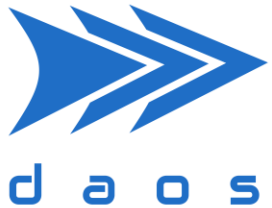
Assessment of DAOS as a backend for ECMWF's FDB

5th DUG, 2021-11-19

Nicolau Manubens, Simon Smart, Tiago Quintino, Adrian Jackson

nicolau.manubens@ecmwf.int



ECMWF epcc

ECMWF's Forecasting Systems

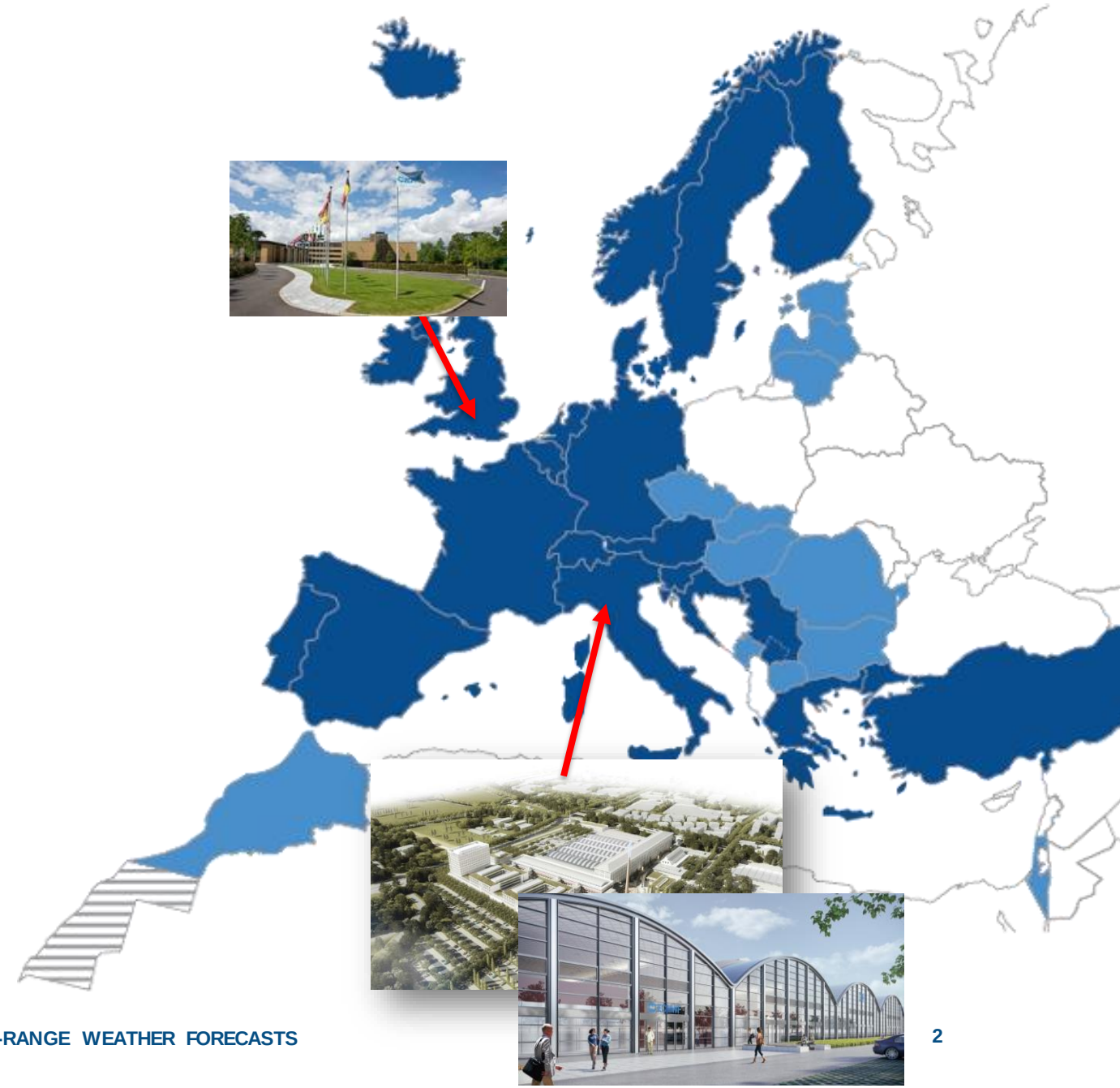
What do we do?

Operations – **Time Critical**

- HRES 0-10 day, 00Z+12Z
  - O1280 (9km) 137 levels
- ENS 0-15 day, 00Z+12Z
  - O640 (18km) 137 levels
- ENS extended 16-46 day, twice weekly
  - O320 (36km) 137 levels
- BC 06Z and 18Z
  - hourly post-processing 0-5 days

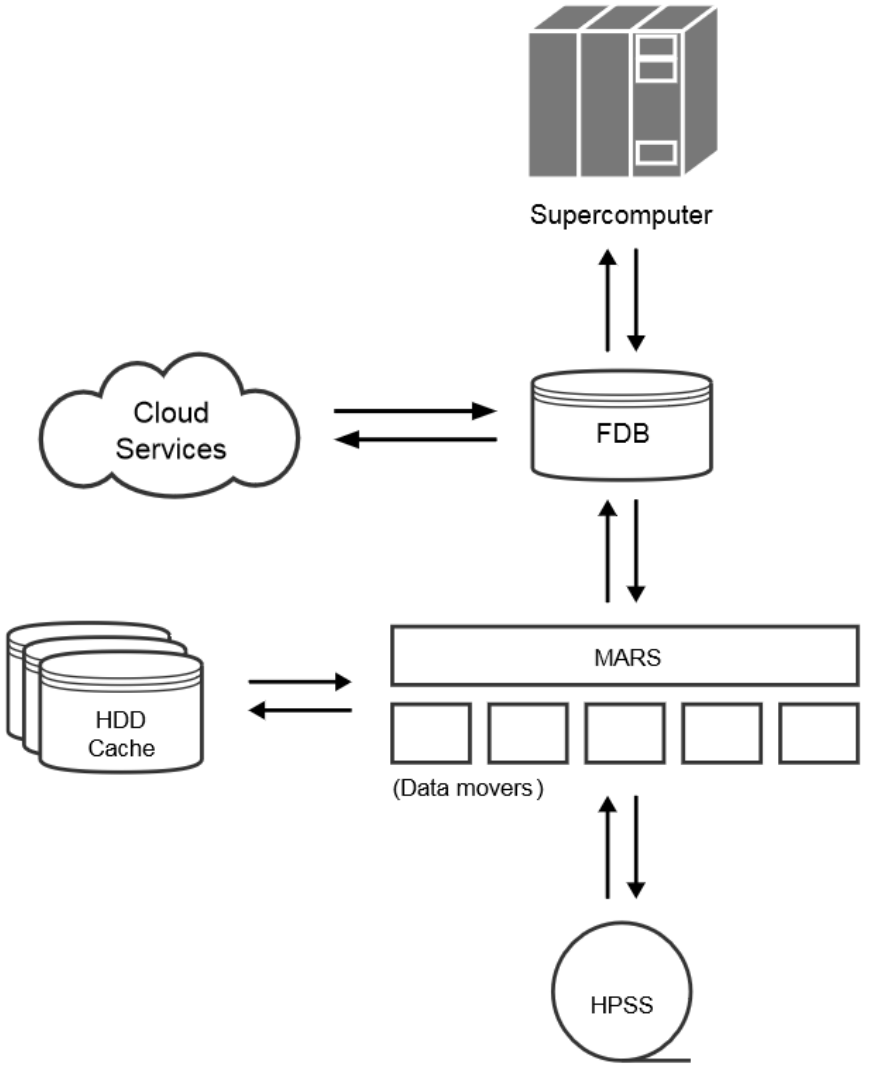
Research – **Non Time Critical**

- Experiments to improving our models
- Reforecasts, Climate reanalysis, etc



FDB and ECMWF's high-performance data infrastructure

- FDB5 is a domain-specific object store
- Software-defined, developed at ECMWF
- Used for storing, indexing and retrieving GRIB data
- Currently runs on Lustre at ECMWF
- Acts as a "hot" storage layer (RAM + HDD)
- Supports custom non-POSIX indexing and storage backends

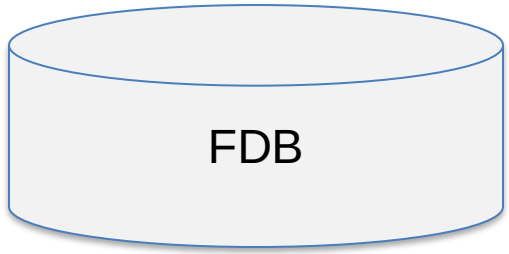


FDB – object store semantics

```
'class=od,
 date=20201224,
 stream=oper,
 levtype=sfc,
 param=10u,
 ...':
```

GRIB binary data

fdb-write



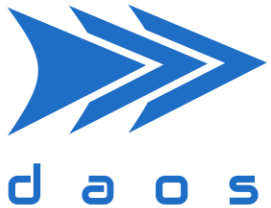
fdb-read

```
'class=od,
 date=20201224,
 stream=oper,
 levtype=sfc,
 param=10u,
 ...'
```

