

# PPA Lustre filesystem 2014 upgrade

Here are some benchmarks for PPA Lustre filesystem that was upgraded in March 2014.

## System specs:

- Lustre server version 2.5.1 on RHEL6.5
- 1 MDS and 8 OSS servers using Dell R610 systems
- Four LSI Engenio (Dell MD3260) arrays with dual-redundant controllers
- 16 OSTs using RAID6 LUNs with 2TB NL SAS drives
- ~344TB of usable capacity
- 10Gb ethernet links to SLAC network
- QDR Infiniband links to "bullet" PPA compute cluster

IOR is a parallel I/O benchmarking utility that runs MPI-coordinated file transfers. The test clients are "bullet" cluster nodes with Lustre mounted over QDR Infiniband. Maximum throughput is > 6GB/sec using multiple nodes or ~3GB/sec with a single node.

## 1 process on 1 node

Max Write: 419.58 MiB/sec (439.96 MB/sec)  
Max Read: 399.31 MiB/sec (418.71 MB/sec)

Command line used: src/C/IOF -F -a POSIX -r -w -b 60g -t 1m -o /lustre/ki/pfs/yemi/IOFFile -O lustreStripeCount=0 -v -i 2

api = POSIX  
test filename = /lustre/ki/pfs/yemi/IOFFile  
access = file-per-process  
pattern = segmented (1 segment)  
ordering in a file = sequential offsets  
ordering inter file = no tasks offsets  
clients = 1 (1 per node)  
repetitions = 2  
xfersize = 1 MiB  
blocksize = 60 GiB  
aggregate filesize = 60 GiB  
Lustre stripe size = Use default  
stripe count = Use default

Operation Max (MiB) Min (MiB) Mean (MiB) Std Dev Max (OPs) Min (OPs) Mean (OPs) Std Dev Mean (s) Op grep #Tasks tPN reps fPP reord  
reordoff reordrand seed segcnt blksiz xsize aggsz

|                |        |        |        |      |        |        |        |      |           |                   |             |         |             |
|----------------|--------|--------|--------|------|--------|--------|--------|------|-----------|-------------------|-------------|---------|-------------|
| write          | 419.58 | 412.07 | 415.82 | 3.75 | 419.58 | 412.07 | 415.82 | 3.75 | 147.76775 | 1 1 2 1 0 1 0 0 1 | 64424509440 | 1048576 | 64424509440 |
| -1 POSIX EXCEL |        |        |        |      |        |        |        |      |           |                   |             |         |             |
| read           | 399.31 | 391.06 | 395.18 | 4.13 | 399.31 | 391.06 | 395.18 | 4.13 | 155.48960 | 1 1 2 1 0 1 0 0 1 | 64424509440 | 1048576 | 64424509440 |
| -1 POSIX EXCEL |        |        |        |      |        |        |        |      |           |                   |             |         |             |

## 3 processes on 1 node

Max Write: 1133.49 MiB/sec (1188.55 MB/sec)  
Max Read: 1055.04 MiB/sec (1106.29 MB/sec)

Command line used: src/C/IOR -F -a POSIX -r -w -b 20g -t 1m -o /lustre/ki/pfs/yemi/IOFile -O lustreStripeCount=0 -v -i 2

api = POSIX  
test filename = /lustre/ki/pfs/yemi/IOFile  
access = file-per-process  
pattern = segmented (1 segment)  
ordering in a file = sequential offsets  
ordering inter file= no tasks offsets  
clients = 3 (3 per node)  
repetitions = 2  
xfersize = 1 MiB  
blocksize = 20 GiB  
aggregate filesize = 60 GiB  
Lustre stripe size = Use default  
stripe count = Use default

Operation Max (MiB) Min (MiB) Mean (MiB) Std Dev Max (OPs) Min (OPs) Mean (OPs) Std Dev Mean (s) Op grep #Tasks tPN reps fPP reord  
reordoff reordrand seed segcnt blksiz xsize aggsiz

```
-----
write 1133.49 975.77 1054.63 78.86 1133.49 975.77 1054.63 78.86 58.58496 3 3 2 1 0 1 0 0 1 21474836480 1048576
64424509440 -1 POSIX EXCEL
read 1055.04 878.41 966.73 88.32 1055.04 878.41 966.73 88.32 64.08962 3 3 2 1 0 1 0 0 1 21474836480 1048576
64424509440 -1 POSIX EXCEL
```

## 8 processes on 1 node

Max Write: 2373.15 MiB/sec (2488.43 MB/sec)  
Max Read: 2470.09 MiB/sec (2590.08 MB/sec)

Command line used: src/C/IOR -F -a POSIX -r -w -b 20g -t 1m -o /lustre/ki/pfs/yemi/IOFile -O lustreStripeCount=0 -v -i 2

api = POSIX  
test filename = /lustre/ki/pfs/yemi/IOFile  
access = file-per-process  
pattern = segmented (1 segment)  
ordering in a file = sequential offsets  
ordering inter file= no tasks offsets  
clients = 8 (8 per node)  
repetitions = 2  
xfersize = 1 MiB  
blocksize = 20 GiB  
aggregate filesize = 160 GiB  
Lustre stripe size = Use default  
stripe count = Use default

