Why is ScaleIO better than Ceph? ☺

(for block I/O)

Maciej Brzeźniak, PSNC

1st SIG-CISS meeting in SurfSARA,

Sep. 25-26th 2017

ScaleIO VS Ceph

Borrowed from

OpenStack Summit presentation

Battle of the Titans – ScaleIO vs. Ceph at OpenStack Summit Tokyo 2015

By Randy Bias, Jeff Thomas, EMC

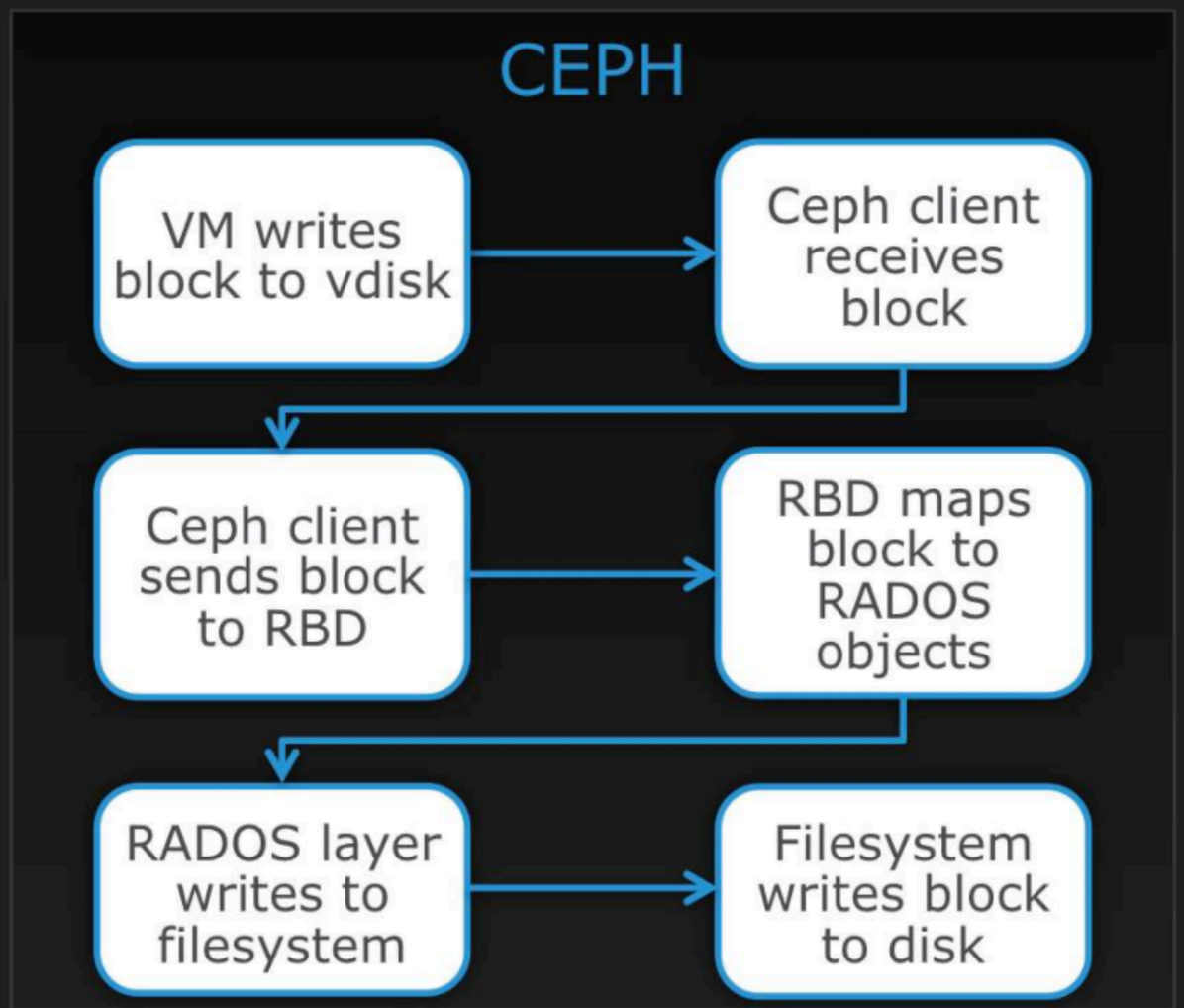
1st SIG-CISS meeting in SurfSARA,

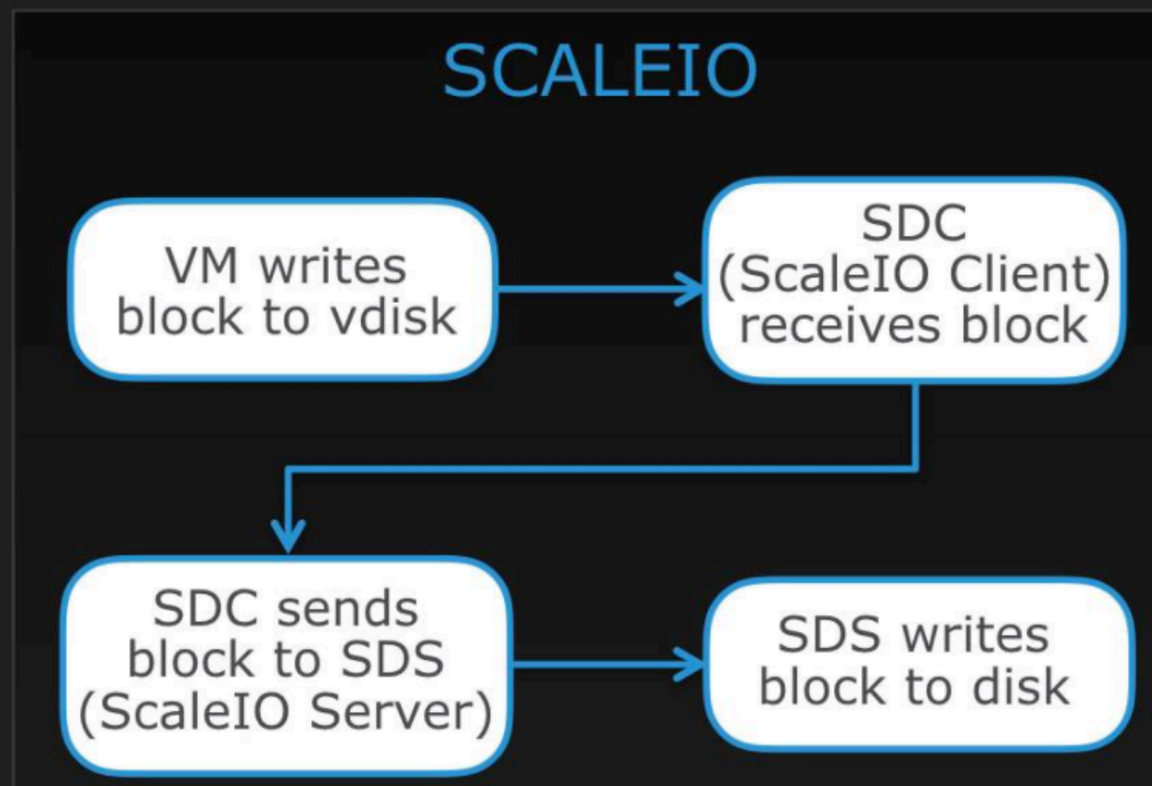
Sep. 25-26th 2017

SCALEIO VS CEPH

CEPH AND SCALEIO APPROACHES

LIMITING OVERHEAD IS A KEY TO UNLOCKING PERFORMANCE





EMC²

More on ScaleIO

from EMC…

Shared by Wojciech Janusz, EMC

Selected by Maciej Brzeźniak, PSNC

Short overview at: https://box.psnc.pl/f/5f9fd84353

More: https://box.psnc.pl/f/72b4964e59/

1st SIG-CISS meeting in SurfSARA,

Sep. 25-26th 2017

ScaleIO at PSNC

PoC in 2016

Maciej Brzeźniak (PSNC),
Krzysztof Wadówka (PSNC)