GRIB binary data

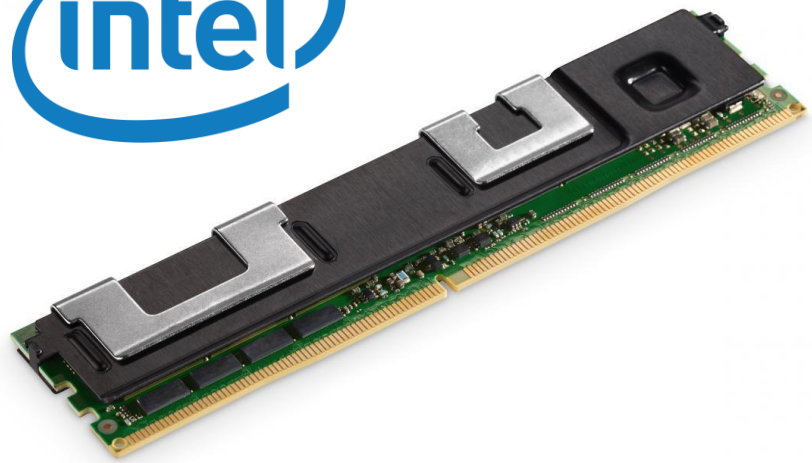
DAOS

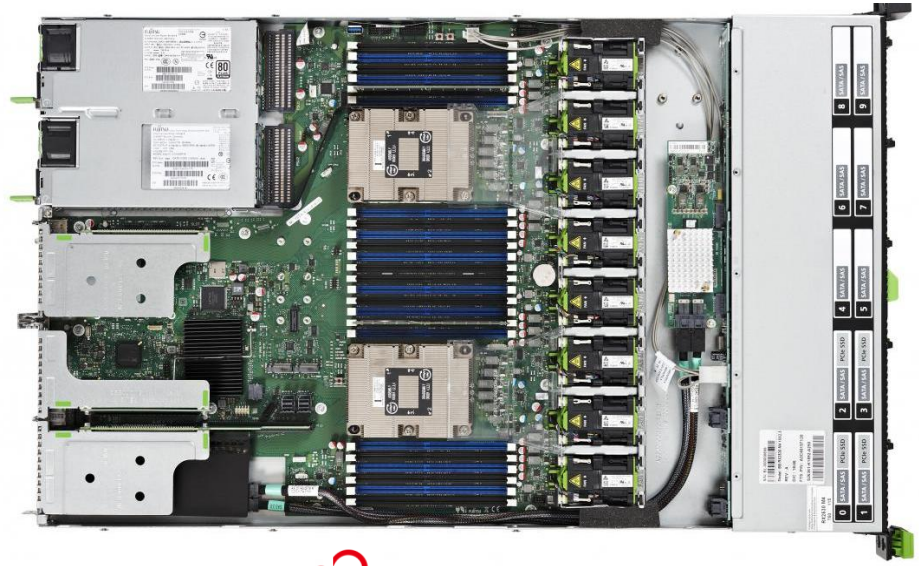
- We are assessing DAOS as a backend for FDB
- Goal: evaluate if DAOS is a good replacement for Lustre in our NWP operations
- Depends on:
  - Performance
  - Scalability
  - Robustness/reliability
  - Viability of porting
    - Software porting complexity
    - Cost – out of scope



NEXTGenIO

- NEXTGenIO is the platform we test DAOS on
  - Dual-CPU Intel® Xeon® SP nodes (48 cores)
  - OmniPath network
  - 192GB DRAM
  - **3TiB / node of NVRAM DIMMs – Intel Optane DCPMM**
  - **34 compute nodes**
  - Hosted @ EPCC, Edinburgh





34 x 3 TiB Byte Addressable Storage

ECMWF

nextgenio

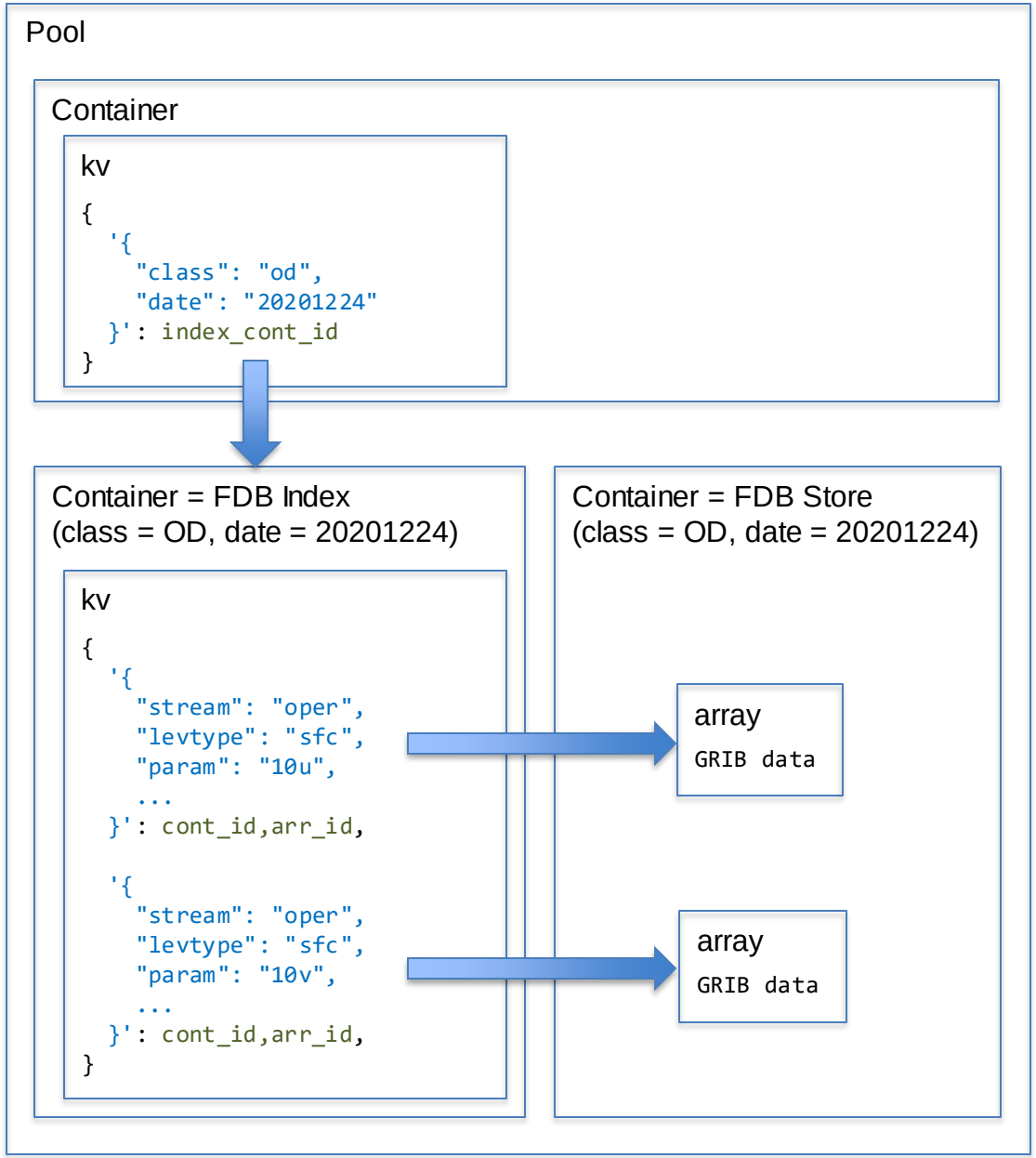
FUJITSU

DAOS deployments

- Test deployments in local VM
  - Installed from RPMs
  - Single server, dual engine
  - Emulated SCM with 'ram', NVME with 'file'
  - Using OFI sockets provider
  - Used for initial developments + consistency (reliability) and robustness tests
- NEXTGenIO deployments
  - Installed from RPMs
  - Multi-server, dual engine
  - 1.6TB SCM per engine, no NVME
  - Using OFI TCP provider over OPA for now
  - Used for performance and scalability tests

Analysis framework – field IO library

- Developed C functions for field IO from/to DAOS
  - Proof of concept for preliminary tests
  - Using DAOS high-level APIs (KV and array)
  - Having FDB's architecture in mind for easy integration later



Analysis framework – test domain

- Planned and ongoing tests, and basic metrics to collect

Test name	Access patterns "0A" and "0B"	Access pattern "1"	Access pattern "2B"
iperf / MPI			
self_test			
IOR with segments			
IOR with repetitions		(repetitions)	
field IO simplified & synchronised			
field IO simplified			
field IO			
FDB/DAOS backend + FDB hammer			

Metric name	Event measured
latency	Individual IO – array open-to-close
	Field IO library call (field IO only)
wall-clock time	Parallel IO – array open-to-close
	Parallel field IO library call (field IO only)
	Benchmark first array open to last close
	Benchmark first to last lib call (field IO only)

  - Field IO and NWP use case (FDB hammer) have no IO synchronisation across clients
  - The metrics to be collected can be used to compute various scores/throughputs

Analysis framework – access patterns

Scenarios "0A" and "0B"

- All clients perform a single write
  - to a same object / index entry (0A)
  - to a separate object / index entry (0B)

Scenario "1"

- All clients perform 100 writes to separate objects / index entries
- Barrier
- All clients perform 100 reads from corresponding object / index entries

Scenario "2B"

- Half of the clients perform 100 writes, each from a same object / index entry
- Simultaneously, other half clients perform 100 reads, each from the same corresponding object / index entry

Analysis framework – test domain (2)

