Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
INFN

A user's perspective on Disk Tiering and caching

Lucio Anderlini lucio.anderlini@fi.infn.it
Lucia Morganti lucia.morganti@cnaf.infn.it
Rosa Petrini rosa.petrini@fi.infn.it
Andrea Rendina andrea.rendina@cnaf.infn.it
Vladimir Sapunenko vladimir.sapunenko@cnaf.infn.it

AI_INFN
Artificial Intelligence technologies for INFN research

LHCb

ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Outline

The problem:
Faster and faster processing units (CPU, GPU, FPGAs) are now moving the data-processing bottleneck on storage.
Larger and larger disks are reducing the throughput per TB.

WLCG (LHCb)
Increasing usage of the computing farms for data reduction and statistical analysis (Analysis Production) reading a large amount of data, for relatively short CPU time.

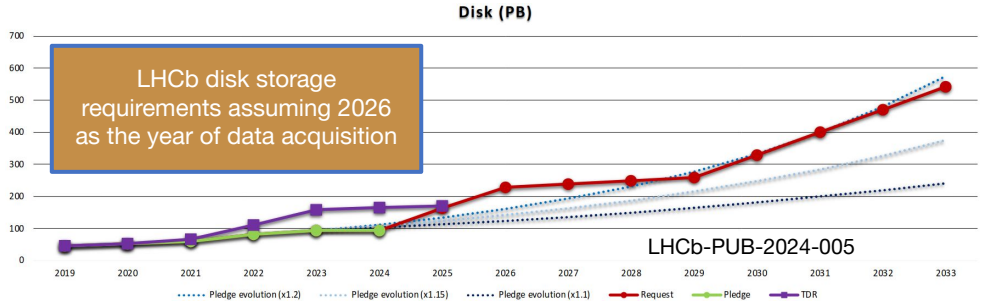
(Cloud) AI_INFN
Increasing usage of virtual environments and containerization leads to wild duplication of the software stack. Notebooks including images are also relatively large. User homes are larger and larger and must be reactive to ease development.

AI_INFN
Artificial Intelligence technologies for INFN research

The problems at WLCG:

Disk access bandwidth as a bottleneck and the lack of fast resources.

- **Drastic increase in data volumes:**
  - The upgrade of LHCb for Run3 has dramatically increased data volumes, resulting in high throughput on HLT2 (High-Level Trigger 2)
- **Evolution in storage usage:**
  - Data reconstruction occurs earlier in the processing pipeline.
  - Storage resources need to withstand multiple read cycles, for example in statistical analysis or training machine learning algorithms.
- **Changing market trends:**
  - Trend towards higher capacity drives to reduce operating costs is leading to a **reduction in throughput** per terabyte.
- Significant **increase** in the demand for **bandwidth** to access disk resources.
- **Disk access bandwidth** has become the primary **bottleneck** for data processing and analysis.
- **Critical resource** for scheduling**: CPU → Disk**

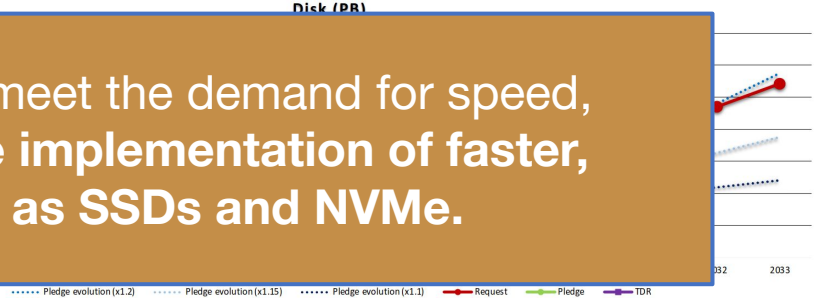


The problems at WLCG:

Disk access bandwidth as a bottleneck and the lack of fast resources.

- **Drastic increase in data volumes:**
  - The upgrade of LHCb for Run3 has dramatically increased data volumes, resulting in high throughput on HLT2 (High-Level Trigger 2)
- **Evolution in storage usage:**
  - Data reconstruction occurs earlier in the processing pipeline.
  - Storage resources need to withstand multiple read cycles, for example in statistical analysis or training machine learning algorithms.
- **Changing market trends:**
  - Trend towards higher capacity drives to reduce operating costs is leading to a **reduction in throughput** per terabyte.