1st SIG-CISS meeting in SurfSARA,

Sep. 25-26th 2017

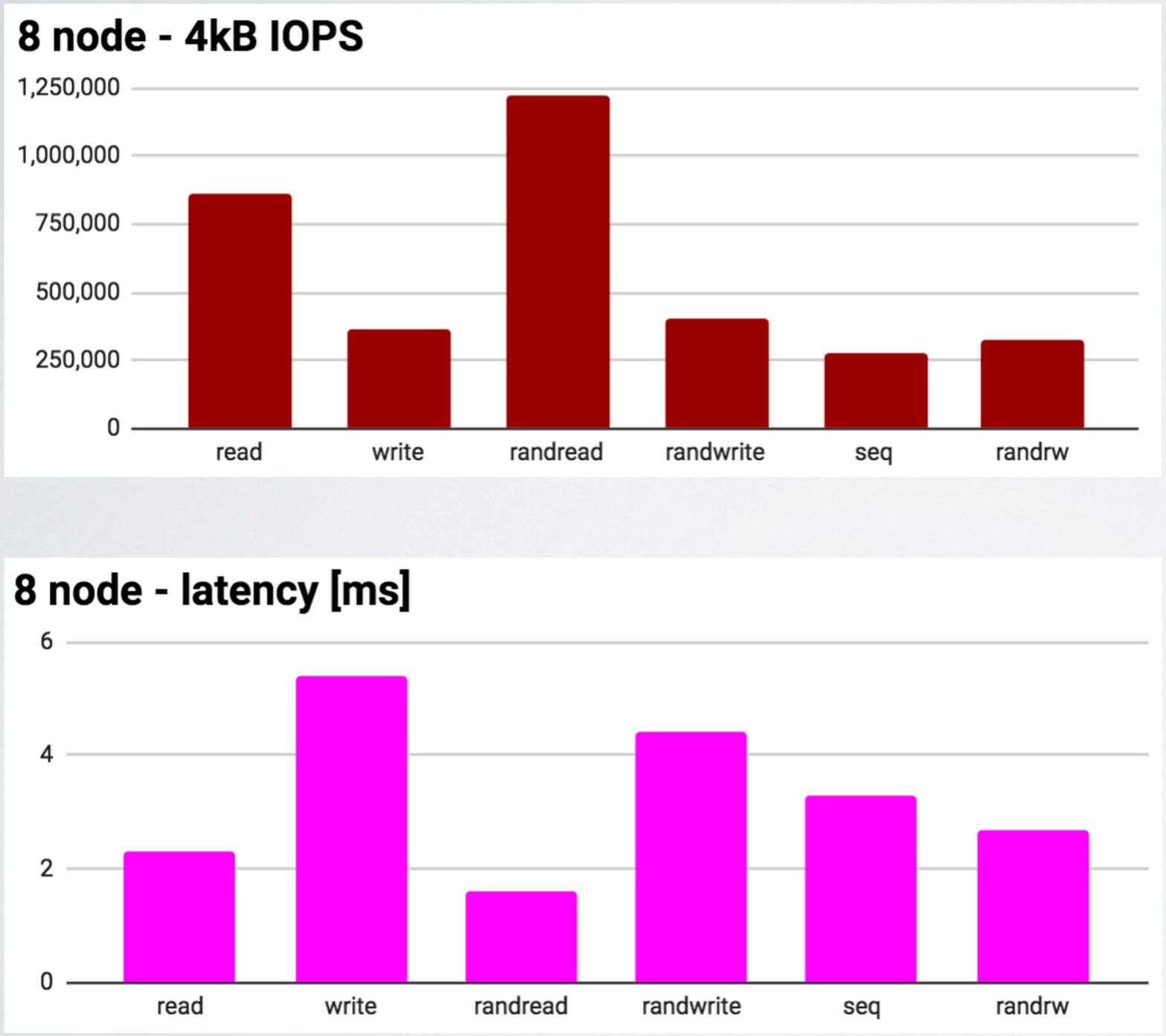
SERVERS FOR SDS – 1U, HIGH-PERFORMANCE SERVERS

