Data in Motion, Energy in Mind – Challenges for AI-Driven HPC

PROF. DR. SARAH M. NEUWIRTH

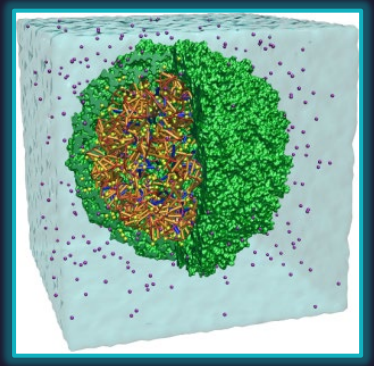
NHR SOUTH-WEST HPC CENTER

JOHANNES GUTENBERG UNIVERSITY MAINZ

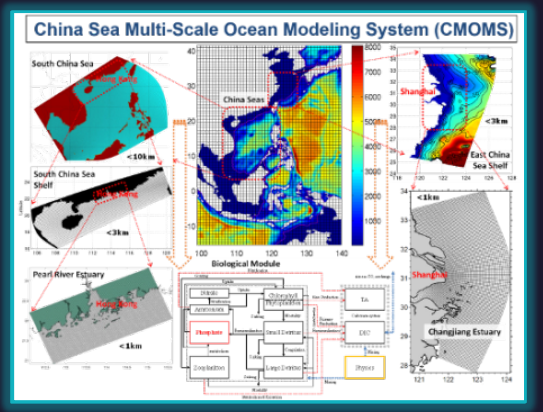
NEUWIRTH@UNI-MAINZ.DE

HPC ROUNDTABLE GERMANY 2025, MUNICH, OCTOBER 2025

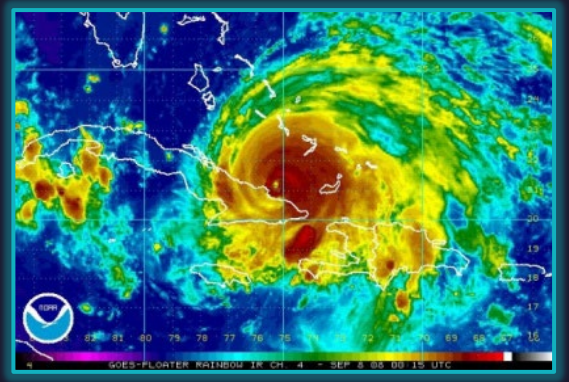
Why is Parallel I/O so difficult?



Molecular dynamics modeling of viruses.



Ocean modeling (HKUST).



Weather forecasting (NOAA).

Injection process in combustion engines (ORNL).

Source: Rob Latham et al., "Parallel I/O in Practice", SC Tutorial.

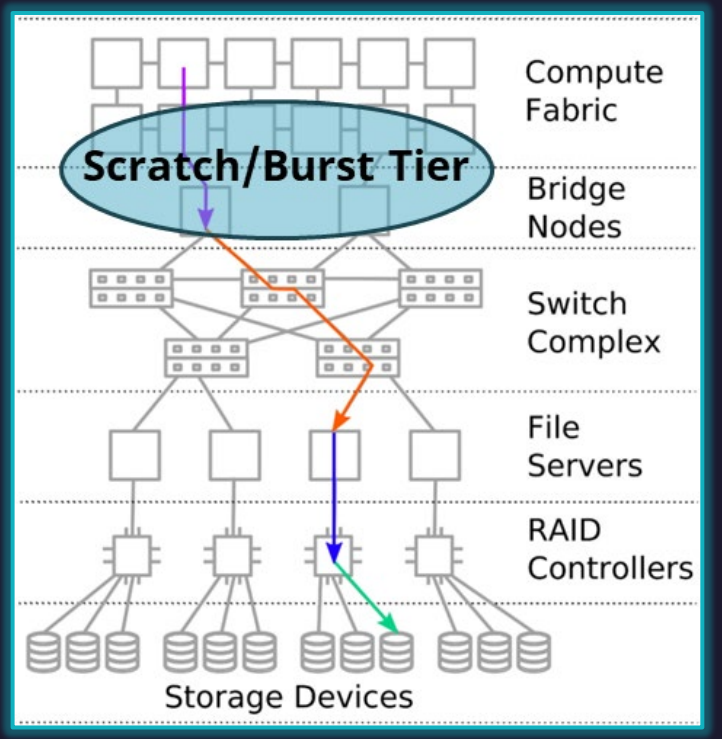
Scientists think about their data in terms of their science problem: molecules, atoms, grid cells, particles.