- Each test can be repeated varying the following:
  - Object class: [S1, S2, SX]
    - oclass can vary separately for the 3 kinds of DAOS objects involved
  - Clients per client node: [1, …, 244]
  - Number of client nodes: [1, 2, 4, 8, 16]
  - Number of server nodes: [1, 2, 4, 8]
- Plus server adjustments: huge test domain!
- Some parameters have been fixed after testing with 2 server nodes and 4 client nodes

General setup

OFI provider

- PSM2 not working well. Selected TCP
  - we will try Cornelis Networks recommendations
  - we will wait for OPX in OFI

DAOS_MD_CAP

- Increasing it to 2048 seems to reduce frequency of some of the crashes in tests with large number of clients
  - was not effective in DAOS 1.2

Configuring dual-rail

Server

- two engines, bound to one interface each
- fabric_iface_port s have to be distant
- tried combinations of "hyperthreads", "first_core" and "pinned_numa_node"
  - pinned_numa_node has to be used for proper server pinning and best performance

Client

- pinning half the clients to each socket results in substantial performance increase: read increases by x10
- having OFI_INTERFACE set to one of the two interfaces is perjudicial

Configuring dual-rail (2)

- IOR access pattern "1", with segments, OC_S1, varying number of engines and ifaces

# sn	# eng per sn	# iface per cn	Maximum write/read bandwidth (GiB/s) observed with as many client nodes				
			1	2	4	8	16
1	1 (ib0)	1 (ib0)	**3.0w / 4.2r**	2.6w / 6.2r			
1	1 (ib0)	2	3.0w / **7.4r**	2.9w / 7.7r			
1	1 (ib1)	1 (ib1)	**1.5w / 4.0r**	1.2w / 4.0r			
1	1 (ib1)	2	1.5w / **3.8r**	1.5w / 2.9r			
1	2	2	5.5w / 7.5r	**5.5w / 9.5r**	5.5w / 10.0r		
2	2	2		9.0w / 12.0r	**9.0w / 18.0r**	8.0w / 16.0r	
4	2	2			15.3w / 32.7r	**16.2w / 32.4r**	21.0w / 35.1r
8	2	2				31.0w / 62.4r	**35.3w / 61.5r**

  - Each test was repeated 9 times with clients per cn = 24, 48, 96. Maximum among 9 repetitions shown
  - Lower performance when using ib1. Verified with MPI test. Possibly due to network configuration issue
  - Bandwidth scales with number of servers

Parameter variation – object class

- IOR, with repetitions, with 2 servers (dual rail), 4 client nodes, 48 clients per client node

Access pattern	read/write bandwidth (MiB/s) observed with OC...		
	S1	S2	SX
IOR test "0A"	390 w	862 w	3545 w
IOR test "0B"	4049 w	5074 w	4826 w
IOR test "1"	4065 w / 7516 r	4128 w / 2169 r	68 w / 97 r
IOR test "2B"	3301 w / 3498 r	2895 w / 3227 r	76 w / 80 r

  - Each test was repeated 10 times. The average of the 10 repetitions for each test is shown in each cell.
  - OC_S1 is suitable for concurrent acces to same object
  - OC_SX is suitable for concurrent access to separate objects (access pattern "1")

Parameter variation – object class (2)

- Exploring achievable bandwidth for different OCs varying number of client processes
- IOR, with repetitions, with 2 servers (dual rail), varying client nodes and clients per client node

Access pattern	read/write bandwidth (MiB/s) observed with OC...		
	S1	S2	SX
IOR test "0B"	5600 w	(pending)	(pending)
IOR test "1"	7300 w / 11600 r	7700 w / 10700 r	8200 w / 3200 r
IOR test "2B"	(pending)	(pending)	(pending)

  - Each test was repeated 10 times with 1, 2, 4 and 8 client nodes, with several (>10) different numbers of clients per client node. The average of the 10 repetitions for each (cn, cpcn) was calculated. The maximum among the several resulting averages is shown in each cell

Parameter variation – object class (3)

- Field IO, with repetitions, with 2 servers (dual rail), 4 client nodes, 48 clients per client node

Access pattern	read/write global bandwidth (MiB/s) observed with OC...				
	S1 S1 S1	S2 S2 S2	SX SX SX	SX S2 S1	SX SX S1
Field IO "0A"	172 w	160 w	171 w	140 w	172 w
Field IO "0B"	158 w	160 w	134 w	138 w	135 w
Field IO "1"	600 w / 233 r	593 w / 231 r	612 w / 691 r	587 w / 218 r	584 w / 577 r
Field IO "2B"	187 w / 155 r	188 w / 153 r	192 w / 165 r	189 w / 156 r	179 w / 156 r

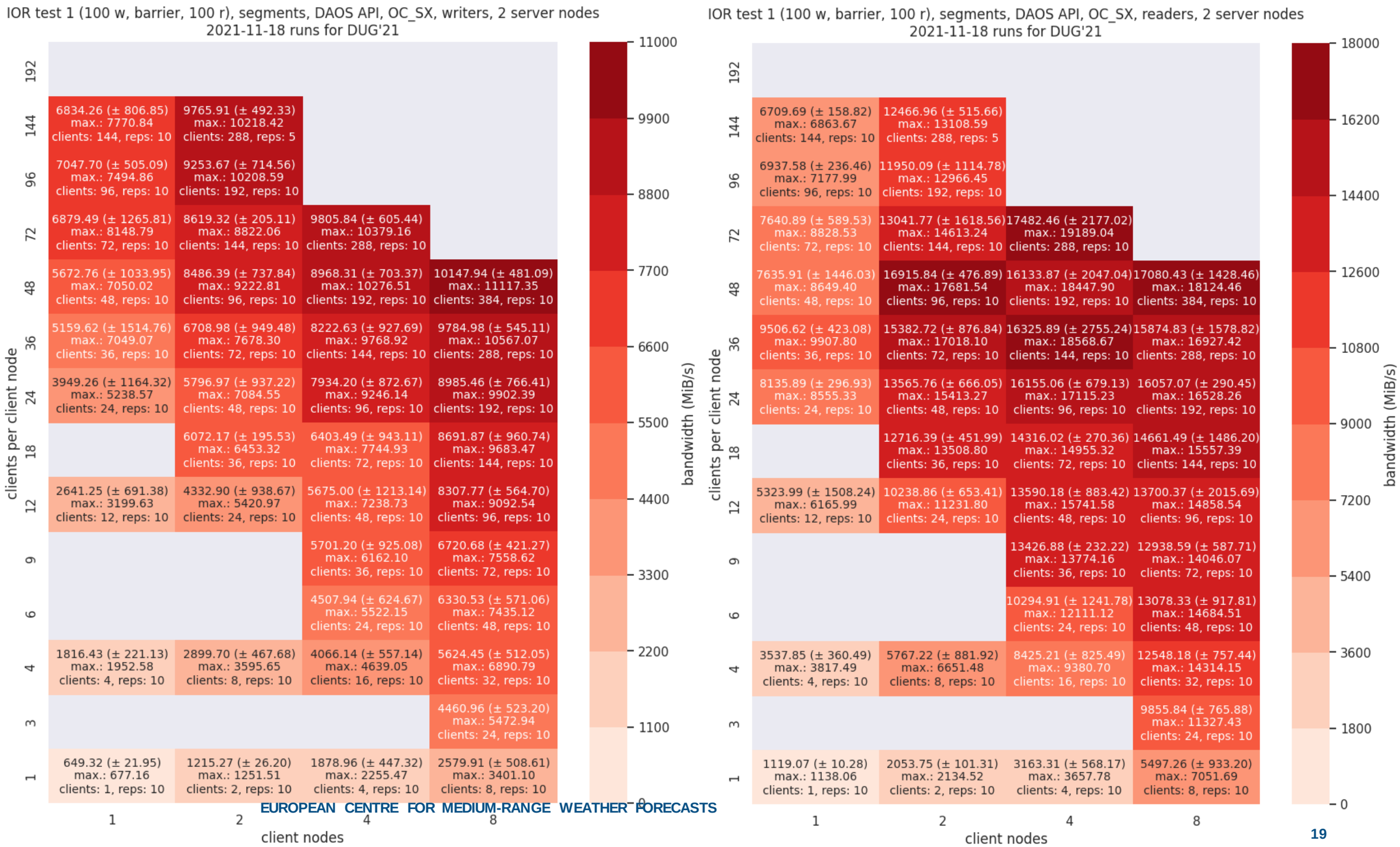
  - Each test was repeated 10 times. The average of the 10 repetitions for each test is shown in each cell.
  - OC_SX for all objects seems to result in best performance for "1" and "2B"
  - "Global bandwidths" are low if compared to IOR bandwidths
    - In access patterns "0A" and "0B", container opening from each client can have an impact
    - In patterns "1" and "2":
      - latency of IO operations (field IO library call) is ~ 0.06 s
      - in IOR, latency of IO (array open to array close) is ~0.04 s

