

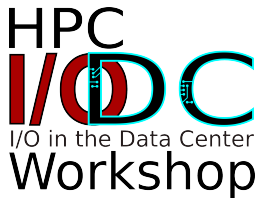
1st HPC I/O in the Data Center Workshop

Julian M. Kunkel, Jay Lofstead, Colin McMurtrie

kunkel@dkrz.de, gflfst@sandia.gov, cmurtrie@cscs.ch

DKRZ, Sandia National Lab, CSCS

07-16-2015



Agenda

9:00 Welcome

9:15- **3 Presentations**

11:00 Coffee break

11:30- **2 Presentations**

12:30 **Discussion round 1 (hot topics)**

13:00 Lunch

14:00- **4 Presentations**

16:00 Coffee break

16:30 **1 Presentation**

17:00 **Discussion round 2 (hot topics)**

17:30 Organization of the paper, Feedback and farewell

Motivation for the Workshop

- Perspective of centers is often ignored – focus on applications
- Data centers aim to provide optimal service and performance

Providing a good storage strategy is challenging

- Though there are few HPC file systems: Lustre, GPFS, (BeeGFS)
 - Management of large volume/file numbers of data is difficult
 - Performance is often suboptimal: HDF5, NetCDF, small files
 - Shared storage and quality of service?
- Middleware to fix file system *issues* present in all file systems
 - PLFS, SIONlib, ADIOS, ...
 - Domain/Application-specific “solutions”, e.g. XIOS, CDI-PIO, ...
- Zoo of emerging storage approaches
 - Burst buffers, specialized storage for small files, ...
 - Alternative storage paradigms from BigData

Understanding Systems and Users

Knowing the behavior would allow to provide a better system

- Users often don't know their I/O patterns
- The I/O stack is challenging even for experts
- A perfect monitoring of usage and efficiency would allow for
 - selection of the right storage technology
 - gearing optimization effort towards mostly used I/O libraries
 - understanding the requirements for the procurement
 - optimizing the data center's efficiency as a whole

Maybe I/O experts from data centers can make a difference

- From **individual** activity towards **community** effort and ultimately useful de-facto **standards**

About the HPC-IODC Workshop

Goal: Bring together I/O experts from data centers

- Regardless of file system
- Foster information exchange
- Opportunity for networking

Topics of interest

- Scientific workload
- Usage characteristics (file, folders, scientific libraries)
- System perspective
- Architecture
- Performance aspects
- Monitoring
- Issues during production
- Problems and potential solutions

Workshop results

- Presentations will be made available on webpage
- Joint journal paper about current practise in the data center
 - Everybody is invited to participate
 - Submission to OpenAccess Journal (superfri.org)
 - Basic structure according to the topics of interest

Workshop results

- Presentations will be made available on webpage
- Joint journal paper about current practise in the data center
 - Everybody is invited to participate
 - Submission to OpenAccess Journal (superfri.org)
 - Basic structure according to the topics of interest

Check if everybody agrees...

Organization of the paper

- Proposed title: „Data Center Perspectives on HPC-IO“
- Basic structure according to the topics of interest
- Summary of the workshop results across the data centers
- Contributions
 - Contemporary survey of data centers
 - Strength and weaknesses of current approaches
 - Perspectives/approaches to overcome issues
- LaTeX in a GIT repository
- Schedule
 - 20 August: Organizers prepared the repository and a draft with sections
 - 20 Sept.: Everybody checks the content and extends it
 - 15 Oct.: Organizers make a last pass over the content
 - 22 Oct.: Submission to the journal