SC22
Dallas, TX | hpc accelerates.

# IO500: The High-Performance Storage Community

## IO500 Committee

- Dean Hildebrand - Google
- Andreas Dilger - Whamcloud/DDN
- Julian Kunkel - Georg-August-Universität Göttingen/GWDG
- Jay Lofstead - Sandia National Laboratories
- George Markomanolis - AMD

# BoF Agenda

1. **Welcome** – Dean Hildebrand
2. **The New IO500 List Analysis** – Jay Lofstead
3. **Award Presentations** – Dean Hildebrand
4. **Community Presentation** – Kevin Harms, Argonne National Laboratory
5. **Roadmap**
   - **Website Update and Demo** - Jean Luca Bez
   - **Benchmark Phases and Extended Access Patterns -** Andreas Dilger
   - **List Split and Reproducibility -** Dean Hildebrand
6. **Community Discussion**

# IO500 Organization Status

- A US non-profit organization IO500 Foundation
    - Domain, mailing list, servers, GitHub belongs to IO500 Foundation
- Website contains results with links to details, CFS, BoF slides, etc.
    - io500.org
    - Contribute fixes at github.com/IO500/webpage
- Please join our mailing list for announcements:
    - io500.org/contact
- Please join our Slack for discussions:
    - io500workspace.slack.com/
    - Join link: rb.gy/sn8esm

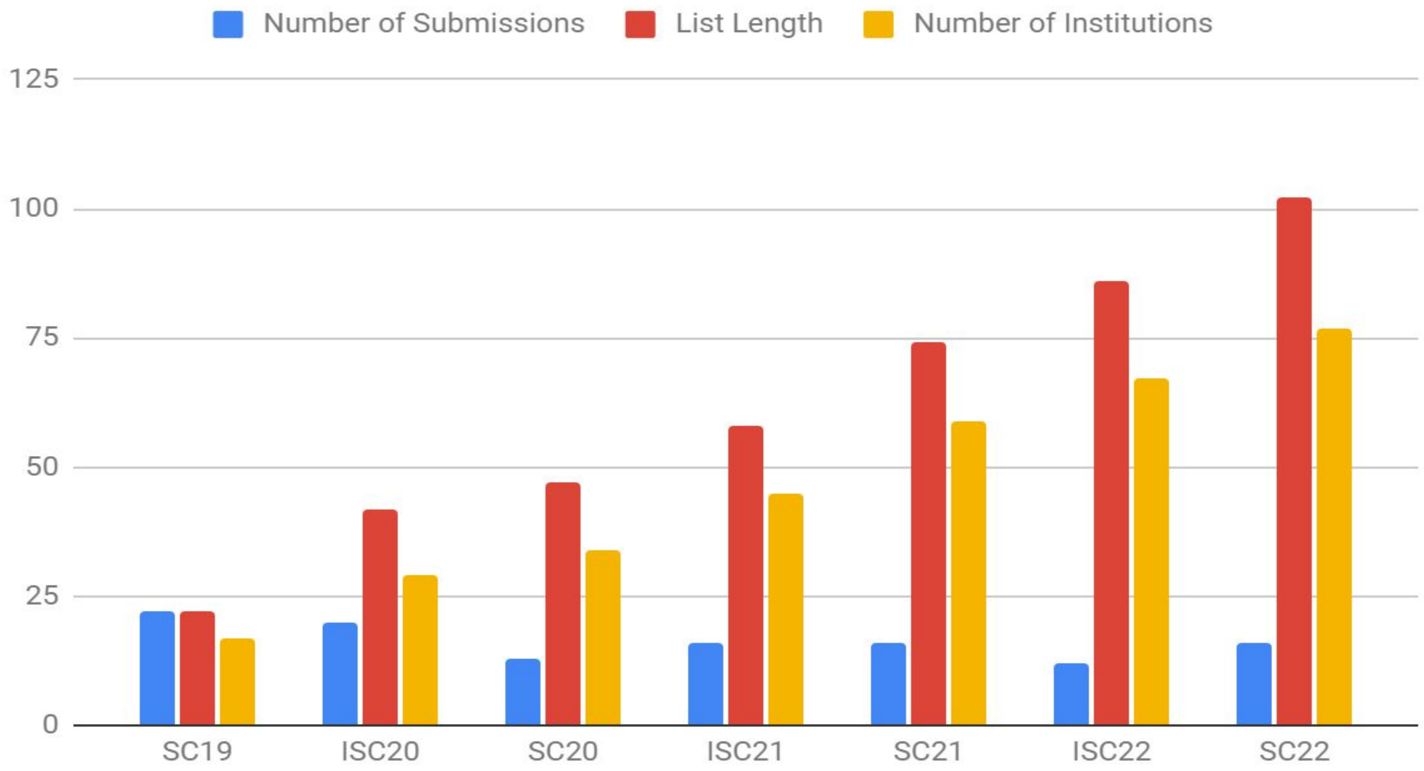
# IO500 List Analysis

# Growth in Entries and Institutions

## IO500 List

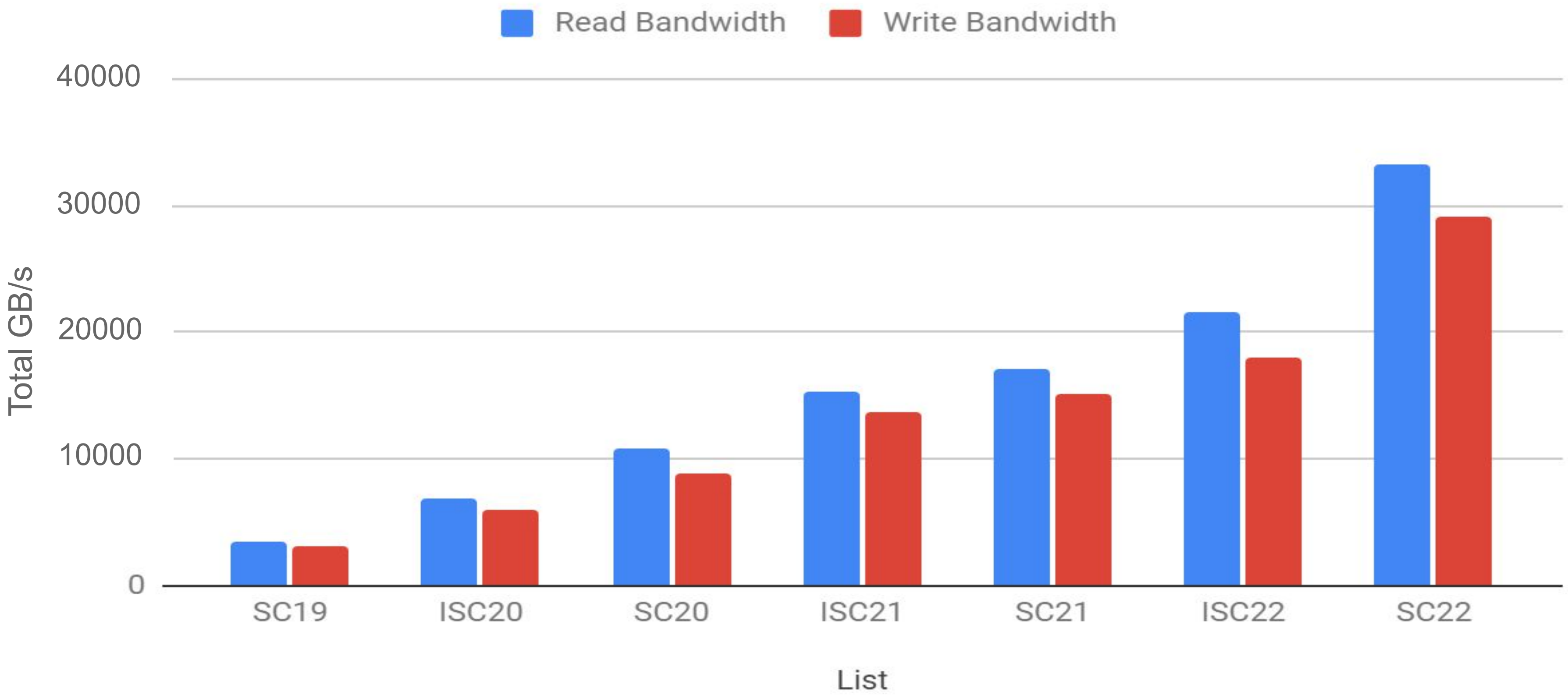
SC22

- 16 submissions
  - 5 for IO500
  - 1 for 10-client
  - 10 for both
- 102 list entries
- 77 institutions



