

# Cold Storage Data Archives: More Than Just a Bunch of Tapes

Bunjamin Memishi<sup>1</sup>, Raja Appuswamy<sup>2</sup>, Marcus Paradies<sup>1</sup>

<sup>1</sup>DLR Institute of Data Science, Jena, Germany <sup>2</sup>EURECOM, Sophia Antipolis, France



## Data-intensive Sciences: A Data Volume Generator Problem



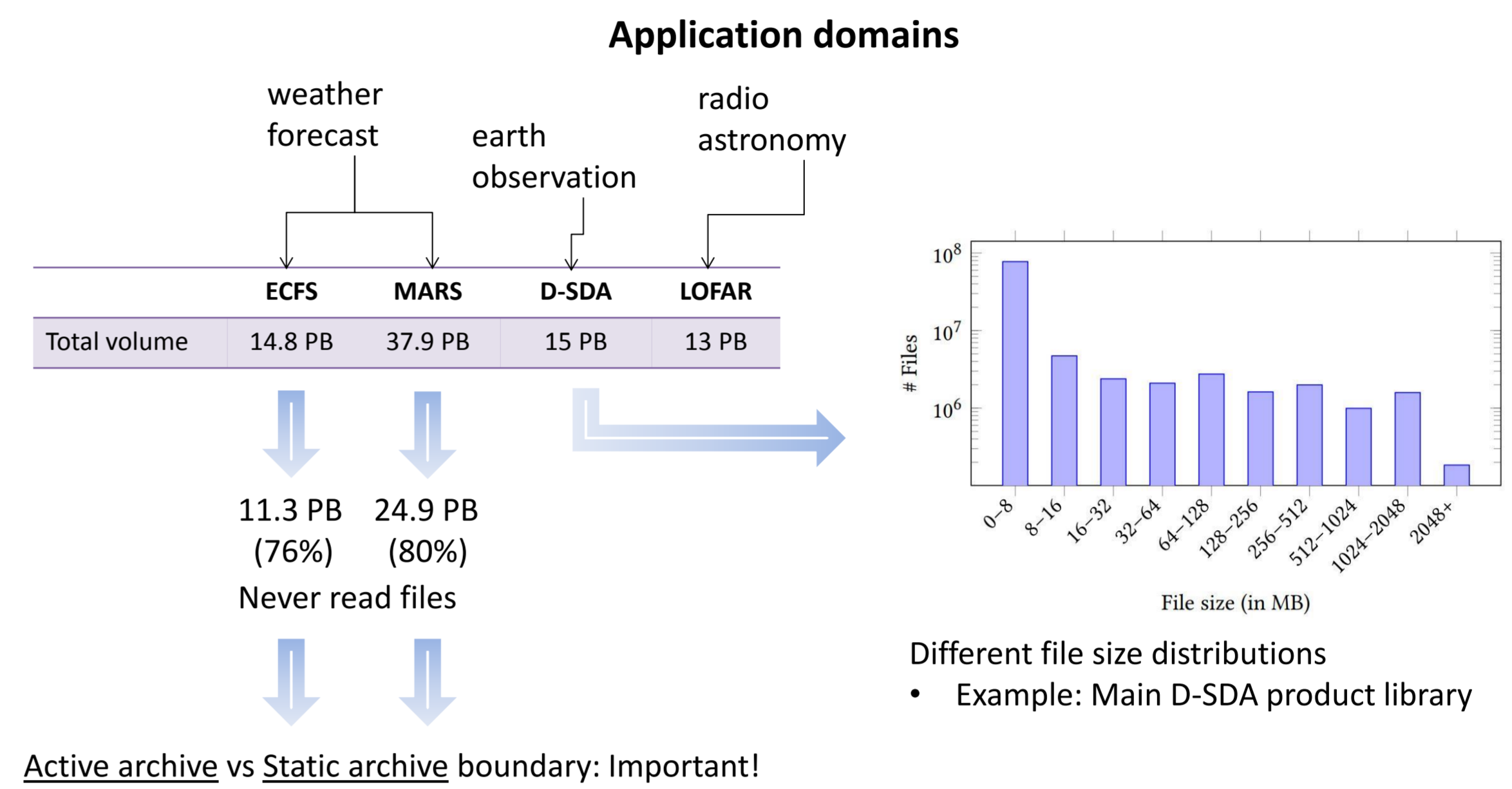
**Challenge:**  
What to store and where?