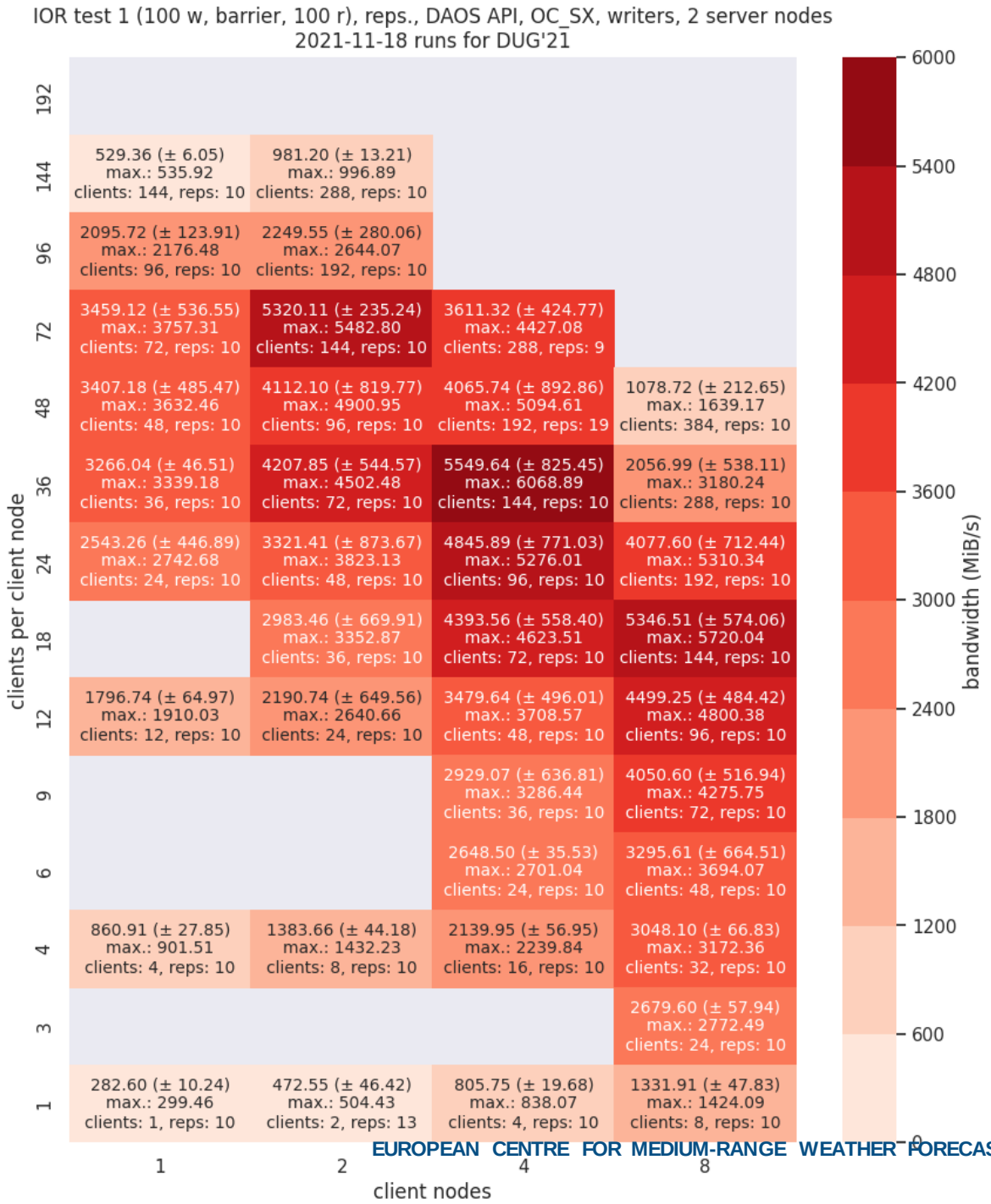
Parameter variation – IOR API

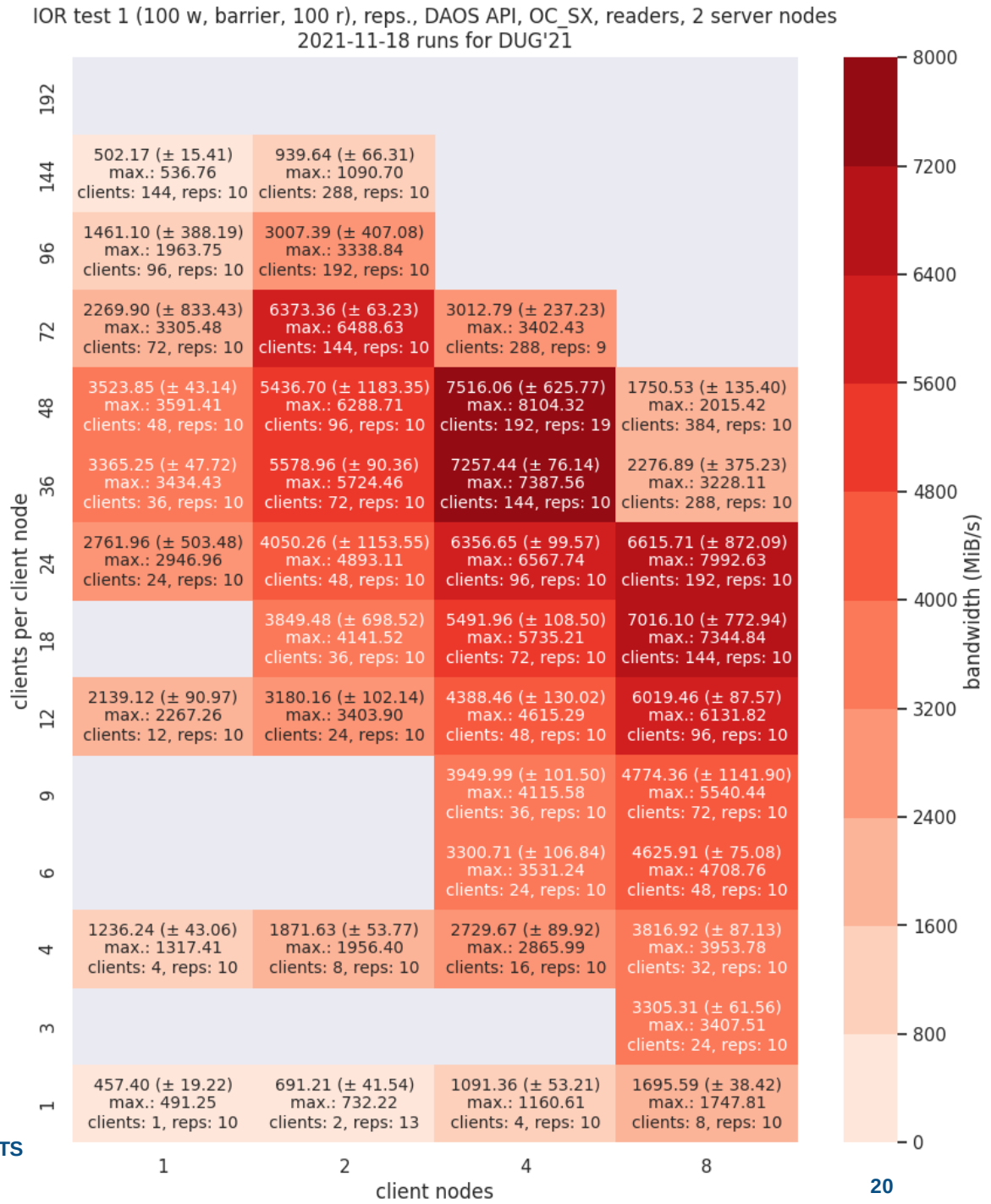
- IOR pattern "1", with repetitions, 2 server nodes (dual-rail), 4 client nodes, 48 clients per cn

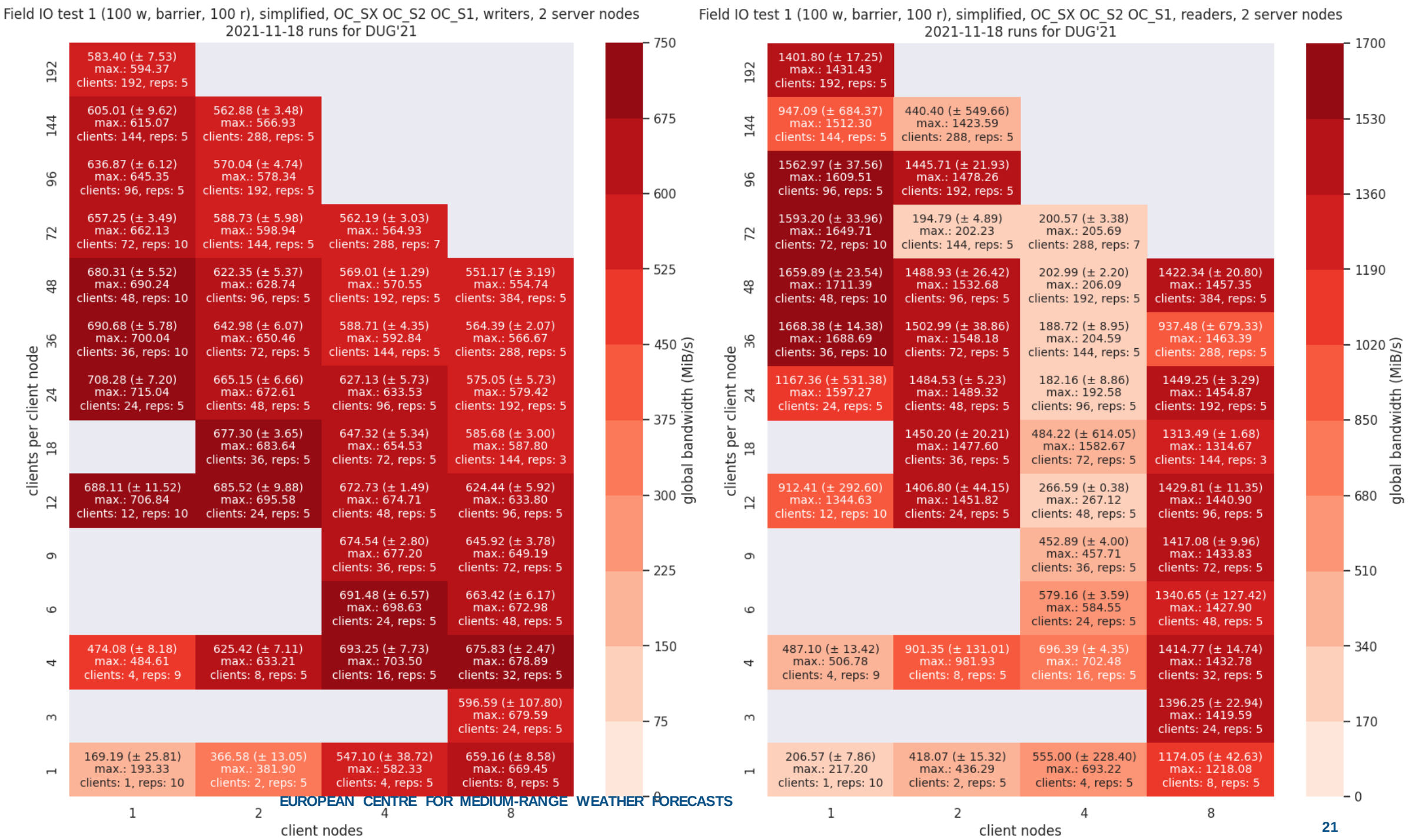
IOR API	read/write bandwidth (MiB/s) observed with OC...		
	S1	S2	SX
DAOS	4065 w / 7516 r	4128 w / 2169 r	68 w / 97 r
DFS	4928 w / 1641 r	4712 w / 563 r	85 w / 88 r
MPIIO	1472 w / 865 r	1394 w / 495 r	78 w / 93 r