Ultimately, physical disks store bytes of data.

Layers in between, the application and physical disks are at various levels of sophistication.

Complex HPC Infrastructures

- Complexity and scope increase the urgency
  - *New computational paradigms* (AI/ML apps vs. BSP-style HPC)
  - *New architectural directions* (e.g., IPU, RISC-V, data flow)
  - *Heterogeneity overall:* node architectures, within the system, storage and parallel file system during application design (e.g., ML within HPC applications)
  - *New operations paradigms* (e.g., cloud, container)
  - Simplistic approaches to increasing compute demand result in *unacceptable power costs*
- Difficult for humans to optimally adapt applications to systems and to detect and diagnose vulnerabilities



B. Settlemyer, G. Amvrosiadis, P. Carns and R. Ross, 2021. *It's Time to Talk About HPC Storage: Perspectives on the Past and Future*, in Computing in Science & Engineering, vol. 23, no. 6, pp. 63-68.

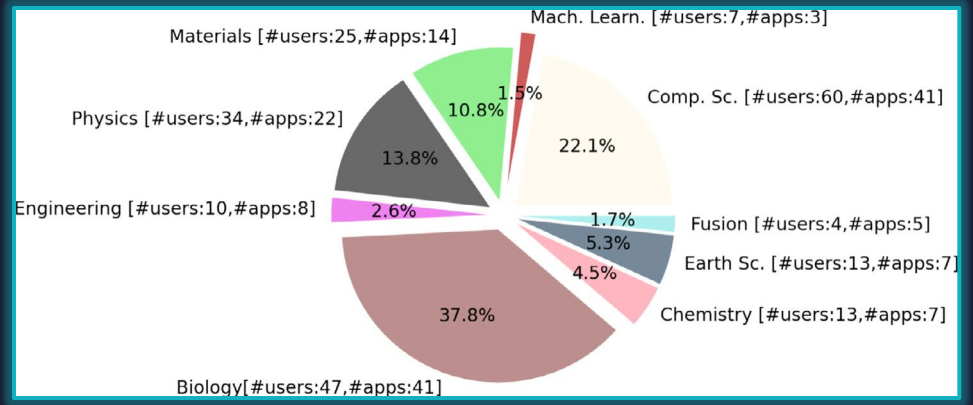
Ciorba, F., 2023. *Revolutionizing HPC Operations and Research*. Keynote at HPCMASPA'23 Workshop.

Carns, P., 2023. *HPC Storage: Adapting to Change*. Keynote at REX-IO'23 Workshop.

Emerging HPC Workloads

- Traditional HPC workloads:
  - Dominated by bulk-synchronous I/O phases
  - Simulation workloads
  - Checkpoint / Restart Files
- Emerging HPC workloads:
  - Artificial Intelligence
  - Data Analytics and Big Data
  - Deep Learning
  - Multi-step Workflows
  - In-situ analysis

Karimi, A.M., Paul, A.K. and Wang, F., 2022. *I/O performance analysis of machine learning workloads on leadership scale supercomputer*. Performance Evaluation, 157, p.102318.