**Recent emerged opportunity:**  
Cold-storage-as-a-service

### Our contribution

- Highlight key **workload characteristics** of scientific data storage
- Describe **design trade-offs** in building cold data storage systems for scientific data archival
- Explore **deployment trade-offs** with respect to migration to the public cloud

## Data Analysis: Volume, Variety, Liveliness



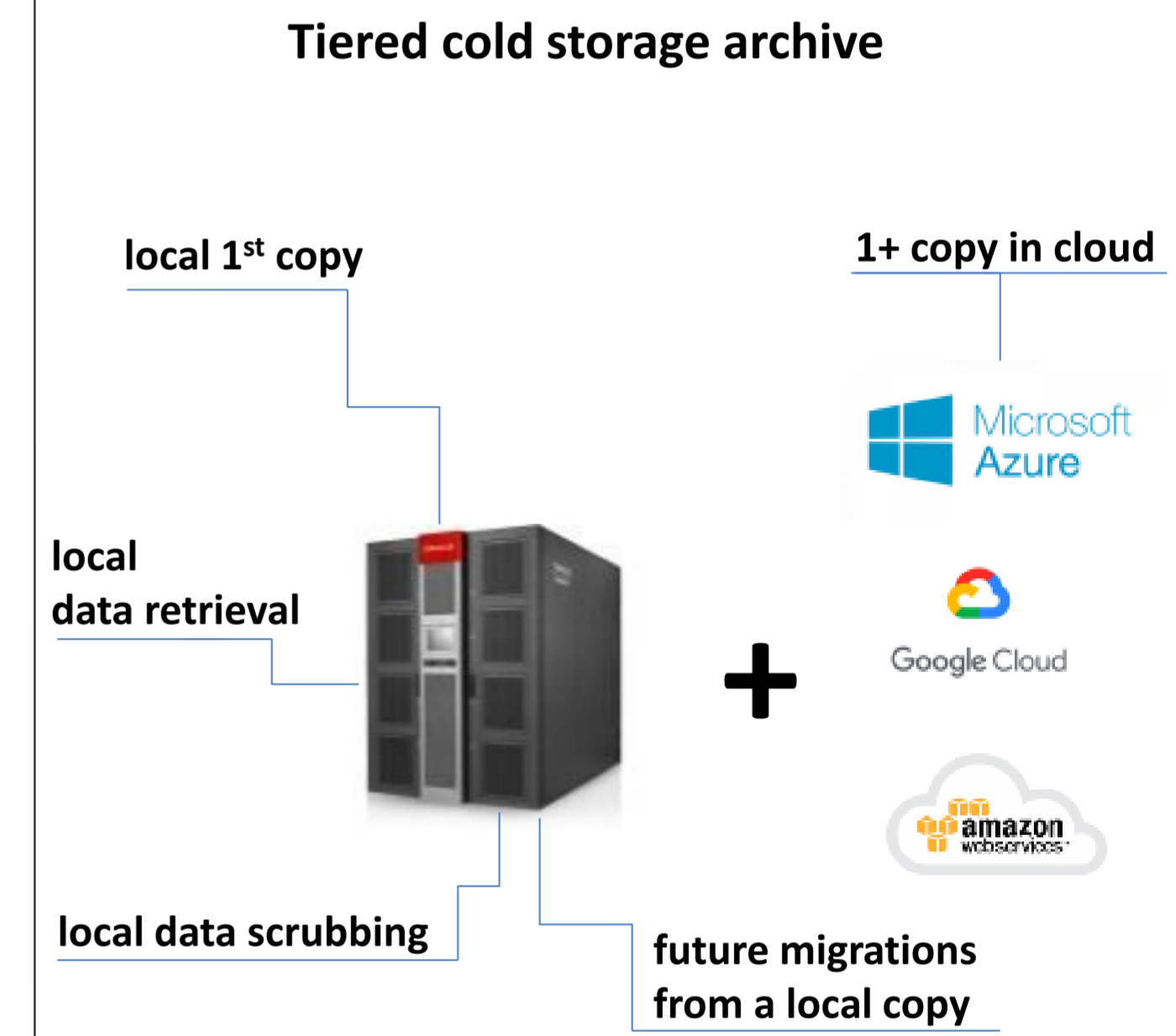
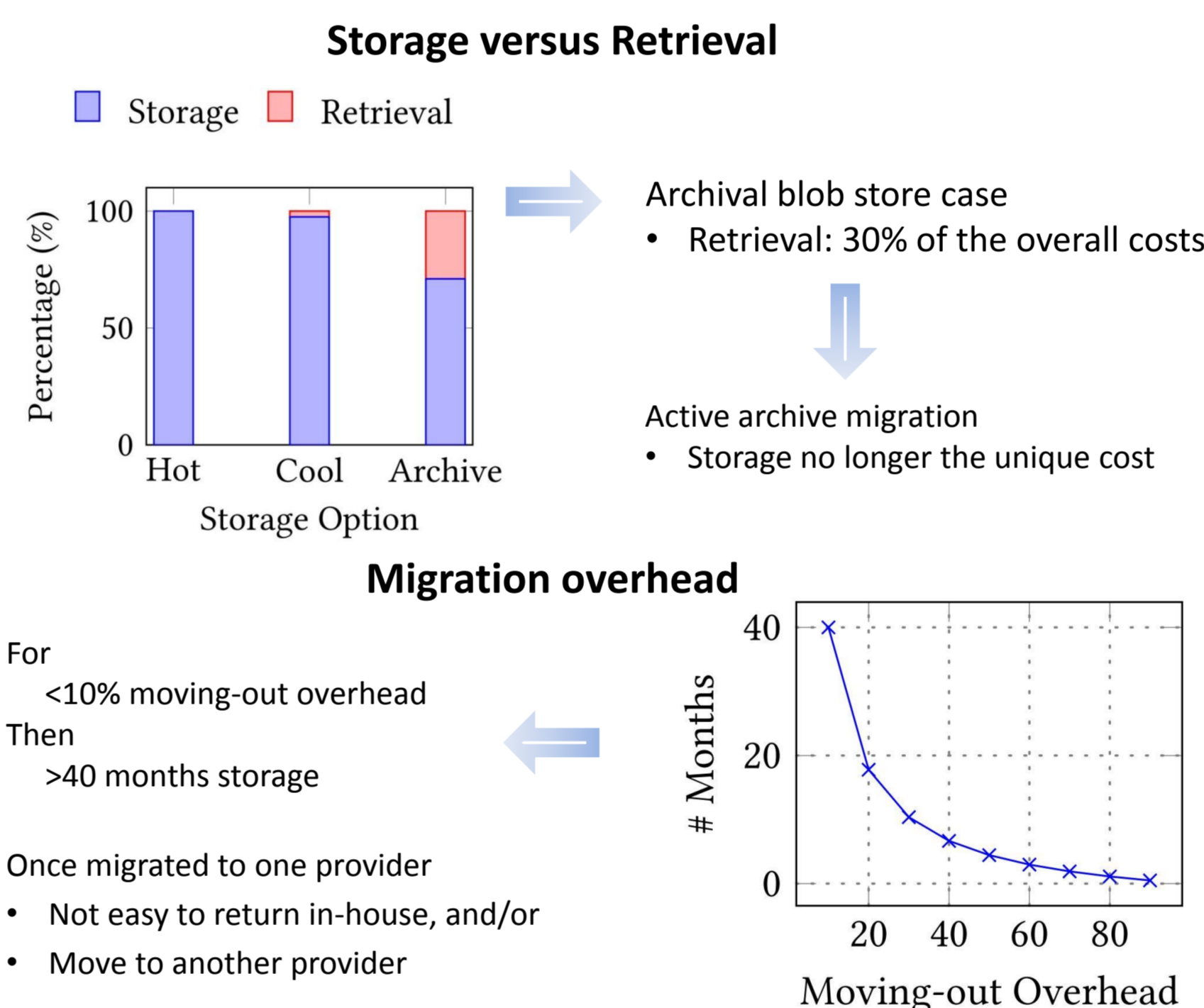
## Deployment Analysis

	Storage (GB/Month)	Retrieval (per GB)	GET Requests (per 10,000)	Latency
Azure Archival Blob (deep)	\$0.0045	\$0.02	\$0.5	Several hours
Azure Cool Blob (nearline)	\$0.0334	\$0.01	\$0.01	61.4 ms
Azure Hot Blob (online)	\$0.0422	\$0	\$0.004	5.3 ms