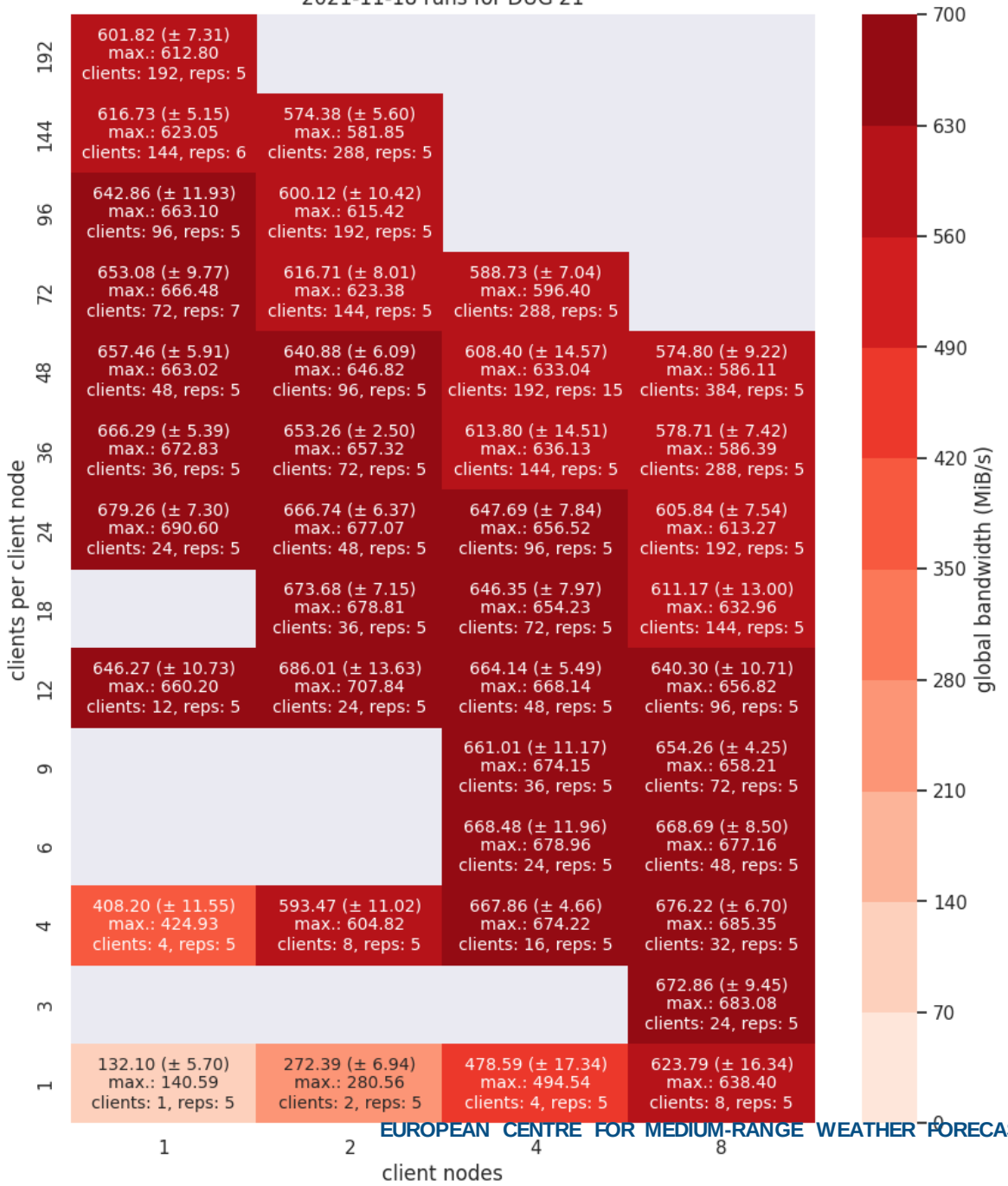
  - Each test was repeated 10 times. The average of the 10 repetitions for each test is shown in each cell.
  - We are possibly hitting issues with dual-rail with DFS and MPIIO
  - Tests with single-rail showed smaller gaps