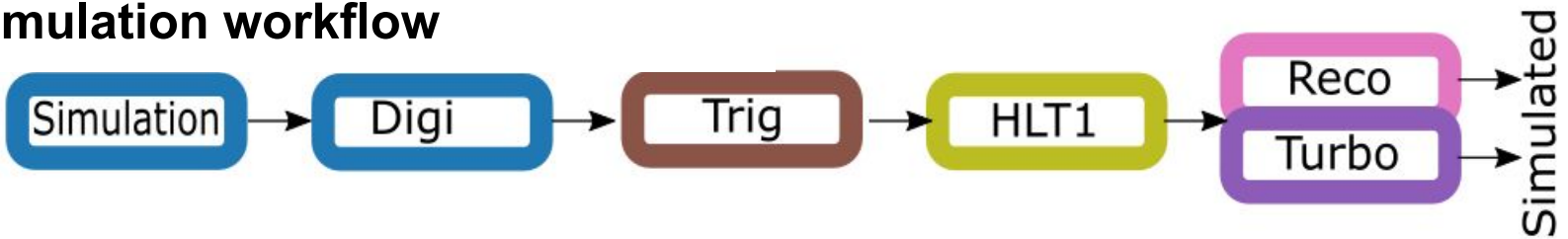
To overcome these limitations and meet the demand for speed, **The WLCG sites are exploring the implementation of faster, drive-based solutions, such as SSDs and NVMe.**



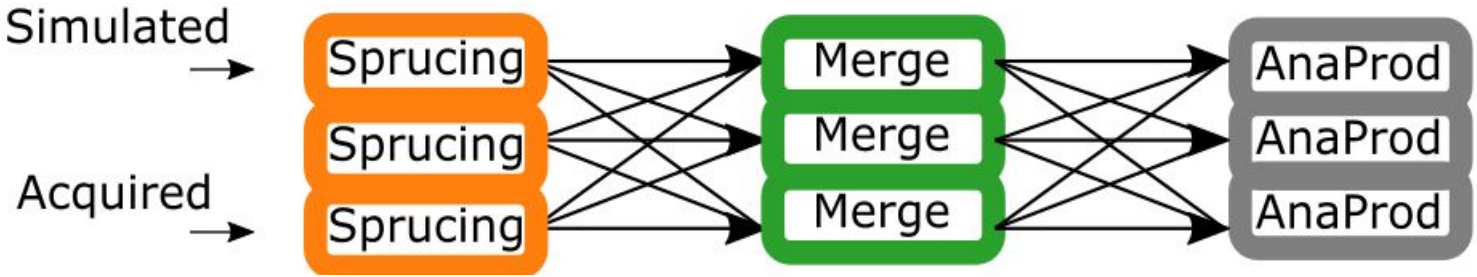
Points of attention in the context of LHCb

The workflow of LHCb

Simulation workflow



Data Processing



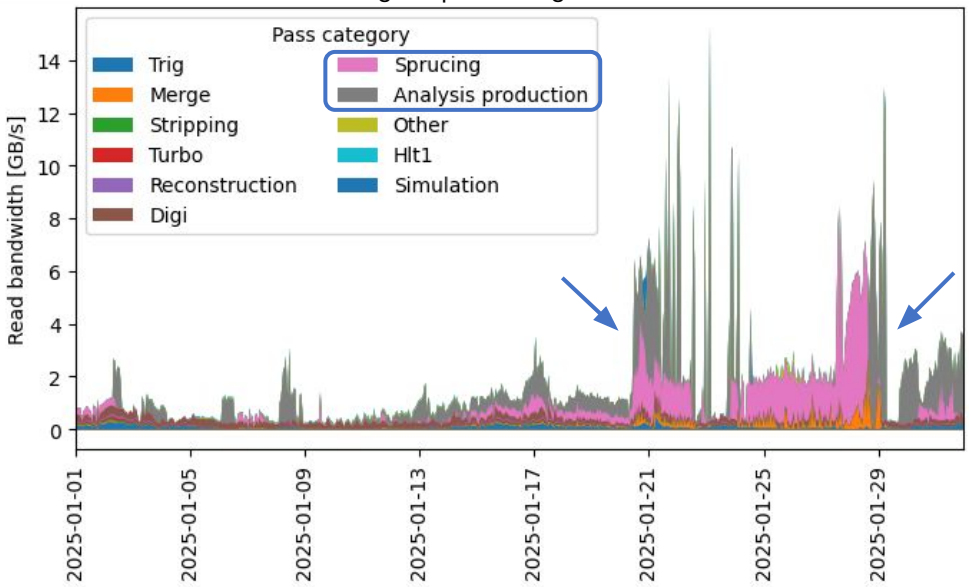
Points of attention in the context of LHCb

Average Throughput per job at LHCb

- How long a job will take to generate an output
- How large is their input and output size
- How many jobs are running in a given instant

Estimate the average bandwidth required for the LHCb production.