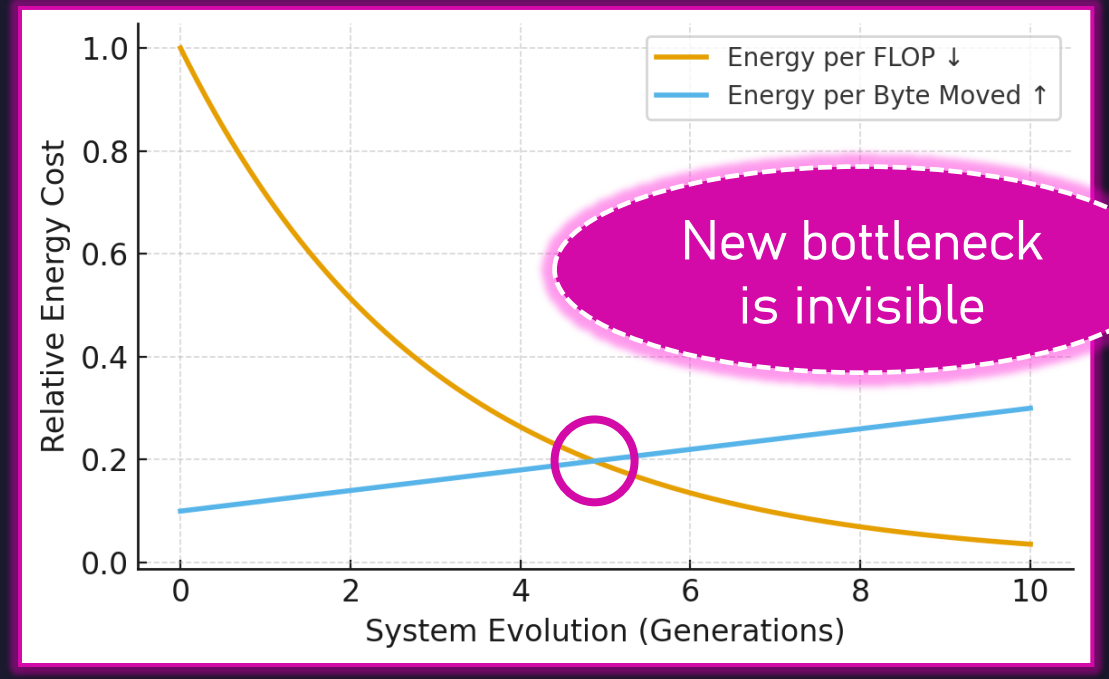
Classification of 23,389 ML jobs on Summit by science domains.

- Vastly different I/O and performance characteristics (random small file accesses, non-sequential, metadata-intensive, and small-transaction reads and writes)

=> HPC I/O has evolved beyond just checkpointing and bulk-synchronous parallel I/O phases

The Hidden Cost of Motion & Data Gravity

- Energy per FLOP ↓ – energy per byte moved ↑
  => *a silent inversion of the performance equation*
- Crossover point:
  Data movement energy ≈ compute energy
- Foundation models and large AI datasets create data gravity wells anchoring compute
- Moving data or checkpoints multiplies both energy and latency costs
- *Energy gravity emerges* — systems self-organize around the cheapest motion path



N. Lewis, J. L. Bez, and S. Byna. 2025. *I/O in Machine Learning Applications on HPC Systems: A 360-degree Survey*. ACM Comput. Surv.

E. Suarez et al. 2025. *Energy-aware operation of HPC systems in Germany*. Front. High Perform. Comput.

"We have optimized compute to exhaustion — yet the energy to move data keeps rising."

Storage Hierarchies Are Out of Sync

"HPC I/O stacks were built for checkpointing, not for AI-driven workflows."

