

Performance Analysis

Synthetic benchmarks: IOR, bonnie++, mdtest

Report contains performance studies based on IOR, mdtest and bonnie++ benchmarks applied on the NOTUR systems: hexagon, titan, njord, stallo.

THIS REPORT IS FOR INFORMATIONAL PURPOSES ONLY, AND MAY CONTAIN TYPOGRAPHICAL ERRORS AND TECHNICAL INACCURACIES. THE CONTENT IS PROVIDED AS IS, WITHOUT EXPRESS OR IMPLIED WARRANTIES OF ANY KIND.
2010
Alexander Oltu UniBCCS

Content

1	Introduction	5
2	Performance studies	5
2.1	Bonnie++	5
2.2	Mdtest	7
2.3	IOR	8
3	Appendix A – applied commands and results	9
3.1	Bonnie++	9
3.2	Mdtest	10
3.3	IOR	11

1 Introduction

These performance studies can be used to select NOTUR system for the type of application based on the file system access pattern. It contains two metadata benchmarks and one read-write test.

The metadata benchmarks measures how fast file system can create, update and delete files.

The read-write test shows how fast clients can write and read. N-to-1 file access method has been used. In this test all clients will write into a single file. IOR, the benchmark used in this test, supports MPIIO, HDF5 and POSIX interfaces. The POSIX interface has been selected to eliminate the performance impact of additional layers like MPIIO or HDF5.

All benchmarks have been executed on live systems. This means that the system load during the test can have affected the results. Therefore, these numbers reflect the possible performance during regular usage of the machines and not the peak performance.

2 Performance studies

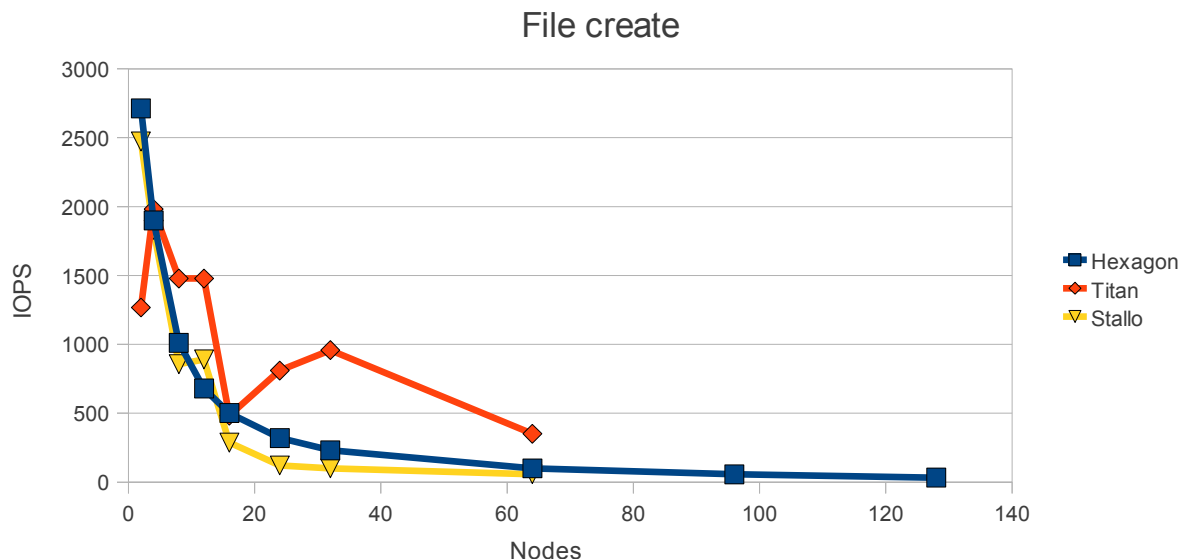
The tests performed included two metadata benchmarks, bonnie++ and mdtest, as well as one file system read-write test using the IOR application. The metadata performance is measured in IOPS (number of operations per second).

2.1 Bonnie++

This test was not executed on njord, due to lack of a way to start multiple non-MPI applications at the same time.