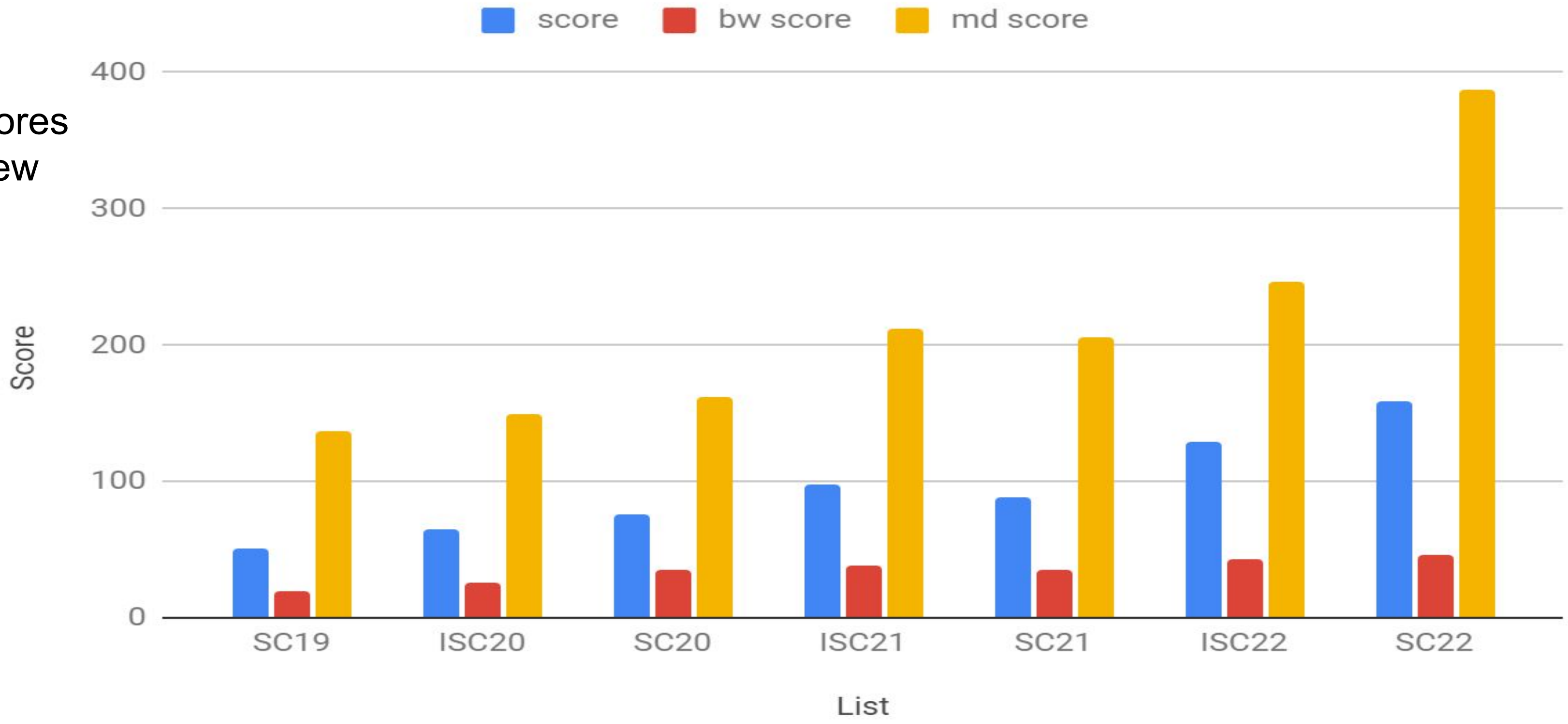
# Aggregate List Bandwidth

## IO500 List



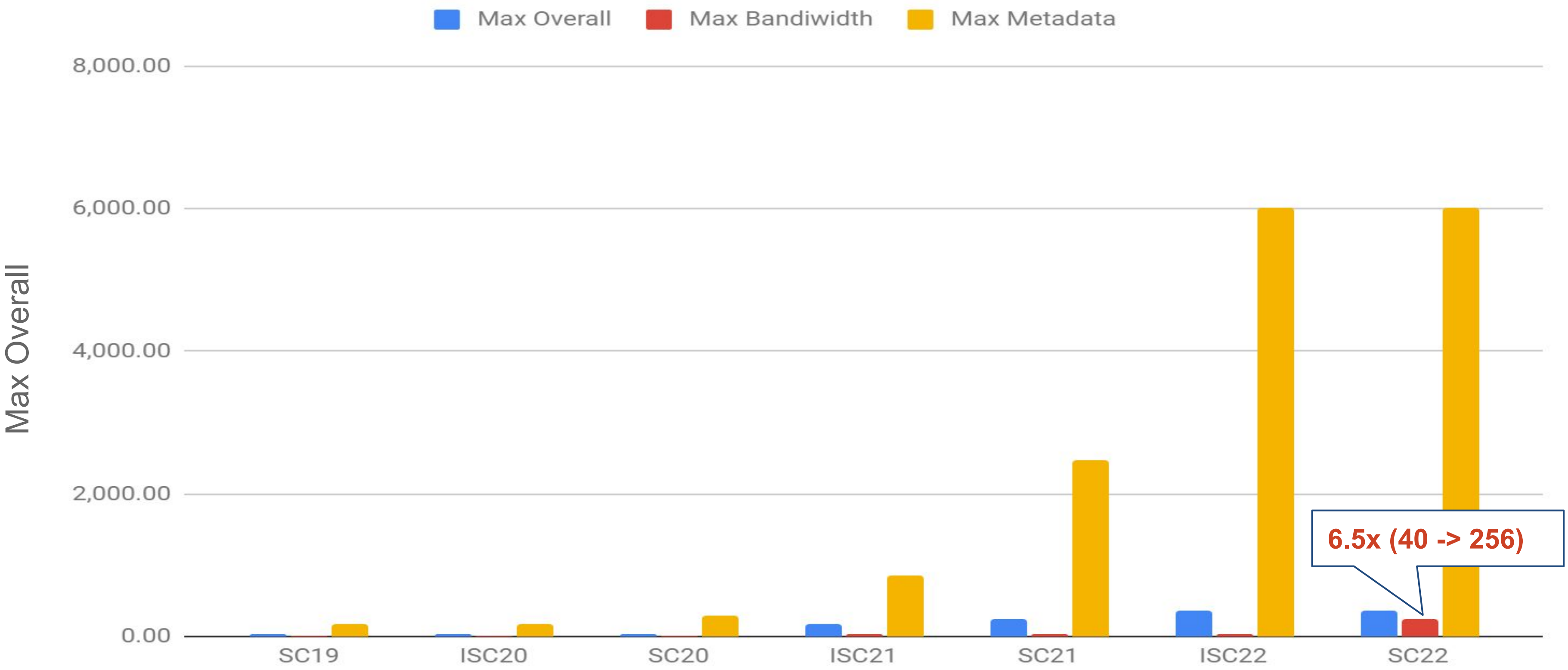
# Median Scores
## IO500 List

Median scores reached new highs



# Growth in Max Score per Client

## IO500 List



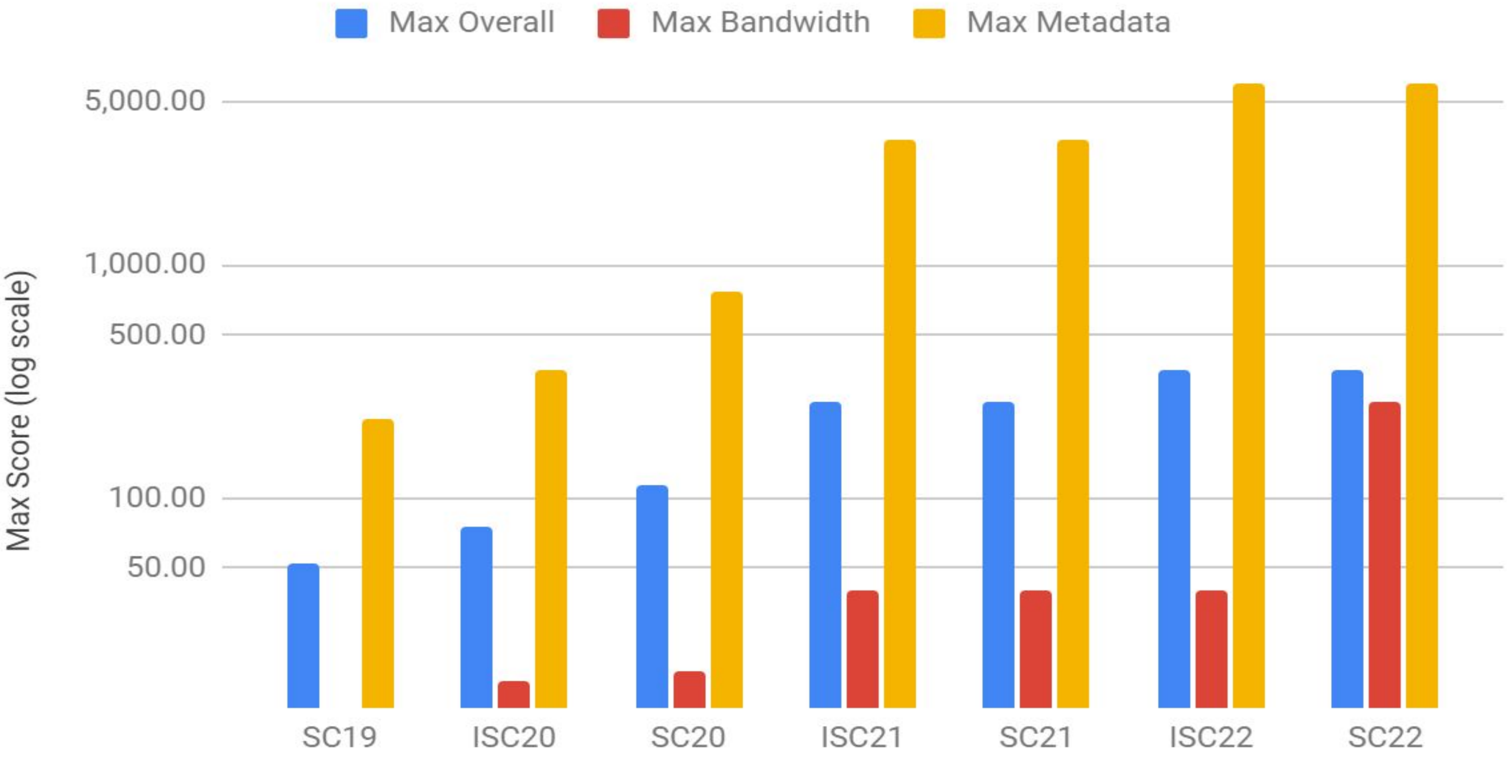
6.5x (40 -> 256)

