ISC'18

Second IO500 List

Why an IO-500

- Honesty
  - Sick of sites only advertising hero numbers
- Expectations
  - Users only see hero numbers and don't know what realistic IO performance is
- Community
  - Collect a repository of results along with info on how tuning was done
- More balanced systems
  - Make procurement of systems pay more attention to storage
- Better storage
  - Force parallel file system developers to focus on the anti-hero workloads
- Easier RFP writing
  - We have already seen an RFP specifying the mdtest_hard parameters.

What is IO-500

- Bandwidth
  - IOR easy run: user's choice
  - IOR hard run: single-shared file, small unaligned, POSIX
- Metadata
  - mdtest easy run: user's choice
  - mdtest hard run: single-shared directory, 3901 byte files, POSIX
  - "find" functionality
    - Of all the files created in the last ~20 minutes, all those that were created
      - in the last ~10 minutes
      - with size 3901
      - matching string "01"

How the IO500

IO⁵⁰⁰

> git clone https://github.com/VI4IO/io-500-dev

> cd io-500-dev

> ./utilities/prepare.sh

> ./io500.sh

> # tune (write/create phases must last for 5 minutes)

> # email submit@io500.org

First IO 500 List

- Nine submissions

- Five file systems
  - BeeGFS, DataWarp, IME, Lustre, Spectrum Scale

- Eight institutions
  - DKRZ, Fraunhofer, IBM, JCAHPC, JSC, Kaust, PNNL, SNL

- 101.48 high score

~~First~~ *Second* IO 500 List

- ~~Nine~~ *Fourteen* submissions
- ~~Five~~ **Six** file systems
  - BeeGFS, DataWarp, IME, Lustre, **PanFS**, Spectrum Scale
- ~~Eight~~ **Ten** institutions
  - DKRZ, Fraunhofer, IBM, JCAHPC, **JINR**, JSC, Kaust, PNNL, SNL, **STFC**
- ~~101.48~~ **137.78** high score

A Warm Welcome to Govorun!

IO⁵⁰⁰

#	information				io500			ior			
	system	institution	filesystem	client nodes	score	bw	md	easy write	easy read	hard write	hard read
						GiB/s	kIOP/s	GiB/s	GiB/s	GiB/s	GiB/s
1	Oakforest-PACS	JCAHPC	IME	2048	101.48	471.25	21.85	742.38	427.41	600.28	258.93
2	Shaheen	Kaust	DataWarp	300	70.90	151.53	33.17	969.45	894.76	15.55	39.09
3	Shaheen	Kaust	Lustre	1000	41.00	54.17	31.03	333.03	220.62	1.44	81.38
4	JURON	JSC	BeeGFS	8	35.77	14.24	89.83	30.42	48.36	1.46	19.16
5	Mistral	DKRZ	Lustre	100	32.15	22.77	45.39	158.19	163.62	1.53	6.79
6	Sonasad	IBM	Spectrum Scale	10	21.63	4.57	102.38	34.13	32.25	0.17	2.33
7	Seislab	Fraunhofer	BeeGFS	24	18.75	5.13	68.58	18.79	22.34	0.89	1.86
8	EMSL Cascade	PNNL	Lustre	126	11.17	4.88	25.57	17.81	30.19	0.39	2.72
9	Serrano	SNL	Spectrum Scale	16	4.25	0.65	27.98	1.08	1.03	0.22	0.71
	Govorun	Joint Institute for Nuclear Research	Lustre	RSC	24	zip	12.08		3.34		43.65

A Warm Welcome to PanFS!

#	information				io500			ior			
	system	institution	filesystem	client nodes	score	bw	md	easy write	easy read	hard write	hard read
						GiB/s	kIOP/s	GiB/s	GiB/s	GiB/s	GiB/s
1	Oakforest-PACS	JCAHPC	IME	2048	101.48	471.25	21.85	742.38	427.41	600.28	258.93
2	Shaheen	Kaust	DataWarp	300	70.90	151.53	33.17	969.45	894.76	15.55	39.09
3	Shaheen	Kaust	Lustre	1000	41.00	54.17	31.03	333.03	220.62	1.44	81.38
4	JURON	JSC	BeeGFS	8	35.77	14.24	89.83	30.42	48.36	1.46	19.16
5	Mistral	DKRZ	Lustre	100	32.15	22.77	45.39	158.19	163.62	1.53	6.79
6	Sonasad	IBM	Spectrum Scale	10	21.63	4.57	102.38	34.13	32.25	0.17	2.33
7	Seislab	Fraunhofer	BeeGFS	24	18.75	5.13	68.58	18.79	22.34	0.89	1.86
8	EMSL Cascade	PNNL	Lustre	126	11.17	4.88	25.57	17.81	30.19	0.39	2.72
9	Serrano	SNL	Spectrum Scale	16	4.25	0.65	27.98	1.08	1.03	0.22	0.71
	Jasmin/Lotus	STFC	PanFS	Panasas	64	zip	2.33	0.26	20.93		