The bonnie++ source code was slightly modified to support the creation of directories with randomly generated names.

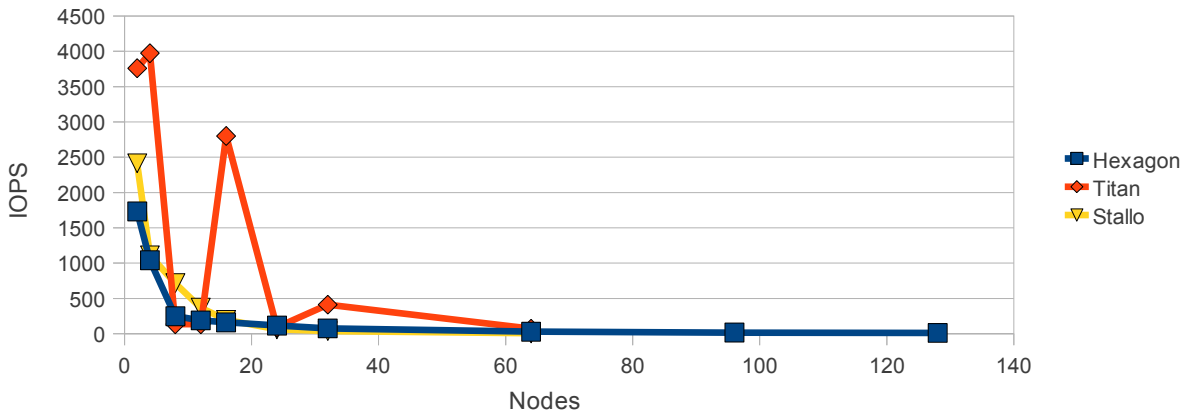
The test has been performed using 16000 zero-sized files, they were created, 'stat'ed and deleted by each node. The default striping pattern have been used on the Lustre filesystems (hexagon and stallo).



For a low amount of nodes the Lustre based machines had the best file create performance. In addition, results show that the performance on the Lustre based machines decreases proportionally with increasing number of clients. Due to the dedicated MDS (meta data server) the Lustre file system for each file create

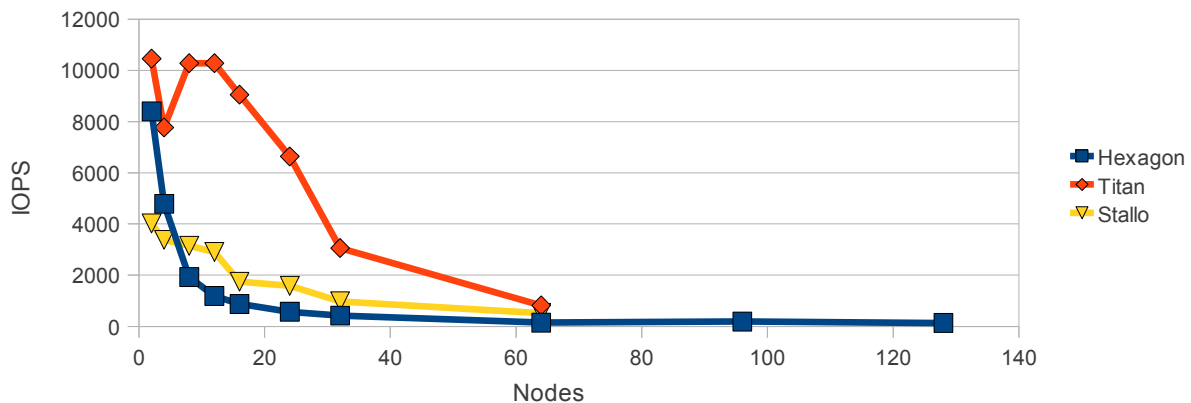
and file delete operation has to perform two operations: the first on the metadata server and the second on the object storage server.

File delete



The titan file system has a performance drop in create and delete operations when using 16 nodes, this result can have been affected by ATLAS jobs running on the titan system or possibly by the specifics of the GPFS file system.