# Growth in Max Scores per Client

## 10-Node Challenge List

No growth in metadata

Good growth in bandwidth this year



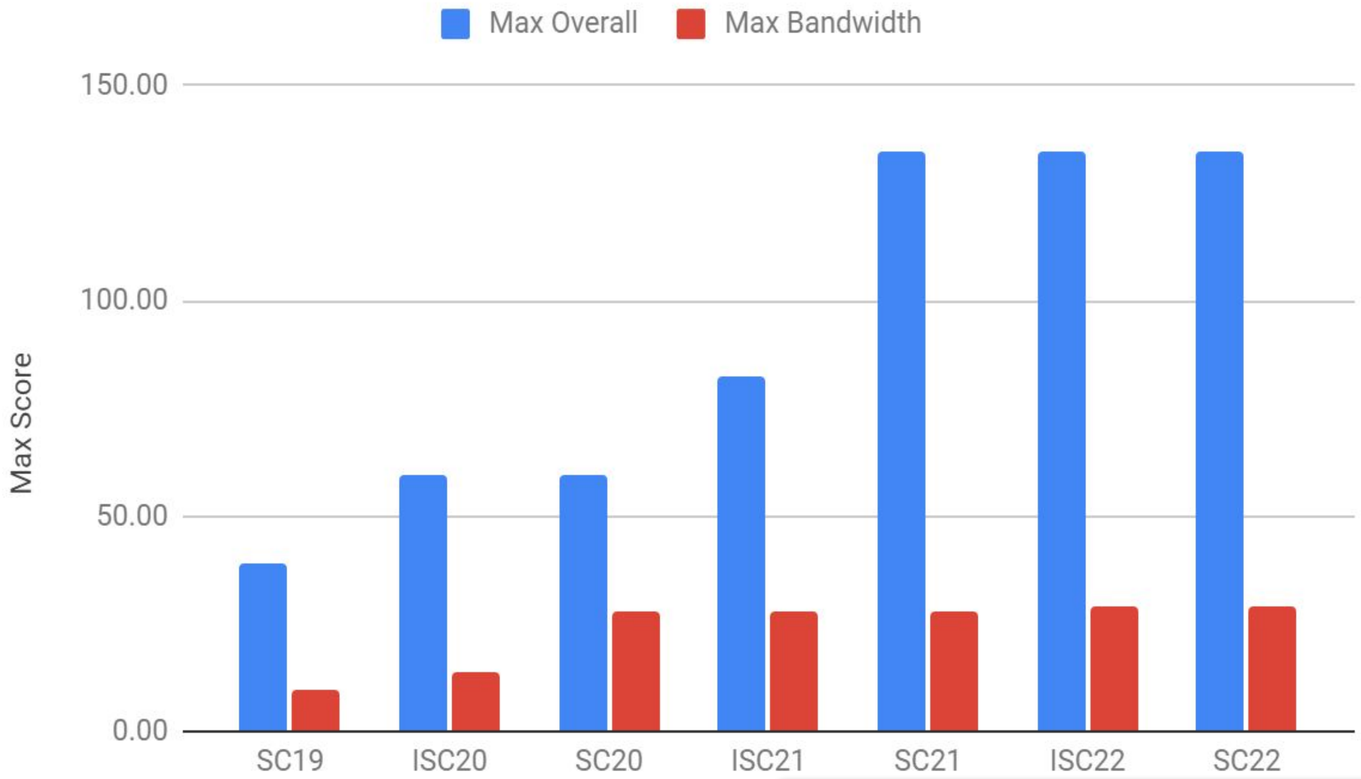
# Growth in Max Score per Storage Server

## IO500 - List

Per-client scores are flat

Per-storage server scores are growing much slower

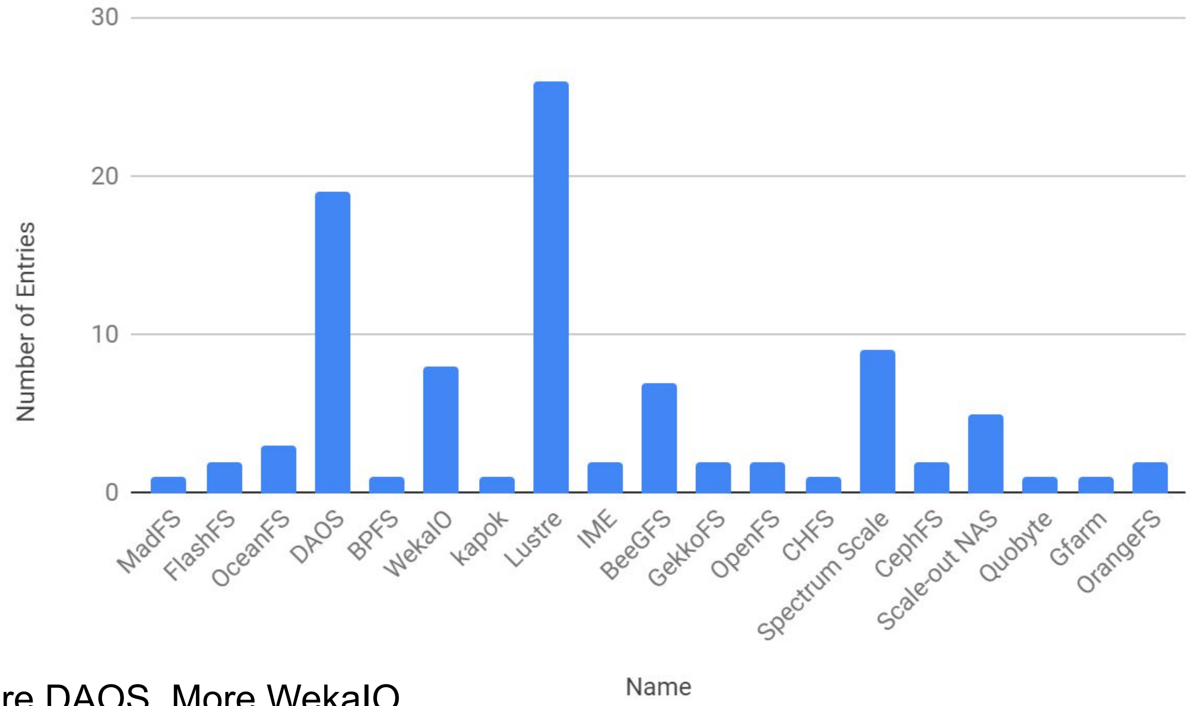
- bandwidth flat for 5 lists
- metadata flat for 3 lists



