

# **EOFS-Workshop 2025 "Offene Parallele Dateisysteme: Status, Entwicklungstrends und Aspekte der Ausbildung und des Betriebs"**

Thursday, February 27, 2025 - Friday, February 28, 2025

NatFak

## **Book of Abstracts**



# Contents

Fraunhofer IML: A flexible storage toolbox. . . . .	1
ETP4HPC SRA 6 Update and the Path to Explainable I/O . . . . .	1
Performanceverhalten von CephFS für HPC und AI . . . . .	1
Beschleunigte Datenpfade mit SmartNICs in verteilten Dateisystemen . . . . .	1
Status Lustre am DKRZ . . . . .	2
EOFS-Workshop 2025 - Begrüßung, Zielsetzung und Agenda . . . . .	2
Data staging between different storage systems (BeeGFS) . . . . .	2
Wird KI uns die Studierenden wegnehmen? . . . . .	3
Lustre at GSI . . . . .	3
DAOS Community Update . . . . .	3
Lustre-Installation an der JGU Mainz: Status Quo und Ausblick . . . . .	3
Transfer von grossen Datenmengen zwischen entfernten parallelen Filesystemen mit Globus . . . . .	4
Lustre status and the path forward . . . . .	4
Machine Learning for I/O: Challenges and Opportunities . . . . .	4



1

## **Fraunhofer IML: A flexible storage toolbox.**

**Author:** Mirko Rahn<sup>1</sup>

<sup>1</sup> *Fraunhofer ITWM*

**Corresponding Author:** mirko.rahn@itwm.fraunhofer.de

The talk gives an overview over the architecture of the Fraunhofer Infinite Memory Layer (IML), a toolbox for software-defined storage and in-memory computing. Some examples of how the IML facilitates the construction of domain-specific distributed storage layers are given. Furthermore the talk will outline how the IML is used as a backend in the Memory-Centric Storage for Exascale (MCSE) project.

2

## **ETP4HPC SRA 6 Update and the Path to Explainable I/O**

**Author:** Sarah Neuwirth<sup>1</sup>

<sup>1</sup> *Johannes Gutenberg University Mainz*

**Corresponding Author:** neuwirth@uni-mainz.de

3

## **Performanceverhalten von CephFS für HPC und AI**

**Author:** Sebastian Krey<sup>1</sup>

<sup>1</sup> *GWDG*

**Corresponding Author:** sebastian.krey@gwdg.de

Im Bereich der Enterprise IT ist Ceph als Open Source Storage Plattform, die Block-, File- und Object Storage bereitstellen kann, seit Jahren etabliert und wird dort insb. in OpenSource Virtualisierungsumgebungen wie OpenStack, Proxmox, etc. eingesetzt. In den letzten Jahren hat sich aber auch das Performanceverhalten insb. von CephFS sehr positiv entwickelt, sodass der NHR Standort GWDG/Uni Göttingen es als Storage Plattform für den Capacity Tier seiner HPC Systeme etabliert, um damit seine Storalandschaft zu vereinheitlichen. In diesem Vortrag wird das Performanceverhalten der installierten Storage Cluster mit der IO500 Benchmarksuite in der Performanceoptimierungsphase der Inbetriebnahme vorgestellt und mit anderen Storalandschaften am Standort verglichen.

4

## **Beschleunigte Datenpfade mit SmartNICs in verteilten Dateisystemen**

**Author:** Jan Saß<sup>1</sup>

<sup>1</sup> *Johannes Gutenberg-Universität*

**Corresponding Author:** sass@uni-mainz.de

Steigende Datenmengen und funktionale Anforderungen an verteilte Dateisysteme erschweren die Entwicklung und Administration verteilter Dateisysteme. SmartNICs bieten eine geeignete Alternative, die es erlaubt eine Aktivitäten in den Datenpfad der Client-Server-Architektur zu verlagern ohne die bestehende Architektur zusätzlich zu belasten. Die NVIDIA BlueField3 DPU (Data Processing Unit) ist eine solche SmartNIC. BlueField3 verfügt neben eigenen ARM-Prozessoren über dedizierte Beschleunigungshardware in Form von sogenannten Data Path Accelerators (DPA) sowie Lookaside Accelerators. DPAs sind direkt an den Datenpfad eingebunden und bieten 256 Threads für latenzkritische Anwendungen. Lookaside Accelerators werden über den ARM-Prozessor gesteuert und erlauben die Abarbeitung dedizierter Aufgaben, bspw. Ver- und Entschlüsselung sowie Dekompression. Wir diskutieren erste Erkenntnisse bei der Verwendung der BlueField3 auf dem Datenpfad und mögliche Einsatzmöglichkeiten in Lustre. Anwendungsszenarien beinhalten die client-seitige transparente Verschlüsselung und Entschlüsselung des Datenverkehr von Lustre.

