CS597 Benchmarking Storage Systems

AN EVALUATION OF DISTRIBUTED STORAGE SYSTEMS

Benchmarking Storage Systems by Akash Mahakode & Sandip Lakshminarasaiah

Table of Contents

1. Introduction

This document describes the performance evaluations done on multiple Distributed File Systems. The aim of this project was to add more value to the "High-Performance Storage Support for Scientific Applications on the Cloud" technical paper. Multiple distributed file systems were evaluated for read and write performance with varying datasets. The evaluations were performed on Amazon EC2 and the behavior of each system was analyzed. The installation, setup and the configuration parameters of each of the system is also documented.

2. Evaluation Systems

The following distributed File Systems were evaluated.

1. S3 -- Amazon S3 is easy to use object storage, with a simple web service interface to store and retrieve any amount of data from anywhere on the web. With Amazon S3, you pay only for the storage you actually use. There is no minimum fee and no setup cost.

2. S3FS – S3FS is a FUSE filesystem application backed by amazon web services simple storage service

3. CephFS -- The Ceph Filesystem (Ceph FS) is a POSIX-compliant filesystem that uses a Ceph Storage Cluster to store its data. The Ceph filesystem uses the same Ceph Storage Cluster system as Ceph Block Devices, Ceph Object Storage with its S3 and Swift APIs, or native bindings (librados).

4. FusionFS -- FusionFS is a new distributed filesystem that will co-exist with current parallel filesystems in High-End Computing, optimized for both a subset of HPC and Many-Task Computing workloads. FusionFS is a user-level filesystem that runs on the compute resource infrastructure, and enables every compute node to actively participate in the metadata and data management. Distributed metadata management is implemented using ZHT, a zero-hop distributed hashtable. ZHT has been tuned for the specific requirements of high-end computing (e.g. trustworthy/reliable hardware, fast networks, non-existent "churn", low latencies, and scientific computing data-access patterns). The data is partitioned and spread out over many nodes based on the data access

patterns. Replication is used to ensure data availability, and cooperative caching delivers high aggregate throughput. Data is indexed, by including descriptive, provenance, and system metadata on each file. FusioFS supports a variety of data-access semantics, from POSIX-like interfaces for generality, to relaxed semantics for increased scalability.

5. HDFS -- HDFS is the primary distributed storage used by Hadoop applications. A HDFS cluster primarily consists of a NameNode that manages the file system metadata and DataNodes that store the actual data. Clients contact NameNode for file metadata or file modifications and perform actual file I/O directly with the DataNodes.

3. Data Set and Hardware

The evaluation was performed for read, write and metadata operations. The nodes on Amazon EC2 was varied from 1, 4, 16 & 64. We used the Amazon EC2 m3.large instances for evaluation. This family provides a balance of compute, memory, and network resources, and it is a good choice for many applications.

3.1 Features

- High Frequency Intel Xeon E5-2670 v2 (Ivy Bridge) Processors
- SSD-based instance storage for fast I/O performance
- Balance of compute, memory, and network resources

3.2 Hardware configuration

Model	vCPU	Mem (GiB)	SSD Storage (GB)
m3.large	2	7.5	1 x 32

3.3 Use Cases

Small and mid-size databases, data processing tasks that require additional memory, caching fleets, and for running backend servers for SAP, Microsoft SharePoint, cluster computing, and other enterprise applications. M3 instances may also launch as an Intel Xeon E5-2670 (Sandy Bridge) Processor running at 2.6 GHz.

3.4 Data sets

For metadata evaluation we used touch command to touch 10000 files and tracked the time. We used the following dataset for read and write operations.

No. of Files per Node	Size per file	Total Storage Needed
10000	0	0
10000	1024	10240000
1000	10240	10240000
1000	102400	102400000
1000	1024000	1024000000
1000	10240000	10240000000
100	102400000	10240000000
10	1024000000	10240000000
1	10240000000	10240000000

4. Evaluation methodology

We aimed to split the benchmarking into 2 parts. Evaluation of the metadata management and the performance of read and write for various file sizes.

4.1 Metadata Evaluation

In order to arrive at the best way to evaluate metadata management on the distributed filesystems, we tested multiple options. We evaluated using touch, a bash script and a java program to create blank / empty files. The time taken for 1000 and 1000 files was captured as shown below.

Metadata Evaluation		
Files	1000	10000
touch	0m0.013s	0m0.132s
bash script	0m1.283s	0m8.276s
java program	0m0.015s	0m0.221s

From the data, it is evident that touch was the fastest way of creating empty files. There are 2 ways of using the touch command. One is using touch in a loop as shown below.

- mkdir some_dir_in_s3
- cd some_dir_in_s3
- time for i in {1..10000}; do touch $i.txt; done

The second way of using touch is to create all files using a single touch command as shown below.

- for i in {1..10000}; do echo $i.txt >> tempFile; done
- tr '\n' ' ' < tempFile > tempFile1
- var=`cat tempFile1`
- mkdir some_dir_in_s3; cd some_dir_in_s3; time touch $var

Both mechanisms of using touch was evaluated on 1, 2 and 4 nodes with cephFS and the results were as show below.

Touch Option 1				
Instances	**node-1**	**node-2**	**node-3**	**node-4**
1	0m14.599s	-	-	-
2	0m15.642s	0m15.794s	-	-
4	0m30.792s	0m30.953s	0m25.762s	0m30.795s

Touch Option 2				
Instances	**node-1**	**node-2**	**node-3**	**node-4**
1	0m6.985s	-	-	-
2	0m10.570s	0m10.552s	-	-
4	0m28.758s	0m29.086s	0m26.923s	0m28.763s

From the data, it is evident that touch with option 2 of creating all files with a single command was the fastest way of creating empty files.

4.2 Read / Write Evaluation

In order to arrive at the best way to evaluate write performance. We evaluated two options.

1. Write using dd. This was done using the command >> dd if=/dev/zero of=file-name bs=1G count=10 conv=fdatasync
2. Write using perl. This was done using the command >> perl -e 'print "a" x '$size'' >> file$i
3. Read was performed using the command cat $1 > /home/ubuntu/tempFile