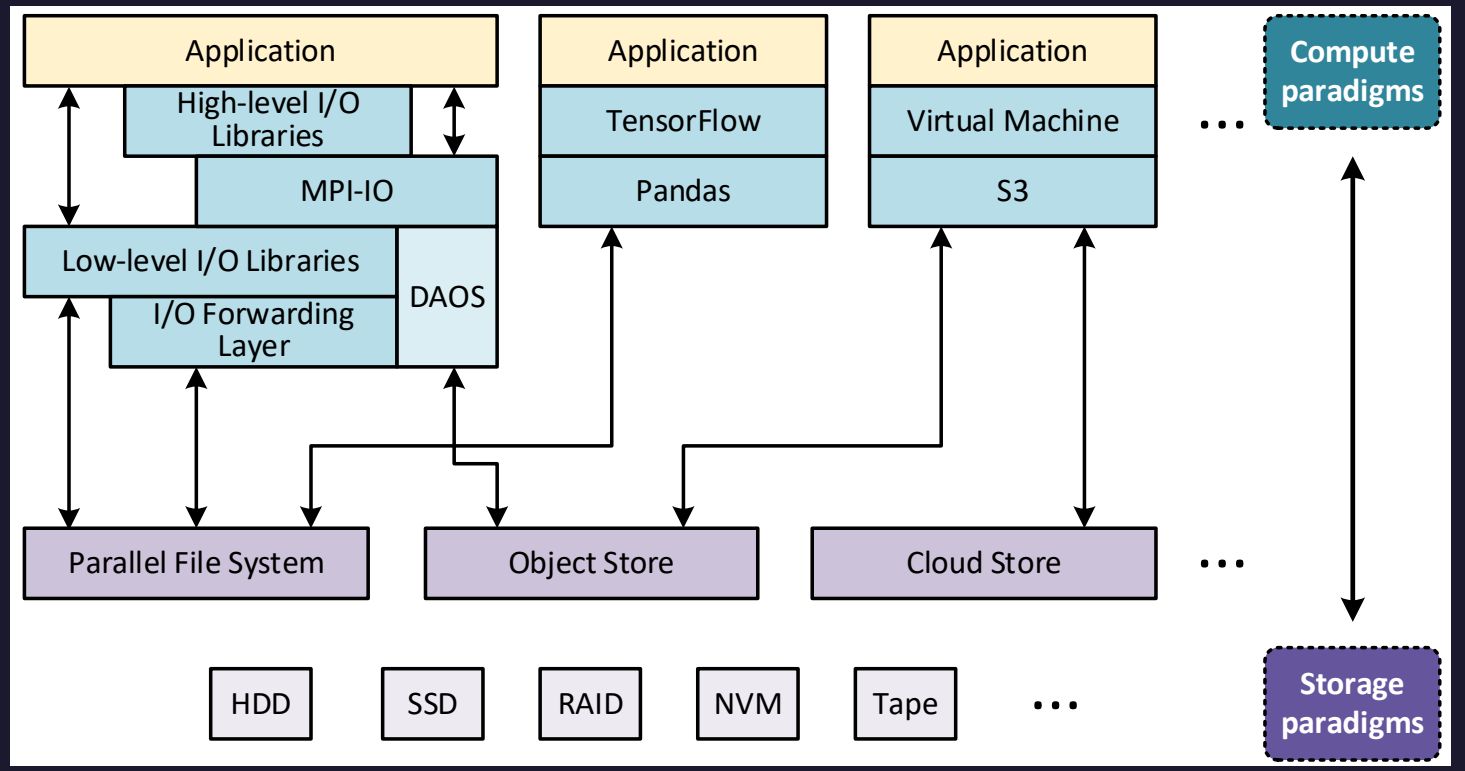
"Random access, metadata-heavy, streaming patterns break our assumptions."

Diverse Workloads

Diverse Software

Diverse Architecture

Diverse Hardware



Application

- Number of processes
- Request sizes
- Access patterns
- I/O operation
- Data volume

Network

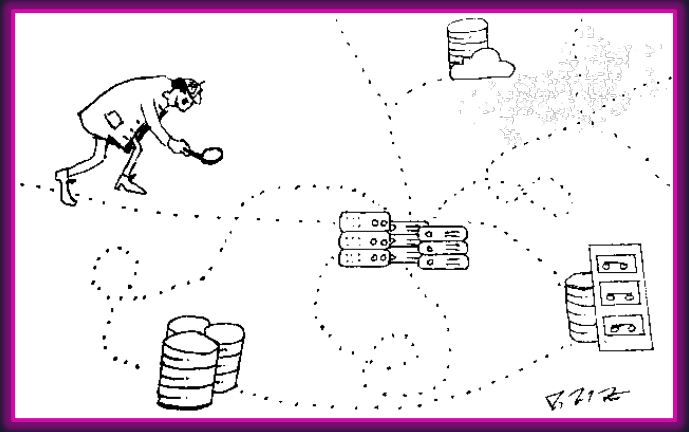
- Message sizes
- Network topology
- Network paths
- Network type