Note: metadata score per server growth reflected in overall score

# Number of File System Entries

## IO500 - List



More Lustre, More DAOS, More WekaIO
some new systems

# Award Ceremony

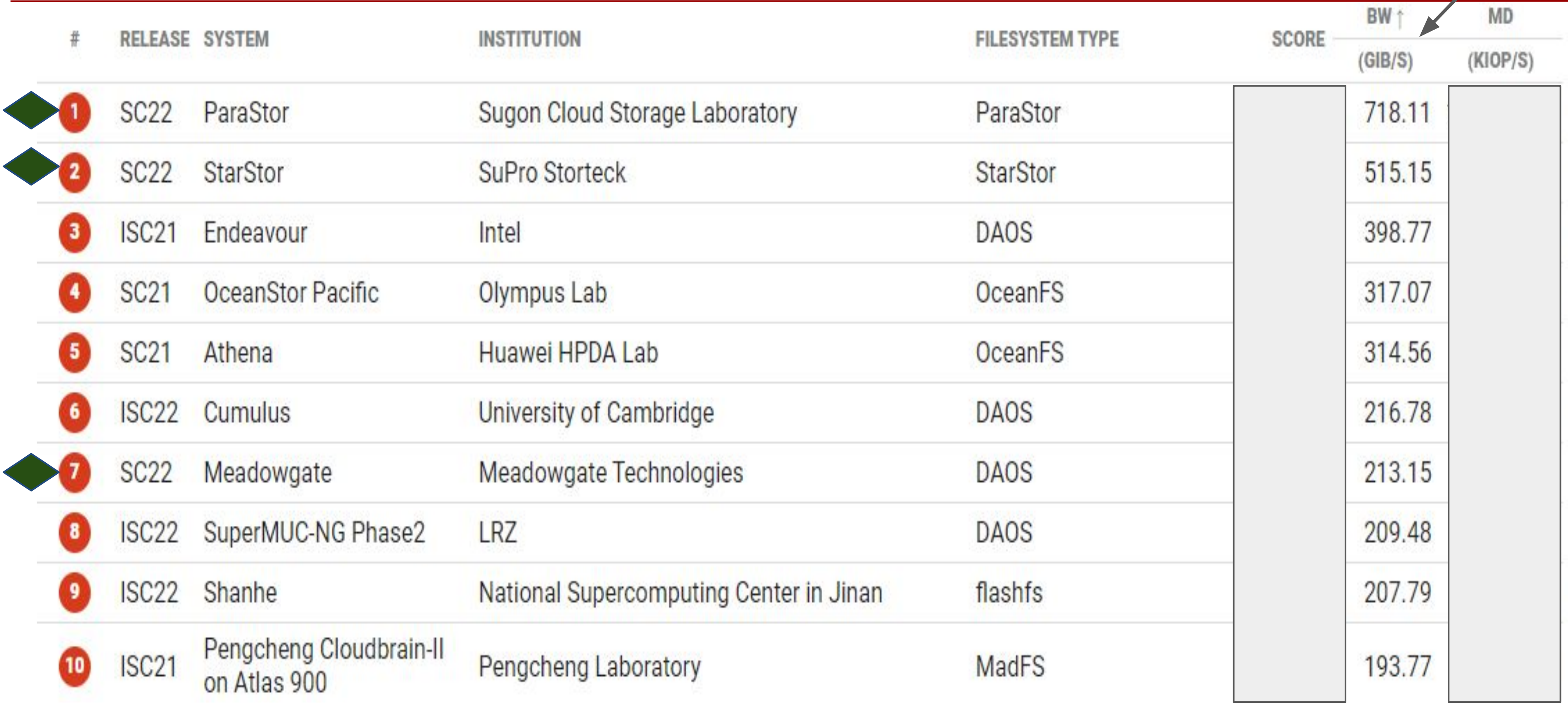
IO500

# Six Awards

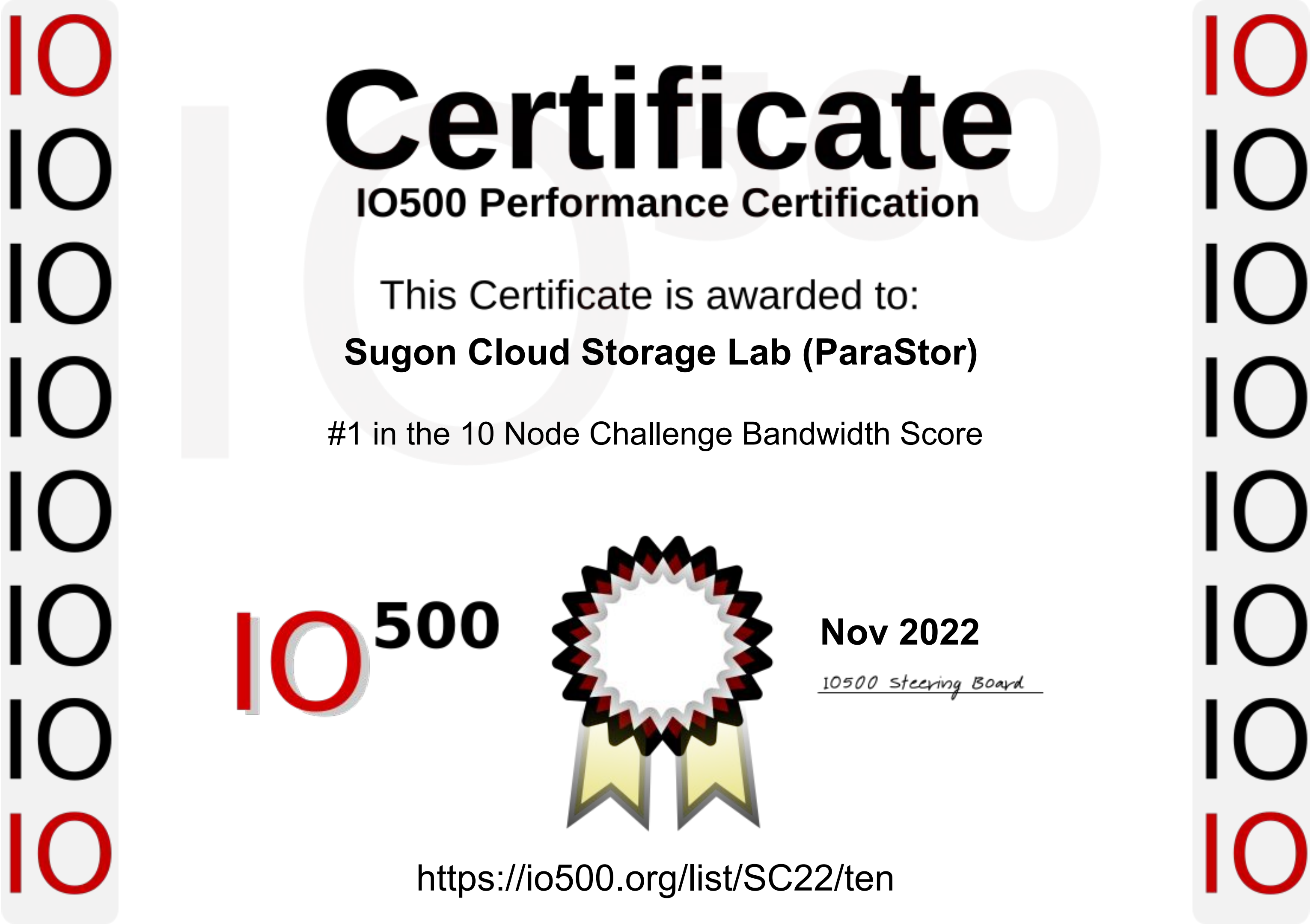
- Full List
  - Bandwidth
  - Metadata
  - Overall
- 10-Node (client) Challenge List
  - Bandwidth
  - Metadata
  - Overall

# 10 node challenge - Bandwidth Winner

**Sorted by BW**



| # | RELEASE | SYSTEM | INSTITUTION | FILESYSTEM TYPE | SCORE | BW ↑<br>(GIB/S) | MD<br>(KIOP/S) |
|---|---|---|---|---|---|---|---|
| ◆ 1 | SC22 | ParaStor | Sugon Cloud Storage Laboratory | ParaStor | [redacted] | 718.11 | [redacted] |
| ◆ 2 | SC22 | StarStor | SuPro Storteck | StarStor | [redacted] | 515.15 | [redacted] |
| 3 | ISC21 | Endeavour | Intel | DAOS | [redacted] | 398.77 | [redacted] |
| 4 | SC21 | OceanStor Pacific | Olympus Lab | OceanFS | [redacted] | 317.07 | [redacted] |
| 5 | SC21 | Athena | Huawei HPDA Lab | OceanFS | [redacted] | 314.56 | [redacted] |
| 6 | ISC22 | Cumulus | University of Cambridge | DAOS | [redacted] | 216.78 | [redacted] |
| ◆ 7 | SC22 | Meadowgate | Meadowgate Technologies | DAOS | [redacted] | 213.15 | [redacted] |
| 8 | ISC22 | SuperMUC-NG Phase2 | LRZ | DAOS | [redacted] | 209.48 | [redacted] |
| 9 | ISC22 | Shanhe | National Supercomputing Center in Jinan | flashfs | [redacted] | 207.79 | [redacted] |
| 10 | ISC21 | Pengcheng Cloudbrain-II on Atlas 900 | Pengcheng Laboratory | MadFS | [redacted] | 193.77 | [redacted] |



# Certificate

## IO500 Performance Certification

This Certificate is awarded to:

**Sugon Cloud Storage Lab (ParaStor)**

#1 in the 10 Node Challenge Bandwidth Score

**Nov 2022**

IO500 Steering Board

https://io500.org/list/SC22/ten

# Certificate

## IO500 Performance Certification

This Certificate is awarded to:

**SuPro Storteck (StarStor)**

#2 in the 10 Node Challenge Overall Score

IO500

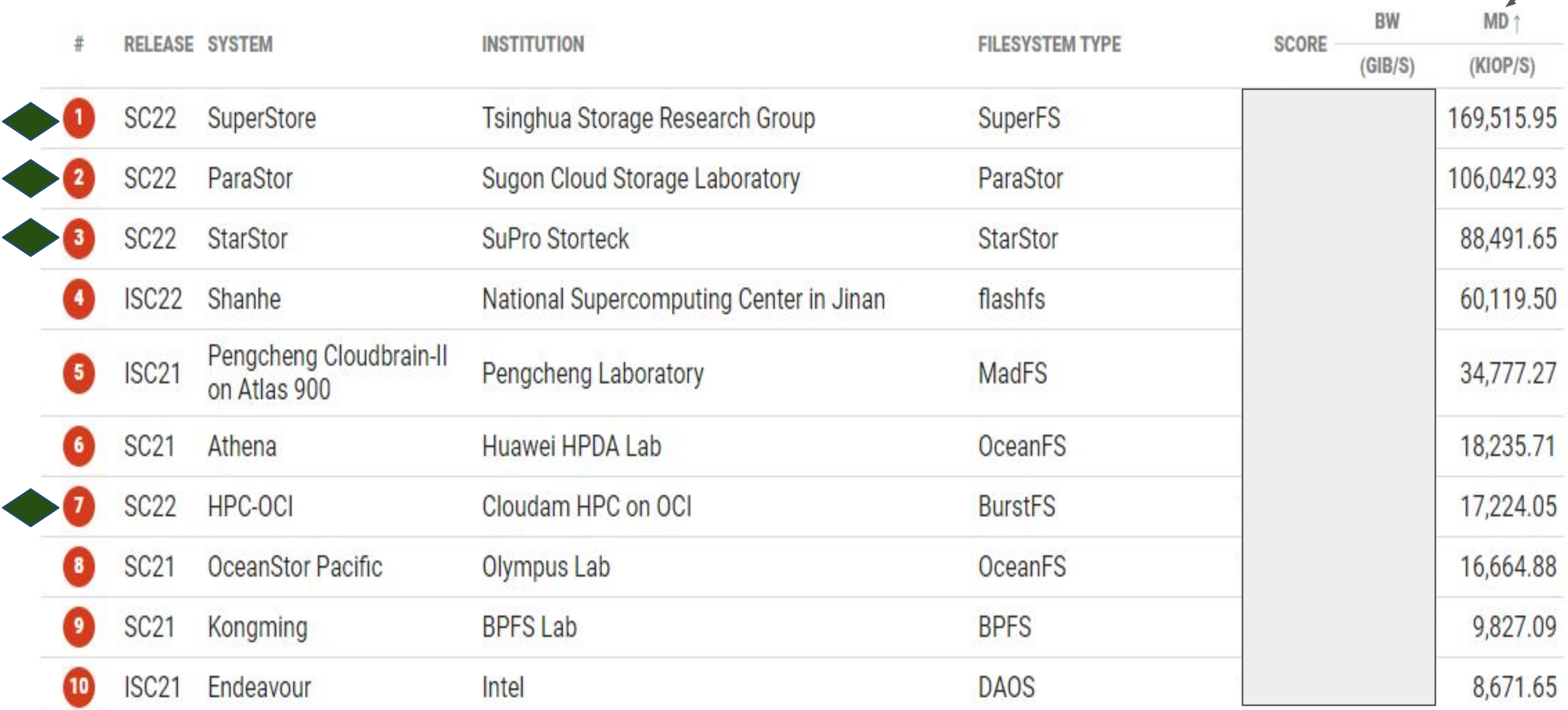
**Nov 2022**

IO500 Steering Board

https://io500.org/list/SC22/ten

# 10-Node Challenge - Metadata Winner

**Sorted by MD**



| # | RELEASE | SYSTEM | INSTITUTION | FILESYSTEM TYPE | SCORE | BW (GIB/S) | MD ↑ (KIOP/S) |
|---|---|---|---|---|---|---|---|
| ◆ 1 | SC22 | SuperStore | Tsinghua Storage Research Group | SuperFS | | | 169,515.95 |
| ◆ 2 | SC22 | ParaStor | Sugon Cloud Storage Laboratory | ParaStor | | | 106,042.93 |
| ◆ 3 | SC22 | StarStor | SuPro Storteck | StarStor | | | 88,491.65 |
| 4 | ISC22 | Shanhe | National Supercomputing Center in Jinan | flashfs | | | 60,119.50 |
| 5 | ISC21 | Pengcheng Cloudbrain-II on Atlas 900 | Pengcheng Laboratory | MadFS | | | 34,777.27 |
| 6 | SC21 | Athena | Huawei HPDA Lab | OceanFS | | | 18,235.71 |
| ◆ 7 | SC22 | HPC-OCI | Cloudam HPC on OCI | BurstFS | | | 17,224.05 |
| 8 | SC21 | OceanStor Pacific | Olympus Lab | OceanFS | | | 16,664.88 |
| 9 | SC21 | Kongming | BPFS Lab | BPFS | | | 9,827.09 |
| 10 | ISC21 | Endeavour | Intel | DAOS | | | 8,671.65 |

# Certificate

## IO500 Performance Certification

This Certificate is awarded to:

**Tsinghua Storage Research Group (SuperStor)**

#1 in the 10 Node Challenge Metadata Score

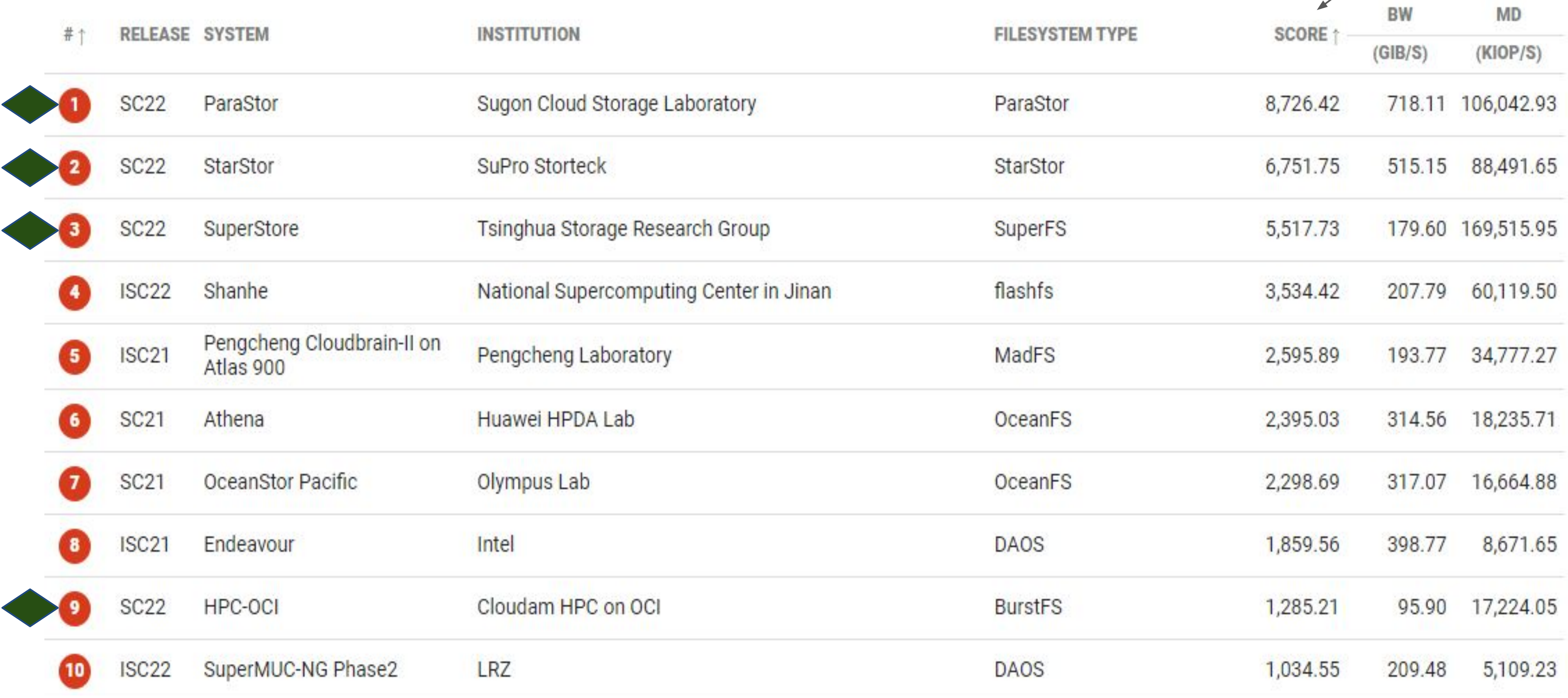
**Nov 2022**

IO500 Steering Board

https://io500.org/list/SC22/ten

# 10-Node Challenge - Winner

Sorted by score



| # ↑ | RELEASE | SYSTEM | INSTITUTION | FILESYSTEM TYPE | SCORE ↑ | BW | MD |
|---|---|---|---|---|---|---|---|
| | | | | | | (GIB/S) | (KIOP/S) |
| ◆ 1 | SC22 | ParaStor | Sugon Cloud Storage Laboratory | ParaStor | 8,726.42 | 718.11 | 106,042.93 |
| ◆ 2 | SC22 | StarStor | SuPro Storteck | StarStor | 6,751.75 | 515.15 | 88,491.65 |
| ◆ 3 | SC22 | SuperStore | Tsinghua Storage Research Group | SuperFS | 5,517.73 | 179.60 | 169,515.95 |
| 4 | ISC22 | Shanhe | National Supercomputing Center in Jinan | flashfs | 3,534.42 | 207.79 | 60,119.50 |
| 5 | ISC21 | Pengcheng Cloudbrain-II on Atlas 900 | Pengcheng Laboratory | MadFS | 2,595.89 | 193.77 | 34,777.27 |
| 6 | SC21 | Athena | Huawei HPDA Lab | OceanFS | 2,395.03 | 314.56 | 18,235.71 |
| 7 | SC21 | OceanStor Pacific | Olympus Lab | OceanFS | 2,298.69 | 317.07 | 16,664.88 |
| 8 | ISC21 | Endeavour | Intel | DAOS | 1,859.56 | 398.77 | 8,671.65 |
| ◆ 9 | SC22 | HPC-OCI | Cloudam HPC on OCI | BurstFS | 1,285.21 | 95.90 | 17,224.05 |
| 10 | ISC22 | SuperMUC-NG Phase2 | LRZ | DAOS | 1,034.55 | 209.48 | 5,109.23 |

# Certificate

## IO500 Performance Certification

This Certificate is awarded to:

**Sugon Cloud Storage Lab (ParaStor)**

#1 in the 10 Node Challenge Overall Score

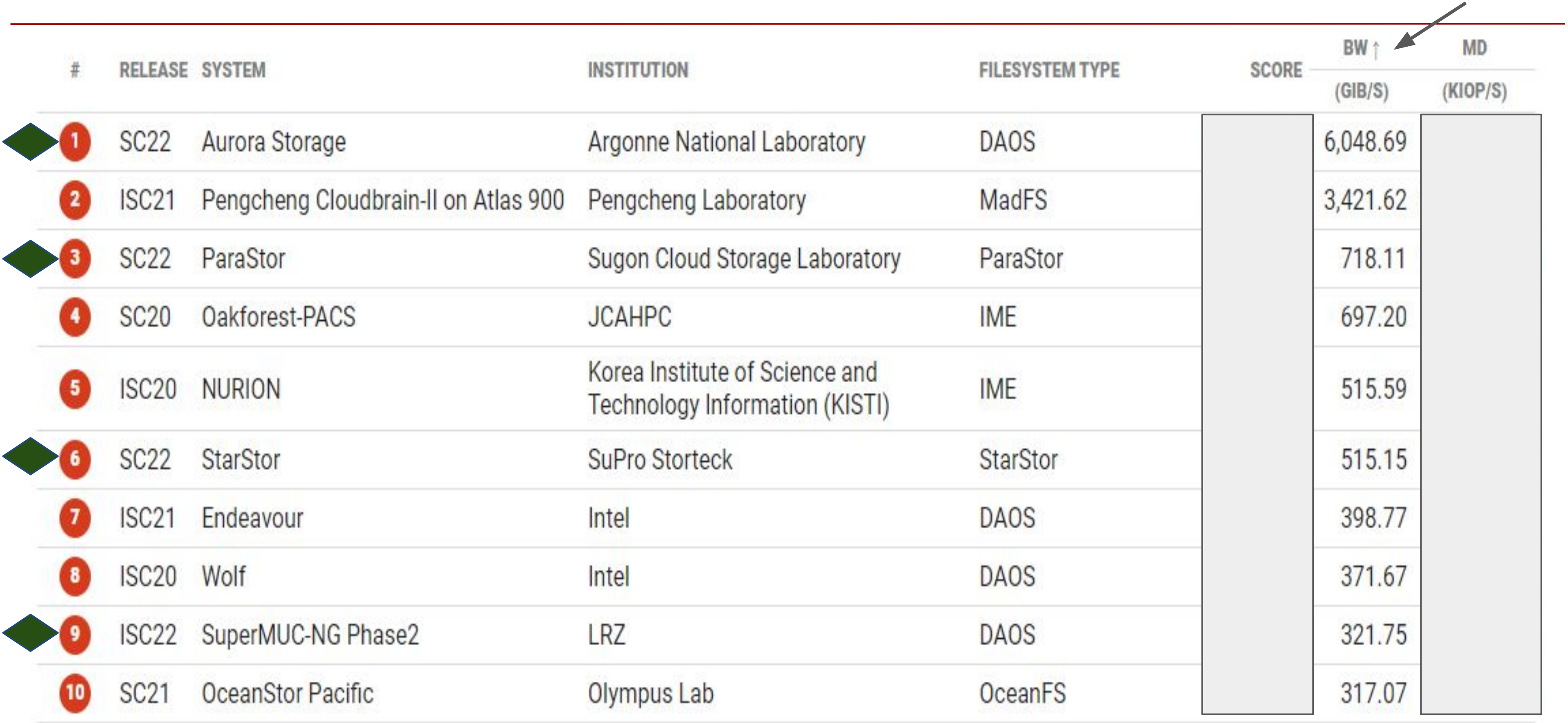
**Nov 2022**

IO500 Steering Board

https://io500.org/list/SC22/ten

# Full list - Bandwidth Winner

Sorted by BW



| # | RELEASE | SYSTEM | INSTITUTION | FILESYSTEM TYPE | SCORE | BW ↑ (GIB/S) | MD (KIOP/S) |
|---|---|---|---|---|---|---|---|
| ◆ 1 | SC22 | Aurora Storage | Argonne National Laboratory | DAOS | | 6,048.69 | |
| 2 | ISC21 | Pengcheng Cloudbrain-II on Atlas 900 | Pengcheng Laboratory | MadFS | | 3,421.62 | |
| ◆ 3 | SC22 | ParaStor | Sugon Cloud Storage Laboratory | ParaStor | | 718.11 | |
| 4 | SC20 | Oakforest-PACS | JCAHPC | IME | | 697.20 | |
| 5 | ISC20 | NURION | Korea Institute of Science and Technology Information (KISTI) | IME | | 515.59 | |
| ◆ 6 | SC22 | StarStor | SuPro Storteck | StarStor | | 515.15 | |
| 7 | ISC21 | Endeavour | Intel | DAOS | | 398.77 | |
| 8 | ISC20 | Wolf | Intel | DAOS | | 371.67 | |
| ◆ 9 | ISC22 | SuperMUC-NG Phase2 | LRZ | DAOS | | 321.75 | |
| 10 | SC21 | OceanStor Pacific | Olympus Lab | OceanFS | | 317.07 | |