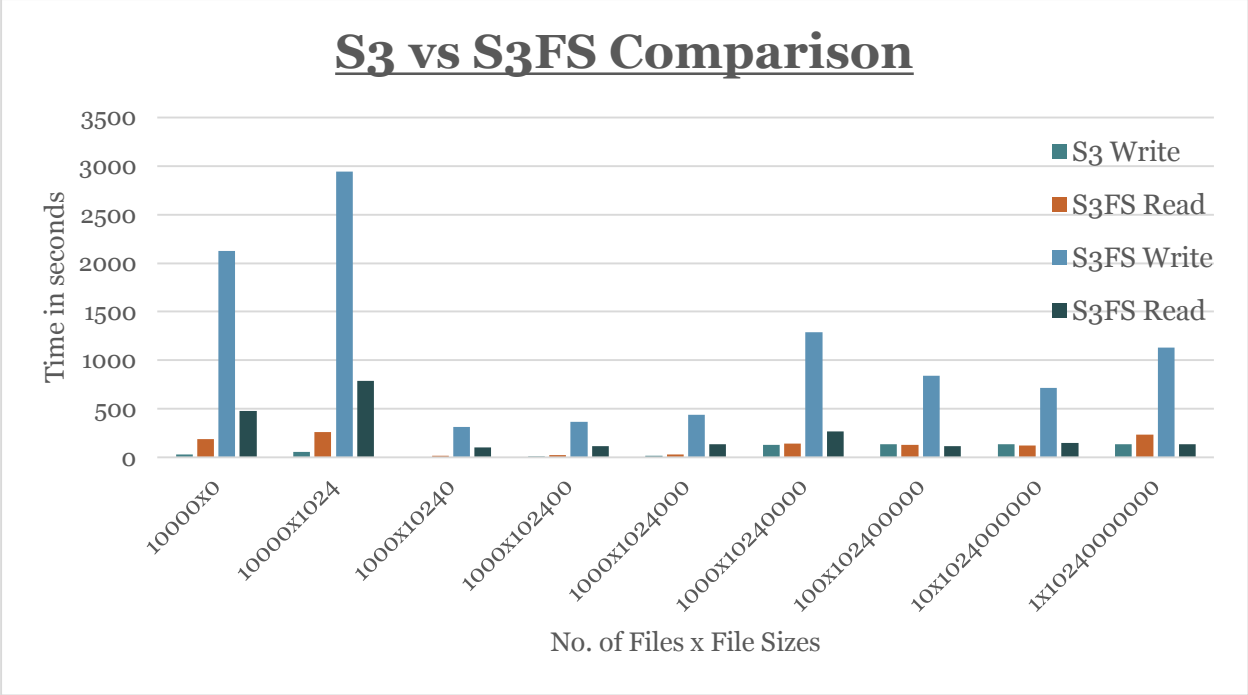
The two methods of write were evaluated and the results were as below. As the Perl method was faster, we opted to use it for all evaluations. For HDFS we used haddop fs -touchz file1 file2 file3 .... file10000

	Perl	dd
1x10GB	1m55.088s	2m4.124s
10x1GB	1m45.610s	2m7.663s
100x100MB	1m43.564s	2m0.689s
1000x10MB	1m43.566s	2m0.993s
1000x1MB	0m2.462s	0m12.188s
1000x100KB	0m1.431s	0m4.086s
1000x10KB	0m1.325s	0m3.201s
10000x1KB	0m13.147s	0m30.612s

5. S3 vs S3FS

We evaluated S3 and S3FS on a single node and found the data not comparable. S3FS was significantly slower than S3 and hence we decided to not pursue it further in the evaluation. Below is the data for comparison of S3 and S3FS.

Files per Node	Size per file	S3		S3FS	
Operation		Write	Read	Write	Read
10000	0	30.129	191.233	2127.356	476.448
10000	1024	54.533	263.495	2939.896	788.488
1000	10240	6.466	21.373	312.25	104.212
1000	102400	11.893	23.592	369.643	118.061
1000	1024000	17.459	30.909	439.747	138.381
1000	10240000	132.97	143.714	1285.966	267.105
100	102400000	135.198	129.698	839.861	120.132
10	1024000000	139.35	126.597	714.952	152.89
1	10240000000	134.146	233.083	1128.058	136.799



6. S3 Evaluation

6.1 Read / Write Evaluation

Amazon S3 was evaluated on 1, 4, 16 and 64 nodes for Read and Write operations. Below is the data for the same.

1 Node:

No. of Files per Node	Size per file	node-1	
		Write	Read
10000	0	30.129	191.233
10000	1024	54.533	263.495
1000	10240	6.466	21.373
1000	102400	11.893	23.592
1000	1024000	17.459	30.909
1000	10240000	132.97	143.714
100	102400000	135.198	129.698
10	1024000000	139.35	126.597
1	10240000000	134.146	233.083

4 Nodes:

No. of Files per Node	Size per file	node-1		node-2		node-3		node-4	
		Write	Read	Write	Read	Write	Read	Write	Read
10000	0	26.423	195.158	28.141	193.59	27.977	198.802	28.992	183.757
10000	1024	51.148	219.568	51.68	203.249	51.019	211.965	54.827	210.167
1000	10240	5.937	17.237	6.504	20.789	5.604	18.361	6.666	19.291
1000	102400	12.686	25.411	12.415	27.981	12.043	20.518	12.119	29.485
1000	1024000	17.044	86.712	16.495	86.951	17.436	85.243	18.463	90.717
1000	10240000	121.893	341.163	122.931	341.841	121.569	337.63	127.992	327.543
100	102400000	132.392	186.047	126.895	188.923	150.087	185.132	141.787	187.317
10	1024000000	129.965	126.534	126.411	126.935	125.705	126.361	156.278	127.009
1	10240000000	133.006	201.439	126.512	204.552	124.609	234.156	138.816	219.785

16 Nodes:

Size (byte)	1024.0		10240.0		102400.0		1024000.0		10240000.0		102400000.0		1024000000.0		10240000000.0	
Count	10000.0		1000.0		1000.0		1000.0		1000.0		100.0		10.0		1.0	
Nodes	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read
1.0	59.6	241.0	7.4	22.0	16.1	27.1	19.0	81.5	125.6	155.3	159.8	130.0	136.9	126.0	142.6	308.5
2.0	61.2	242.9	10.4	20.8	14.2	26.0	19.3	81.8	123.8	155.3	137.5	128.8	134.0	126.6	138.6	335.8
3.0	62.9	235.3	8.0	24.2	14.0	39.9	19.3	65.9	131.6	161.0	151.1	131.2	138.2	126.5	136.3	290.8
4.0	62.1	230.7	9.0	22.2	15.2	27.2	19.1	78.5	124.2	187.8	133.6	128.9	126.4	125.8	126.5	309.5
5.0	60.2	232.3	9.0	23.7	14.5	27.2	19.4	80.2	129.7	151.0	153.3	127.5	140.5	126.1	157.7	291.0
6.0	63.9	242.8	6.7	23.8	14.1	25.0	21.9	78.7	126.4	153.4	161.7	130.6	141.7	127.0	136.9	280.4
7.0	65.8	239.2	8.1	25.0	14.1	27.8	19.8	64.1	132.1	153.6	161.5	128.3	139.6	126.0	143.3	287.0
8.0	61.9	238.0	6.6	22.4	14.8	26.3	20.2	78.7	147.9	165.7	136.6	133.7	137.6	129.5	154.0	290.1
9.0	62.3	235.7	7.5	22.4	15.0	28.2	22.8	59.0	125.2	157.4	131.3	125.5	127.2	126.2	138.0	313.6
10.0	60.3	237.4	7.0	21.9	14.5	29.8	19.2	58.1	129.1	158.8	165.3	127.6	154.7	126.5	150.3	285.6
11.0	62.5	231.5	6.3	20.2	14.4	28.8	22.6	77.6	125.8	158.6	140.4	138.7	129.5	127.1	138.1	277.9
12.0	64.2	236.7	8.0	21.0	16.0	33.9	20.1	91.4	132.1	163.8	152.3	134.3	135.2	127.3	137.6	292.4
13.0	63.5	231.7	7.1	21.0	14.5	26.8	21.7	79.8	125.9	160.2	131.0	131.7	130.1	126.6	139.7	292.0
14.0	63.1	236.9	8.2	22.5	15.8	32.6	20.6	95.0	143.3	168.7	149.2	126.5	126.1	129.8	136.9	288.9