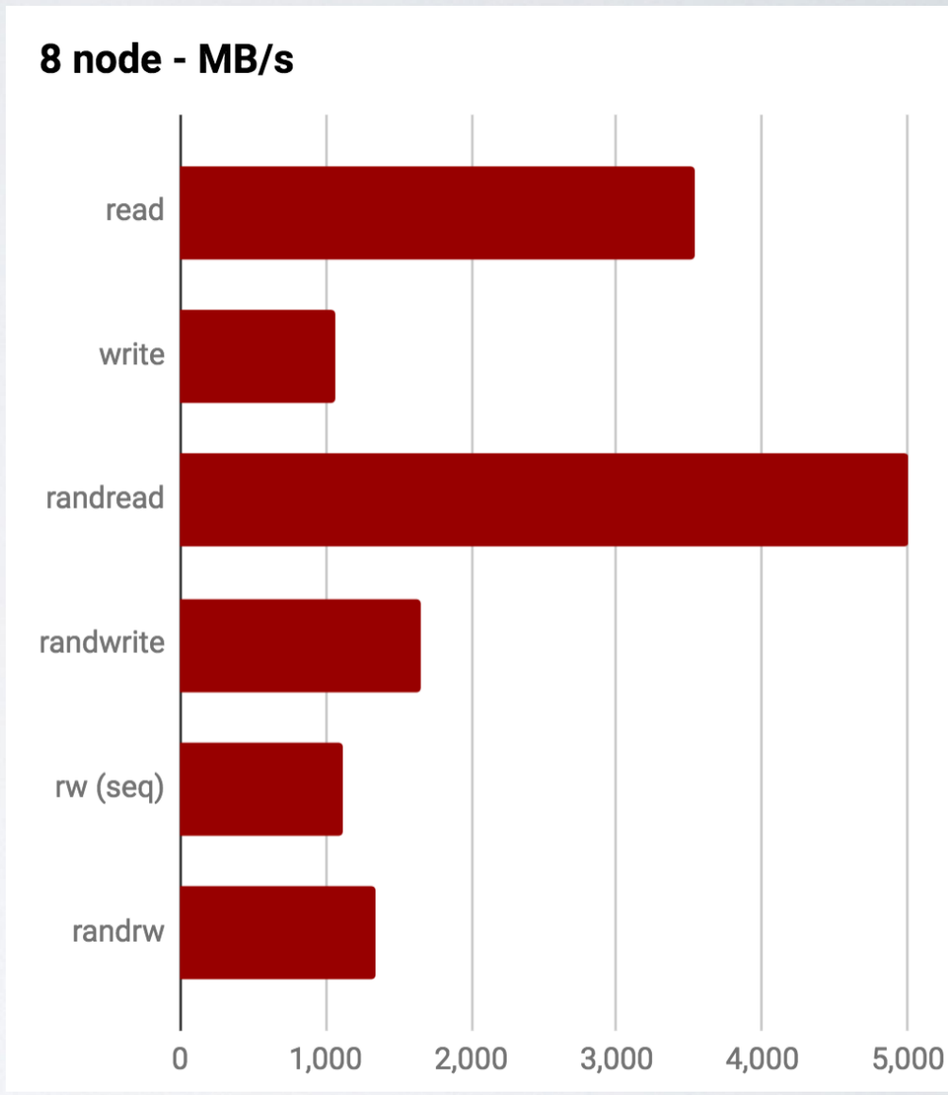
QuantaGrid D51PH-1ULH:

- 1U server – options:
  - 12xHDD 3,5" – all hot swappable
  - 4x SSD 2,5" – loadable from front
- CPUs & RAM:
  - 2 socket mainboard
  - Intel E5 v3/4
  - 16x DIMM slots: up to 512GB
- Network:
  - 2x10GbE or 1x 40GbE now
- Usage:
  - Ceph for RBD on HDDs
  - ScaleIO on SSDs

  possibly in a 'hyper-convergent setup – running OpenStack in the same time

TESTS OF SCALEIO AT PSNC (RESULTS)





TESTS OF SCALEIO AT PSNC (RESULTS)

1 NODE	FIO test name (bs=4K)	Bandwidth (MB/s)	IOPS	Latency (mSec)	FIO test name (bs=1024K)	Bandwidth (MB/s)	IOPS	Latency (mSec)
read	**1_1**	465	116,321	2.1	**1_1b**	1,118	1,118	226
write	**1_2**	318	79,551	3.2	**1_2b**	1,118	1,118	227
randread	**1_3**	709	177,268	1.4	**1_3b**	1,118	1,118	226
randwrite	**1_4**	705	176,352	1.4	**1_4b**	1,118	1,118	226
rw (r/w)	**1_5**	198	49,530	2.3 / 2.8	**1_5b**	1,024	1,024	124
randrw	**1_6**	338	84,531	1.5 / 1.5	**1_6b**	1,117	1,116	123

8 NODEs	FIO test name (bs=4K)	Bandwidth (MB/s)	IOPS	Latency (mSec)	FIO test name (bs=1024K)	Bandwidth (MB/s)	IOPS	Latency (mSec)
read	**2_1**	3534	862,844	2.3	**2_1b**	9,369	8,934	227
write	**2_2**	1059	368,652	5.4	**2_2b**	5,420	5,169	392
randread	**2_3**	4997	1,220,200	1.6	**2_3b**	9,373	8,939	227
randwrite	**2_4**	1660	405,467	4.4	**2_4b**	4,558	4,347	407
rw	**2_5**	1123	274,270	3.3	**2_5b**	4,884	4,659	239
randrw	**2_5**	1351	329,997	2.7	**2_6b**	4,044	3,859	250

