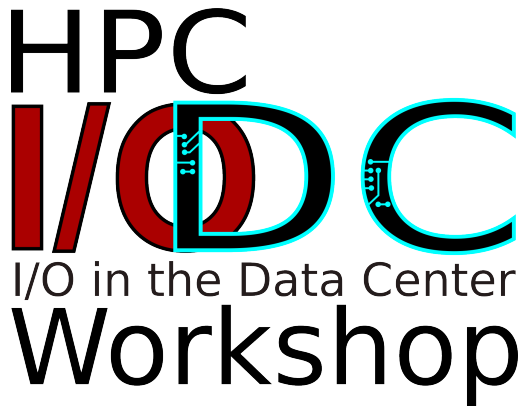
2nd HPC I/O in the Data Center Workshop

Julian M. Kunkel, Jay Lofstead, Colin McMurtrie

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

DKRZ, Sandia National Lab, CSCS

06-23-2016



Agenda

9:00	*Welcome*
9:10	**Keynote – Robert Ross, Argonne National Laboratory**
10:00	**2 Research Papers**
11:00	*Coffee break*
11:30	I/O Community
12:00	**2 Storage Expert Presentations**
12:30	Discussion round 1 (hot topics)
13:00	*Lunch*
14:00	**3 Storage Expert Presentations**
15:30	Discussion round: NVM impacts
16:00	*Coffee break*
16:30	**2 Research Papers**
17:30	Discussion round, Farewell
18:00	*End*

Motivation for the Workshop

- I/O perspective of centers is often ignored
- 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

- A perfect understanding 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
- But users often don't know their I/O patterns
- The I/O stack is challenging even for experts

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

- From **individual** activity towards **community** effort and ultimately useful **conventions**

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

- Scientic 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 our webpage
- Research Papers are published in Springer LNCS
- We will write a preface and summarize the workshop results