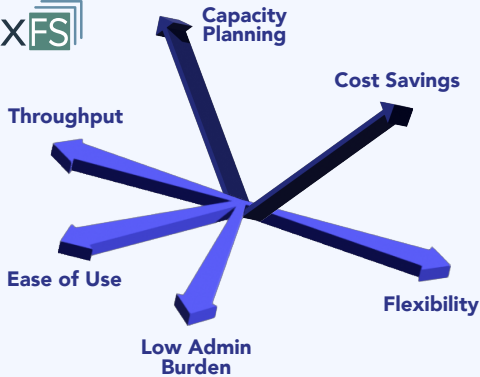**Capacity Planning** · **Cost Savings** · **Throughput** · **Flexibility** · **Ease of Use** · **Low Admin Burden**

- flexFS is a high-performance, POSIX-compliant elastic file system designed to overcome scalability and throughput limitations

- Built on an object store (e.g., AWS S3), flexFS easily serves 1000s of concurrently mounted hosts with minimal infrastructure

- Significantly cheaper than mainstream alternatives, get faster time-to-results & better capacity planning

Cryo-EM Analysis: Scale Your Science, Not Your Spend

Why pay more for less?

- Lustre costs 3.5 times flexFS for 21% slower end-to-end performance!

- EFS costs at least 1.75 times flexFS—not including per-GB read and write costs—and is 29% slower than flexFS

- flexFS is faster than either Lustre or EFS at every step of the analysis workflow.

- flexFS does not charge for data ingest and egress in contrast to EFS

- flexFS is truly elastic – up and down. You only pay for storage you use. Lustre needs pre-provisioning in increments of 2.4 TB, is difficult to shrink capacity, and you end up paying for unused storage

Benchmark Details

- Time and costs for analysis and storage of an EMPIAR cryo-EM dataset (10288, CB1) computed using three different file systems on AWS: flexFS, FSx for Lustre, and EFS

- EMPIAR 10288 contains 476 GB raw data comprising 2,756 multi-frame TIFF images. This dataset was chosen due to existing literature and benchmarks in the field.

- Cryo-EM analysis typically performed using CryoSPARC or RELION. We chose RELION and executed a typical single particle analysis workflow: Motion Correction, CTF Estimation, Particle Picking, 2D Classification, and 3D Reconstruction.

- Two sets of analyses done for each file system – i) single instance; ii) concurrent runs on 8 instances.

- g5.12xlarge EC2 instances used each featuring 4 A10G TensorCore GPUs and 48 vCPUs.

- Cost comparison in figure 1 is for 100 TB of storage - easily reached with a few cryo-EM projects

- Reprocessing with different parameters or validation between algorithms requires accessing data multiple times resulting in increased costs for I/O and compute.

- Figure 4 highlights cost savings at scale, typical for large organizations in the life sciences industry.

Results

- The I/O pattern in cryo-EM analysis has both throughput and latency challenges

- While EFS and FSx for Lustre support a balance of throughput and latency, achieving high throughput performance becomes increasingly costly for both options.

- flexFS demonstrates high performance at low costs for throughput & latency sensitive workflows.

- Lustre costs** 3.5x flexFS with 21% slower performance end-to-end.

- EFS costs** 1.75-3x[1] flexFS with 29% slower performance end-to-end.

- flexFS shows better performance at each step of the cryo-EM workflow

- Cost relief with flexFS increases as data volume grows. Results calculated for storage of 100 – 800 TB of data.

*** Example pricing in AWS us-east-1 region (early 2025). Actual prices vary based on configuration*

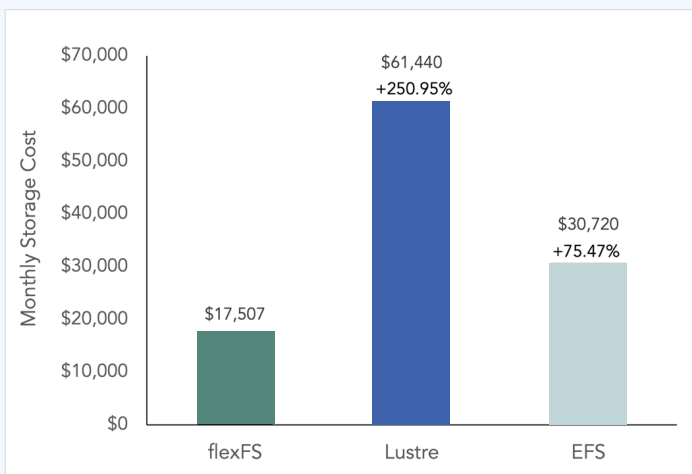*[1]AWS charges each time data is transferred to and from EFS storage*