# Certificate

## IO500 Performance Certification

This Certificate is awarded to:

**Argonne National Laboratory (Aurora Storage)**

#1 in the IO500 Bandwidth Score

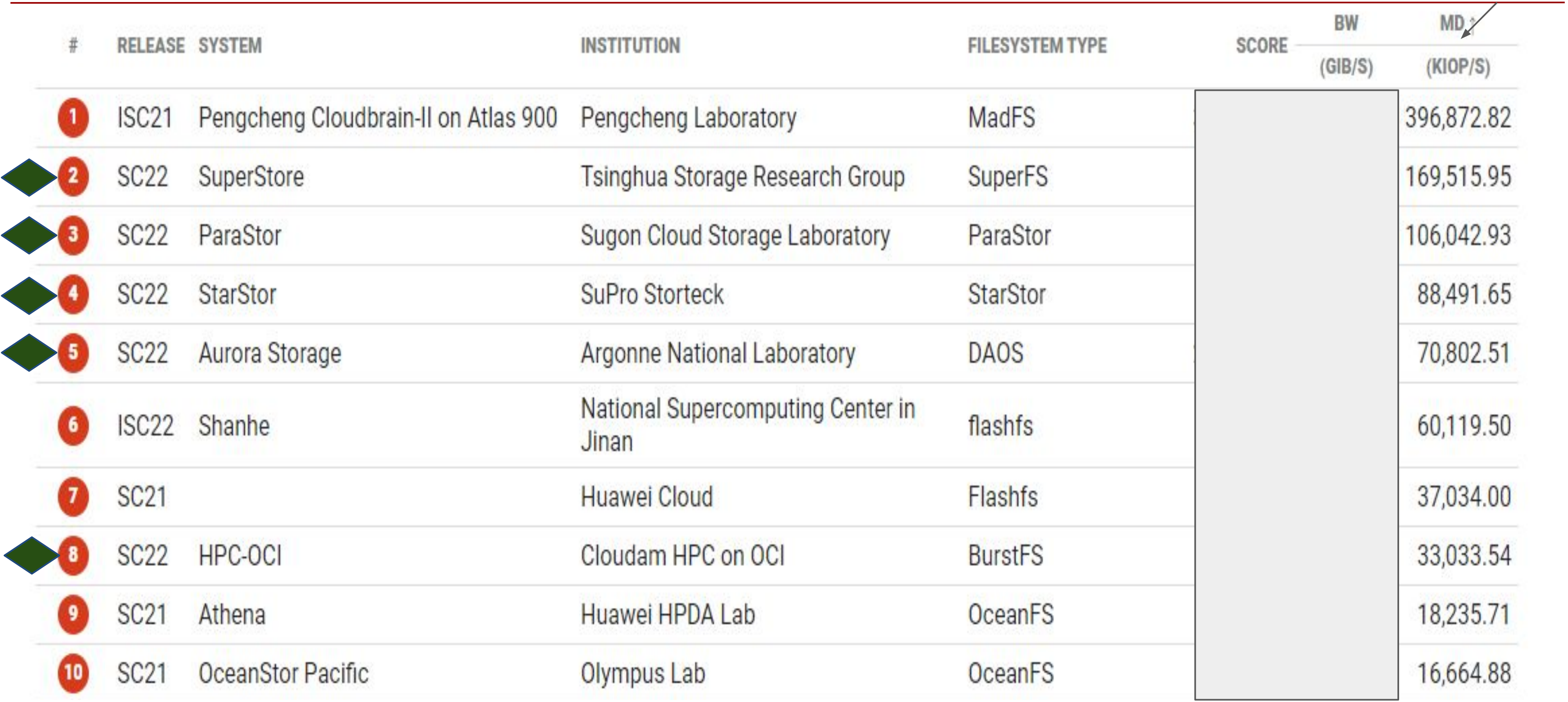
**Nov 2022**

IO500 Steering Board

https://io500.org/list/SC22/io500

# Full list - Metadata Winner

Sorted by MD



| # | RELEASE | SYSTEM | INSTITUTION | FILESYSTEM TYPE | SCORE | BW | MD ↕ |
|---|---|---|---|---|---|---|---|
| | | | | | | (GIB/S) | (KIOP/S) |
| 1 | ISC21 | Pengcheng Cloudbrain-II on Atlas 900 | Pengcheng Laboratory | MadFS | [redacted] | [redacted] | 396,872.82 |
| ◆ 2 | SC22 | SuperStore | Tsinghua Storage Research Group | SuperFS | [redacted] | [redacted] | 169,515.95 |
| ◆ 3 | SC22 | ParaStor | Sugon Cloud Storage Laboratory | ParaStor | [redacted] | [redacted] | 106,042.93 |
| ◆ 4 | SC22 | StarStor | SuPro Storteck | StarStor | [redacted] | [redacted] | 88,491.65 |
| ◆ 5 | SC22 | Aurora Storage | Argonne National Laboratory | DAOS | [redacted] | [redacted] | 70,802.51 |
| 6 | ISC22 | Shanhe | National Supercomputing Center in Jinan | flashfs | [redacted] | [redacted] | 60,119.50 |
| 7 | SC21 | | Huawei Cloud | Flashfs | [redacted] | [redacted] | 37,034.00 |
| ◆ 8 | SC22 | HPC-OCI | Cloudam HPC on OCI | BurstFS | [redacted] | [redacted] | 33,033.54 |
| 9 | SC21 | Athena | Huawei HPDA Lab | OceanFS | [redacted] | [redacted] | 18,235.71 |
| 10 | SC21 | OceanStor Pacific | Olympus Lab | OceanFS | [redacted] | [redacted] | 16,664.88 |

# Certificate

## IO500 Performance Certification

This Certificate is awarded to:

**Pengcheng Laboratory (Cloudbrain-II)**

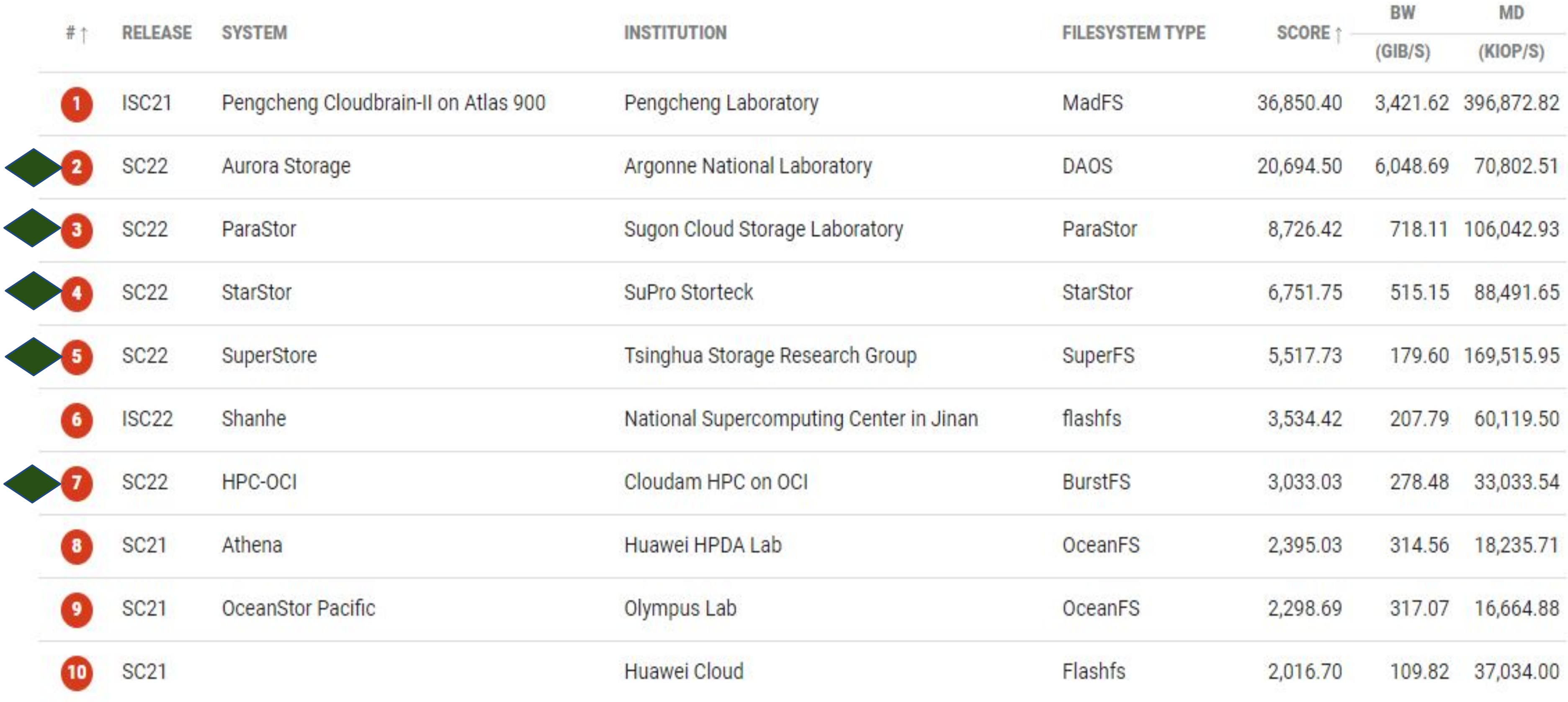
#1 in the IO500 Metadata Score

**Nov 2022**

IO500 Steering Board

https://io500.org/list/SC22/io500

# Full list - Winner



| # ↑ | RELEASE | SYSTEM | INSTITUTION | FILESYSTEM TYPE | SCORE ↑ | BW | MD |
|---|---|---|---|---|---|---|---|
| | | | | | | (GIB/S) | (KIOP/S) |
| 1 | ISC21 | Pengcheng Cloudbrain-II on Atlas 900 | Pengcheng Laboratory | MadFS | 36,850.40 | 3,421.62 | 396,872.82 |
| ◆ 2 | SC22 | Aurora Storage | Argonne National Laboratory | DAOS | 20,694.50 | 6,048.69 | 70,802.51 |
| ◆ 3 | SC22 | ParaStor | Sugon Cloud Storage Laboratory | ParaStor | 8,726.42 | 718.11 | 106,042.93 |
| ◆ 4 | SC22 | StarStor | SuPro Storteck | StarStor | 6,751.75 | 515.15 | 88,491.65 |
| ◆ 5 | SC22 | SuperStore | Tsinghua Storage Research Group | SuperFS | 5,517.73 | 179.60 | 169,515.95 |
| 6 | ISC22 | Shanhe | National Supercomputing Center in Jinan | flashfs | 3,534.42 | 207.79 | 60,119.50 |
| ◆ 7 | SC22 | HPC-OCI | Cloudam HPC on OCI | BurstFS | 3,033.03 | 278.48 | 33,033.54 |
| 8 | SC21 | Athena | Huawei HPDA Lab | OceanFS | 2,395.03 | 314.56 | 18,235.71 |
| 9 | SC21 | OceanStor Pacific | Olympus Lab | OceanFS | 2,298.69 | 317.07 | 16,664.88 |
| 10 | SC21 | | Huawei Cloud | Flashfs | 2,016.70 | 109.82 | 37,034.00 |