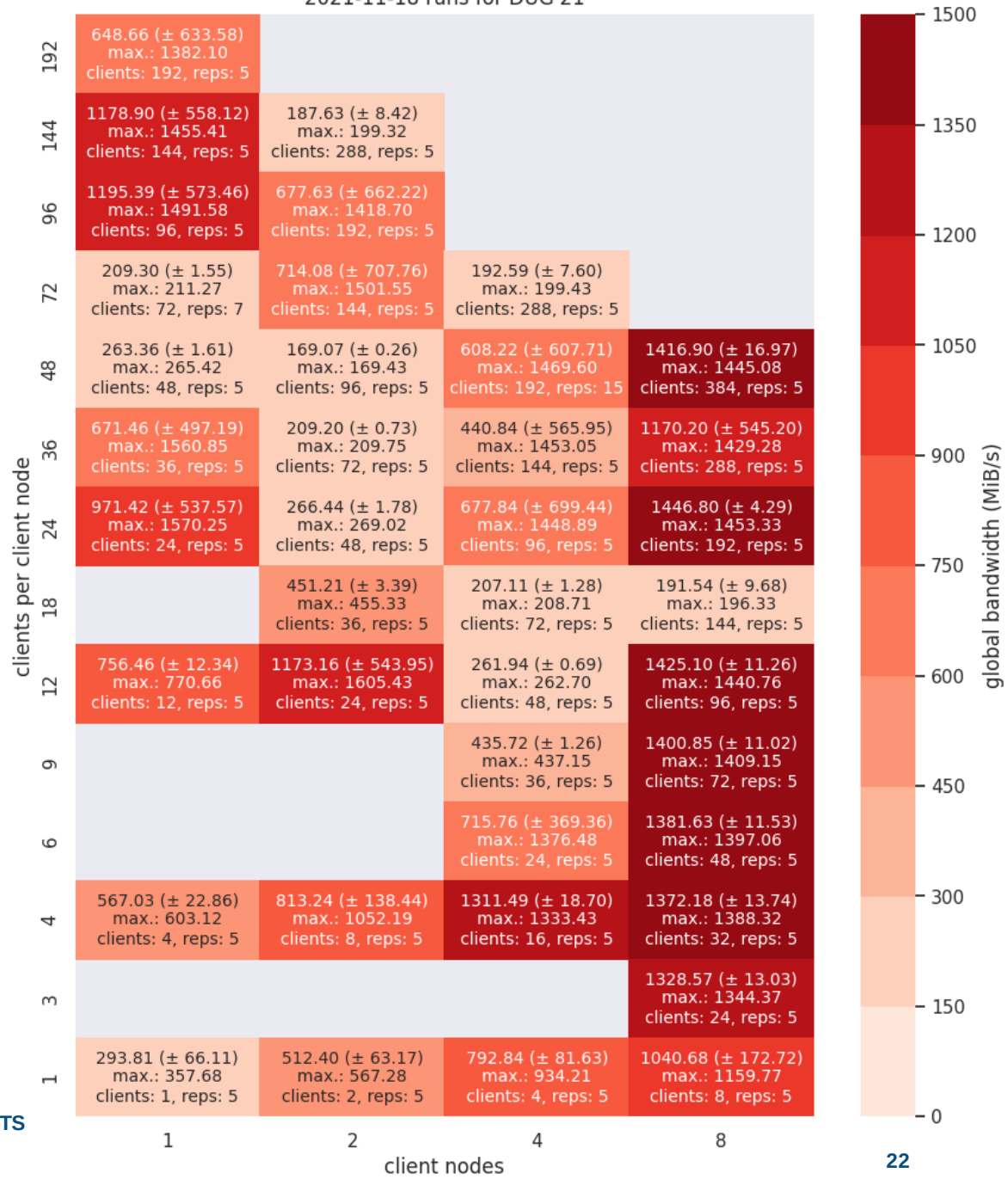

IOR test 1 (100 w, barrier, 100 r), segments, DAOS API, OC_SX, writers, 2 server nodes
2021-11-18 runs for DUG'21
clients per client node
client nodes
bandwidth (MiB/s)
1 2 4 8
1 3 4 6 9 12 18 24 36 48 72 96 144 192
6834.26 (± 806.85) max.: 7770.84 clients: 144, reps: 10
9765.91 (± 492.33) max.: 10218.42 clients: 288, reps: 5
7047.70 (± 505.09) max.: 7494.86 clients: 96, reps: 10
9253.67 (± 714.56) max.: 10208.59 clients: 192, reps: 10
6879.49 (± 1265.81) max.: 8148.79 clients: 72, reps: 10
8619.32 (± 205.11) max.: 8822.06 clients: 144, reps: 10
9805.84 (± 605.44) max.: 10379.16 clients: 288, reps: 10
5672.76 (± 1033.95) max.: 7050.02 clients: 48, reps: 10
8486.39 (± 737.84) max.: 9222.81 clients: 96, reps: 10
8968.31 (± 703.37) max.: 10276.51 clients: 192, reps: 10
10147.94 (± 481.09) max.: 11117.35 clients: 384, reps: 10
5159.62 (± 1514.76) max.: 7049.07 clients: 36, reps: 10
6708.98 (± 949.48) max.: 7678.30 clients: 72, reps: 10
8222.63 (± 927.69) max.: 9768.92 clients: 144, reps: 10
9784.98 (± 545.11) max.: 10567.07 clients: 288, reps: 10
3949.26 (± 1164.32) max.: 5238.57 clients: 24, reps: 10
5796.97 (± 937.22) max.: 7084.55 clients: 48, reps: 10
7934.20 (± 872.67) max.: 9246.14 clients: 96, reps: 10
8985.46 (± 766.41) max.: 9902.39 clients: 192, reps: 10
6072.17 (± 195.53) max.: 6453.32 clients: 36, reps: 10
6403.49 (± 943.11) max.: 7744.93 clients: 72, reps: 10
8691.87 (± 960.74) max.: 9683.47 clients: 144, reps: 10
2641.25 (± 691.38) max.: 3199.63 clients: 12, reps: 10
4332.90 (± 938.67) max.: 5420.97 clients: 24, reps: 10
5675.00 (± 1213.14) max.: 7238.73 clients: 48, reps: 10
8307.77 (± 564.70) max.: 9092.54 clients: 96, reps: 10
5701.20 (± 925.08) max.: 6162.10 clients: 36, reps: 10
6720.68 (± 421.27) max.: 7558.62 clients: 72, reps: 10
4507.94 (± 624.67) max.: 5522.15 clients: 24, reps: 10
6330.53 (± 571.06) max.: 7435.12 clients: 48, reps: 10
1816.43 (± 221.13) max.: 1952.58 clients: 4, reps: 10
2899.70 (± 467.68) max.: 3595.65 clients: 8, reps: 10
4066.14 (± 557.14) max.: 4639.05 clients: 16, reps: 10
5624.45 (± 512.05) max.: 6890.79 clients: 32, reps: 10
4460.96 (± 523.20) max.: 5472.94 clients: 24, reps: 10
649.32 (± 21.95) max.: 677.16 clients: 1, reps: 10
1215.27 (± 26.20) max.: 1251.51 clients: 2, reps: 10
1878.96 (± 447.32) max.: 2255.47 clients: 4, reps: 10
2579.91 (± 508.61) max.: 3401.10 clients: 8, reps: 10
11000 9900 8800 7700 6600 5500 4400 3300 2200 1100 0
IOR test 1 (100 w, barrier, 100 r), segments, DAOS API, OC_SX, readers, 2 server nodes
2021-11-18 runs for DUG'21
clients per client node
client nodes
bandwidth (MiB/s)
1 2 4 8
1 3 4 6 9 12 18 24 36 48 72 96 144 192
6709.69 (± 158.82) max.: 6863.67 clients: 144, reps: 10
12466.96 (± 515.66) max.: 13108.59 clients: 288, reps: 5
6937.58 (± 236.46) max.: 7177.99 clients: 96, reps: 10
11950.09 (± 1114.78) max.: 12966.45 clients: 192, reps: 10
7640.89 (± 589.53) max.: 8828.53 clients: 72, reps: 10
13041.77 (± 1618.56) max.: 14613.24 clients: 144, reps: 10
17482.46 (± 2177.02) max.: 19189.04 clients: 288, reps: 10
7635.91 (± 1446.03) max.: 8649.40 clients: 48, reps: 10
16915.84 (± 476.89) max.: 17681.54 clients: 96, reps: 10
16133.87 (± 2047.04) max.: 18447.90 clients: 192, reps: 10
17080.43 (± 1428.46) max.: 18124.46 clients: 384, reps: 10
9506.62 (± 423.08) max.: 9907.80 clients: 36, reps: 10
15382.72 (± 876.84) max.: 17018.10 clients: 72, reps: 10
16325.89 (± 2755.24) max.: 18568.67 clients: 144, reps: 10
15874.83 (± 1578.82) max.: 16927.42 clients: 288, reps: 10
8135.89 (± 296.93) max.: 8555.33 clients: 24, reps: 10
13565.76 (± 666.05) max.: 15413.27 clients: 48, reps: 10
16155.06 (± 679.13) max.: 17115.23 clients: 96, reps: 10
16057.07 (± 290.45) max.: 16528.26 clients: 192, reps: 10
12716.39 (± 451.99) max.: 13508.80 clients: 36, reps: 10
14316.02 (± 270.36) max.: 14955.32 clients: 72, reps: 10
14661.49 (± 1486.20) max.: 15557.39 clients: 144, reps: 10
5323.99 (± 1508.24) max.: 6165.99 clients: 12, reps: 10
10238.86 (± 653.41) max.: 11231.80 clients: 24, reps: 10
13590.18 (± 883.42) max.: 15741.58 clients: 48, reps: 10
13700.37 (± 2015.69) max.: 14858.54 clients: 96, reps: 10
13426.88 (± 232.22) max.: 13774.16 clients: 36, reps: 10
12938.59 (± 587.71) max.: 14046.07 clients: 72, reps: 10
10294.91 (± 1241.78) max.: 12111.12 clients: 24, reps: 10
13078.33 (± 917.81) max.: 14684.51 clients: 48, reps: 10
3537.85 (± 360.49) max.: 3817.49 clients: 4, reps: 10
5767.22 (± 881.92) max.: 6651.48 clients: 8, reps: 10
8425.21 (± 825.49) max.: 9380.70 clients: 16, reps: 10
12548.18 (± 757.44) max.: 14314.15 clients: 32, reps: 10
9855.84 (± 765.88) max.: 11327.43 clients: 24, reps: 10
1119.07 (± 10.28) max.: 1138.06 clients: 1, reps: 10
2053.75 (± 101.31) max.: 2134.52 clients: 2, reps: 10
3163.31 (± 568.17) max.: 3657.78 clients: 4, reps: 10
5497.26 (± 933.20) max.: 7051.69 clients: 8, reps: 10
18000 16200 14400 12600 10800 9000 7200 5400 3600 1800 0


IOR test 1 (100 w, barrier, 100 r), reps., DAOS API, OC_SX, writers, 2 server nodes
2021-11-18 runs for DUG'21
clients per client node
client nodes
bandwidth (MiB/s)
1 2 4 8
192 144 96 72 48 36 24 18 12 9 6 4 3 1
529.36 (± 6.05) max.: 535.92 clients: 144, reps: 10
981.20 (± 13.21) max.: 996.89 clients: 288, reps: 10
2095.72 (± 123.91) max.: 2176.48 clients: 96, reps: 10
2249.55 (± 280.06) max.: 2644.07 clients: 192, reps: 10
3459.12 (± 536.55) max.: 3757.31 clients: 72, reps: 10
5320.11 (± 235.24) max.: 5482.80 clients: 144, reps: 10
3611.32 (± 424.77) max.: 4427.08 clients: 288, reps: 9
3407.18 (± 485.47) max.: 3632.46 clients: 48, reps: 10
4112.10 (± 819.77) max.: 4900.95 clients: 96, reps: 10
4065.74 (± 892.86) max.: 5094.61 clients: 192, reps: 19
1078.72 (± 212.65) max.: 1639.17 clients: 384, reps: 10
3266.04 (± 46.51) max.: 3339.18 clients: 36, reps: 10
4207.85 (± 544.57) max.: 4502.48 clients: 72, reps: 10
5549.64 (± 825.45) max.: 6068.89 clients: 144, reps: 10
2056.99 (± 538.11) max.: 3180.24 clients: 288, reps: 10
2543.26 (± 446.89) max.: 2742.68 clients: 24, reps: 10
3321.41 (± 873.67) max.: 3823.13 clients: 48, reps: 10
4845.89 (± 771.03) max.: 5276.01 clients: 96, reps: 10
4077.60 (± 712.44) max.: 5310.34 clients: 192, reps: 10
2983.46 (± 669.91) max.: 3352.87 clients: 36, reps: 10
4393.56 (± 558.40) max.: 4623.51 clients: 72, reps: 10
5346.51 (± 574.06) max.: 5720.04 clients: 144, reps: 10
1796.74 (± 64.97) max.: 1910.03 clients: 12, reps: 10
2190.74 (± 649.56) max.: 2640.66 clients: 24, reps: 10
3479.64 (± 496.01) max.: 3708.57 clients: 48, reps: 10
4499.25 (± 484.42) max.: 4800.38 clients: 96, reps: 10
2929.07 (± 636.81) max.: 3286.44 clients: 36, reps: 10
4050.60 (± 516.94) max.: 4275.75 clients: 72, reps: 10
2648.50 (± 35.53) max.: 2701.04 clients: 24, reps: 10
3295.61 (± 664.51) max.: 3694.07 clients: 48, reps: 10
860.91 (± 27.85) max.: 901.51 clients: 4, reps: 10
1383.66 (± 44.18) max.: 1432.23 clients: 8, reps: 10
2139.95 (± 56.95) max.: 2239.84 clients: 16, reps: 10
3048.10 (± 66.83) max.: 3172.36 clients: 32, reps: 10
2679.60 (± 57.94) max.: 2772.49 clients: 24, reps: 10
282.60 (± 10.24) max.: 299.46 clients: 1, reps: 10
472.55 (± 46.42) max.: 504.43 clients: 2, reps: 13
805.75 (± 19.68) max.: 838.07 clients: 4, reps: 10
1331.91 (± 47.83) max.: 1424.09 clients: 8, reps: 10
6000 5400 4800 4200 3600 3000 2400 1800 1200 600 0