File System

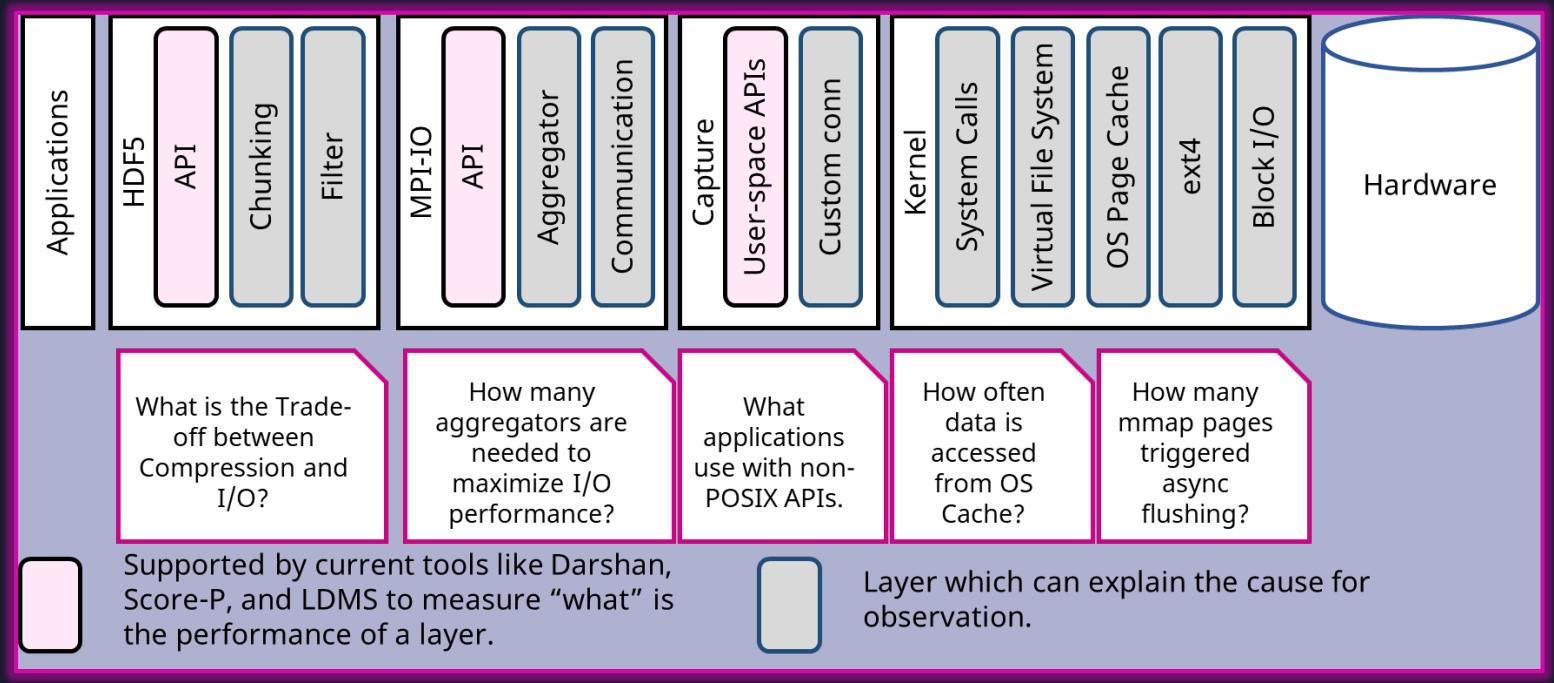
- Type of file system
- Disk types
- Stripe sizes
- File hierarchy
- Shared access

"Storage tiers must evolve toward AI-native, energy-aware data paths."