# Certificate

## IO500 Performance Certification

This Certificate is awarded to:

**Pengcheng Laboratory (Cloudbrain-II)**

#1 in the IO500 Overall Score

**Nov 2022**

IO500 Steering Board

https://io500.org/list/SC22/io500

# List of Awarded Systems in the Ranked Lists

| | | | | | |
|---|---|---|---|---|---|
| 10-Node | Bandwidth | Sugon Cloud Storage Lab | ParaStor | 718.11 | GiB/s |
| | Metadata | Tsinghua Storage Research | SuperFS | 169,515.95 | kIOPS |
| | **Overall** | Sugon Cloud Storage Lab | ParaStor | **8,726.42** | **score** |
| IO500 | Bandwidth | Argonne National Laboratory | DAOS | 6,048.69 | GiB/s |
| | Metadata | Pengcheng Cloudbrain-II | MadFS | 396,872.82 | kIOPS |
| | **Overall** | Pengcheng Cloudbrain-II | MadFS | **36,850.37** | **score** |

# Community Presentation

Argonne NATIONAL LABORATORY

# DAOS on Aurora

## IO500 BoF SC22

**Kevin Harms**
**Argonne Leadership Computing Facility**

IO500
https://io500.org

# Acknowledgements

- Mohamad Chaarawi - Intel
  - Did all the IO500 runs
  - Assisted with much of the performance testing

- Intel Aurora test team
  - Doing all the hardware work of testing and evaluation of the new hardware

Argonne NATIONAL LABORATORY
U.S. DEPARTMENT OF ENERGY
intel
Hewlett Packard Enterprise
Aurora

# Aurora

Leadership Computing Facility
Exascale Supercomputer

Peak Performance
**≧ 2 Exaflops DP**

Intel GPU
**Intel® Data Center GPU Max**

Intel Xeon Processor
**Intel® Xeon® CPU Max**

Platform
**HPE Cray-Ex**

**Compute Node**
2 Xeon Intel® Xeon® CPU Max processors
6 Intel® Data Center GPU Max
Node Unified Memory Architecture
8 fabric endpoints

**GPU Architecture**
Intel XeHPC architecture
High Bandwidth Memory Stacks

**Node Performance**
>130 TF

**System Size**
>9,000 nodes

**Aggregate System Memory**
>10 PB aggregate System Memory

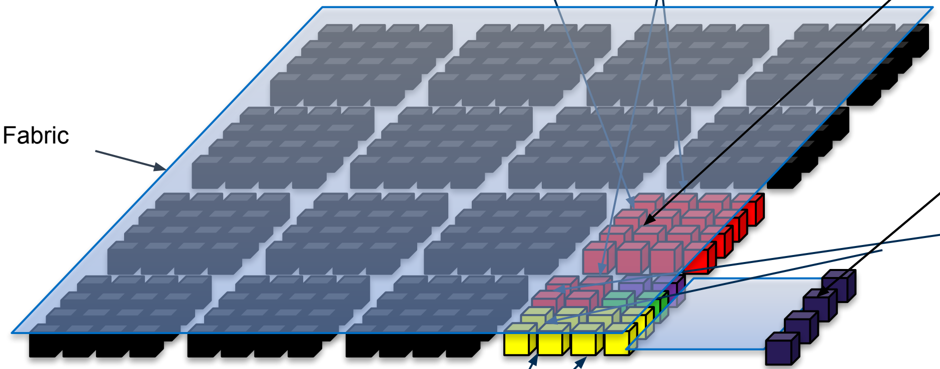
**System Interconnect**
HPE Slingshot 11
Dragonfly topology with adaptive routing

**Network Switch**
25.6 Tb/s per switch (64 200 Gb/s ports)
Links with 25 GB/s per direction

**High-Performance Storage**
220 PB
≧25 TB/s DAOS bandwidth

**Software Environment**
- C/C++
- Fortran
- SYCL/DPC++
- OpenMP offload
- Kokkos
- RAJA
- Intel Performance Tools

# Aurora Storage Overview



**DAOS Performance**
220 PB capacity @ EC16+2
≥ 25 TB/s

## DAOS Nodes (DNs)

Xeon servers
NVRAM and NVMe attached storage
DAOS service

## Lustre Performance

Grand – 100 PB @ 650 GB/s
Eagle – 100 PB @ 650 GB/s

Slingshot Fabric

System Service Nodes (SSNs)

User Access Nodes (UANs)

## Scalable Storage Cluster (SSC)

Xeon servers connected to JBOD
Lustre OSSs & MDSs

## Gateway Nodes (GNs)

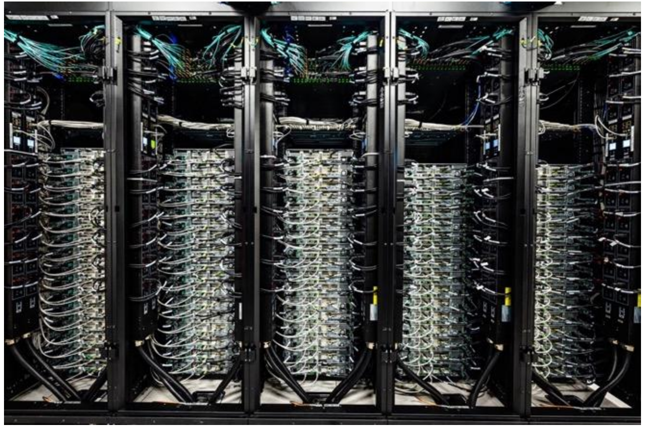
Xeon servers with no local storage
Access to external storage

**Gateway nodes**

# DAOS Node Details

- Intel Coyote Pass System
  - (2) Xeon 5320 CPU (Ice Lake)
  - (16) 32GB DDR4 DIMMs
  - (16) 512GB Intel Optane Persistent Memory 200
  - (16) 15.3TB Samsung PM1733
  - (2) HPE Slingshot NIC

- 1024 Total Servers
  - Each node will run 2 DAOS engines
  - 2048 DAOS engines



# IO-500 Results

https://io500.org

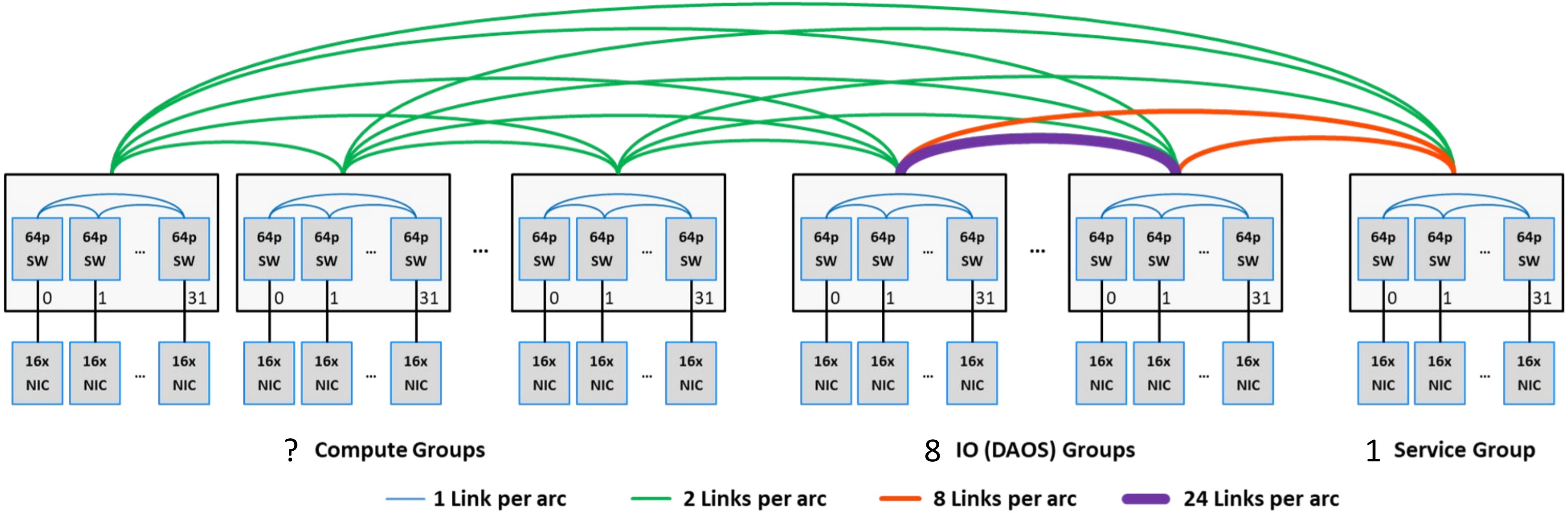
- Overall Score
  - Bandwidth 6048.687 GiB/s
  - IOPS 70802.506 Kiops
  - TOTAL 20694.496

```
IO500 version io500-sc22_v2 (standard)
[RESULT]       ior-easy-write     8121.216166 GiB/s : time 315.290 seconds
[RESULT]    mdtest-easy-write   145696.813787 kIOPS : time 330.344 seconds
[      ]            timestamp        0.000000 kIOPS : time 0.005 seconds
[RESULT]       ior-hard-write     4795.290353 GiB/s : time 327.292 seconds
[RESULT]    mdtest-hard-write    53017.993976 kIOPS : time 386.900 seconds
[RESULT]                 find    15489.721171 kIOPS : time 4303.172 seconds
[RESULT]        ior-easy-read     7902.422629 GiB/s : time 323.396 seconds
[RESULT]     mdtest-easy-stat   101911.220782 kIOPS : time 464.149 seconds
[RESULT]        ior-hard-read     4349.587799 GiB/s : time 359.606 seconds
[RESULT]     mdtest-hard-stat    80885.419297 kIOPS : time 256.336 seconds
[RESULT]   mdtest-easy-delete   113779.205446 kIOPS : time 416.927 seconds
[RESULT]     mdtest-hard-read    63803.314017 kIOPS : time 322.842 seconds
[RESULT]   mdtest-hard-delete    88201.041190 kIOPS : time 236.203 seconds
[SCORE ] Bandwidth 6048.686604 GiB/s : IOPS 70802.506017 kiops : TOTAL 20694.496120
```