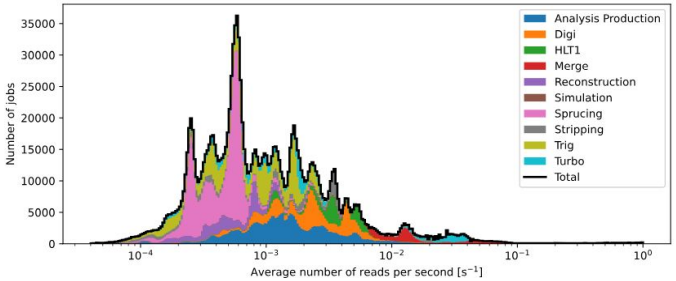
January 2025: after several migrations and fixes of the CNAF infrastructure, and during a reprocessing of the LHCb data.

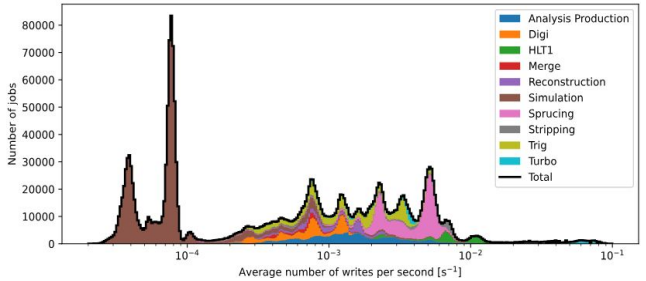


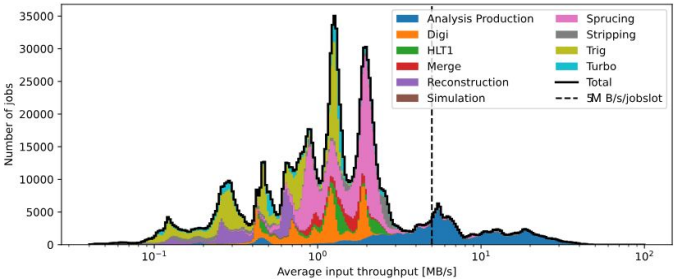
Points of attention in the context of LHCb

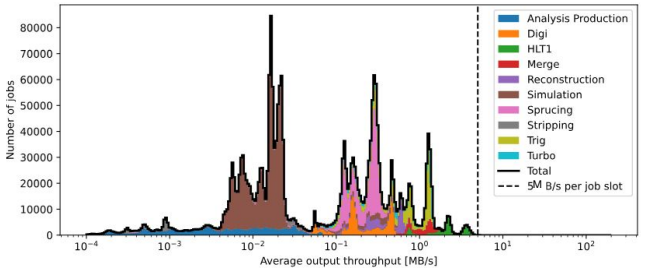
Average Throughput per job at LHCb

Estimated average throughput per job type → $throughput = \frac{number\ of\ bytes}{lifetime}$