TESTS OF SCALEIO AT PSNC (TEST CONFIGURATION)

10 x QuantaGrid servers

- 4x SSDs per server
- 10Gbit links (dual)
- Converged:

  SDCs and SDSs

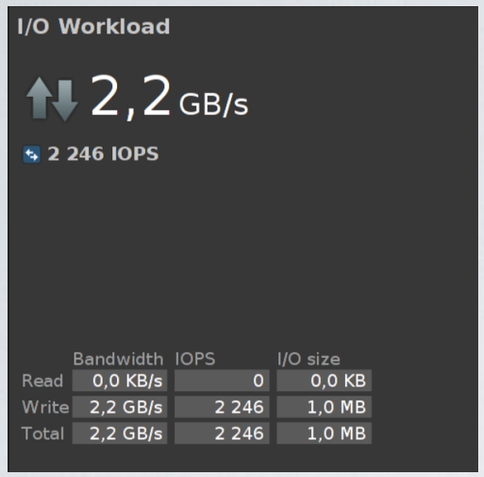
  on the same servers

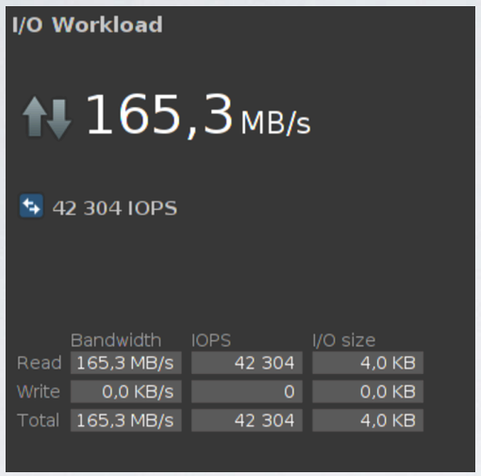
System	436.5 TB	14.0 TB (3.2 %)	6.0 GB/s	6,188
default	436.5 TB	14.0 TB (3.2 %)	6.0 GB/s	6,188
SDS_[10.20.21.21]	43.7 TB	1.4 TB (3.2 %)	609.0 MB/s	609
SDS_[10.20.21.22]	43.7 TB	1.4 TB (3.2 %)	615.8 MB/s	616
SDS_[10.20.21.23]	43.7 TB	1.4 TB (3.3 %)	640.2 MB/s	640
SDS_[10.20.21.24]	43.7 TB	1.4 TB (3.2 %)	615.8 MB/s	616
SDS_[10.20.21.25]	43.7 TB	1.4 TB (3.2 %)	630.2 MB/s	630
SDS_[10.20.21.26]	43.7 TB	1.4 TB (3.2 %)	611.4 MB/s	611
SDS_[10.20.21.27]	43.7 TB	1.4 TB (3.2 %)	616.8 MB/s	617
SDS_[10.20.21.28]	43.7 TB	1.4 TB (3.2 %)	615.8 MB/s	616
SDS_[10.20.21.29]	43.7 TB	1.4 TB (3.2 %)	615.4 MB/s	615
SDS_[10.20.21.30]	43.7 TB	1.4 TB (3.2 %)	617.6 MB/s	618