15.0	62.8	235.2	8.2	20.3	13.8	26.4	19.8	72.5	151.7	159.5	141.1	134.5	135.4	126.1	135.8	292.4
16.0	66.5	230.7	8.5	24.4	15.1	35.0	20.1	68.6	146.3	195.2	136.8	136.9	140.3	126.6	150.7	302.2

64 Nodes:

Size	1024 (1KB)		10240 (10 KB)		102400 (100KB)		1024000 (1MB)		10240000 (10 MB)		102400000 (100 MB)		1024000000 (1 GB)		10240000000 (10 GB)	
Count	10000.0		1000.0		1000.0		1000.0		1000.0		100.0		10.0		1.0	
Nodes	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read
1.0	71.2	249.4	5.4	22.9	14.7	22.5	19.5	32.6	141.5	187.5	132.8	128.9	132.5	126.5	132.5	458.9
2.0	63.3	257.9	5.9	28.5	15.4	21.5	18.4	32.5	145.9	277.5	138.7	301.3	130.3	126.8	155.7	292.3
3.0	70.1	255.8	6.4	27.5	15.4	24.4	19.5	31.3	143.3	165.4	136.6	170.8	136.7	126.6	136.3	275.8
4.0	62.2	258.6	7.4	27.4	16.4	26.6	21.5	32.4	146.4	198.3	153.3	143.9	136.4	127.6	145.3	294.4
5.0	61.3	247.4	3.6	29.5	17.5	27.5	22.4	33.5	156.4	146.3	161.5	263.6	137.8	126.9	159.7	270.2
6.0	64.3	246.3	4.3	29.5	16.4	26.5	23.5	33.6	162.4	198.5	143.2	132.0	148.4	126.4	137.9	283.6
7.0	58.5	253.2	5.5	27.5	17.4	24.4	22.5	32.5	183.0	156.1	143.2	131.5	158.7	127.1	158.0	322.1
8.0	74.0	257.6	6.9	21.2	15.7	23.9	19.1	34.6	124.5	143.4	131.8	129.2	123.1	126.6	130.2	312.3
9.0	62.0	243.8	5.4	25.3	17.5	25.6	19.5	32.8	132.4	320.4	135.4	187.4	133.0	126.9	132.6	342.5
10.0	67.9	237.4	7.9	26.4	16.5	26.4	18.4	33.8	129.1	339.6	138.7	175.3	137.0	127.7	136.4	267.8
11.0	66.8	248.8	8.4	27.5	17.5	27.6	19.5	33.8	132.5	333.3	145.7	255.2	152.2	128.3	134.9	277.9
12.0	70.4	251.7	8.4	28.4	15.4	28.6	19.5	32.5	156.4	195.4	153.4	126.4	132.3	125.9	133.8	265.6
13.0	67.0	250.0	8.3	25.2	15.3	21.3	20.8	37.0	156.5	150.1	144.2	128.8	139.0	126.4	138.9	318.8
14.0	72.4	251.1	7.4	29.5	16.6	25.4	21.5	32.6	142.5	176.5	140.7	132.2	136.8	127.4	135.4	281.5
15.0	61.5	243.4	6.4	22.7	16.5	27.5	22.5	33.3	141.6	198.4	137.3	132.8	128.5	255.3	140.2	283.9
16.0	71.2	266.4	9.6	28.5	15.2	21.1	20.1	41.8	125.6	159.0	137.2	135.0	137.0	129.7	137.3	262.1
17.0	68.5	264.5	7.4	26.4	16.5	22.5	21.4	38.6	126.4	151.9	139.3	129.4	156.1	126.6	133.7	292.4
18.0	64.1	248.4	6.4	29.5	17.5	23.6	18.5	33.6	125.4	178.3	136.5	288.2	129.2	130.2	132.0	427.6
19.0	64.1	241.3	7.3	29.5	15.4	24.5	19.5	32.9	124.5	153.5	134.1	161.1	130.6	129.1	156.2	292.8
20.0	63.6	272.3	4.9	28.5	16.5	29.5	20.2	91.7	130.0	227.7	138.7	149.9	134.3	132.8	139.4	266.7
21.0	65.3	256.3	8.7	24.4	16.6	30.6	20.9	128.5	145.5	128.5	140.2	131.7	145.2	126.5	156.8	292.6
22.0	62.6	239.5	7.5	22.4	14.7	28.5	21.4	34.3	139.4	265.4	132.6	137.4	134.3	132.9	137.3	295.7
23.0	73.5	256.3	9.5	21.6	14.8	24.4	20.5	36.2	138.2	146.9	128.0	128.8	131.3	127.0	125.6	299.5
24.0	64.5	247.9	6.5	29.6	15.5	21.5	22.4	38.6	132.4	236.5	142.4	133.8	134.1	126.4	131.1	291.0
25.0	63.8	246.7	10.3	28.5	15.4	23.0	21.5	28.5	150.8	154.7	159.2	134.5	139.7	126.7	139.1	264.4
26.0	65.0	252.8	8.5	26.8	16.3	24.8	21.4	38.5	144.1	146.8	132.8	132.6	136.1	126.4	134.7	292.3