File read

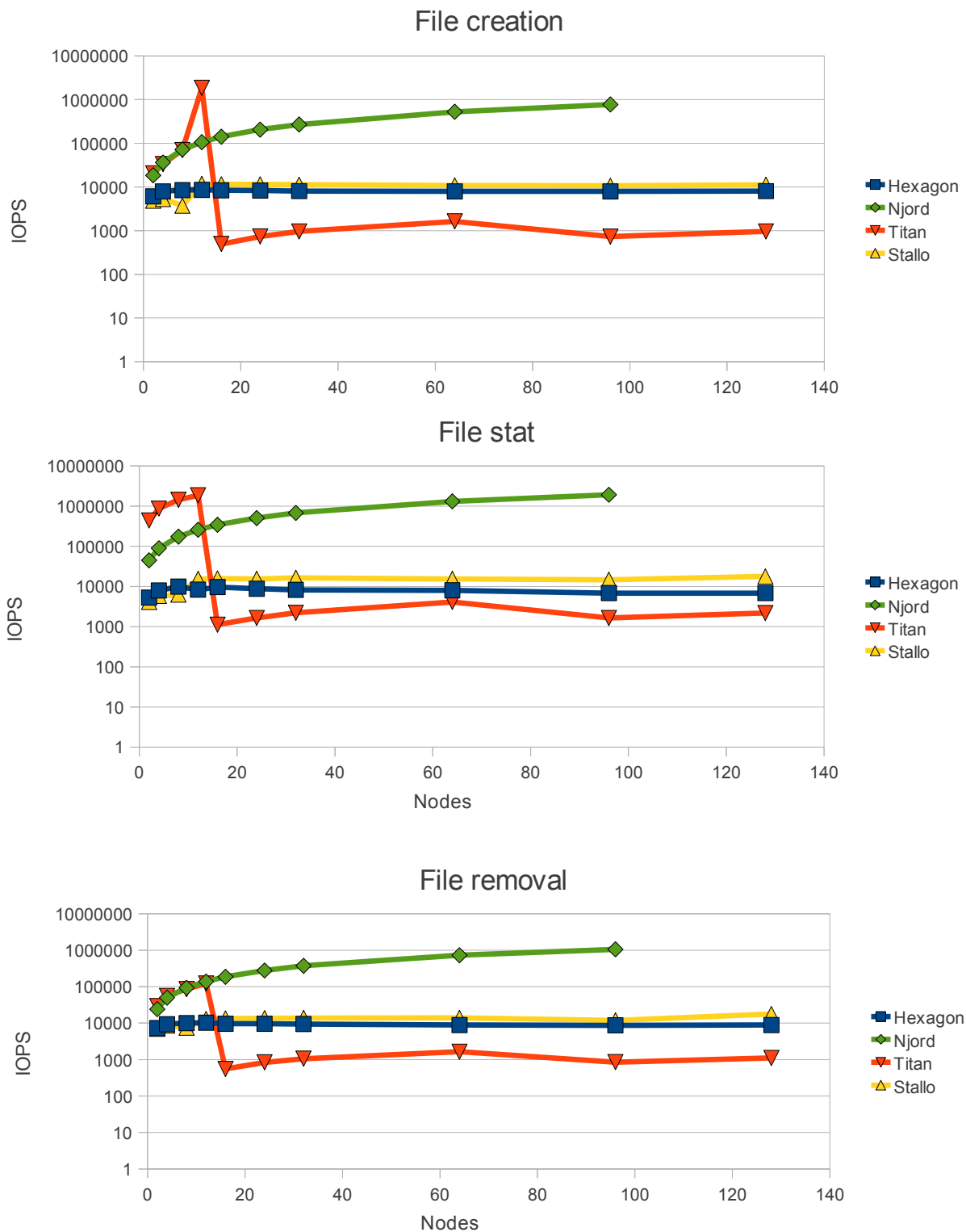


The best read performance in IOPS was on the titans GPFS file system.