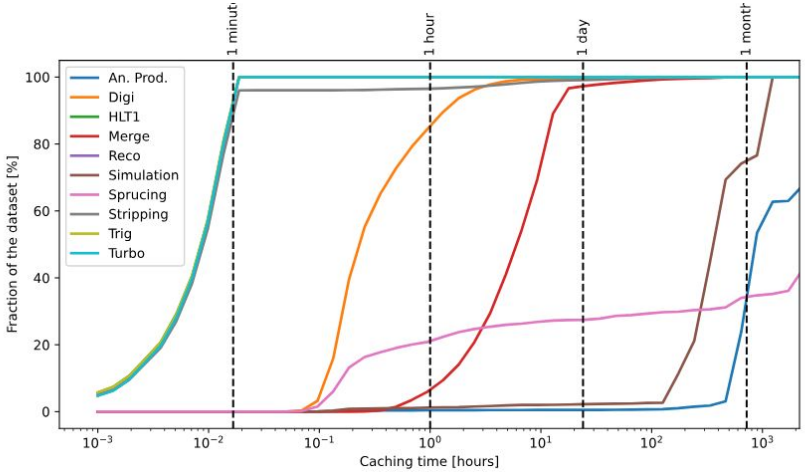




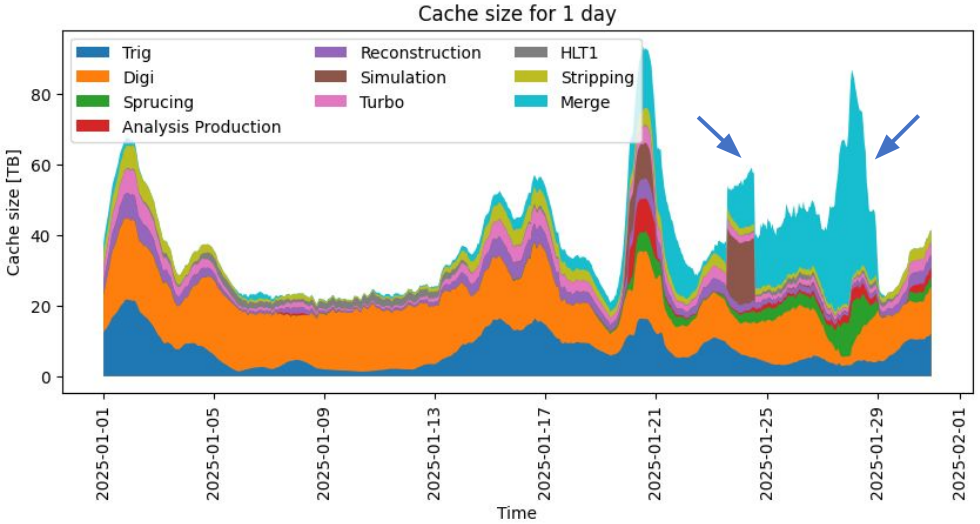


Points of attention in the context of LHCb

Caching time for different types of jobs



- **Analysis Production** and **merge** jobs make an intensive usage of the storage
- Merge run immediately after sprucing and delete temporary files right after merging.



Estimate of the disk storage cost of caching input data for different processing passes.

CNAF's solution

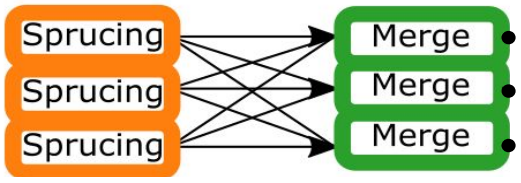
Hardware Overview:

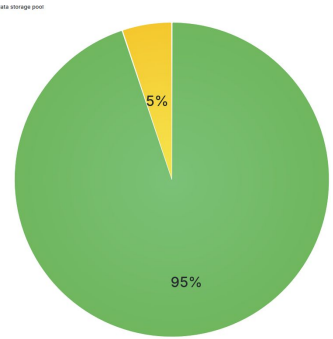
- System: DDN SFA 200NVX2
- Storage: 77 TB on NVMe (24 TB still available on request)

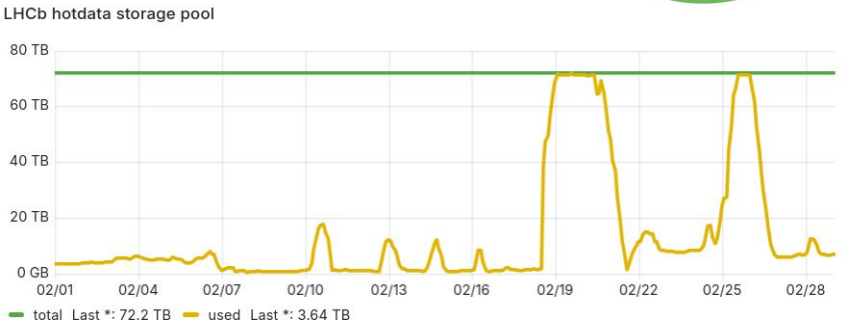
Objective: Fast and efficient temporary buffer for merge files produced by *Sprucing jobs*, that are deleted after processing by *Merge jobs.*

Configuration:

- Created new buffer fileset to isolate files from Sprucing jobs and enabled specific GPFS policy for this buffer only
- hot migration from previous setup (rsync to move old data to the new fileset)
- Monitoring configured for “HOTDISK” usage







Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

INFN

Storage challenges with problems on the AI_INFN platform

AI_INFN

Artificial Intelligence technologies for INFN research

The AI_INFN platform solution

AI_INFN Artificial Intelligence technologies for INFN research

Platform design and user needs

Provide a **centrally maintained cloud-based infrastructure** for interactive and batch ML fast prototyping, with access to modern hardware accelerators (GPU, FPGA…) and systems tuned for ML performance

→ Solution used to train PINN (Physics-Informed Neural Networks) in the ENI-PIML project

ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

- **Different hardware accelerators** physically connected to **different servers**
- Need for software that runs **symmetrically** across the heterogeneous platform: **users' homes** must be identical when accessing different servers.
- **Shared** disk volumes accessible by multiple users on different machines: **READ-WRITE MANY**

→ Request for a network FS

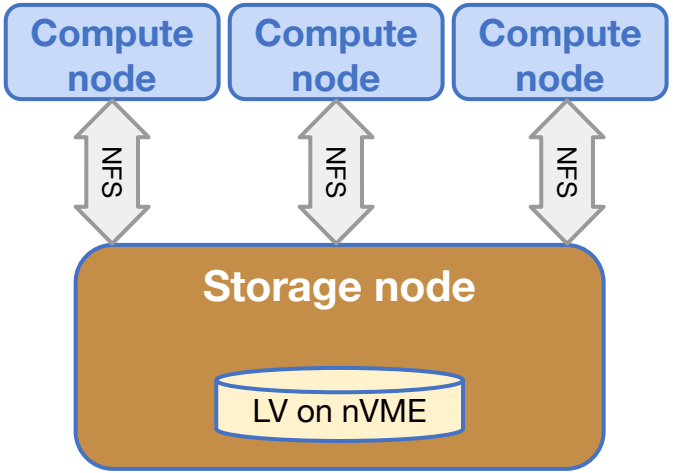
The AI_INFN platform solution

Configuration & Benchmark

AI_INFN Artificial Intelligence technologies for INFN research

Three different configurations were compared.

1. **Single NVMe volume on bare metal**



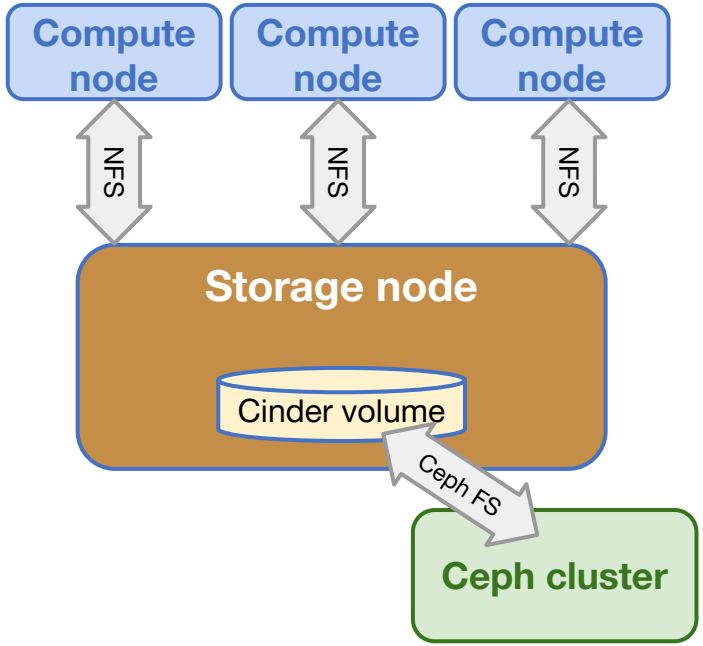
The AI_INFN platform solution

AI_INFN Artificial Intelligence technologies for INFN research

Configuration & Benchmark

Three different configurations were compared.

1. **Single NVMe volume on bare metal**
2. **CEPH with Cinder volumes**
   a. Significant increase in filesystem latency