Figure 1: Comparison of monthly storage cost for 100 TB of cryo-EM data for flexFS, Lustre, and EFS

Lustre costs 3.5 times flexFS

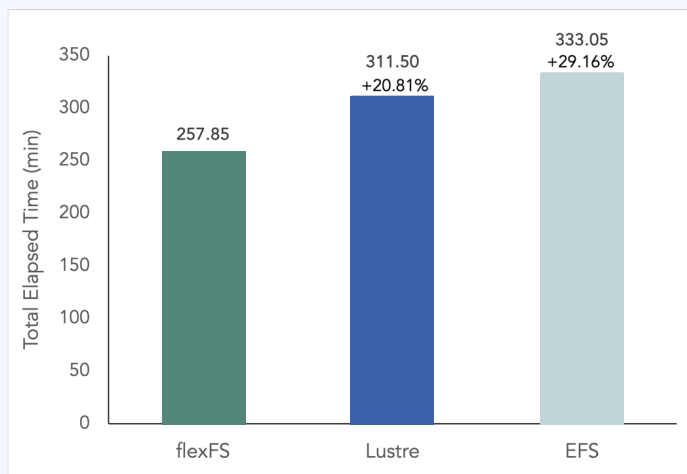EFS costs at least 1.75 times flexFS depending on access patterns

Figure 2: Comparison of total elapsed time to run the cryo-EM analysis for flexFS, Lustre, and EFS

Lustre is 21% **slower** than flexFS
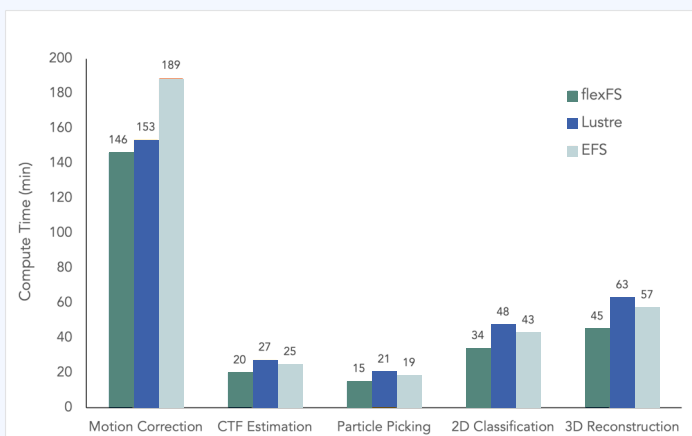
EFS is 29% **slower** than flexFS

Figure 3: Comparison of elapsed compute time for each step of the cryo-EM analysis workflow

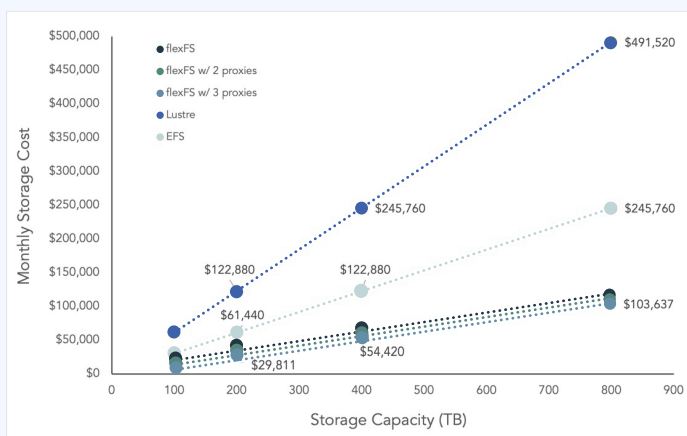flexFS is **faster** than Lustre and EFS at each step of the cryo-EM analysis workflow

Figure 4: Comparison of monthly cost for different storage capacities for flexFS, Lustre, and EFS

EFS costs 2x and Lustre costs 4x flexFS as data volume grows

Increasing proxy servers for flexFS to improve low latency throughput does not impact costs significantly

Summary

- Cryo-EM data analysis is complex and involves processing large volumes of data and running compute-intensive algorithms.

- Cost of storing cryo-EM datasets and running the analysis, often multiple times for a given project, can be a significant burden on an organization's drug discovery budget

- Typically, organizations make a trade-off between performance and costs of running cryo-EM workflows and storage of cryo-EM data

- In this benchmark, we demonstrated that flexFS outperformed FSx for Lustre and EFS, while being significantly cheaper

- Lustre was 21% slower & 3 – 4x more expensive than flexFS, depending on data volume

- EFS was also 29% slower and 1.75 – 3x more expensive than flexFS, depending on access patterns

CONTACT

lifesciences@paradigm4.com to get started with using flexFS for your cryo-EM analysis!

CHECK OUT

flexFS.io or **docs.flexfs.io** for additional information.