# DERIVING STORAGE INSIGHTS FROM THE IO500

Luke Logan, Jay Lofstead, Anthony Kougkas, and Xian-He Sun
llogan@hawk.iit.edu ◦ gflofst@sandia.gov ◦ {akougkas,sun}@iit.edu

SCALABLE COMPUTING SOFTWARE LABORATORY
ILLINOIS INSTITUTE OF TECHNOLOGY

# OVERVIEW

- Data-intensive applications bottlenecked by storage performance
- The IO-500 [1] is a community benchmark that stresses storage systems
  - Collects details of the storage system (e.g., OS and # of nodes)
- We want to analyze this data to gain insights on:
  - Storage system designs
  - Purchasing decisions
  - Potential bottlenecks
  - Benefits/drawbacks of different hardware compositions

[1] IO500, August 2020. [Online]. Available: https://www.vi4io.org/io500/about/start

# THE IO-500 DATASET

- 57 Columns and 115 rows
- Submissions range from Nov. 2019 to July 2020
- Intel, NVIDIA, and Red Hat have made submissions
- Tianhe-2E, Oracle Cloud, Oakforest-PACS, and Frontera
- System information
  - OS and kernel used for metadata/storage/client nodes
  - Amount of RAM, Storage type/interface, and interconnect used for metadata/storage nodes
  - Filesystem used for storing data (e.g., Lustre)

# DATA CLEANING

- Entries in various fields were not standardized
  - Multiple phrasings for the same thing
- Multiple submissions skipped information or provided less detail than expected
  - For example, some wrote AWS and did not specify the instance
- Meaning of some fields were interpreted differently by different users
  - For example, amount of volatile memory
    - Per-node or in total?
    - Include NVRAM or not?
    - Total amount of storage?

# RESEARCH QUESTIONS

- Which FS/OS is better for different workloads?
- What is the best hardware composition of the nodes for different workloads?
  - RAM, Storage Type (e.g., NVRAM), Network (e.g., 100 GbE), CPU, etc.
- How many MD and DS nodes are necessary to get maximal performance?
- What is the minimum (energy/financial) cost needed to perform a certain workload?
- Suggestions from others are welcome!

# LIMITATIONS

- Should collect in the future:
  - The types of CPU used in the nodes (e.g., model #, cores, frequency)
  - The model # of the storage devices used and the amount of storage per-node
  - The topology of the storage system
  - The interconnect used to connect Client nodes with Metadata and Storage nodes
- Would be ideal, but not required:
  - The power consumption of the nodes
  - The financial cost of the nodes
    - This is considered private information
- Suggestions are welcome here too!

# FUTURE STEPS

- Include the results of SC'20 in the IO-500
- Improve the data collection for the IO-500
  - More guidance on the format of inputs
  - Better definitions of the different fields
  - New fields (CPU, model numbers, etc.)
- Gather community input on additional research questions and data collection
- Investigating the different research questions