Exascale I/O for Unstructured Grids (EIUG) Workshop

Julian M. Kunkel (DKRZ)
Olaf Ippisch (TU Clausthal)
Sebastian Oeste (TU Dresden)

2017-09-25
About DKRZ

German Climate Computing Center

DKRZ – Partner for Climate Research

High Performance Computing.
Sophisticated Data Management.
Competent Service.
Group Wissenschaftliches Rechnen (Scientific Computing)

Composed of DKRZ research division and Universität Hamburg research group

Research
- Analysis of parallel I/O
- I/O & energy tracing tools
- Middleware optimization
- Alternative I/O interfaces
- Data reduction techniques
- Cost & energy efficiency
Goal of the Workshop

Identify strategies for the efficient access of large data sets

- Specifically targeting scientific data along (unstructured) grids

Approach of the workshop:

1. Information exchange between experts, vendors, and users
2. Mostly 35+5 minute slots for direct questions
3. Discussion slots at the end of each day
   - Monday: identify and discuss issues
   - Tuesday: potential solutions (emerging questions)
Support

This workshop is supported by:

And powered by:
Community Activity: The Virtual Institute for I/O

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

Philosophical cornerstones of the institute

- To allow participation of everybody without a membership fee
- To treat every member and participant equally
- To be an independent organization
  - Independent of vendors and research facilities
Open Organization

- The organization uses a wiki as central hub
  - Everybody (registered users) can edit the content
  - Major 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
  - IO-500 (development of a benchmark for the IO-500 list)
- Major changes should be discussed on the contribute mailing list
- Members can vote for changes

*Everybody is welcome to participate*
The Data Center List (DCL)

- Tracks characteristics of data centers over time
- Extends High-Performance Storage List (HPSL)
  - Component model including site, supercomputer, storage
  - Covers costs, energy, etc.
  - Schema is extensible based on feedback
- Provides tools to explore data
- Community maintained (currently 39 sites)
Demo

http://www.vi4io.org
Some More Analysis: Relationship Storage/Memory Capacity

- Correlation storage cap. vs.
  - memory capacity = 0.64
  - compute peak = 0.057
- Mean(storage/mem capacity) = 59
Introduction

Workshop

Community

Agenda Structure

Agenda

Monday

- Talks
- 12:00-13:00 Lunch (sponsored)
- Talks
- 15:00-15:30 Break
- Talks
- Discussion
- 18:00 Guided tour
- 19:00 Social event

Tuesday

- Talks
- 10:20-10:50 Break
- Talks
- 12:00-13:00 Lunch
- Talks
- Discussion
- 16:00 Farewell
Social Event

- Location: Scandic Hamburg Emporio, in walking distance
- When: 19:00
- How: We will walk to the location at 18:40 from Mercure Hotel
- Sponsored, excepted for stronger alcoholics (beer is OK)
A moderated discussion slot at the end of each day
Add questions and relevant issues to our Google Doc
Follow the discussion link under the agenda on the web page
Don’t be shy ... add your thoughts!