**Storage versus Retrieval cost**  
Main implications for deploying scientific applications in public cloud

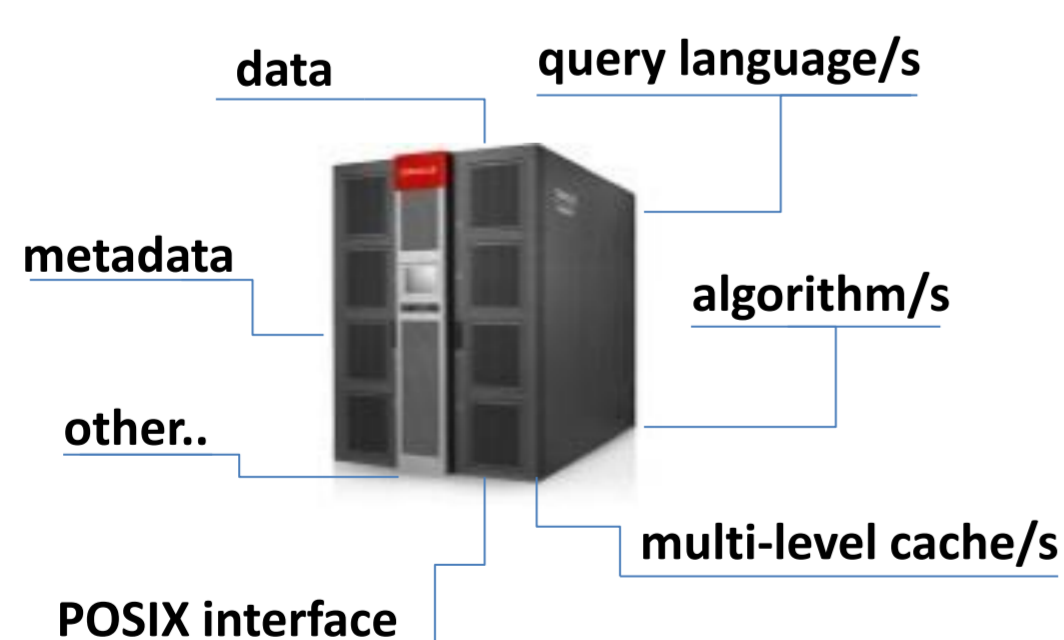
**Use case**

	Cost
1 PB scientific archive	Azure Archival Blob \$79K
Stored for 1 year	Azure Cool Blob \$430K
Read once during that year	Azure Hot Blob \$531K