IOR test 1 (100 w, barrier, 100 r), reps., DAOS API, OC_SX, readers, 2 server nodes
2021-11-18 runs for DUG'21
clients per client node
client nodes
bandwidth (MiB/s)
1 2 4 8
192 144 96 72 48 36 24 18 12 9 6 4 3 1
502.17 (± 15.41) max.: 536.76 clients: 144, reps: 10
939.64 (± 66.31) max.: 1090.70 clients: 288, reps: 10
1461.10 (± 388.19) max.: 1963.75 clients: 96, reps: 10
3007.39 (± 407.08) max.: 3338.84 clients: 192, reps: 10
2269.90 (± 833.43) max.: 3305.48 clients: 72, reps: 10
6373.36 (± 63.23) max.: 6488.63 clients: 144, reps: 10
3012.79 (± 237.23) max.: 3402.43 clients: 288, reps: 9
3523.85 (± 43.14) max.: 3591.41 clients: 48, reps: 10
5436.70 (± 1183.35) max.: 6288.71 clients: 96, reps: 10
7516.06 (± 625.77) max.: 8104.32 clients: 192, reps: 19
1750.53 (± 135.40) max.: 2015.42 clients: 384, reps: 10
3365.25 (± 47.72) max.: 3434.43 clients: 36, reps: 10
5578.96 (± 90.36) max.: 5724.46 clients: 72, reps: 10
7257.44 (± 76.14) max.: 7387.56 clients: 144, reps: 10
2276.89 (± 375.23) max.: 3228.11 clients: 288, reps: 10
2761.96 (± 503.48) max.: 2946.96 clients: 24, reps: 10
4050.26 (± 1153.55) max.: 4893.11 clients: 48, reps: 10
6356.65 (± 99.57) max.: 6567.74 clients: 96, reps: 10
6615.71 (± 872.09) max.: 7992.63 clients: 192, reps: 10
3849.48 (± 698.52) max.: 4141.52 clients: 36, reps: 10
5491.96 (± 108.50) max.: 5735.21 clients: 72, reps: 10
7016.10 (± 772.94) max.: 7344.84 clients: 144, reps: 10
2139.12 (± 90.97) max.: 2267.26 clients: 12, reps: 10
3180.16 (± 102.14) max.: 3403.90 clients: 24, reps: 10
4388.46 (± 130.02) max.: 4615.29 clients: 48, reps: 10
6019.46 (± 87.57) max.: 6131.82 clients: 96, reps: 10
3949.99 (± 101.50) max.: 4115.58 clients: 36, reps: 10
4774.36 (± 1141.90) max.: 5540.44 clients: 72, reps: 10
3300.71 (± 106.84) max.: 3531.24 clients: 24, reps: 10
4625.91 (± 75.08) max.: 4708.76 clients: 48, reps: 10
1236.24 (± 43.06) max.: 1317.41 clients: 4, reps: 10
1871.63 (± 53.77) max.: 1956.40 clients: 8, reps: 10
2729.67 (± 89.92) max.: 2865.99 clients: 16, reps: 10
3816.92 (± 87.13) max.: 3953.78 clients: 32, reps: 10
3305.31 (± 61.56) max.: 3407.51 clients: 24, reps: 10
457.40 (± 19.22) max.: 491.25 clients: 1, reps: 10
691.21 (± 41.54) max.: 732.22 clients: 2, reps: 13
1091.36 (± 53.21) max.: 1160.61 clients: 4, reps: 10
1695.59 (± 38.42) max.: 1747.81 clients: 8, reps: 10
8000 7200 6400 5600 4800 4000 3200 2400 1600 800 0


Field IO test 1 (100 w, barrier, 100 r), simplified, OC_SX OC_S2 OC_S1, writers, 2 server nodes
2021-11-18 runs for DUG'21
583.40 (± 7.53) max.: 594.37 clients: 192, reps: 5
605.01 (± 9.62) max.: 615.07 clients: 144, reps: 5
562.88 (± 3.48) max.: 566.93 clients: 288, reps: 5
636.87 (± 6.12) max.: 645.35 clients: 96, reps: 5
570.04 (± 4.74) max.: 578.34 clients: 192, reps: 5
657.25 (± 3.49) max.: 662.13 clients: 72, reps: 10
588.73 (± 5.98) max.: 598.94 clients: 144, reps: 5
562.19 (± 3.03) max.: 564.93 clients: 288, reps: 7
680.31 (± 5.52) max.: 690.24 clients: 48, reps: 10
622.35 (± 5.37) max.: 628.74 clients: 96, reps: 5
569.01 (± 1.29) max.: 570.55 clients: 192, reps: 5
551.17 (± 3.19) max.: 554.74 clients: 384, reps: 5
690.68 (± 5.78) max.: 700.04 clients: 36, reps: 10
642.98 (± 6.07) max.: 650.46 clients: 72, reps: 5
588.71 (± 4.35) max.: 592.84 clients: 144, reps: 5
564.39 (± 2.07) max.: 566.67 clients: 288, reps: 5
708.28 (± 7.20) max.: 715.04 clients: 24, reps: 5
665.15 (± 6.66) max.: 672.61 clients: 48, reps: 5
627.13 (± 5.73) max.: 633.53 clients: 96, reps: 5
575.05 (± 5.73) max.: 579.42 clients: 192, reps: 5
677.30 (± 3.65) max.: 683.64 clients: 36, reps: 5
647.32 (± 5.34) max.: 654.53 clients: 72, reps: 5
585.68 (± 3.00) max.: 587.80 clients: 144, reps: 3
688.11 (± 11.52) max.: 706.84 clients: 12, reps: 10
685.52 (± 9.88) max.: 695.58 clients: 24, reps: 5
672.73 (± 1.49) max.: 674.71 clients: 48, reps: 5
624.44 (± 5.92) max.: 633.80 clients: 96, reps: 5
674.54 (± 2.80) max.: 677.20 clients: 36, reps: 5
645.92 (± 3.78) max.: 649.19 clients: 72, reps: 5
691.48 (± 6.57) max.: 698.63 clients: 24, reps: 5
663.42 (± 6.17) max.: 672.98 clients: 48, reps: 5
474.08 (± 8.18) max.: 484.61 clients: 4, reps: 9
625.42 (± 7.11) max.: 633.21 clients: 8, reps: 5
693.25 (± 7.73) max.: 703.50 clients: 16, reps: 5
675.83 (± 2.47) max.: 678.89 clients: 32, reps: 5
596.59 (± 107.80) max.: 679.59 clients: 24, reps: 5
169.19 (± 25.81) max.: 193.33 clients: 1, reps: 10
366.58 (± 13.05) max.: 381.90 clients: 2, reps: 5
547.10 (± 38.72) max.: 582.33 clients: 4, reps: 5
659.16 (± 8.58) max.: 669.45 clients: 8, reps: 5
clients per client node
192
144
96
72
48
36
24
18
12
9
6
4
3
1
client nodes
1
2
4
8
global bandwidth (MiB/s)
750
675
600
525
450
375
300
225
150
75
0
Field IO test 1 (100 w, barrier, 100 r), simplified, OC_SX OC_S2 OC_S1, readers, 2 server nodes
2021-11-18 runs for DUG'21
1401.80 (± 17.25) max.: 1431.43 clients: 192, reps: 5
947.09 (± 684.37) max.: 1512.30 clients: 144, reps: 5
440.40 (± 549.66) max.: 1423.59 clients: 288, reps: 5
1562.97 (± 37.56) max.: 1609.51 clients: 96, reps: 5
1445.71 (± 21.93) max.: 1478.26 clients: 192, reps: 5
1593.20 (± 33.96) max.: 1649.71 clients: 72, reps: 10
194.79 (± 4.89) max.: 202.23 clients: 144, reps: 5
200.57 (± 3.38) max.: 205.69 clients: 288, reps: 7
1659.89 (± 23.54) max.: 1711.39 clients: 48, reps: 10
1488.93 (± 26.42) max.: 1532.68 clients: 96, reps: 5
202.99 (± 2.20) max.: 206.09 clients: 192, reps: 5
1422.34 (± 20.80) max.: 1457.35 clients: 384, reps: 5
1668.38 (± 14.38) max.: 1688.69 clients: 36, reps: 10
1502.99 (± 38.86) max.: 1548.18 clients: 72, reps: 5
188.72 (± 8.95) max.: 204.59 clients: 144, reps: 5
937.48 (± 679.33) max.: 1463.39 clients: 288, reps: 5
1167.36 (± 531.38) max.: 1597.27 clients: 24, reps: 5
1484.53 (± 5.23) max.: 1489.32 clients: 48, reps: 5
182.16 (± 8.86) max.: 192.58 clients: 96, reps: 5
1449.25 (± 3.29) max.: 1454.87 clients: 192, reps: 5
1450.20 (± 20.21) max.: 1477.60 clients: 36, reps: 5
484.22 (± 614.05) max.: 1582.67 clients: 72, reps: 5
1313.49 (± 1.68) max.: 1314.67 clients: 144, reps: 3
912.41 (± 292.60) max.: 1344.63 clients: 12, reps: 10
1406.80 (± 44.15) max.: 1451.82 clients: 24, reps: 5
266.59 (± 0.38) max.: 267.12 clients: 48, reps: 5
1429.81 (± 11.35) max.: 1440.90 clients: 96, reps: 5
452.89 (± 4.00) max.: 457.71 clients: 36, reps: 5
1417.08 (± 9.96) max.: 1433.83 clients: 72, reps: 5
579.16 (± 3.59) max.: 584.55 clients: 24, reps: 5
1340.65 (± 127.42) max.: 1427.90 clients: 48, reps: 5
487.10 (± 13.42) max.: 506.78 clients: 4, reps: 9
901.35 (± 131.01) max.: 981.93 clients: 8, reps: 5
696.39 (± 4.35) max.: 702.48 clients: 16, reps: 5
1414.77 (± 14.74) max.: 1432.78 clients: 32, reps: 5
1396.25 (± 22.94) max.: 1419.59 clients: 24, reps: 5
206.57 (± 7.86) max.: 217.20 clients: 1, reps: 10
418.07 (± 15.32) max.: 436.29 clients: 2, reps: 5
555.00 (± 228.40) max.: 693.22 clients: 4, reps: 5
1174.05 (± 42.63) max.: 1218.08 clients: 8, reps: 5
clients per client node
192
144
96
72
48
36
24
18
12
9
6
4
3
1
client nodes
1
2
4
8
global bandwidth (MiB/s)
1700
1530
1360
1190
1020
850
680
510
340
170
0