27.0	66.5	255.3	11.4	29.4	17.7	27.5	23.3	39.6	132.5	219.9	129.0	129.2	128.4	126.2	125.8	290.0
28.0	58.2	249.1	12.6	28.4	15.3	22.6	21.4	32.5	132.3	210.4	139.5	129.6	138.3	126.1	141.1	272.3
29.0	68.6	239.4	7.3	27.4	15.5	21.5	20.4	33.6	142.4	287.4	142.3	314.5	151.9	128.5	150.7	269.3
30.0	68.9	242.3	9.4	28.5	16.8	25.9	19.5	38.5	132.6	256.5	135.3	130.7	134.1	126.8	150.7	300.8
31.0	59.0	240.9	9.5	29.5	15.3	28.5	18.5	33.5	136.0	293.6	141.7	132.4	133.3	126.6	135.2	271.3
32.0	63.6	260.0	9.5	27.5	16.5	28.6	18.9	31.8	142.5	310.4	135.6	128.1	137.0	127.0	139.2	268.7
33.0	61.3	256.2	10.4	26.6	17.5	29.6	20.0	38.3	132.5	323.3	151.8	127.4	140.3	126.7	135.3	290.5
34.0	63.4	255.5	7.0	30.6	16.4	23.9	17.5	35.4	134.4	303.2	146.0	159.6	138.2	126.9	140.7	282.7
35.0	67.5	240.9	7.9	29.5	16.5	27.5	19.5	38.5	137.5	298.3	140.1	301.3	135.5	127.6	131.2	292.9
36.0	60.2	233.6	9.9	26.7	17.5	28.0	19.3	32.5	136.5	265.5	132.5	145.3	130.8	314.3	132.5	325.5
37.0	71.2	263.5	11.8	32.5	16.2	28.6	20.0	33.4	139.3	287.5	206.7	133.8	211.1	128.5	142.5	321.5
38.0	60.4	251.3	10.5	31.6	16.5	29.5	18.5	39.6	132.4	198.4	142.5	130.5	136.2	126.0	142.9	272.1
39.0	66.0	271.3	9.5	32.5	17.4	29.5	19.5	31.3	135.4	153.0	143.7	127.9	159.8	126.6	139.1	269.9
40.0	64.4	254.4	9.4	29.6	16.5	27.6	19.5	34.0	131.3	310.4	146.1	255.6	145.8	129.0	140.2	269.0
41.0	63.9	238.4	8.0	29.8	17.5	28.4	20.5	37.5	133.3	354.3	136.5	142.7	135.8	129.1	135.3	287.0
42.0	61.4	253.2	7.5	31.3	15.5	29.5	21.4	39.5	130.7	339.5	148.2	180.3	137.2	126.8	142.0	384.7
43.0	60.0	238.7	6.5	27.4	15.5	28.9	18.5	36.6	142.3	342.2	133.8	315.7	135.1	127.2	130.4	292.4
44.0	61.2	245.0	8.5	28.5	15.5	29.5	19.4	33.5	143.3	334.0	143.9	127.7	143.5	127.7	137.8	270.7
45.0	65.1	234.9	9.5	26.5	17.5	28.5	19.5	31.4	132.1	151.7	143.6	126.1	145.9	126.7	136.8	272.5
46.0	63.9	243.4	7.5	24.5	16.4	29.5	19.5	28.5	133.3	187.5	148.9	127.5	152.1	126.7	136.6	253.5
47.0	61.8	256.0	9.6	28.5	17.5	29.5	17.4	27.5	137.5	198.5	144.9	287.6	135.3	126.6	137.1	290.6
48.0	70.2	250.0	9.5	22.5	16.4	27.4	18.5	39.5	125.9	337.0	130.7	177.7	149.4	126.7	124.7	289.9
49.0	66.3	256.4	8.5	21.4	15.5	26.5	20.4	33.1	129.6	146.8	139.3	127.9	137.4	126.3	153.8	296.2
50.0	68.5	263.8	9.4	28.4	16.5	27.0	21.4	33.3	129.6	154.0	147.3	127.6	136.1	126.3	147.8	302.3
51.0	59.1	257.4	9.2	29.5	17.5	27.6	19.5	32.5	142.4	310.3	135.5	131.4	138.1	126.1	144.6	292.7
52.0	56.8	246.7	9.5	31.4	16.4	29.5	19.5	32.5	132.5	324.0	139.8	129.4	134.5	126.5	135.2	269.8
53.0	68.0	252.6	8.5	27.5	15.5	26.6	18.5	38.5	131.4	145.4	139.2	129.8	140.9	126.7	135.6	303.2
54.0	72.5	235.4	7.4	28.5	17.4	22.5	17.5	33.6	128.3	149.2	158.2	126.2	136.0	126.6	148.3	265.8
55.0	70.4	257.2	6.5	30.5	16.5	25.5	21.3	33.2	131.2	149.3	139.1	128.9	138.0	126.6	137.7	314.1
56.0	66.6	267.4	8.6	33.0	17.5	27.6	19.3	37.1	142.4	157.6	139.2	130.0	136.4	126.6	136.7	271.5
57.0	65.1	231.2	8.1	21.6	16.5	28.5	22.6	34.3	141.3	187.4	148.9	229.7	136.9	126.1	137.2	289.5
58.0	70.4	235.4	9.5	31.5	16.4	29.5	21.9	33.3	142.5	176.4	140.7	253.2	157.3	126.9	136.5	318.5
59.0	70.4	253.4	6.5	32.3	17.4	27.5	19.5	35.4	158.5	314.8	150.1	161.4	137.7	126.3	140.1	278.8

60.0	62.8	246.6	12.9	26.9	17.5	26.5	18.5	34.2	133.3	198.4	131.9	130.3	126.9	127.2	127.3	303.4
61.0	60.3	247.7	11.5	24.6	17.5	29.5	20.6	33.7	123.9	147.6	135.8	129.0	137.0	127.6	146.8	275.4
62.0	66.8	263.0	12.6	27.5	16.5	29.4	17.5	32.2	128.3	147.4	141.1	128.2	139.8	126.5	136.5	291.1
63.0	60.3	249.4	5.4	29.6	15.4	27.5	18.4	32.6	129.3	321.8	132.6	128.9	137.6	126.5	154.2	458.9
64.0	73.8	248.4	9.6	30.4	16.4	25.4	19.5	37.3	130.4	333.3	146.4	277.3	141.6	126.4	138.1	290.7
Average	**65.4**	**250.5**	**8.3**	**27.8**	**16.4**	**26.6**	**20.0**	**36.8**	**137.6**	**226.0**	**142.0**	**164.1**	**139.7**	**132.1**	**139.3**	**296.6**

6.2 Metadata Evaluation

Amazon S3 was evaluated on 4, 16 and 64 nodes for metadata operations. Below is the data for the same.

4 Nodes:

No. of Files per Node	**Size per file**	node-1		node-2		node-3		node-4	
		Write	**Read**	**Write**	**Read**	**Write**	**Read**	**Write**	**Read**
10000	0	26.423	195.158	28.141	193.59	27.977	198.802	28.992	183.757

16 Nodes:

0 Bytes	
10000.0	
Write	**Read**
31.0	215.6
27.5	216.4
26.9	200.4
31.0	224.1
33.2	220.1
25.6	209.3
31.5	210.4
28.1	203.0
32.3	217.9
29.0	206.5
28.7	217.7
28.2	203.6
33.8	212.0

28.7	210.8
29.6	193.2
30.7	208.8

64 Nodes:

0 (0 KB)	
10000.0	
Write	**Read**
48.3	222.9
46.5	220.4
52.4	224.9
53.9	216.9
47.4	229.2
89.2	275.7
53.9	264.8
91.1	288.6
42.1	239.7
46.0	227.4
43.1	222.8
50.1	249.3
70.9	251.7
37.3	224.4
47.3	216.6
55.4	307.2
54.9	227.9
46.6	233.8
74.4	228.6
45.2	228.6
62.5	225.9
48.7	222.9
78.9	238.7
105.2	238.0
48.0	241.7

