The Virtual Institute for I/O and the IO-500

Julian Kunkel¹, Jay Lofstead², John Bent³, George Markomanolis⁴

¹ University of Reading ² Sandia National Laboratories ³ DDN ⁴ Oak Ridge National Lab.

Contact: julliankunkel@gmail.com

The research community in high-performance computing is organized loosely. There are many distinct resources such as homepages of research groups and benchmarks. The Virtual Institute for I/O aims to provide a hub for the community and particularly newcomers to find relevant information in one single entry point. It hosts the comprehensive data center list (CDCL). Similarly to the top500, it contains information about supercomputers and their storage systems.

I/O benchmarking, particularly, the intercomparison of measured performance between sites is tricky as there are more hardware components involved and configurations to take into account. Therefore, together with the community, we standardized an HPC I/O benchmark, the IO-500 benchmark, for which the first list had been released during supercomputing in Nov 2017. Such a benchmark is also useful to assess the impact of system issues like the Meltdown and Spectre bugs.

This poster introduces the Virtual Institute for I/O, the high-performance storage list and the effort for the IO-500 which are in progress with community projects.

The Virtual Institute for I/O

The goals of the Virtual Institute for I/O (VI4IO) are:

- Provide a platform for I/O researchers and enthusiasts for exchanging information
- Foster training and international collaboration in the field of high-performance I/O
- Track/encourage the development of large storage systems by hosting information about high-performance storage systems

The philosophical cornerstones of VI4IO are:

- Treat contributors/participants equally
- Allow free participation without any fee inclusive to all
- Independent of vendors/research facilities

Open Organization

The organization uses a wiki as central hub

- Registered users can edit the content
- Mayors changes should be discussed on the contribute mailing list
- Tag clouds link between similar entities
- Supported by mailing lists, e.g.
  - Call-for-papers
  - Announcements
  - Contributions / suggestions

Community Content

The wiki covers all worldwide research groups that address high-performance I/O including:

- A taglist for available knowledge
- Research projects such as file systems
- Ongoing research projects

Everyone is welcome to add (own) group(s)!

B) Relevant I/O related tools and benchmarks

- io500
- http://io-500.org
- https://vi4io.org
- Derived schema and alternative views
- Extended schema and alternative views
- CDCL list
- Support training and teaching for storage

IO-500 Effort

Together with the community, we created the IO-500 benchmark to compare storage systems.

Goals for the benchmark:

- Capture user-experienced performance
- Reported performance is representative for:
  - IOEasy: Applications with well optimized I/O patterns
  - IOHard: Applications that require a random workload
  - MDEasy: Metadata/small objects
  - MDHard: Small files (901 bytes) in a shared directory
  - Find: Finding relevant objects based on patterns

Challenges:

- Representative: for optimized, naive I/O-heavy workloads, and small objects
- Inclusive: cover various storage technology and non-POSIX APIs
- Trustworthy: representative results and prevent cheating
- Cheap: easy to run and short benchmarking time (in the order of minutes)

Benefits for the community beyond the IO-500:

- Support the development of benchmarks that are used (IO-500 builds on standard benchmarks)
- Feed back best practices of tool usage (e.g., find) and benchmarks
- Aid detailed comparison of individual system characteristics while having a ranked list
- Share best-practices to obtain good performance

IO-500 List Nov 2018

There are several lists available, the full list contains all the results submitted for comparison while entries can enter a ranked list upon user choice and only one solution per system.

There are several ranked lists with awards that stimulate aspects of I/O system development:

- The IO-500 award for the fastest system in the ranked IO-500 list
- The 10 node challenge fosters performance for small-scale runs

Further awards may follow.

The ranked IO-500 list:

Flexible equations

It supports equations to compute derived metrics, here flexible equation

client_nodes

The system displays a net graph for further distinguishing the best systems:

As we can see, this can be used to create arbitrary new rankings and investigate the data.

All results are available

The individual submission scripts and results for the benchmarks are preserved and can be accessed. The data is also available as CSV file for offline analysis.

Data Center List

The comprehensive data center list with its system model describes how characteristics are assigned to components. Storage is difficulty to assign to a single component as it is often shared across supercomputers, therefore, a flexible component based model is used.

Supported components:

- Site: Describes the facility
- Supercomputer: A system
- Storage system
- Nodes
- Network
- Building

The schema is under active development—we aim to describe data center characteristics. The web page allows the creation of a topology for the facility to indicate the relation between the components—ultimately multiple views will be created to show, e.g.:

- Logical network connectivity
- Physical layout in racks
- Building organization

Metrics: Most metrics can be determined with out measurement and describe hardware and software characteristics that should be known to the site and vendor. A few metrics cover actually observed metadata and I/O performance, in this case the measurement procedure must be defined. The list stores data entered in the wiki into a database and converts data to a base unit.

The following is an example of the schema for the DKRZ system:

The rules for determining performance are relaxed due to the complexity of I/O measurements, but this is augmented by the IO-500.

Derived Analysis

With the collected data many in-depth analysis becomes possible, for example, the relationship between storage and memory capacity:

Ongoing Work

- IO-500:
  - Clarified execution rules
  - Procedures to adapt IO-500
  - Integration of optional benchmarks
  - Continuous integration deployment including performance regression
  - Finalize vendor engagement program
- VI4IO standardization efforts
- Data center representation
- Next-generation interfaces (NGI)
- CDCL list
- Extended schema and alternative views
- More CDCL sites
- Better link between IO-500 and CDCL
- Support training and teaching for storage

VI4IO, IO-500, and You

You are welcome to join the mailing lists or our slack channel and participate!

Join us on Slack:

The content is under open licenses.

More details on:

- https://vi4io.org
- http://io-500.org