Explainable I/O?!

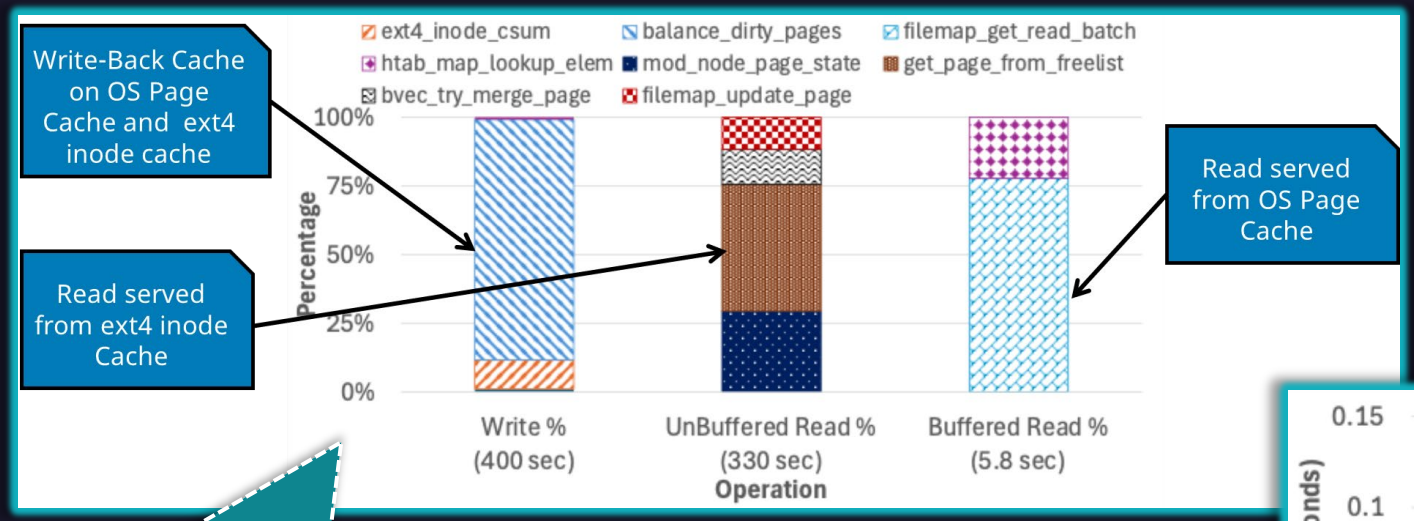


- Modern HPC & AI workloads traverse deep, multi-layered I/O stacks
- Existing profilers show *what* happened (calls, bytes, rates) but *not why* it happened
- *Explainable I/O (XIO)* correlates performance events across layers to reveal:
  - causal relationships,
  - inter-layer interference, and
  - true sources of variation



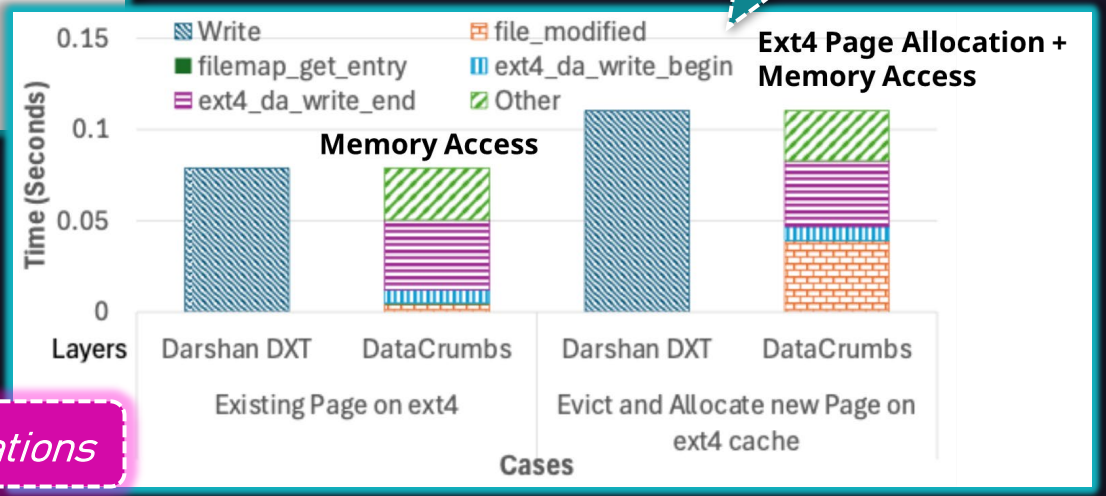
Neuwirth, S. and Devarajan, H., Wang, C., and Lofstead, J., 2025. *XIO: Toward eXplainable I/O for HPC Systems*. SSDBM'25.

