

Optimizing Scratch Filesystems for HPC and AI Workloads

Overview:

This half-day technical workshop examines the types of scratch filesystems in supporting high-performance computing (HPC) and AI workloads and how to troubleshoot and optimise their performance.

Attendees will gain insight into multiple scratch filesystem and storage technologies through a set of three real-world scenarios utilizing IO-intensive (challenging) HPC algorithms.

Vendors will each describe their filesystem via IO 500 Benchmark, then show their results for up to three application types via a highlight of performance before tuning and then after.

Highlighting Methods relating to:

1. Gathering information pertaining to performance (logging and what to log)
2. Interpreting the results and identifying issues
3. Remediating issues on that filesystem
4. Other ways to address issues (ie non-scratch FS alternatives).

How the after result was achieved will be shown via demonstrations and led exercises.

The scratch technologies included are Lustre, DAOS, VAST, Weka, and HDS.

The session will highlight strengths, use cases, and trade-offs of different technology and tuning and look at the approach of each vendor and technology in how it would address these cases and the needs of each use case.

Participants will walk away with a deeper understanding of scratch storage strategies aligned with domain-specific computing challenges.

Target Audience:

HPC systems architects, storage administrators, AI infrastructure engineers, data scientists, and research computing professionals.

Preparation Technical description:

The plan has been to provide three use cases ahead of time to presenters (as well as the requirement to IO500 to baseline each test scratch system to help frame the test environment)

The three subject domain algorithm examples and data sets:

1. Radio Astronomy (Pawsey reference data set two examples - radio astronomy prerequisite processing to ready an intermediate product for use with Cotter).
2. AI/MLPerf - Cosmoflow 3D convolutional neural network written in TensorFlow with the Keras API and uses [Horovod](#) for distributed training.
3. Bioinformatics IO Example from Monash University.

They are to be found at:

https://drive.google.com/drive/folders/1LZq4ocO3ue4hEY2a18GSGetlqawzr1L0?usp=drive_link

There will be co-presenters from companies and technical leaders to facilitate this workshop.

The following people are confirmed to be ready to contribute:

- Evan Koutsandreou - Weka IO
- Ed Habjan - HPE DAOS
- Gary Mancuso – Hitachi Vantara
- Kyle Lamb - VAST Data
- Shuichi Ihara - Performance Engineering Manager, Principal Engineer - DataDirect Networks Japan. - Lustre

Considerations/inclusions to speak to:

- Technology fit or tuning for different domains
- Scalability, manageability, and economics
- Informal discussions with technology leaders
- Demos and vendor materials
- Future collaboration opportunities

Recording and sharing info - Volunteer website provider?

Workshop Organising Committee members

Chair and primary workshop leader: Chris Schlipalius - Storage Manager, The Pawsey Supercomputing Centre

Workshop Vice Chairs:

- Gin Tan - Associate Director, Monash University
- Sarah Walters - Research Systems Projects and Delivery Manager, University of Queensland
- Gary Mancuso - HPC and AI Specialist. Hitachi Vantara (Industry)
- Chris Hines - Senior HPC Consultant, Monash University

Location/Venue

The workshop will be held in Boulevard B2, located on the Boulevard level of the BCEC. The room will be set cabaret style (open ended rounds).

Audio Visual facilities will be available for the workshop including data projector and screen, microphone, flipchart, and white board if needed. Wireless internet will be available with login details to be provided on the day. Power boards will also be available for attendees to plug in their laptops.

An A/V technician will be available on the day to assist with the uploading of presentation slides. We recommend meeting with the technician at least 15 minutes prior to the commencement of your session. The A/V technician will remain floating between workshops rooms throughout the day and is available to assist with any problems in the unlikely event these arise.

Note there will be no chair allocated to the session so please facilitate the day yourselves.

Catering

Arrival tea/coffee and morning tea will be provided

Catering will be served in the Boulevard 1-3 Concourse – please check the online program for break times. <https://conference.eresearch.edu.au/program/>

Schedule (Half-Day Workshop) Timetable

8:30AM Registration | Boulevard 1-3 Concourse

9:00 AM – 9:05AM Welcome - Presenters: Chris Schlipalius (Pawsey) and Gin Tan (Monash)
Recording and sharing information. Volunteer website provider?

9:05- 9:30 AM Workshop Objectives- Presenters: Sarah Walters (UQ) and Gin Tan (Monash) or Chris Hines (Monash) or Chris Schlipalius (7 mins per example?)

1. List the goals for the session.
2. Provide an introduction to scratch storage in HPC and AI
3. Provide an overview of technologies and algorithms and datasets used - the three examples

- Sarah will cover ? Cosmoflow/ML Bench?
- Gin Tan will cover? the Monash University Bio example
- Do you need Chris Schlipalius to cover the Pawsey example?

9:30 AM to 10:00 AM Presenters: Sarah Walters and Gary Mancuso

1. Why do we do all this?
2. Overview & Baseline Results Explainer and IO500
(The goal here is to explain the benchmarking of the environments used for the scratch and why this matters. ie Apples with Apples.)
3. Gary's presentation on these items

Other possible inclusions/mentions:

- One may look at other options like improving the Application code, use of local memory or NMVME on nodes say.
- IO500 methodology for evaluating scratch performance - is it accurate or useful?
- How Baseline results look across filesystems - how they look in general (Sarah and Gary)
- How benchmarks relate to real-world workloads IO500 Benchmarking vs the real world.
- Why good methodology and practice is important.
- How a multi user system is difficult to profile - so what can be done?

10:00 AM – 10:30 AM Key International Presenter - Shuichi Ihara (DDN) Lustre

10:30 AM – 11:00 AM Morning Tea | Boulevard 1-3 Concourse

11:00 AM – 11:40 AM Presenter - Evan Koutsandreou WekaIO

11:40 AM - 12:10 PM Presenter - Kyle Lamb VAST Data

12:10 PM- 12:25 PM Presenter - Ed Habjan HPE DAOS

12:25 PM - 12:55PM (or until they kick us out) Interaction for all - questions and answers session and followups. Do attendees wish to have an open discussion with a panel of presenters as we should have time.

1:00PM Hard Close (Room must be vacant for cleaning at 1:15pm)