AG Computing
christian.boehme@gwdg.de

Analyzing Parallel I/O

Data analysis news from GWDG and analysis tasks in the Scalable Storage Competition

Christian Boehme



November 13, 2024

Table of contents

I/O Monitoring

Hardware Landscape

Ceph

IO500 results

SSC - Scalable storage competition

GWDG
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

- open competition
- teams and individuals allowed
- active phase (completing tasks) starts Jan 2025
- predefined tasks
- planing for 2025/2026 started

SCC Tasks

- IO500
  - Submission to IO500 Webpage (Research section)
  - Optimize for 10 Client setup
  - Description of the configurations measured and performance improvement made
- MD Workbench - Lowest (maximum latency) for the fixed configuration
- Elbencho
  - arbitrary number of client nodes, 100 KByte files (same file size as DLIO)
  - 10 Client nodes, 100 KByte file vs. 100 KByte access in shared files
  - 1 client node, single 100 MByte block, ensuring cache blows out

SCC Tasks cont’d

- Quantum Espresso
  - Tune a standard use case
  - Sanity check: provide image from the visualization pipeline
- NVIDIA DALI Pipeline - Large file access pipeline with prepared python code
- Conda environment - Startup time of prepared environment, uncached
- Performance analysis (report)
- Secret Task