A different striping pattern has been tested on the hexagons Lustre file system. Using a single stripe showed slightly higher performance for up to 96 nodes. For the 36 stripe pattern (the maximum for hexagon) the performance was very low. This can be explained by how the Lustre file system creates files (two object operations). Increasing the number of stripes therefore increases the overall number of operations required.

2.2 Mdtest

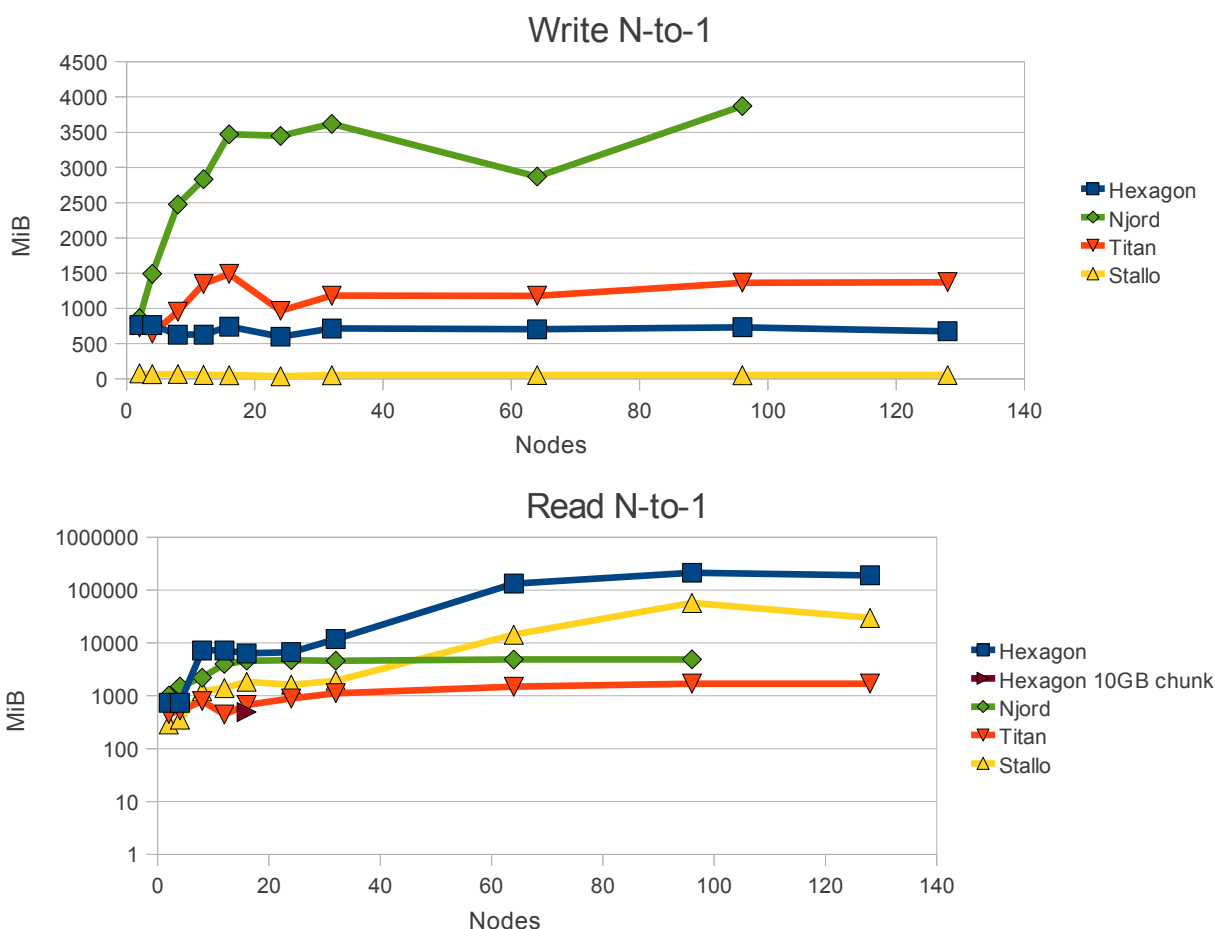
The mdtest had lower and more typical load pattern. Each client created, 'stat'ed and deleted 10 files. The test has been rerun 200 times for increased accuracy. All results are given in IOPS.