Operation Max (MiB) Min (MiB) Mean (MiB) Std Dev Max (OPs) Min (OPs) Mean (OPs) Std Dev Mean (s) Op grep #Tasks tPN reps fPP reord  
reordoff reordrand seed segcnt blksiz xsize aggsiz

```
-----
write 2373.15 2116.37 2244.76 128.39 2373.15 2116.37 2244.76 128.39 73.22739 8 8 2 1 0 1 0 0 1 21474836480 1048576
171798691840 -1 POSIX EXCEL
read 2470.09 2289.77 2379.93 90.16 2470.09 2289.77 2379.93 90.16 68.94123 8 8 2 1 0 1 0 0 1 21474836480 1048576
171798691840 -1 POSIX EXCEL
```

## 16 processes on 1 node

Max Write: 2959.63 MiB/sec (3103.40 MB/sec)  
Max Read: 2456.74 MiB/sec (2576.08 MB/sec)

Command line used: src/C/IOF -F -a POSIX -r -w -b 20g -t 1m -o /lustre/ki/pfs/yemi/IOFFile -O lustreStripeCount=0 -v -i 2

api = POSIX  
test filename = /lustre/ki/pfs/yemi/IOFFile  
access = file-per-process  
pattern = segmented (1 segment)  
ordering in a file = sequential offsets  
ordering inter file= no tasks offsets  
clients = 16 (16 per node)  
repetitions = 2  
xfersize = 1 MiB  
blocksize = 20 GiB  
aggregate filesize = 320 GiB  
Lustre stripe size = Use default  
stripe count = Use default

Operation Max (MiB) Min (MiB) Mean (MiB) Std Dev Max (OPs) Min (OPs) Mean (OPs) Std Dev Mean (s) Op grep #Tasks tPN reps fPP reord  
reordoff reordrand seed segcnt blksize xsize aggsz

|                 |         |         |         |       |         |         |         |       |           |    |    |   |   |   |   |   |   |   |             |         |
|-----------------|---------|---------|---------|-------|---------|---------|---------|-------|-----------|----|----|---|---|---|---|---|---|---|-------------|---------|
| write           | 2959.63 | 2767.42 | 2863.53 | 96.10 | 2959.63 | 2767.42 | 2863.53 | 96.10 | 114.56132 | 16 | 16 | 2 | 1 | 0 | 1 | 0 | 0 | 1 | 21474836480 | 1048576 |
| 343597383680 -1 | POSIX   | EXCEL   |         |       |         |         |         |       |           |    |    |   |   |   |   |   |   |   |             |         |
| read            | 2456.74 | 2430.77 | 2443.76 | 12.98 | 2456.74 | 2430.77 | 2443.76 | 12.98 | 134.09235 | 16 | 16 | 2 | 1 | 0 | 1 | 0 | 0 | 1 | 21474836480 | 1048576 |
| 343597383680 -1 | POSIX   | EXCEL   |         |       |         |         |         |       |           |    |    |   |   |   |   |   |   |   |             |         |

## 8 processes across 4 nodes

Max Write: 3101.05 MiB/sec (3251.68 MB/sec)

Max Read: 2917.03 MiB/sec (3058.73 MB/sec)

Command line used: src/C/IOF -F -a POSIX -r -w -b 30g -t 1m -o /lustre/ki/pfs/yemi/IOFFile -O lustreStripeCount=0 -v -i 2

api = POSIX  
test filename = /lustre/ki/pfs/yemi/IOFFile  
access = file-per-process  
pattern = segmented (1 segment)  
ordering in a file = sequential offsets  
ordering inter file= no tasks offsets  
clients = 8 (2 per node)  
repetitions = 2  
xfersize = 1 MiB  
blocksize = 30 GiB  
aggregate filesize = 240 GiB  
Lustre stripe size = Use default  
stripe count = Use default

Operation Max (MiB) Min (MiB) Mean (MiB) Std Dev Max (OPs) Min (OPs) Mean (OPs) Std Dev Mean (s) Op grep #Tasks tPN reps fPP reord  
reordoff reordrand seed segcnt blksize xsize aggsz

|                 |         |         |         |        |         |         |         |        |          |   |   |   |   |   |   |   |   |   |             |         |
|-----------------|---------|---------|---------|--------|---------|---------|---------|--------|----------|---|---|---|---|---|---|---|---|---|-------------|---------|
| write           | 3101.05 | 2873.49 | 2987.27 | 113.78 | 3101.05 | 2873.49 | 2987.27 | 113.78 | 82.38866 | 8 | 2 | 2 | 1 | 0 | 1 | 0 | 0 | 1 | 32212254720 | 1048576 |
| 257698037760 -1 | POSIX   | EXCEL   |         |        |         |         |         |        |          |   |   |   |   |   |   |   |   |   |             |         |
| read            | 2917.03 | 2749.41 | 2833.22 | 83.81  | 2917.03 | 2749.41 | 2833.22 | 83.81  | 86.81822 | 8 | 2 | 2 | 1 | 0 | 1 | 0 | 0 | 1 | 32212254720 | 1048576 |
| 257698037760 -1 | POSIX   | EXCEL   |         |        |         |         |         |        |          |   |   |   |   |   |   |   |   |   |             |         |

## 32 processes across 8 nodes

Command line used: src/C/lor -F -a POSIX -r -w -b 20g -t 1m -o /lustre/ki/pfs/yemi/lorfile -O lustreStripeCount=0 -v -i 2

[illegible][illegible]