46.0	244.2
37.9	238.5
36.9	228.1
45.1	219.8
85.9	222.5
40.0	245.1
35.8	235.8
32.6	233.8
48.3	239.6
49.9	223.4
44.5	220.8
53.9	305.7
38.9	231.5
103.5	239.7
42.6	220.8
35.5	222.7
49.0	239.5
47.6	233.2
44.2	238.5
104.1	243.8
34.4	235.3
48.0	227.2
45.2	215.8
58.7	241.3
75.4	234.6
35.5	224.3
40.8	232.9
45.8	229.2
65.1	234.6
56.0	227.5
58.4	230.4
38.1	218.9
43.4	224.0

47.3	221.2
36.0	227.8
64.7	254.1
64.8	220.4
38.2	222.9
47.8	225.6

7. CephFS Evaluation

7.1 Read / Write Evaluation

CephFS was evaluated on 4, 16 and 64 nodes for Read and Write operations. Below is the data for the same.

4 Nodes:

No. of Files per Node	Size per file in bytes	node-1		node-2		node-3		node-4	
		Write	Read	Write	Read	Write	Read	Write	Read
10000	0	32.156	30.338	32.041	32.404	32.097	31.342	32.314	33.893
10000	1024	44.836	34.647	42.822	35.859	43.801	35.98	45.585	37.658
1000	10240	4.756	3.269	4.736	3.665	4.582	3.235	4.578	3.203
1000	102400	5.036	3.341	3.799	3.306	4.959	3.423	5.173	3.272
1000	1024000	16.975	32.272	14.184	31.49	14.883	32.271	16.685	36.054
1000	10240000	382.861	217.889	266.884	259.457	366.754	241.676	368.559	244.021
100	102400000	310.661	323.723	270.221	319.888	316.947	322.672	329.216	328.034
10	1024000000	300.064	394.704	317.364	380.637	299.299	403.587	286.129	388.145
1	10240000000	330.356	299.211	343.261	292.477	333.934	285.257	301.753	289.583

16 Nodes:

Size (byte)	1kb		10kb		100kb		1mb		10mb		100mb		1gb		10gb	
Count	10000.0		1000.0		1000.0		1000.0		1000.0		100.0		10.0		1.0	
Nodes	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read

1.0	122.2	49.7	8.8	2.0	9.8	3.2	47.1	23.4	387.5	357.7	371.1	340.6	310.8	475.2	366.8	386.8
2.0	110.9	50.5	9.0	1.8	9.2	4.3	46.1	26.1	421.1	353.9	278.7	352.4	316.7	473.3	374.8	396.4
3.0	120.9	49.7	8.8	1.9	11.1	3.6	47.9	26.4	421.1	353.0	378.2	348.2	349.9	486.9	325.7	392.7
4.0	121.7	52.1	8.9	2.3	10.9	3.7	46.5	26.9	406.0	358.2	356.3	342.4	308.8	473.9	338.7	403.3
5.0	120.0	50.5	8.2	2.0	9.9	3.8	46.3	23.6	405.3	360.0	377.9	346.2	348.6	477.1	362.5	395.8
6.0	117.1	48.6	8.5	1.9	10.1	3.2	46.7	25.6	387.4	359.0	378.8	349.1	344.1	466.5	372.5	389.2
7.0	125.5	48.0	8.4	1.9	9.3	3.0	46.3	23.3	386.2	361.8	356.5	336.4	344.0	448.2	370.2	371.6
8.0	109.7	49.8	8.6	2.4	8.6	3.8	17.8	22.6	295.9	362.3	368.0	336.9	341.6	453.6	370.3	373.5
9.0	124.1	48.9	9.9	1.9	9.8	3.7	44.6	24.4	410.6	361.2	377.5	355.4	316.6	486.0	371.1	399.1
10.0	116.5	49.2	8.9	1.8	10.0	3.2	46.4	25.1	406.4	360.6	356.3	374.6	340.6	471.2	370.4	388.9
11.0	122.7	49.6	7.9	2.2	10.0	5.2	46.0	24.4	387.8	360.4	377.8	344.1	313.0	476.2	373.7	380.4
12.0	115.8	47.0	8.6	1.9	9.5	3.6	24.8	28.0	409.8	358.5	374.7	345.1	314.2	463.8	372.8	388.0
13.0	114.9	48.6	8.1	1.9	9.7	3.5	46.2	25.4	407.7	351.3	379.2	346.7	344.5	433.5	356.5	384.9
14.0	124.7	49.7	9.7	1.9	9.6	3.6	46.4	23.9	405.7	352.1	378.4	356.3	321.7	476.2	369.1	399.4
15.0	121.2	50.1	9.2	1.9	10.0	3.2	45.8	24.4	405.7	351.3	381.3	342.8	343.8	456.9	321.8	410.3
16.0	122.1	50.1	8.6	1.8	9.5	3.4	46.9	24.6	417.6	361.5	355.5	348.5	309.8	484.0	373.6	369.4
Average (in sec.)	**119.4**	**49.5**	**8.8**	**2.0**	**9.8**	**3.6**	**43.2**	**24.9**	**397.6**	**357.7**	**365.4**	**347.9**	**329.3**	**468.9**	**361.9**	**389.4**

64 Nodes:

Size	1kb		10kb		100kb		1mb		10mb		100mb		1gb		10gb	
Count	10000.0		1000.0		1000.0		1000.0		1000.0		100.0		10.0		1.0	
Nodes	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read
1.0	730.2	258.2	47.5	14.0	62.6	14.6	53.0	30.3	503.8	248.9	384.0	357.9	340.0	465.6	327.2	436.8
2.0	729.5	255.7	44.9	11.8	63.3	13.1	67.1	31.8	504.3	250.1	386.1	351.2	391.0	461.7	375.6	396.6
3.0	721.4	258.5	50.8	14.1	61.9	12.7	50.9	31.9	504.7	253.2	388.7	354.5	360.4	446.9	329.0	415.8
4.0	722.5	253.6	49.1	14.7	62.1	14.5	50.2	29.1	507.2	221.3	383.5	354.1	398.9	438.8	333.4	429.3
5.0	729.5	259.4	47.6	14.2	60.9	8.7	81.8	29.5	503.7	238.8	383.4	352.6	369.1	470.0	405.3	426.3
6.0	738.1	257.2	51.7	19.1	60.8	11.7	77.2	27.7	506.2	257.2	382.1	356.4	364.6	437.3	404.1	399.3
7.0	726.2	259.0	56.0	7.5	59.0	15.4	53.5	27.0	489.8	227.9	383.5	354.7	385.1	448.1	297.5	423.5
8.0	731.5	256.4	49.2	13.1	62.9	15.4	59.2	30.3	485.3	212.1	358.8	345.6	364.4	425.7	367.9	420.4
9.0	740.7	253.9	48.1	14.2	60.7	9.0	64.8	29.4	505.1	228.5	359.0	350.4	361.3	433.7	319.5	424.2
10.0	740.0	258.6	54.4	7.1	60.9	8.1	75.7	25.7	507.4	227.6	381.3	340.8	388.6	430.8	378.8	402.7
11.0	733.4	258.6	54.6	6.7	63.3	13.5	63.4	26.2	477.9	253.9	383.0	351.0	392.2	479.6	331.7	424.7