It is assumed that the GPFS local caching had a higher influence during this lower load (operation on 10 files per client). This can be clearly seen in the results from the njord system. For the titan system GPFS cache has less influence on the performance when using more than 16 clients. This surprisingly different behavior for the two GPFS file systems, can be related to the Linux and AIX implementation, version number or hardware type.

2.3 IOR

In this test each client wrote 1GB into the same file. This operation has then been repeated three times and the average used. All results are represented in MiB/s (mebibytes per second), to get MB/s the number must be multiplied by 1.048.



Due to file system space-constraints, the size for each client was selected to be 1GB, this will affect read performance due to caching effects, especially on hexagon. To eliminate this effect the file size must be bigger than the available memory on each node.

In the N-to-1 test njord shows very good write performance. To be sure that this is not caused by the caching effect, two reruns have been performed: one with two nodes and 25GB chunk size for each client and another with 16 nodes and 25GB chunk size for each client. In these reruns the results were nearly identical, showing that caching had minimal effect.

On hexagon, one rerun using 16 nodes and a 10GB chunk size were executed and showed a lower performance due to the reduced cache effect.

The poor write performance on the stallo system can be related to the disk upgrade, but a retest using two nodes on October 18th gave the same result.

3 Appendix A – applied commands and results

3.1 Bonnie++

The version used has patch to create random directory names.

```
$submit_pref ./$EXE -q -s 0 -n 16:0:0:1 >>run.csv 2>>${stderr}
```

Results:

Version	N of files K	Hexagon						Titan						Stallo					
		Sequential create	Read		Delete		Sequential create	Read		Delete		Sequential create	Read		Delete				
	Create IOPS	%CP	IOPS	%CP	IOPS	%CP	Create IOPS	%CP	IOPS	%CP	IOPS	%CP	Create IOPS	%CP	IOPS	%CP	IOPS	%CP	
2	Nodes																		
Avg	2713	11.5	8397	39	1733.5	1267.5	3	10456	14	3759.5	9.5	2472.5	17.5	4017	26.5	2414.5	19.5		
4	Nodes																		
Avg	1899	8.25	4786	21	1040.5	1982.3	4.75	7770	8.75	3972.5	12	1831.8	13.25	3381	26	1110	9		
8	Nodes																		
Avg	1011.38	4.125	1930	8.5	248.13	1477.4	3.5	10279	11.75	134.75	0	855.88	5.875	3148	24.25	715.25	5.625		
12	Nodes																		
Avg	679.583	2.75	1184	5.3333	188.58	1477.4	3.5	10279	11.75	134.75	0	890.92	6.083	2900	21.17	367.5	2.917		
16	Nodes																		
Avg	501	2	880.1	3.75	161.19	482	1	9052	10	2801	8	286.81	2	1745	13.13	196.81	1		
24	Nodes																		
Avg	319.583	1	568.3	2.0833	116.33	809.17	1.708	6638	6.875	89.542	0	121.08	0.042	1585	11.5	63.818	3		
32	Nodes																		
Avg	230.938	0.969	420.2	1.2813	76.031	958.28	2.031	3062	2.9688	412.19	0.969	99.5	0	976.47	6.969	37.656	0		
64	Nodes																		
Avg	99.4688	0	147	0	29.547	350.88	0.516	821.1	0.4375	71.375	0.063	57.938	0	519.72	3.063	5.6563	0		
96	Nodes																		
Avg	56.6146	0	187	0.1042	17.115														
128	Nodes																		
Avg	31.9141	0	131.1	0	11.398														

3.3 IOR

```
$submit_pref ./$EXE -a POSIX -i 3 -d 32 -E -o "$IORFILE" -s 1 -b 1G -t 1m
>>$stderr
```

Results:

Operati on	Hexagon				Hexagon				Titan				Stallo				Njord																				
	Max (MiB)	Min (MiB)	Mean (MiB)	Std Dev (OPs)	Max (MiB)	Min (MiB)	Mean (MiB)	Std Dev (OPs)	Max (MiB)	Min (MiB)	Mean (MiB)	Std Dev (OPs)	Max (MiB)	Min (MiB)	Mean (MiB)	Std Dev (OPs)	Max (MiB)	Min (MiB)	Mean (MiB)	Std Dev (OPs)																	
2 nodes																																					
write	763.17	759.98	761.11	1.46	763.17	759.98	761.11	1.46	2.6908	859.92	556.9	726.44	126.31	859.92	556.9	726.44	126.31	2.915	83.53	75	78.57	3.62	83.53	75	78.57	3.62	26.12	978.6	799.02	859.5	84.25	978.6	799.02	859.5	84.25	2.4045	
read	862.66	640.64	741.87	91.69	862.66	640.64	741.87	91.69	2.8021	518.89	345.63	454.57	77.45	518.89	345.63	454.57	77.45	4.658	356.69	156.7	287	92.21	356.69	156.69	287	92.21	8.2346	1377	791.37	1009	261.3	1377	791.37	1009	261.3	2.1521	
4																																					
write	763.17	759.98	761.11	1.46	763.17	759.98	761.11	1.46	2.6908	747.73	488.99	643.39	111.41	747.73	488.99	643.39	111.41	6.587	71.57	63.31	66.23	3.78	71.57	63.31	66.23	3.78	62.044	1578	1352	1491	99.57	1578	1352	1491	99.57	2.7802	
read	862.66	640.64	741.87	91.69	862.66	640.64	741.87	91.69	2.8021	638.2	406.5	524.5	94.64	638.2	406.5	524.5	94.64	8.08	364.66	334.3	354.49	14.26	364.66	334.32	354.49	14.26	11.574	2627	915.12	1507	792.3	2627	915.12	1507	792.3	3.4065	
8																																					
write	760.01	368.74	627.68	183.11	760.01	368.74	627.68	183.11	14.619	1005.25	911.69	952.09	39.24	1005.3	911.69	952.09	39.24	8.619	69.56	65.34	68.11	1.96	69.56	65.34	68.11	1.96	120.37	2873	2052	2475	335.8	2873	2052	2475	335.8	3.3733	
read	15585	2417.52	7211.25	5942	15585	2417.52	7211.25	5942	2.0568	882.98	701.03	797.52	74.69	882.98	701.03	797.52	74.69	10.36	1271.83	1123	1220.95	69.11	1271.83	1123.24	1221	69.11	6.732	3256	1363.4	2215	784	3256	1363.4	2215	784	4.1892	
12																																					
write	760.01	368.74	627.68	183.11	760.01	368.74	627.68	183.11	14.619	1403.95	1286.5	1346.84	48.03	1404	1286.5	1347	48.03	9.135	57.67	52.92	56	2.18	57.67	52.92	56	2.18	219.77	3022	2676	2833	143.1	3022	2676	2833	143.1	4.3478	
read	15585	2417.52	7211.25	5942	15585	2417.52	7211.25	5942	2.0568	456.83	426.06	444.03	13.08	456.83	426.06	444.03	13.08	27.7	1848.22	1113	1402.67	319.78	1848.22	1112.82	1402.7	319.78	9.1817	4446	3705	4029	309.7	4446	3705	4029	309.7	3.0673	
16																																					
write	767.42	701.8	741.12	28.32	767.42	701.8	741.12	28.32	22.14	593.9	1527.32	1420.8	1489.19	48.48	1527.3	1420.8	1489	48.48	11.01	55.1	48.94	51.69	2.56	55.1	48.94	51.69	2.56	317.76	3493	3435	3470	25.53	3493	3435	3470	25.53	4.7212