Field IO test 1 (100 w, barrier, 100 r), OC_SX OC_SX OC_SX, writers, 2 server nodes
2021-11-18 runs for DUG'21
clients per client node
client nodes
global bandwidth (MiB/s)
1 2 4 8
192 144 96 72 48 36 24 18 12 9 6 4 3 1
601.82 (± 7.31) max.: 612.80 clients: 192, reps: 5
616.73 (± 5.15) max.: 623.05 clients: 144, reps: 6
574.38 (± 5.60) max.: 581.85 clients: 288, reps: 5
642.86 (± 11.93) max.: 663.10 clients: 96, reps: 5
600.12 (± 10.42) max.: 615.42 clients: 192, reps: 5
653.08 (± 9.77) max.: 666.48 clients: 72, reps: 7
616.71 (± 8.01) max.: 623.38 clients: 144, reps: 5
588.73 (± 7.04) max.: 596.40 clients: 288, reps: 5
657.46 (± 5.91) max.: 663.02 clients: 48, reps: 5
640.88 (± 6.09) max.: 646.82 clients: 96, reps: 5
608.40 (± 14.57) max.: 633.04 clients: 192, reps: 15
574.80 (± 9.22) max.: 586.11 clients: 384, reps: 5
666.29 (± 5.39) max.: 672.83 clients: 36, reps: 5
653.26 (± 2.50) max.: 657.32 clients: 72, reps: 5
613.80 (± 14.51) max.: 636.13 clients: 144, reps: 5
578.71 (± 7.42) max.: 586.39 clients: 288, reps: 5
679.26 (± 7.30) max.: 690.60 clients: 24, reps: 5
666.74 (± 6.37) max.: 677.07 clients: 48, reps: 5
647.69 (± 7.84) max.: 656.52 clients: 96, reps: 5
605.84 (± 7.54) max.: 613.27 clients: 192, reps: 5
673.68 (± 7.15) max.: 678.81 clients: 36, reps: 5
646.35 (± 7.97) max.: 654.23 clients: 72, reps: 5
611.17 (± 13.00) max.: 632.96 clients: 144, reps: 5
646.27 (± 10.73) max.: 660.20 clients: 12, reps: 5
686.01 (± 13.63) max.: 707.84 clients: 24, reps: 5
664.14 (± 5.49) max.: 668.14 clients: 48, reps: 5
640.30 (± 10.71) max.: 656.82 clients: 96, reps: 5
661.01 (± 11.17) max.: 674.15 clients: 36, reps: 5
654.26 (± 4.25) max.: 658.21 clients: 72, reps: 5
668.48 (± 11.96) max.: 678.96 clients: 24, reps: 5
668.69 (± 8.50) max.: 677.16 clients: 48, reps: 5
408.20 (± 11.55) max.: 424.93 clients: 4, reps: 5
593.47 (± 11.02) max.: 604.82 clients: 8, reps: 5
667.86 (± 4.66) max.: 674.22 clients: 16, reps: 5
676.22 (± 6.70) max.: 685.35 clients: 32, reps: 5
672.86 (± 9.45) max.: 683.08 clients: 24, reps: 5
132.10 (± 5.70) max.: 140.59 clients: 1, reps: 5
272.39 (± 6.94) max.: 280.56 clients: 2, reps: 5
478.59 (± 17.34) max.: 494.54 clients: 4, reps: 5
623.79 (± 16.34) max.: 638.40 clients: 8, reps: 5
700 630 560 490 420 350 280 210 140 70 0


Field IO test 1 (100 w, barrier, 100 r), OC_SX OC_SX OC_SX, readers, 2 server nodes
2021-11-18 runs for DUG'21
clients per client node
client nodes
global bandwidth (MiB/s)
1 2 4 8
192 144 96 72 48 36 24 18 12 9 6 4 3 1
648.66 (± 633.58) max.: 1382.10 clients: 192, reps: 5
1178.90 (± 558.12) max.: 1455.41 clients: 144, reps: 5
187.63 (± 8.42) max.: 199.32 clients: 288, reps: 5
1195.39 (± 573.46) max.: 1491.58 clients: 96, reps: 5
677.63 (± 662.22) max.: 1418.70 clients: 192, reps: 5
209.30 (± 1.55) max.: 211.27 clients: 72, reps: 7
714.08 (± 707.76) max.: 1501.55 clients: 144, reps: 5
192.59 (± 7.60) max.: 199.43 clients: 288, reps: 5
263.36 (± 1.61) max.: 265.42 clients: 48, reps: 5
169.07 (± 0.26) max.: 169.43 clients: 96, reps: 5
608.22 (± 607.71) max.: 1469.60 clients: 192, reps: 15
1416.90 (± 16.97) max.: 1445.08 clients: 384, reps: 5
671.46 (± 497.19) max.: 1560.85 clients: 36, reps: 5
209.20 (± 0.73) max.: 209.75 clients: 72, reps: 5
440.84 (± 565.95) max.: 1453.05 clients: 144, reps: 5
1170.20 (± 545.20) max.: 1429.28 clients: 288, reps: 5
971.42 (± 537.57) max.: 1570.25 clients: 24, reps: 5
266.44 (± 1.78) max.: 269.02 clients: 48, reps: 5
677.84 (± 699.44) max.: 1448.89 clients: 96, reps: 5
1446.80 (± 4.29) max.: 1453.33 clients: 192, reps: 5
451.21 (± 3.39) max.: 455.33 clients: 36, reps: 5
207.11 (± 1.28) max.: 208.71 clients: 72, reps: 5
191.54 (± 9.68) max.: 196.33 clients: 144, reps: 5
756.46 (± 12.34) max.: 770.66 clients: 12, reps: 5
1173.16 (± 543.95) max.: 1605.43 clients: 24, reps: 5
261.94 (± 0.69) max.: 262.70 clients: 48, reps: 5
1425.10 (± 11.26) max.: 1440.76 clients: 96, reps: 5
435.72 (± 1.26) max.: 437.15 clients: 36, reps: 5
1400.85 (± 11.02) max.: 1409.15 clients: 72, reps: 5
715.76 (± 369.36) max.: 1376.48 clients: 24, reps: 5
1381.63 (± 11.53) max.: 1397.06 clients: 48, reps: 5
567.03 (± 22.86) max.: 603.12 clients: 4, reps: 5
813.24 (± 138.44) max.: 1052.19 clients: 8, reps: 5
1311.49 (± 18.70) max.: 1333.43 clients: 16, reps: 5
1372.18 (± 13.74) max.: 1388.32 clients: 32, reps: 5
1328.57 (± 13.03) max.: 1344.37 clients: 24, reps: 5
293.81 (± 66.11) max.: 357.68 clients: 1, reps: 5
512.40 (± 63.17) max.: 567.28 clients: 2, reps: 5
792.84 (± 81.63) max.: 934.21 clients: 4, reps: 5
1040.68 (± 172.72) max.: 1159.77 clients: 8, reps: 5
1500 1350 1200 1050 900 750 600 450 300 150 0

Messages to take home

- Preliminary results are encouraging
  - The API seems fit-for-purpose
  - Performance results obtained so far with IOR are encouraging
- Effort required to optimise configuration
- Ongoing work to understand and reduce performance gap between IOR and weather field IO