HPS

University of
Reading

BoF: The IO-500 and the Virtual Institute of I/O

IO^{500} **vI4Io**

Limitless Storage
Limitless Possibilities
https://hps.vi4io.org

Julian M. Kunkel, Jay Lofstead, John Bent, George Markomanolis

ISC HPC

2018-06-27

LIMITLESS **POTENTIAL** | LIMITLESS **OPPORTUNITIES** | LIMITLESS **IMPACT**

Outline

1 Overview

2 Comprehensive Data Center List (CDCL)

3 Roadmap

4 Summary

Introduction

University of
Reading

Goals of the Virtual Institute for I/O

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

https://www.vi4io.org

Introduction

University of Reading

Philosophical cornerstones of the institute

- Treat every member and participant equally
- Allow free participation without any membership fee inclusive to all
- Be independent of vendors and research facilities

Open Organization

University of Reading

- The organization uses a wiki as central hub
  - Everybody (registered users) can edit the content
  - Mayor changes should be discussed (see below)
  - The wiki uses tag clouds to link between similar entities
- Supported by mailing lists
  - Call-for-papers
  - Announce list for relevant information
  - Contribute list to discuss and steer organizational issues
- Mayor changes should be discussed on the contribute mailing list
- Members can vote for changes

Everybody is welcome to participate

Wiki Content

University of **Reading**

- Groups involved in high-performance storage
  Overview of research groups (evtl. companies involved in research)
  - Product development the group is involved in
  - Research projects (with links to their source)
  - Tags for layers, products and knowledge
- Tools: *Overview of relevant tools with small descriptions*
  - Types of tools: analysis, benchmarking, I/O middleware
  - Tags for layers and features
- Data Comprehensive Center List (CDCL) / High-Performance storage list
  Characteristics of data center systems
  - Editable and owned by the community
- Internal section
  Provides templates and describes rules for editing the page

Comprehensive Data Center List (CDCL)

University of Reading

The CDCL contains system characteristics for sites, supercomputer and storage

System Model

- The system model has been refined since ISC
    - ▶ Now based on an extensible JSON schema, optimized editor
    - ▶ Supports now (all) logical components and subcomponents
- Characteristics and peak values
- Measured values *-500

Components with characteristics

- Site, supercomputer, online storage, tape archives
- Compute nodes, storage nodes, local storage, accelerators, ...
- Supporting: e.g., CPU type, memory available, ...

Overview
OOOO

Comprehensive Data Center List (CDCL)
OO●

Roadmap
OO

Summary
O

CDCL Storage List 2018

University of
Reading

Features

■ Table view with selectable columns

■ Flexible metrics and aggregation
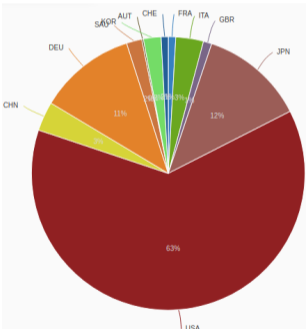
2018

#	site.institution	site.storage system.net capacity	site.supercomputer.compute peak	site.supercomputer.memory capacity
		in PiB	in PFLOPS	in TB
1	Oak Ridge National Laboratory	250.04	220.64	3511.66
2	Los Alamos National Laboratory	72.83	11.08	2110.00
3	German Climate Computing Center	52.00	3.69	983.60
4	Lawrence Livermore National Laboratory	48.85	20.10	1500.00
5	RIKEN Advanced Institute for Computational Science	39.77	10.62	1250.00
6	National Center for Atmospheric Research	37.00	5.33	202.75
7	National Energy Research Scientific Computing Center	30.00	4.90	224.30
8	National Center for Supercomputing Applications	27.60	13.40	1649.27
9	Global Scientific Information and Computing Center	25.84	17.89	275.98
10	Joint Center for Advanced HPC	24.10	24.91	919.29
11	Cineca	23.71	12.93	455.17
12	Argonne National Laboratory	21.32	10.00	768.00
13	Forschungszentrum Jülich	20.30	6.25	454.15
14	Japan Agency for Marine-Earth Science and Technology	19.62	1.31	320.00
15	Korea Meteorological Administration	19.27	2.90	0.00
16	National Supercomputing Center in Wuxi	17.76	125.00	1310.00
17	Maryland Advanced Research Computing Center	17.00	0.87	92.67
18	King Abdullah University of Science and Technology	16.96	7.20	790.00
19	Air Force Research Laboratory	15.54	5.61	447.00
20	Leibniz Supercomputing Centre	15.00	3.58	194.00
21	National Supercomputing Center in Guangzhou	14.40	59.60	1286.00
22	National Aeronautics and Space Administration	14.21	4.97	664.00
23	Texas Advanced Computing Center	12.43	9.60	270.00
24	Engineer Research and Development Center - US Army Corps	10.66	4.57	441.60
25	Sandia National Laboratories	9.93	0.50	22.10
26	Karlsruhe Institute of Technology (KIT)	9.57	1.61	222.00
27	High-Performance Computing Centre Stuttgart	8.88	7.40	964.00
28	Total Exploration Production	8.17	6.71	54.00
29	Swiss National Supercomputing Centre	7.73	25.32	521.00
30	Eni S.p.A.	6.66	4.60	0.00
31	Nagoya University	5.33	3.20	92.00
32	PGS	5.33	5.37	584.00
33	European Centre for Medium-Range Weather Forecasts	5.33	4.25	0.00
34	Army Research Laboratory DoD Supercomputing Resource	4.09	3.70	424.00
35	University of Edinburgh	3.91	2.55	0.00
36	Pacific Northwest National Laboratory	2.40	3.40	184.00
37	Navy DoD Supercomputer Resource Center	2.11	2.05	0.00
38	Vienna Scientific Cluster	1.81	0.68	42.18
39	Center for Scientific Computing	0.75	0.51	77.57

Capacity grouped by country

Roadmap for 2018

University of Reading

Web page development

■ Provide a Javascript for embedding into any data center web page

- ▶ Allowing the site to describe and visualize their system
- ▶ Hosted by the site directly
- ▶ Allowing a simple export into VI4IO data center list
- ⇒ Towards a **standardized presentation** of systems !

■ Polish presentation of site's information

Roadmap for 2018

University of Reading

Supported community activities

- ■ Roadmaps for community benchmarks (ior, mdtest, ...)
- ■ Standardized presentation of systems
- ■ Standardization of lossy compression specifications
- ■ Stabilization of IO-500 presentation of its results
- ■ New training page linking resources for learning high-performance storage

Summary

University of Reading

- ■ The Virtual Institute for I/O is a community hub
    - ▶ Open to everybody and free to join
- ■ It contains information about
    - ▶ Tools, benchmarks
    - ▶ Research groups
    - ▶ Standardization efforts
- ■ It hosts the Comprehensive Data Center List (CDCL)
    - ▶ Covers many metrics and allows flexible visualization
    - ▶ Will track metrics across years
    - ▶ Can be updated by members
- ■ Contact me if you are interested in **standardized** system presentation
- ■ ***We need you to participate!***