SEAGATE

IO 500?

John Bent,
Seagate Government Solutions

Why not an IO 500?

Not fair.
Never apples-to-apples.

Don’t repeat the
horror of Linpack.

Impossible to design a good benchmark.

A single number is the only way it will work.

A single number means it’s a horrible benchmark.

Why not #1: Not fair. Never apples to apples.

Why not #2: Don’t repeat Linpack

Don’t repeat Linpack

- Linpack skewed supercomputers away from theoretical ideal architecture
- A second skew can only pull them back towards theoretical!

Why not #3: Impossible to design a good benchmark

This is a **Challenge**

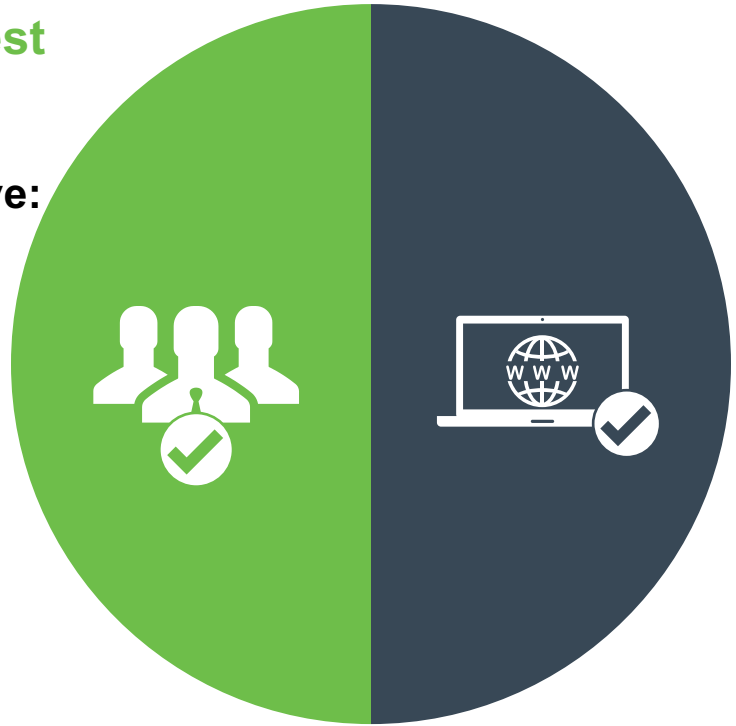
I find **Interesting**

Thanks for helping!

Why an IO 500?

Make vendors be honest

- **I recently asked three people how much bandwidth per disk drive:**
  - *Marketer: “230 MB/s”*
  - *Sales: “100 MB/s”*
  - *User: “15 MB/s”*
- **I recently asked three people how much bandwidth per supercomputer:**
  - *Marketer: “230 MB/s times the number of disks”*
  - *Sales: “1 TB/s”*
  - *User: “10 GB/s”*



Make sites be honest

Well-aligned N-N is irrelevant.

Strawperson Benchmark Proposal

Five minutes to do as much as possible

- All data must go to persistent storage
- Reads must be different client than write

Auditor must validate.