The DataWarp Score Evolution

...ation			io500			ior				mdtest								find
filesystem	client nodes	score	bw	md	easy write	easy read	hard write	hard read	easy create	easy stat	easy delete	hard create	hard read	hard stat	hard delete		hard	
			GiB/s	kIOP/s	GiB/s	GiB/s	GiB/s	GiB/s	kIOP/s	kIOP/s	kIOP/s	kIOP/s	kIOP/s	kIOP/s	kIOP/s		kIOP/s	
DataWarp	300	70.90	151.53	33.17	969.45	894.76	15.55	39.09	50.71	49.38	48.89	11.40	0.00	38.73	18.92		43.20	
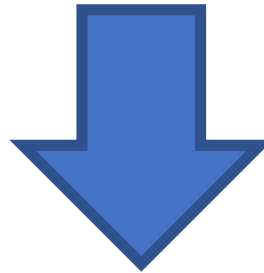
	77.37	496.81	12.05	982.36	1130.59	139.59	392.93	14.25	18.83	16.96	3.44	17.94	17.65	16.17	5.54

The IME Score Evolution

io500			ior				mdtest							find
score	bw	md	easy write	easy read	hard write	hard read	easy create	easy stat	easy delete	hard create	hard read	hard stat	hard delete	hard
	GiB/s	kIOP/s	GiB/s	GiB/s	GiB/s	GiB/s	kIOP/s	kIOP/s	kIOP/s	kIOP/s	kIOP/s	kIOP/s	kIOP/s	kIOP/s
101.48	471.25	21.85	742.38	427.41	600.28	258.93	28.29	54.20	35.88	1.51	57.38	61.50	0.95	186.69
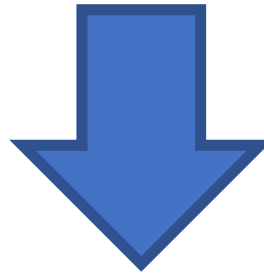
137.78	560.10	33.89	744.76	664.47	692.74	287.09	37.02	202.01	22.18	1.96	266.71	67.24	0.76	394.98

The Second List!

IO⁵⁰⁰

2018-06

This is the official list from ISC-HPC 2018. The list shows the best result for a given combination of system/institution/filesystem.

#	information						io500		
	system	institution	filesystem	storage vendor	client nodes	data	score	bw	md
								GiB/s	kIOP/s
1	Oakforest-PACS	JCAHPC	IME	DDN	2048	zip	137.78	560.10	33.89
2	ShaheenII	KAUST	DataWarp	Cray	1024	zip	77.37	496.81	12.05
3	ShaheenII	KAUST	Lustre	Cray	1000		41.00*	54.17	31.03*
4	JURON	JSC	BeeGFS	ThinkparQ	8		35.77*	14.24	89.81*
5	Mistral	DKRZ	Lustre2	Seagate	100		32.15	22.77	45.39
6	Sonasad	IBM	Spectrum Scale	IBM	10	zip	24.24	4.57	128.61
7	Seislab	Fraunhofer	BeeGFS	ThinkparQ	24		16.96	5.13	56.14
8	Mistral	DKRZ	Lustre1	Seagate	100	zip	15.47	12.68	18.88
9	Govorun	Joint Institute for Nuclear Research	Lustre	RSC	24	zip	12.08	3.34	43.65
10	EMSL Cascade	PNNL	Lustre		126		11.12	4.88	25.33
11	Serrano	SNL	Spectrum Scale	IBM	16		4.25*	0.65	27.98*
12	Jasmin/Lotus	STFC	PanFS	Panasas	64	zip	2.33	0.26	20.93

Congrats to Oakforest-PACS Again!!