read	7080.59	5426.38	6342.61	686.99	7080.59	5426.38	6342.61	686.99	2.6153	493.22	676	667.66	671.22	3.51	676	667.66	671.22	3.51	24.41	2209.17	1431	1940.24	318.82	2209.17	1431.31	1840.2	318.82	9.1923	4694	4542	4621	62.27	4694	4542	4621	62.27	3.5464
24																																					
write	732.07	387.03	597.85	150.91	732.07	387.03	597.85	150.91	44.503	1000	927.14	964.99	29.81	1000	927.14	964.99	29.81	25.49	51.9	10.12	37.03	19.06	51.9	10.12	37.03	19.06	1134.2	3762	3213	3447	231.4	3762	3213	3447	231.4	7.1604	
read	9478.96	3207.79	6711.27	2612.53	9478.96	3207.79	6711.27	2612.53	4.518	944.02	819.76	895.28	54.15	944.02	819.76	895.28	54.15	27.56	3441.91	151.4	1623.92	1365.4	3441.91	151.39	1623.9	1365.4	62.901	4832	4653	4713	83.99	4832	4653	4713	83.99	5.2162	
32																																					
write	739.72	698.68	714.49	18.03	739.72	698.68	714.49	18.03	45.891	1228.22	1116.9	1179.2	46.4	1228.2	1116.9	1179.2	46.4	27.83	54.18	50.12	51.64	1.81	54.18	50.12	51.64	1.81	635.31	3737	3458	3618	117.7	3737	3458	3618	117.7	9.0667	
read	14837	5860.36	11825.8	4218.26	14837	5860.36	11825.8	4218.26	3.339	1162.24	1066.1	1122.29	40.88	1162.2	1066.1	1122	40.88	29.24	2284.09	1707	1932.83	251.71	2284.09	1707.19	1932.8	251.71	17.224	4635	4600	4615	14.46	4635	4600	4615	14.46	7.0999	
64																																					
write	759.31	612.91	703.18	64.46	759.31	612.91	703.18	64.46	94.041	1196.85	1168.9	1179.14	12.58	1196.9	1168.9	1179	12.58	55.59	55.11	49.61	53.12	2.49	55.11	49.61	53.12	2.49	123EXCEL	3880	1186	2871	1199	3880	1186	2871	1199	30.209	
read	148668	101417	132896	22261	148668	101417	132896	22260.5	0.5093	1693.52	1367.4	1485.44	147.58	1693.5	1367.4	1485	147.58	44.53	19375	11353	14368	3565	19375	11353.4	14368	3565	4.8169	4888	4796	4852	39.93	4888	4796	4852	39.93	13.509	
96																																					
write	752.31	717.46	731.48	15.02	752.31	717.46	731.48	15.02	134.45	1398.1	1335.7	1364.19	25.76	1398.1	1335.7	1364	25.76	72.09	54.85	50.25	52.2	1.94	54.85	50.25	52.2	1.94	188EXCEL	3929	3810	3872	48.85	3929	3810	3872	48.85	25.395	
read	221130	202537	213188	7828.41	221130	202537	213188	7828.41	0.4618	1741.65	1672.2	1702.14	29.16	1741.7	1672.2	1702	29.16	57.77	127859	21418	57329	49675	127859	21418	57329	49675	3.229	4989	4816	4889	73.16	4989	4816	4889	73.16	20.114	
128																																					
write	688.3	656.74	674.51	13.19	688.3	656.74	674.51	13.19	194.4	1394.69	1342.6	1371.8	21.75	1394.7	1342.6	1371.8	21.75	95.57	55.5	50.91	52.62	2.05	55.5	50.91	52.62	2.05	249EXCEL	3929	3810	3872	48.85	3929	3810	3872	48.85	25.395	
read	298150	4408.03	190261	131996	298150	4408.03	190261	131996	10.221	1747.62	1674.4	1705.92	30.75	1747.6	1674.4	1706	30.75	76.86	50241.7	7408	29763	17537	50241.7	7408.29	29763	17537	8.1481										