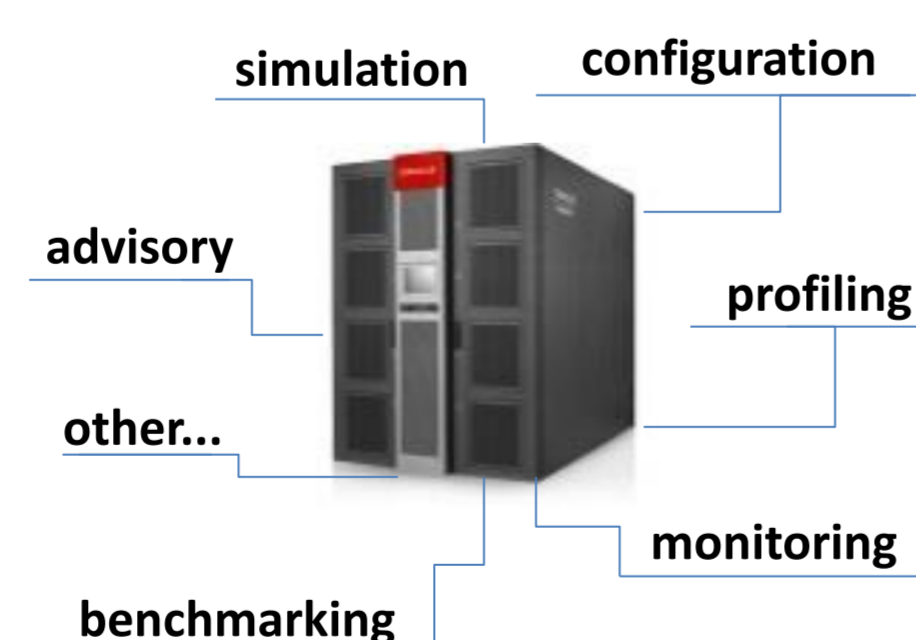
## Cold Storage Challenges

### Active Archive—Polystore



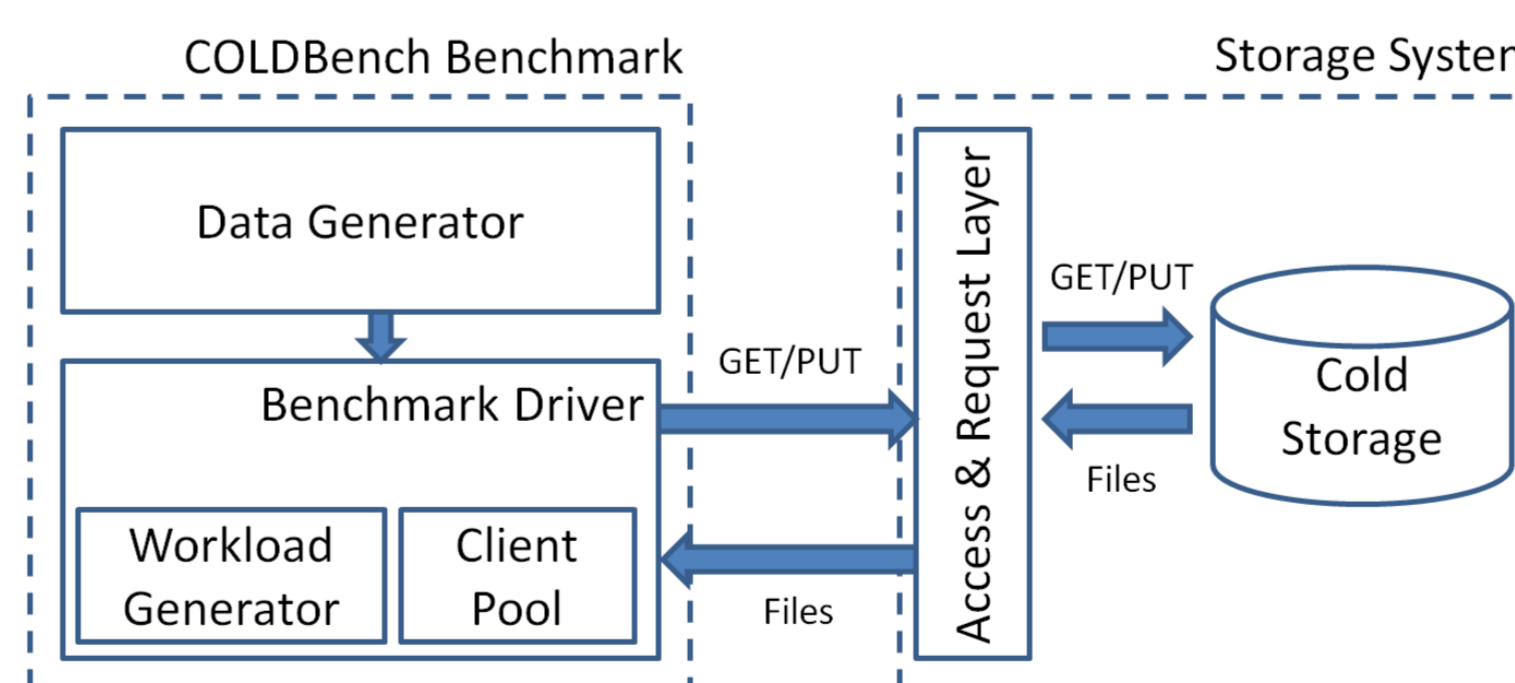
Interesting candidate for adopting polystore architecture

### Provisioning & Configuration



Mandatory to having multiple tools

### ColdBench: Towards a cold storage benchmark



Current design of our ColdBench framework

### Archive Profiling & Monitoring

Current methodologies, inappropriate and time-consuming due to

- Complexity
- Privacy



Full system monitoring exploration concurrently with data archive design

DaMoN'19, Amsterdam, July 01, 2019  
Contact: Bunjamin Memishi, [bunjamin.memishi@dlr.de](mailto:bunjamin.memishi@dlr.de)