12.0	724.4	251.6	54.3	5.9	63.2	9.9	73.5	25.8	489.9	239.2	383.0	350.2	378.8	449.7	332.0	422.2
13.0	739.9	253.6	49.2	11.8	59.1	16.1	84.4	28.4	493.3	231.1	382.9	347.5	367.6	457.5	344.5	417.6
14.0	741.5	257.1	45.5	13.9	60.7	10.1	64.7	27.5	494.4	231.7	385.9	364.9	369.6	436.0	325.1	420.8
15.0	724.4	257.2	48.7	13.9	62.1	12.5	53.7	27.2	507.2	260.3	388.2	361.9	414.7	470.4	360.3	401.8
16.0	730.5	256.4	48.4	14.2	63.1	16.7	65.4	33.1	507.4	257.4	386.0	360.8	370.7	438.2	360.8	421.2
17.0	710.3	258.1	48.6	14.7	49.8	8.0	83.7	26.0	503.5	243.7	382.6	367.1	361.6	453.8	398.9	418.5
18.0	730.4	258.3	53.8	7.6	59.6	15.4	81.3	30.6	505.4	245.8	385.1	364.0	379.9	434.4	403.4	429.9
19.0	724.9	257.9	49.2	11.7	60.8	10.1	76.9	28.6	493.4	238.2	383.9	346.9	367.4	445.0	323.8	433.9
20.0	726.8	256.3	46.5	13.9	61.4	13.0	56.2	29.3	461.4	262.6	386.8	361.2	385.1	430.7	424.9	419.6
21.0	730.5	258.2	55.2	7.6	62.6	14.7	63.2	28.2	507.1	238.6	380.4	353.6	343.6	465.2	377.9	431.3
22.0	708.1	257.2	27.8	15.7	64.0	9.0	49.9	29.8	494.0	245.6	382.6	364.6	391.3	477.1	405.5	432.0
23.0	720.6	257.2	48.9	16.0	63.1	8.2	62.6	25.3	507.9	231.7	384.1	362.5	369.0	452.6	314.1	426.4
24.0	706.4	258.8	54.9	7.0	59.5	9.6	49.1	29.7	493.0	257.4	388.4	368.8	376.0	462.4	353.2	399.8
25.0	720.5	258.7	49.5	13.2	61.3	14.5	49.4	29.2	500.9	247.9	381.5	356.1	385.1	473.5	328.5	422.4
26.0	725.3	251.5	46.2	6.7	63.4	13.8	65.9	27.1	476.8	236.0	383.6	354.6	356.1	442.5	315.2	418.8
27.0	695.2	255.9	22.3	13.0	59.4	13.0	63.3	30.4	488.2	245.7	383.6	357.4	382.3	442.2	396.4	407.6
28.0	731.9	259.6	48.8	15.8	63.7	16.8	47.2	28.7	508.5	240.8	383.8	347.7	386.8	451.9	333.2	415.8
29.0	720.4	257.3	48.6	13.8	61.6	14.1	66.0	27.6	489.4	245.5	382.2	358.2	385.7	457.8	381.5	428.0
30.0	726.1	258.1	49.1	14.0	63.0	14.0	58.7	29.7	480.4	250.5	382.8	341.3	363.5	453.5	341.5	434.7
31.0	726.6	258.5	45.2	11.4	62.2	14.7	78.0	31.6	508.3	276.5	399.6	371.4	384.3	438.5	356.3	414.1
32.0	732.4	254.5	55.4	12.1	63.0	16.9	64.2	31.4	489.6	258.0	384.1	354.6	311.9	461.4	305.4	431.5
33.0	731.0	257.6	54.6	6.6	63.9	14.5	55.5	26.6	508.5	225.3	384.3	351.3	342.6	479.0	393.1	423.2
34.0	729.6	255.1	49.2	16.3	63.1	13.5	75.2	31.0	489.1	238.0	379.4	345.9	366.2	484.1	391.3	423.2
35.0	742.2	258.4	52.1	6.4	63.3	9.8	48.5	27.7	502.1	237.6	359.4	361.8	368.6	459.1	328.1	424.6
36.0	732.6	254.4	49.2	15.6	61.4	8.7	65.7	28.6	490.2	247.5	382.1	353.5	364.5	446.2	336.2	418.7
37.0	726.6	249.9	53.3	11.9	61.2	8.4	65.2	27.5	502.3	254.9	387.7	360.4	356.0	444.7	379.6	423.1
38.0	741.3	249.1	52.9	6.5	59.3	13.3	84.0	27.9	503.4	240.6	383.6	350.5	361.3	450.1	334.6	400.1
39.0	721.1	250.3	54.8	13.9	59.6	10.3	78.5	31.6	494.1	252.0	385.2	351.7	364.6	460.7	430.9	331.5
40.0	706.4	257.6	45.3	11.4	63.0	8.1	49.1	28.3	491.2	229.0	382.6	353.9	343.1	457.6	434.6	343.1
41.0	726.7	257.5	32.4	11.0	60.1	13.9	78.1	32.8	501.5	241.5	382.5	349.8	341.5	466.2	323.2	441.8
42.0	729.9	259.2	48.5	14.3	60.8	10.3	78.7	29.5	479.1	228.4	382.9	352.6	383.6	425.9	319.9	408.0
43.0	731.7	259.8	55.3	7.4	61.0	14.8	67.4	27.4	504.5	254.6	385.4	350.4	369.5	455.7	347.5	410.7
44.0	725.2	250.0	49.1	14.1	63.5	14.1	84.2	31.6	503.5	255.7	385.9	352.9	407.9	427.7	345.7	424.4