Example: XIO with DataCrumbs



Different kernel stack calls can help identify buffered vs unbuffered read calls.

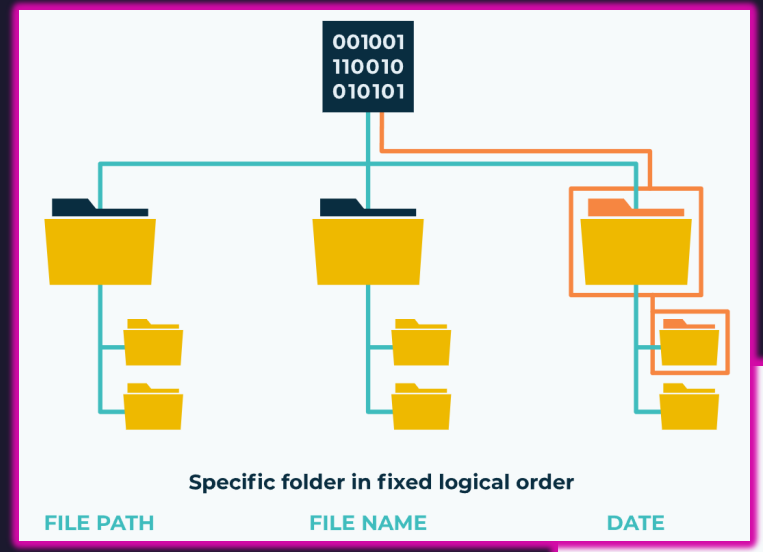
DataCrumbs: Low-Overhead Multi-Layer Profiling for enabling Explainable I/O

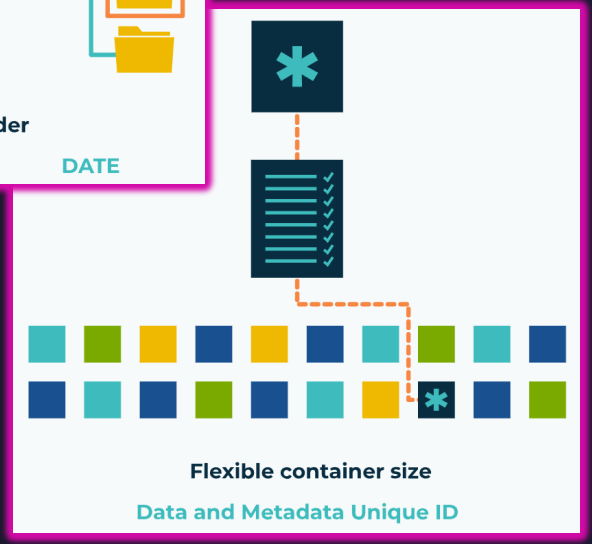


DataCrumbs: eBPF-based, kernel + user tracing *=> causal explanations*

Coordinating Data and Metadata – The Next Storage Frontier

- AI workflows explode metadata: billions of small objects and rapid state changes
- Uncoordinated metadata = hidden latency + energy waste
- Intelligent orchestrators unify data + metadata flows ⟶ semantic placement, ML-guided caching
- Metadata coordination becomes the control plane for efficient, explainable I/O