The AI_INFN platform solution

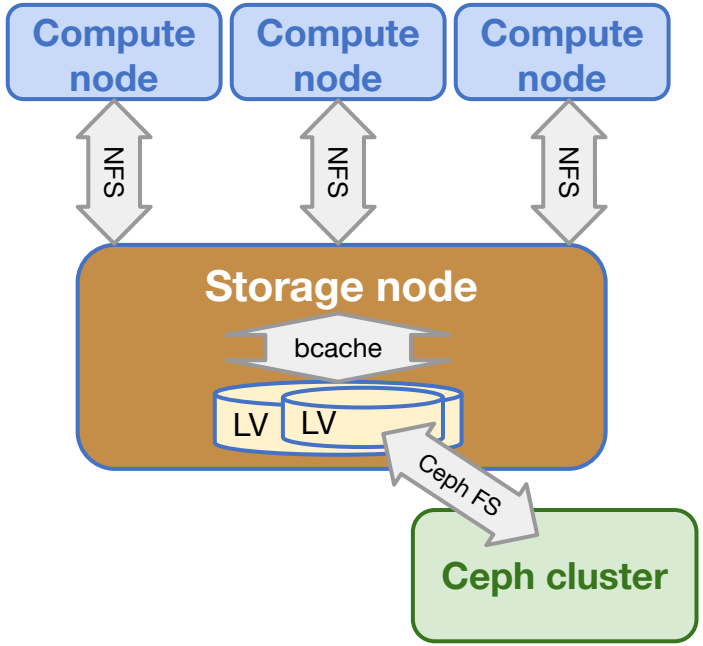
AI_INFN Artificial Intelligence technologies for INFN research

Configuration & Benchmark

Three different configurations were compared.

1. **Single NVMe volume on bare metal**
2. **CEPH with Cinder volumes**
   a. Significant increase in filesystem latency
3. **Hybrid storage approach**
   a. Fast I/O volumes on **NVMe** and Additional data stored on CEPH disk using **bcache**

Add redundancy with incremental backups using Borg

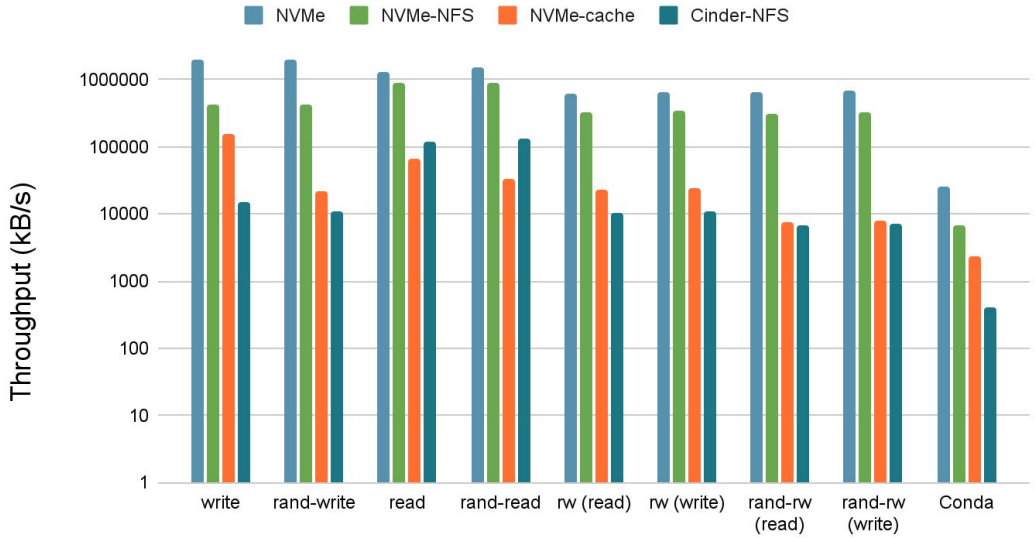


The AI_INFN platform solution

AI_INFN Artificial Intelligence technologies for INFN research

Configuration & Benchmark

Performance Storage: Benchmark



Tipo	write (MB/s)	rand write (MB/s)	read (MB/s)	rand read (MB/s)	rw (MB/s)	rand rw (MB/s)	Conda (MB/s)
NVMe	2002	1958	1329	1512	r = 624 w = 655	r = 643 w = 676	25.9 7.6 s
NVMe NFS	437	433	892	907	r = 320 w = 337	r = 306 w = 322	6.7 29.5 s
NVMe cache	154	21.3	68.3	32.8	r = 23.6 w = 24.8	r = 7.6 w = 8.0	2.3 1 m 23.6 s
Cinder NFS	14.8	11.2	118	134	r = 10.4 w = 10.9	r = 6.8 w = 7.1	0.4 7m 39s

Conclusion & Outcomes

- **LHCb**: For data-processing, I/O has become a more critical bottleneck than CPU usage
  - The introduction of dedicated NVMe buffers and targeted management of temporary files has significantly improved the efficiency of the merging and analysis production workflows.
  - Other experiments have opted for the CNAF's solution proposed for the LHCb.
- **AI_INFN**: the main challenge lies in enabling consistent and fast access to data across multiple nodes simultaneously, tuned for intense metadata operations.
  - Combining logical volumes, NVMe-based caching and distributed file systems (CephFS) has created a performing shared infrastructure, with bcache plays a key role in metadata management.
- Benchmarks:
  - Bare-metal NVMe delivers the best performance.
  - Ceph cluster is the most scalable.
  - Hybrid solutions offer a solid trade-off between speed and scalability.