45.0	720.3	256.9	50.7	13.0	62.4	14.8	53.8	29.0	509.5	258.3	385.0	358.7	388.6	465.6	340.2	431.0
46.0	722.3	256.6	49.3	16.5	61.1	8.8	82.9	28.7	478.7	248.8	384.3	344.2	357.8	470.9	324.4	422.9
47.0	719.2	258.8	49.2	6.6	63.7	16.8	75.3	29.4	491.8	249.5	384.6	342.6	401.2	441.2	417.4	438.2
48.0	724.3	257.2	48.0	6.0	61.2	12.9	47.1	29.8	505.9	262.1	387.8	355.5	381.1	453.5	361.8	422.9
49.0	727.8	253.6	54.4	11.2	63.3	16.3	50.3	32.4	504.8	248.5	384.2	359.1	395.9	476.0	392.2	422.8
50.0	732.4	255.2	48.2	13.7	61.1	13.5	56.4	29.0	506.3	222.5	380.9	346.0	381.9	467.1	373.8	409.7
51.0	721.2	258.4	46.4	14.1	63.5	16.1	56.0	29.8	489.5	257.8	384.2	359.9	406.7	457.8	341.3	389.5
52.0	720.6	256.6	45.6	12.6	62.8	14.0	78.0	30.3	490.2	240.0	383.1	349.1	339.9	434.9	317.4	417.5
53.0	727.3	257.5	31.4	14.2	61.2	13.0	82.1	25.5	490.6	239.4	381.0	341.3	368.5	463.0	391.1	414.2
54.0	705.4	257.4	54.2	6.9	62.6	14.3	54.0	24.6	488.2	222.9	382.9	353.8	342.5	450.1	331.9	421.9
55.0	730.6	250.1	55.5	6.5	61.0	13.6	49.0	30.2	505.2	237.1	383.4	356.6	370.3	435.1	387.8	419.2
56.0	731.2	258.8	48.9	7.4	62.4	15.6	64.3	31.5	495.4	266.9	387.5	370.1	397.9	465.9	423.9	402.6
57.0	730.5	259.3	52.1	6.2	63.0	14.3	55.5	27.7	505.7	248.3	384.4	359.9	362.2	460.9	326.3	424.7
58.0	715.4	255.0	51.3	16.8	59.7	15.9	77.8	28.9	504.6	262.6	391.2	354.9	368.7	440.1	363.5	422.6
59.0	712.2	258.8	51.0	14.6	62.5	8.3	82.4	28.6	486.4	252.7	383.5	348.8	368.4	459.7	335.3	404.5
60.0	701.5	246.3	46.6	12.9	62.0	14.3	75.0	30.4	491.3	254.7	386.2	352.8	357.2	455.8	320.0	416.4
61.0	737.8	258.5	45.2	11.7	60.8	14.1	64.2	29.6	490.5	252.5	383.2	366.7	369.4	471.3	318.5	431.8
62.0	723.8	258.2	51.4	13.8	64.2	13.4	67.1	31.9	479.9	251.3	383.5	355.7	356.2	476.5	325.9	428.1
63.0	692.5	252.5	55.8	7.6	50.1	15.0	74.9	29.6	512.6	295.5	385.5	377.7	341.5	465.2	311.3	443.6
64.0	729.4	256.3	49.3	14.6	61.0	11.0	69.0	33.0	508.1	261.2	387.4	357.0	402.4	464.2	409.2	407.7
Average	**725.2**	**256.3**	**48.9**	**11.8**	**61.5**	**12.9**	**65.6**	**29.1**	**496.9**	**246.0**	**383.3**	**355.2**	**371.5**	**454.1**	**356.7**	**417.4**

7.2 Metadata Evaluation

CephFS was evaluated on 4, 16 and 64 nodes for metadata operations. Below is the data for the same.

4 Nodes:

No. of Files per Node	Size per file	node-1		node-2		node-3		node-4	
		Write	Read	Write	Read	Write	Read	Write	Read
10000	0	32.156	30.338	32.041	32.404	32.097	31.342	32.314	33.893

Instances #	node-1	node-2	node-3	node-4
1	0m6.985s	-	-	-
2	0m10.570s	0m10.552s	-	-

4	0m17.575s	0m16.151s	0m17.188s	0m19.086s

16 Nodes:

metadata	0kb	
0 byte	10000	
Touch	Write	Read
6.369	104.903	106.978
5.508	109.528	107.541
4.911	111.655	104.374
5.312	104.708	104.89
7.196	113.08	104.833
5.131	112.284	104.422
4.751	109.866	103.768
4.76	95.634	102.57
4.905	112.209	104.555
6.317	108.739	108.597
6.202	103.569	103.444
5.489	102.595	103.989
6.353	104.357	105.08
3.696	104.695	107.633
5.008	108.709	103.856
5.646	113.344	107.407
5.472	107.492	105.246

64 Nodes:

metadata	0kb	
0 byte	10000	
Touch	Write	Read
31.638	615.045	755.597
32.546	635.641	752.995
23.729	636.255	771.716
25.168	638.997	755.125
33.014	639.718	748.184

32.13	639.413	749.581
29.254	640.195	757.092
34.055	641.141	752.439
32.561	644.186	749.011
32.472	645.417	769.078
24.678	644.351	755.873
32.134	648.171	751.518
32.105	648.2	756.014
29.362	649.124	734.145
32.712	648.4	747.665
29.026	649.639	747.482
32.412	649.958	748.258
30.63	651.524	752.266
20.71	652.398	711.942
33.732	651.064	748.685
28.39	651.346	752.192
32.482	652.626	770.906
30.585	650.448	749.239
26.367	652.863	756.232
28.242	651.981	752.665
31.558	652.292	755.462
32.776	652.319	739.464
31.542	652.538	756.324
32.486	651.322	755.224
31.258	653.691	758.724
31.396	652.733	755.362
29.798	654.485	756.931
25.099	658.598	757.246
33.188	660.73	756.297
28.856	660.383	756.624
16.116	659.933	753.304
31.724	658.756	756.944
31.65	660.734	752.451

32.853	658.782	756.289
28.643	657.865	747.589
29.6	661.514	756.308
15.327	661.404	748.07
29.465	661.615	748.586
32.344	658.232	737.914
31.744	659.316	755.444
31.637	658.815	756.512
31.898	660.396	749.204
32.428	657.042	761.029
32.69	658.155	756.14
29.008	656.686	756.176
32.362	659.341	760.166
31.964	657.452	757.046
32.539	657.15	821.278
31.878	658.867	756.426
30.729	657.277	759.611
31.242	657.728	756.515
34.689	658.894	755.928
33.097	660.379	734.745
33.323	660.287	751.787
20.466	660.178	756.889
31.495	660.316	756.155
19.933	662.009	757.244
32.689	660.773	755.683
33.844	661.22	755.224
30.0854	653.0986	754.2221

8. HDFS Evaluation

8.1 Read / Write Evaluation

HDFS was evaluated on 4, 16 and 64 nodes for Read and Write operations. Below is the data for the same.