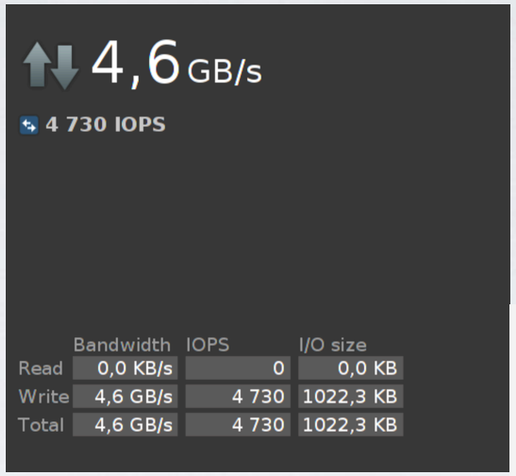
SOME VIEWS ON THE CONSOLE

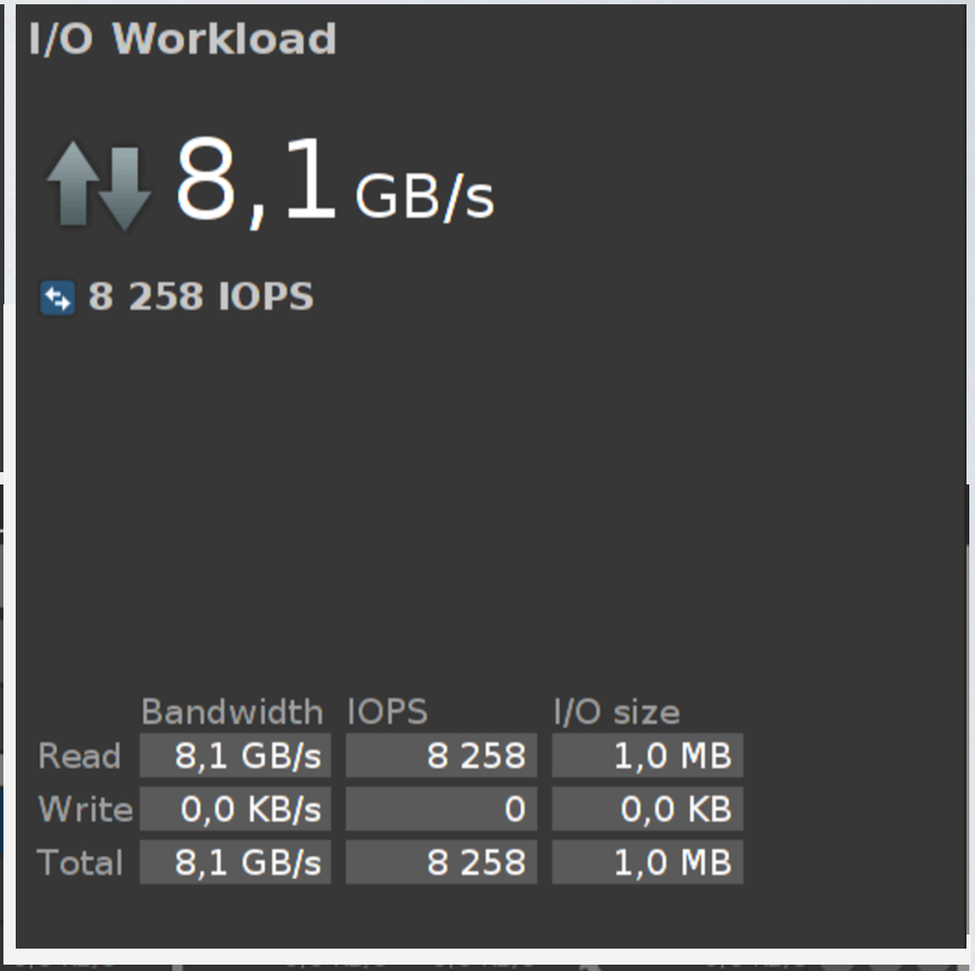
Item	Total Capacit	Capacity In-Use	I/O	Bandwidth	IOPS	Rebuild	Rebalance	Alerts
System	436,5 TB	45,2 TB (10,4 %)		0,0 KB/s	0	1,7 GB/s	0,0 KB/s	2 1
default	436,5 TB	45,2 TB (10,4 %)		0,0 KB/s	0	1,7 GB/s	0,0 KB/s	
SDS_[10.20.21.21]	43,7 TB	4,9 TB (11,2 %)		0,0 KB/s	0	172,8 MB/s → → 157,2 MB/s	0,0 KB/s → → 0,0 KB/s	
SDS_[10.20.21.22]	43,7 TB	4,8 TB (11,1 %)		0,0 KB/s	0	168,8 MB/s → → 246,2 MB/s	0,0 KB/s → → 0,0 KB/s	
SDS_[10.20.21.23]	43,7 TB	4,9 TB (11,1 %)		0,0 KB/s	0	186,4 MB/s → → 454,0 MB/s	0,0 KB/s → → 0,0 KB/s	
SDS_[10.20.21.24]	43,7 TB	4,8 TB (11,1 %)		0,0 KB/s	0	193,6 MB/s → → 397,4 MB/s	0,0 KB/s → → 0,0 KB/s	
SDS_[10.20.21.25]	43,7 TB	4,9 TB (11,2 %)		0,0 KB/s	0	287,0 MB/s → → 58,2 MB/s	0,0 KB/s → → 0,0 KB/s	
SDS_[10.20.21.26]	43,7 TB	4,8 TB (11,1 %)		0,0 KB/s	0	134,8 MB/s → → 172,2 MB/s	0,0 KB/s → → 0,0 KB/s	
SDS_[10.20.21.27]	43,7 TB	4,8 TB (10,9 %)		0,0 KB/s	0	250,0 MB/s → → 76,8 MB/s	0,0 KB/s → → 0,0 KB/s	
SDS_[10.20.21.28]	43,7 TB	4,9 TB (11,3 %)		0,0 KB/s	0	205,0 MB/s → → 160,6 MB/s	0,0 KB/s → → 0,0 KB/s	
SDS_[10.20.21.29]	43,7 TB	4,8 TB (11,0 %)		0,0 KB/s	0	153,4 MB/s → → 33,6 MB/s	0,0 KB/s → → 0,0 KB/s	
SDS_[10.20.21.30]	43,7 TB	1,6 TB (3,6 %)		0,0 KB/s	0	0,0 KB/s → → 0,0 KB/s	0,0 KB/s → → 0,0 KB/s	
default								
/dev/sda	3,6 TB	126,0 GB (3,4 %)		0,0 KB/s	0	0,0 KB/s → → 0,0 KB/s	0,0 KB/s → → 0,0 KB/s	
/dev/sdc	3,6 TB	128,0 GB (3,4 %)		0,0 KB/s	0	0,0 KB/s → → 0,0 KB/s	0,0 KB/s → → 0,0 KB/s	