# IO500 Configuration

- 250 DAOS nodes as servers
  - 2 engines per node
  - 500 engines total
- 260 DAOS nodes used as clients
- Aurora storage resources used
  - 5 total dragonfly IO groups
  - Servers and clients mixed within groups (approximately 50/50)
  - No topology optimizations for object locations
- Run configured for maximum performance using no data protection (rf=0)
- pfind modified to use DAOS libdfs API
  - https://github.com/mchaarawi/pfind/tree/dfs_find
- IOR and mdtest use DAOS backends committed to upstream repos
  - SX object class used for wide striping

# Aurora Network Architecture



- Increased DAOS inter-group bandwidth
  - Support rebuilding and inter-server communication
  - Prevent DAOS server traffic interfering with application communication
- Increased bandwidth to service group
  - Support off-cluster access and data-movement to other storage systems

# Summation

- Projecting bandwidth to Aurora
  - 6 TB/s * 4 = 24 TB/s (96% of advertised)
  - 70M IOps * 4 = 280 M IOps (lower than desired)
- Significantly more clients once we have Aurora
  - Rather than 1:1 ratio
- Significantly more network resources once using all compute groups

# Join the Aurora Team

- Looking for a post-doc to work on DAOS
  - ALCF’s performance engineering group is looking for a Postdoctoral Appointee to perform research and development on the open source DAOS storage system, in the context of the upcoming exascale platforms, and Aurora in particular.
  - Three areas of interest for study are:
    - new opportunities for applications to optimize I/O that isn't oriented around file access. DAOS provides very low latency access and the possibility allows applications to write data in a more "read-optimized" format with minimal penalty versus write-optimized formats.
    - DAOS supports a prototype "active storage" interface, and exploration of some HPC type workloads (like pointer chasing, lookup tables, etc.)
    - With the proliferation of CPUs and accelerators with significant dedicated high performance memory, the DAOS client should provide a mechanism to utilize device memory with direct-to-NIC memory movement bypassing CPU memory.

- https://argonne.wd1.myworkdayjobs.com/Argonne_Careers/job/Lemont-IL-USA/Postdoctoral-Appointee---Exascale-Storage-using-DAOS_414419

# Acknowledgements

This research used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357.

# Roadmap

# Roadmap for the IO500

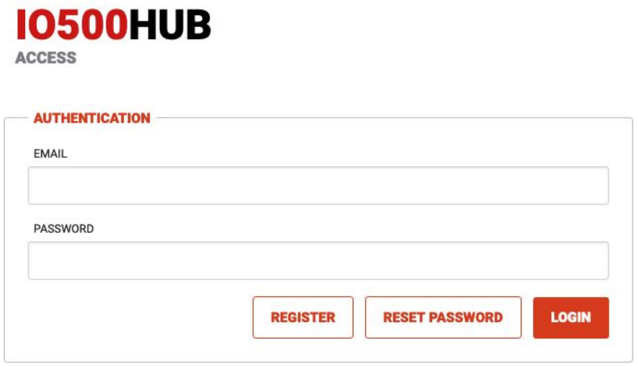
- Still working on splitting lists into Production and Research
- Fill in gaps in IO500 to improve usage patterns
  - Collect and evaluate results for potential new benchmark phases
    - Not part of benchmark score yet
  - Create proposals to give rationale and details of any potential new phase
    - Proposal must gain community consensus before official inclusion
- New `io500.org` submissions page - thanks Jean Luca
  - Will continue adding more mandatory fields and integrate reproducibility questionnaire
  - Please give feedback and be patient in the transition
- Community meeting
  - Skipped a meeting in September due to lack of topics
  - Target Feb 2023 if topics to discuss

# ISC 23 (May 21-25, 2023)

- Call for submission: March ~15th
- Testing phase ends: April ~1st
  - Code freeze, but please test before!
- Submission deadline: May 8th
- List release: BoF date TBD (ISC during May 21-25)

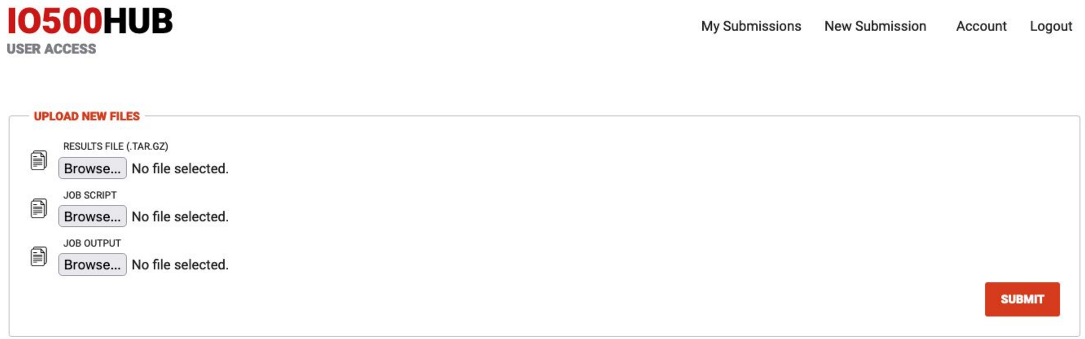
# New IO500 Submission Form

# New IO500 submission platform



**Goals**

- Manage account and submissions
- List all previous submissions
- Make new submissions when calls are open
- Easier submission and results analysis
- Allow users to update metadata of submissions until the deadline
- Integrated workflow for review and publication



RESULTS FILE (.TAR.GZ): No file selected.

JOB SCRIPT: No file selected.

JOB OUTPUT: No file selected.

# Benchmark Phases and Extended Access Patterns

# IO500 Survey Results

- Most users want that the benchmark to evolve to cover more patterns
  - Should test concurrent metadata ops (53%)
  - Should split find into easy/hard (38%)
  - Should add random read 4KB (38%)
  - Should add random write 4KB (35%)
  - Should add random read 1MB (36%)
  - Should add random write 1MB (35%)
  - Benchmark should stay as it is (22%)

Added to `--mode=extended` run

# Benchmark Phases and Extended Access Patterns

- Experimental `extended` mode with extra phases
  - New phases subject to change until final agreement
  - Two submissions for ISC22 with extended data, need more feedback
- Pending issues
  - Comparison of score between standard / extended modes
  - New phases may change the result of existing phases in rare cases
- Request that future submission use extended mode
  - Take only the values of **current** IO500 phases to calculate score
  - Allow to compare new results with historical submissions
- Committee working on specification of all I/O patterns
  - Motivation, use cases, description of actual IO pattern, …
- Code base is there, please give us feedback anytime

# Questions About Extended Access Patterns

- Open questions
  - Should both 4KB and 1MB patterns be added, or only one (which)?
    - Current IOR implementation needs write phase at same IO size as read
    - Pseudo-random IO pattern ensures “dense” files, allows data verify
  - Should random **write** phases be counted in the score, or only reads?
    - Relatively few HPC workloads have purely random writes
  - Should find-hard be “harder” than just “find in mdtest-hard directory”?
    - Extra attributes, something other than filename (string) comparison?
  - Should a directory rename test be added?
    - Is this a hierarchical namespace, or flat names with ‘/’ in them?
- Overall runtime would increase by about 30 minutes if all phases added

# Reproducibility & List Split Progress

# Reproducibility and Production/Research Lists

## Reproducibility Stats

- 3 of 14 submissions completed questionnaire
- No new “Production” systems added to list

## Production list unchanged from ISC22

1. SuperMUC-NG-EC - DAOS
2. Oracle Cloud - WEKA
3. Lenovo-Lenox3-EC - DAOS
4. CTPAI - DAOS (newer test run)

# Reproducibility and Production/Research Lists

## Key Requirements

### Production

- Complete submission metadata
- Complete reproducibility questionnaire
- Highest reproducibility score
  - Storage software availability
- No single point of failure
- Production system running production applications

### Research

- Complete submission metadata
- Complete reproducibility questionnaire (starting in ISC23)

# IO500 Reproducibility and List Split Progress

## Progress

- Mandatory fields being added
- New submissions platform being released for ISC23

## Next Steps for ISC23

- Continue adding mandatory fields and the questionnaire to new submissions page
- Split lists for ISC22, SC22, ISC23 and beyond
  - These lists all have reproducibility questionnaires
  - Prior to ISC22, all entries will by default remain on “Research” list
    - Submitters wanting to move their submissions prior to ISC22 to the “Production” list can file a request with the IO500 committee and fill out a Reproducibility questionnaire
- Determine best method to publish submitted information and reproducibility questionnaire
- Build review committee
  - Please reach out if interested in helping to review submissions

# Voice of the Community & Open Discussion

# backup

# Open Floor

- How to collect storage system metadata more easily?
- Can we encourage vendors to support the tool development and schema development?
- Vote with raised hands
    - random I/O 4KB vs. 1MB, what do people want?

# Collecting Storage System Metadata

- Improved submission schema with templates to simplify collection
  - Supporting storage-system specific schemas
  - Remove uncertainty about the semantics of fields
  - More useful metadata about test system (nodes, storage, network)
- Started integrating tools to automatically collect system configuration
  - Support the capturing of accurate system data with each submission
  - Simplify collection of system details for end users
  - Client scripts to capture kernel, filesystem, node, network, and other info
  - Per-filesystem-type script, can be customized to best collect information
  - Seek contributions from users/vendors for scripts for their filesystems
- Explanations with video: https://www.youtube.com/watch?v=R_Fq_ks4hnM