5

## Status Lustre am DKRZ

**Author:** Carsten Beyer<sup>None</sup>

**Corresponding Author:** beyer@dkrz.de

- Status Lustre / laufender Betrieb
- Probleme mit MultiRail bei zusätzlichem tcp LNET (auf IB) und externen Clients
- Dateisystemanalyse / MDT-Scans

6

## EOFS-Workshop 2025 - Begrüßung, Zielsetzung und Agenda

**Authors:** Frank Baetke<sup>1</sup>; Sarah Neuwirth<sup>2</sup>

<sup>1</sup> *EOFS*

<sup>2</sup> *Johannes Gutenberg University Mainz*

**Corresponding Author:** frank.baetke@gmx.com

7

## Data staging between different storage systems (BeeGFS)

**Author:** Ingo Martini<sup>1</sup>

<sup>1</sup> *ThinkParQ / BeeGFS*

**Corresponding Author:** ingo.martini@thinkparq.com

One of the challenges of dealing with large amounts of data is distributing and moving data between specialized storage systems to take advantage of their individual benefits. The use of different storage technologies for central repositories and scratch space and the staging of data into

and out of fast scratch file systems requires solutions that can move data between these systems efficiently, with high bandwidth and low administration overhead. During the talk, we will show the solutions ThinkParQ is currently working on in and around BeeGFS and discuss them with the participants.

8

## Wird KI uns die Studierenden wegnehmen?

**Author:** Michael Kuhn<sup>1</sup>

<sup>1</sup> *Otto von Guericke University Magdeburg*

**Corresponding Author:** michael.kuhn@ovgu.de

Systemnahe Themengebiete, insbesondere die der Speicher- und Dateisysteme, werden in der Studierendenausbildung häufig vernachlässigt. Notwendige Betriebssystemkonzepte werden häufig nicht mehr als Teil des Grundstudiums vermittelt, stattdessen werden zunehmend Veranstaltungen zu Machine Learning bzw. Künstlicher Intelligenz gewünscht und gefordert. Daher gestalten sich Arbeiten an Speicher- und Dateisystemen aufgrund des benötigten Vorwissens und der teilweise langen Einarbeitungszeiten schwierig. Im Rahmen des Vortrags soll sowohl auf die Entwicklungen bei der Besetzung von Professuren und der Entwicklung von Studiengängen als auch auf die Lehre an der Fakultät für Informatik der OVGU eingegangen werden.

9

## Lustre at GSI

**Author:** Thomas Roth<sup>1</sup>

<sup>1</sup> *GSI Darmstadt*

**Corresponding Author:** t.roth@gsi.de

GSI Helmholtzzentrum in Darmstadt betreibt ein Lustre-Dateisystem zur Speicherung und Analyse von Experimentdaten, Simulationsdaten und "Theorie-Daten" aller verbundener Kollaborationen und Gruppen.

Der Vortrag berichtet von verschiedenen Aspekten und Schwierigkeiten im Aufbau und Betrieb des Dateisystems.

12

## DAOS Community Update

**Author:** Michael Hennecke<sup>1</sup>

<sup>1</sup> *HPE*

**Corresponding Author:** michael.hennecke@hpe.com

13

## Lustre-Installation an der JGU Mainz: Status Quo und Ausblick

**Author:** Sergey Noskov<sup>1</sup>

**Co-author:** Markus Tacke<sup>1</sup>

<sup>1</sup> *ZDV Universität Mainz*

**Corresponding Authors:** noskov@uni-mainz.de, tacke@uni-mainz.de

Es wird die Lustre-Installation für den Cluster Mogon-NHR vorgestellt. Die Präsentation enthält technische Details über die eingesetzte Hardware und die Lustre-Tuning-Parameter. Der Ausblick zeigt geplante Änderungen und Erweiterungen für die Zukunft.

14

## Transfer von grossen Datenmengen zwischen entfernten parallelen Filesystemen mit Globus

**Author:** Christian Bolliger<sup>1</sup>

<sup>1</sup> *ETH Zurich*

**Corresponding Author:** christian.bolliger@id.ethz.ch

Der Transfer von Daten im Tera- und Petabytebereich zwischen parallelen Filesystemen wirft spezifische Fragen auf.

- Wie kann das Netzwerk optimal genutzt werden?
- Wie kann ein abgebrochener Transfer wieder aufgenommen werden?
- Wie kann eine gemeinsame Authentisierung und Authorisierung erreicht werden?
- Wie können umfangreiche Rohdaten der entsprechenden Forschungscommunity zur Verfügung gestellt werden.

Mit der Globus-Plattform, die gridftp mit einem Zertifikatsmanagement verbindet, können diese Herausforderungen ohne übermässigen Aufwand angegangen werden.

15

## Lustre status and the path forward

**Authors:** Marc Vef<sup>1</sup>; Sébastien Buisson<sup>2</sup>

<sup>1</sup> *DDN\Whamcloud*

<sup>2</sup> *DDN*

**Corresponding Author:** mvef@ddn.com

This talk will present the current status of Lustre development, upcoming features, and road map. Further, other topics, such as nodemaps and the evolution of I/O interfaces, will be discussed.



16

## Machine Learning for I/O: Challenges and Opportunities

**Author:** Radita Liem<sup>1</sup>

<sup>1</sup> *RWTH Aachen University*

**Corresponding Author:** liem@itc.rwth-aachen.de

The increasing integration of machine learning and AI into HPC workflows presents both challenges and opportunities for I/O performance analysis. AI workloads, for example, generate I/O patterns that differ significantly from traditional HPC workloads, making it difficult to balance with the current I/O optimization configurations. On the other hand, machine learning also offers powerful tools to address challenges for predicting I/O performance, thus improving scheduling strategy, procurement, and application tuning.

In this talk, I will present two works related to machine learning for I/O. The first work is a benchmark extension to emulate AI workloads, and the second one utilizes a transfer learning workflow to create an effective I/O performance prediction with a fraction of the data and computing power compared to the predecessor works, which require resources not available to small and medium clusters.