/dev/sdd	3,6 TB	140,0 GB (3,8 %)		0,0 KB/s	0	0,0 KB/s → → 0,0 KB/s	0,0 KB/s → → 0,0 KB/s	
/dev/sde	3,6 TB	138,0 GB (3,7 %)		0,0 KB/s	0	0,0 KB/s → → 0,0 KB/s	0,0 KB/s → → 0,0 KB/s	
/dev/sdf	3,6 TB	146,0 GB (3,9 %)		0,0 KB/s	0	0,0 KB/s → → 0,0 KB/s	0,0 KB/s → → 0,0 KB/s	
/dev/sdg	3,6 TB	116,0 GB (3,1 %)		0,0 KB/s	0	0,0 KB/s → → 0,0 KB/s	0,0 KB/s → → 0,0 KB/s	
/dev/sdh	3,6 TB	132,0 GB (3,5 %)		0,0 KB/s	0	0,0 KB/s → → 0,0 KB/s	0,0 KB/s → → 0,0 KB/s	
/dev/sdi	3,6 TB	142,0 GB (3,8 %)		0,0 KB/s	0	0,0 KB/s → → 0,0 KB/s	0,0 KB/s → → 0,0 KB/s	
/dev/sdj	3,6 TB	118,0 GB (3,2 %)		0,0 KB/s	0	0,0 KB/s → → 0,0 KB/s	0,0 KB/s → → 0,0 KB/s	
/dev/sdk	3,6 TB	160,0 GB (4,3 %)		0,0 KB/s	0	0,0 KB/s → → 0,0 KB/s	0,0 KB/s → → 0,0 KB/s	
/dev/sdl	3,6 TB	138,0 GB (3,7 %)		0,0 KB/s	0	0,0 KB/s → → 0,0 KB/s	0,0 KB/s → → 0,0 KB/s	
/dev/sdm	3,6 TB	124,0 GB (3,3 %)		0,0 KB/s	0	0,0 KB/s → → 0,0 KB/s	0,0 KB/s → → 0,0 KB/s	

SOME VIEWS ON THE CONSOLE









ON THE REALITY SIDE OF THINGS ;)
SCALEIO @PSNC NOW

- We're in the process of acquiring licenses (16 TB)
- We're will re-use old servers (former GPFS nodes):
  - 4x IBM x3650 M4
  - CPUs: E5-2643: 4C, 8T
  - RAM: 96 GB / box

- Outlook:
  - We plan tests + production at small scale
  - Possibly upgrade next year

- *Alternatives:?*
  - *In parallel we're going to test Huawei's Fusion Storage*
  - *Results to be reported during following SIG-CISS metings*

ScaleIO vs Ceph

THANK YOU

Maciej Brzeźniak, PSNC

1st SIG-CISS meeting in SurfSARA,

Sep. 25-26th 2017