Thank you for your attention!

The AI_INFN platform solution

AI_INFN Artificial Intelligence technologies for INFN research

Benchmark

Tipo	write (MB/s)	rand-write (MB/s)	read (MB/s)	rand-read (MB/s)	rw (MB/s)	rand-rw (MB/s)	Conda (MB/s) 196MiB
NVMe	2002MB/s	1958MB/s	1329MB/s	1512MB/s	r= 624MB/s w= 655MB/s	r= 643MB/s w= 676MB/s	0m7.610s 25.89 MB/s
NVMe-NFS	437MB/s	433MB/s	892MB/s	907MB/s	r= 320MB/s w= 337MB/s	r= 306MB/s w= 322MB/s	0m29.545s 6668 KB/s
NVMe-cache	154 MB/s	21.3 MB/s	68.3 MB/s	32.8 MB/s	r= 23.6 MB/s w= 24.8 MB/s	r= 7.6MB/s w= 8.0 MB/s	1m23.634s 2343 KB/s
Cinder-NFS	14.8 MB/s	11.2 MB/s 8573 kB/s	876 MB/s	929 MB/s	r= 10.4 MB/s w= 10.9 MB/s	r= 6.8 MB/s w= 7.1 MB/s	7m38.966s 427 KB/s

CNAF's solution

HW details: 77TB on NVMe disks in a DDN SFA 200NVX2 system (still 24 TB available upon request)

Aim: temporary store files produced by Sprucing jobs which, after processing by Merge jobs, are deleted

Steps for configuration:

- Oct 24: define a storage pool and a policy within GPFS for LHCb-disk fileset so that all new *.dst files in /storage/gpfs_lhcb/disk/ end up in NVMe disks
- the buffer filled very quickly, showing 99% of files not accessed in the last 3 days
- Nov 24: wrong path (all disk fileset), wrong policy. Need to define a buffer fileset (/storage/gpfs_lhcb/disk/lhcb/buffer), so that policies can then be defined within GPFS (this step required a process of rsync to move old data to the new fileset while writing data)
- define the following policy: RULE 'HOTDISK' SET POOL 'hotdata' FOR FILESET ('LHCb-Buffer') LIMIT(99) WHERE NAME LIKE '%.dst'
- configure monitoring for used space in the selected "HOTDISK" storage pool
- from the initially allocated 200 TB, reduce to 77 TB, which seem enough for the moment. Originally, LHCb estimated less than 50 TB needed.

The AI_INFN platform solution

AI_INFN
Artificial Intelligence technologies for INFN research

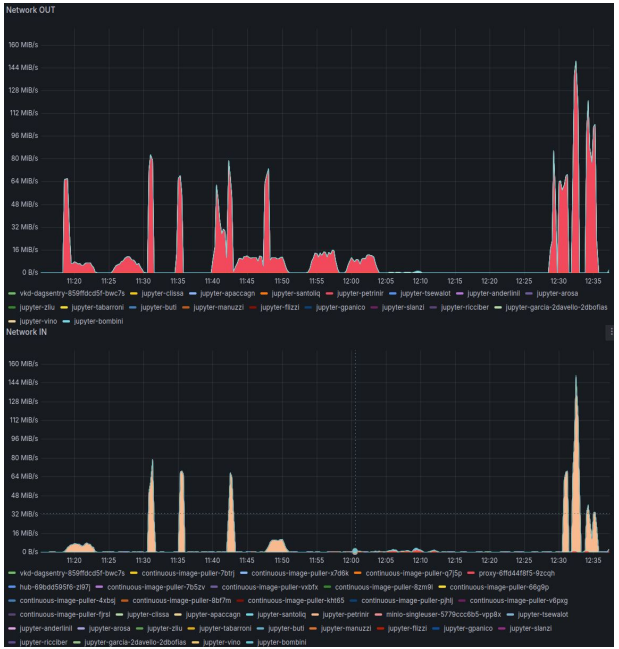
Configuration steps

I. **Single NVMe volume on bare metal:**
   A. Fails beyond 2TB → Kubernetes marks the file system as read-only
   B. Triggers eviction mode → Causes complete system crash