Snope and peek as necessary

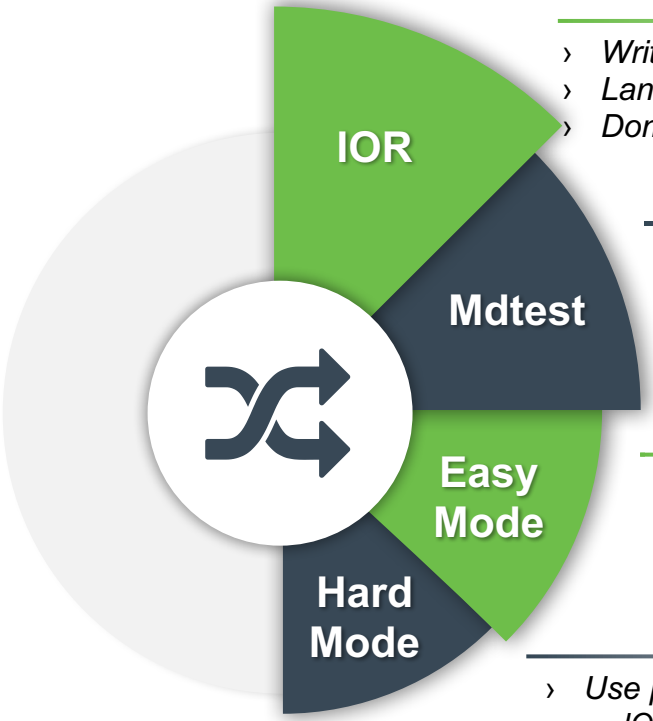
Use well-defined benchmarks

- IOR
- mdtest

Four results

- IOR hard, IOR easy
- mdtest hard, mdtest easy
- Combine into a single value?

Two benchmarks, two modes, four results



› *Write and then cross-client read as much data as possible within five minutes.*
› *Lance Evans: "Use preallocated files to ensure this is a data test not a metadata one"*
› *Dominic Manno: "Each writer must write the same amount of data. No time-based!"*

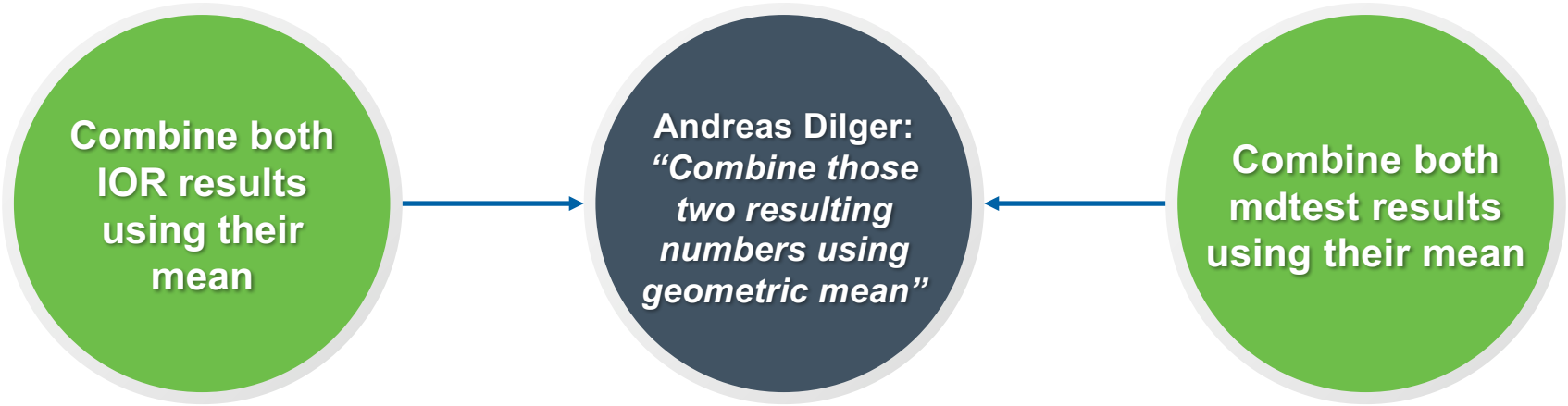
Create and then cross-client stat as many small files as possible within five minutes.

› *Use whatever parameters you want*
› *A custom IO module is allowed*

› *Use pre-specified parameters*
  ○ *IOR: N-1, small, unaligned IO*
  ○ *mdtest: Only one directory*

› *Use standard IO module*
› *All files and data must be accessible via standard Linux tools (e.g. ls, grep)*

E Pluribus Unum? (E Patru Unum?)



Another benchmark idea from Lance:

Y-axis is bandwidth, x-axis is block size, do IOR to preallocated files, compare the resulting curves

Next Steps