W. Yang et al. 2025. *The Impact of Modern AI in Metadata Management*. Human-centric Intelligent Systems. https://doi.org/10.1007/s44230-025-00106-5

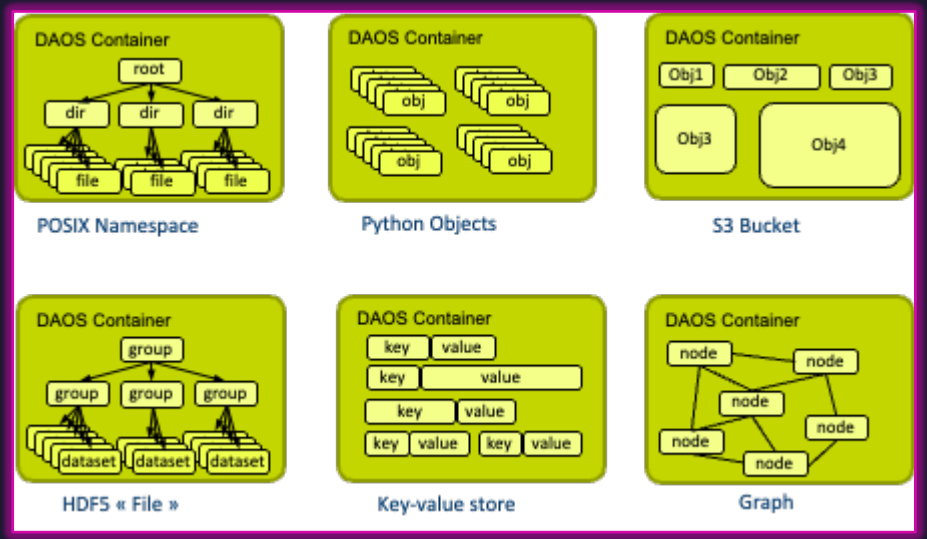
Y. Li et al. 2025. *Parallel Data Object Creation: Towards Scalable Metadata Management in High-Performance I/O Library*. PDSW'25. (to appear)

Object Stores and AI-Ready Storage Solutions for HPC

- Object stores provide a tier-less, scalable foundation for AI and HPC convergence
- Rich metadata and flat namespaces enable fine-grained data placement and lifecycle automation
- Emerging architectures blend parallel file systems + object stores + NVM tiers + cloud gateways
- AI-assisted storage services → predictive caching, semantic data grouping, explainable placement

=> Shift from I/O as interface to storage as intelligence

HPCwire Storage Series:
- The Future of Storage for HPC and AI, October 13, 2025
- Future of Storage: HPC and AI Storage by the Numbers, October 16, 2025
- The Future of Storage for AI and HPC: The Requirements, They Are a Changin', October 20, 2025



DAOS: Datasets as First-class Citizen
Source: https://daos.io/daos-overview#design_principles

Object storage: *Deconstructed database*

Discussion Summary

Data movement energy dominates HPC-AI power budget.

Question: *"Are we optimizing compute while ignoring motion?"*

Move from monitoring *what* happened to explaining *why* it happened.

Question: *"If we can explain AI decisions, why not I/O behavior?"*

Converging file, object, and cloud layers for intelligence.

Question: *"Will storage become the AI engine?*

Cost of Motion	Storage & I/O Stack	Explainable I/O	Data + Metadata	AI-ready Storage

Legacy parallel I/O models misfit AI access.

Question: *"If we rebuilt the HPC stack today, would energy and AI awareness be a design principle?"*

Data + metadata coordination as key to scalable, energy-aware storage.

Question: *"Can metadata become as explainable / optimized as I/O?"*

Thank you!

Prof. Dr. Sarah M. Neuwirth
NHR South-West HPC Center
Johannes Gutenberg University Mainz

neuwirth@uni-mainz.de

https://www.hpca-group.de/
https://nhrsw.de/