II. **CHEP with Cinder volumes:**
   A. Significant increase in filesystem latency
   B. Conda environment installation time increases from 40 seconds to 35 minutes

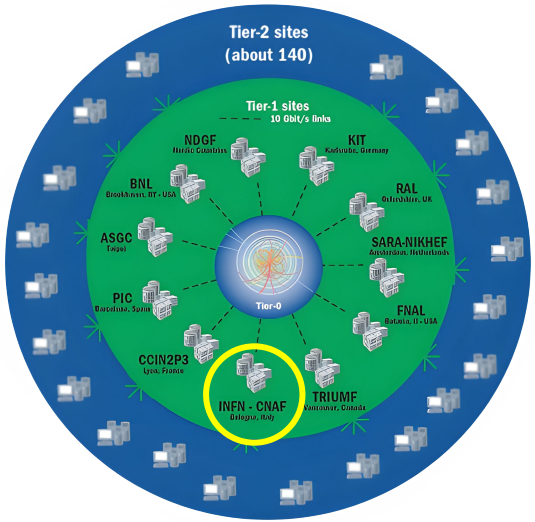
III. **Hybrid storage approach:** Fast I/O volumes on NVMe and Additional data stored on CHEP disk using **bcache**:
   A. ~10% slower file system
   B. NO redundancy, double risk of failure: data loss causes VM inconsistency

IV. **Add redundancy with incremental backups using Borg**
   A. 3 copies of non-cached data on CHEP
   B. 1 copy of cached data
   C. 3 copies of web application



The problems at WLCG:

The need for larger volumes and new data access requirements.

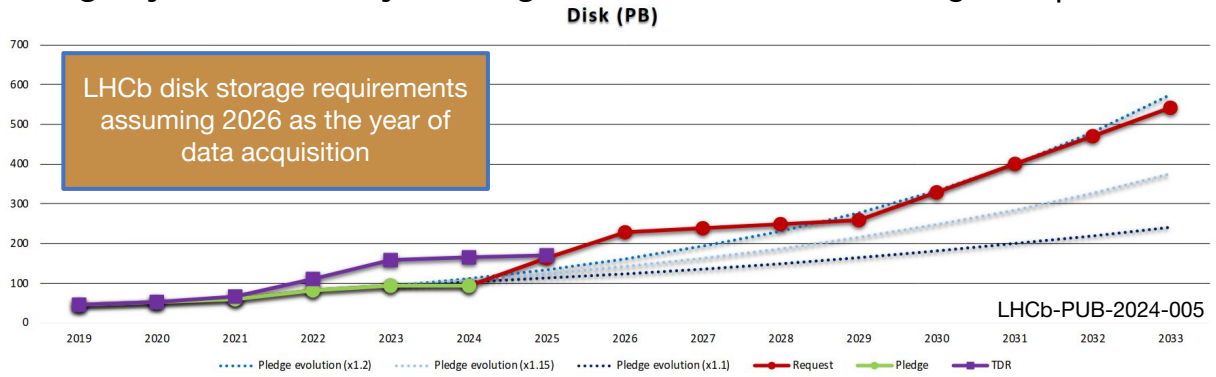


- **Drastic increase in data volumes:**
  - The upgrade of LHCb for Run3 has dramatically increased data volumes, resulting in high throughput on HLT2 (High-Level Trigger 2)
  - Storage of large amounts of data on tape and disk after the "sprucing" process.
- **Evolution in storage usage:**
  - Data reconstruction occurs earlier in the processing pipeline.
  - Storage resources need to withstand multiple read cycles, for example in statistical analysis or training machine learning algorithms.
- **Changing market trends:**
  - The market trend towards higher capacity drives to reduce operating costs is leading to a **reduction in throughput** per terabyte.

The problems at WLCG:

Disk access bandwidth as a bottleneck and the lack of fast resources.

- The WLCG sites have seen a significant **increase** in the demand for **bandwidth** to access disk resources.
- **Disk access bandwidth** has become the primary **bottleneck** for data processing and analysis.
- **I/O of storage systems is a key missing resource** → it is not enough to optimise the CPU.



To overcome these limitations and meet the demand for speed,
it is necessary to implement faster drive-based solutions such as SSDs and NVMe.