4 Nodes:

Size	1024 (1KB)		10240 (10 KB)		102400 (100KB)		1024000 (1MB)		10240000 (10 MB)		102400000 (100 MB)		1024000000 (1 GB)		10240000000 (10 GB)	
Count	10000		1000		1000		1000		1000		100		10		1	
Nodes	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read
1	133	22	17	5	15	7	21	10	141	193	113	162	130	182	135	189
2	118	22	17	5	19	6	22	15	138	196	114	163	129	181	124	178
3	125	22	18	6	19	6	35	12	139	192	113	161	126	184	133	187
4	162	23	17	6	26	7	31	11	147	193	117	161	149	202	138	192

16 Nodes:

Size	1024 (1KB)		10240 (10 KB)		102400 (100KB)		1024000 (1MB)		10240000 (10 MB)		102400000 (100 MB)		1024000000 (1 GB)		10240000000 (10 GB)	
Count	10000		1000		1000		1000		1000		100		10		1	
Nodes	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read
1	351	28	20	6	30	6	25	10	121	166	114	165	117	166	116	164
2	369	27	18	6	20	7	27	10	142	169	130	168	123	171	126	167
3	304	27	26	7	19	7	41	10	131	161	115	166	116	173	118	164
4	321	28	25	6	28	7	37	9	154	169	116	165	120	167	122	173
5	264	28	24	6	19	7	35	9	121	165	114	163	114	169	115	168
6	364	28	23	7	29	7	28	10	123	168	116	169	116	170	122	168
7	424	29	28	7	30	7	46	9	159	162	116	168	117	166	124	173
8	258	27	17	6	19	6	25	10	122	169	115	166	115	164	121	171
9	280	27	18	6	19	6	40	10	168	164	117	166	119	172	118	166
10	297	28	25	6	19	7	27	10	122	167	117	168	106	170	115	165
11	379	28	19	6	31	6	42	9	122	169	117	168	117	170	121	170
12	373	28	22	6	27	7	41	10	122	168	120	168	118	171	119	170
13	391	26	18	6	29	7	24	10	122	168	118	167	115	168	120	171
14	300	27	18	6	19	6	33	10	125	166	117	171	117	167	119	170

15	444	25	33	7	40	8	53	15	148	170	139	227	426	212	152	230
16	348	27	19	6	28	6	24	10	130	169	118	171	117	168	122	171

64 Nodes:

Size	1024 (1KB)		10240 (10 KB)		102400 (100KB)		1024000 (1MB)		10240000 (10 MB)		102400000 (100 MB)		1024000000 (1 GB)		10240000000 (10 GB)	
Count	10000		1000		1000		1000		1000		100		10		1	
Nodes	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read
1	458	7	102	13	25	10	71	23	278	205	133	116	120	166	226	231
2	301	8	93	14	24	12	58	9	273	186	184	169	116	164	128	281
3	466	8	84	14	25	11	56	10	270	207	187	167	119	169	235	245
4	283	8	93	15	51	9	36	9	275	230	132	116	118	165	131	263
5	434	8	122	14	53	9	54	9	283	166	187	168	118	168	235	243
6	480	8	95	14	52	8	37	9	279	224	133	121	117	166	249	283
7	541	10	87	13	55	10	59	10	292	237	134	184	119	167	130	234
8	294	8	118	13	50	8	54	9	285	234	133	126	118	166	224	294
9	496	8	122	14	52	11	52	9	276	239	134	114	116	169	244	264
10	441	8	99	14	15	11	36	10	234	183	135	121	117	166	147	287
11	467	7	115	13	22	11	37	10	283	207	134	108	119	166	245	265
12	468	7	125	14	53	7	50	9	257	227	192	169	138	164	238	245
13	308	8	82	14	54	11	58	9	257	229	132	114	116	217	222	294
14	484	8	91	13	22	12	58	9	231	239	197	170	119	166	243	246
15	284	8	125	14	11	11	38	10	261	232	132	111	117	160	241	243
16	267	8	124	14	16	11	36	9	280	231	134	116	116	168	247	276
17	465	9	91	14	52	8	37	9	290	224	204	120	117	167	226	274
18	467	7	126	13	53	8	36	10	262	219	134	180	121	163	238	274
19	289	7	91	14	51	8	59	9	280	187	185	121	115	165	245	283
20	291	8	98	15	44	11	36	10	274	247	187	162	117	162	250	263
21	293	8	114	13	10	11	54	9	228	200	185	171	121	251	131	257
22	293	8	93	13	10	12	54	10	271	210	132	178	117	169	131	275
23	294	8	99	14	54	6	54	9	277	225	190	119	120	165	139	293
24	350	8	122	15	12	12	41	10	286	222	199	166	120	171	131	243

25	342	7	124	15	53	9	37	9	285	276	133	174	131	167	131	267
26	287	9	93	14	54	9	40	9	295	227	135	111	119	163	116	285
27	302	8	97	14	24	11	37	10	266	199	174	115	118	166	223	264
28	321	8	123	14	20	11	36	9	278	195	134	153	119	172	245	274
29	342	7	91	13	51	8	38	9	291	228	134	112	117	170	131	286
30	470	8	122	13	35	11	39	10	282	230	134	118	117	172	232	294
31	387	8	96	15	27	11	59	8	298	101	132	120	116	166	240	263
32	375	7	124	14	45	8	55	9	296	244	133	111	121	169	132	266
33	395	7	96	14	51	8	38	9	203	216	185	118	119	167	131	271
34	376	7	122	14	53	10	59	9	281	241	134	173	119	173	126	278
35	368	8	123	13	54	9	49	10	280	239	133	111	121	166	131	277
36	410	6	95	13	53	9	53	8	296	192	196	114	118	185	244	263
37	411	9	96	13	50	8	37	9	150	195	132	179	117	165	129	269
38	482	7	127	14	25	11	37	10	283	213	150	114	117	163	243	264
39	398	8	90	14	52	9	59	10	152	228	133	136	115	164	131	265
40	397	7	120	13	53	9	55	9	272	232	189	109	116	164	240	254
41	367	7	84	14	52	8	55	9	276	195	138	162	115	163	132	273
42	512	7	96	14	52	8	58	9	151	210	132	111	116	224	241	232
43	420	7	118	15	53	8	39	10	280	226	134	114	118	166	231	264
44	444	7	90	13	51	8	53	9	282	184	133	122	119	169	130	283
45	397	9	86	13	53	8	38	9	141	214	213	113	124	171	240	264
46	298	8	100	13	27	12	31	9	306	236	134	186	115	168	251	253
47	367	7	94	14	24	11	37	9	244	215	131	113	117	165	130	274
48	438	7	133	14	52	8	37	9	144	173	133	113	119	169	132	262
49	475	7	76	13	25	11	53	9	279	221	133	114	117	173	132	291
50	498	7	96	13	53	9	37	9	290	187	133	120	116	164	236	273
51	465	8	127	13	32	10	51	9	283	160	186	116	125	165	245	263
52	320	7	125	14	52	10	53	10	292	226	131	178	117	165	241	238
53	543	6	96	14	29	11	38	9	291	233	134	112	118	169	248	265
54	378	9	114	14	53	8	56	9	238	207	133	116	116	167	241	273
55	392	9	95	14	40	11	54	9	281	179	133	114	117	168	243	284
56	387	9	119	14	56	8	54	9	271	231	134	113	118	191	231	274
57	375	9	118	15	32	11	52	9	142	188	203	109	119	172	233	273

58	398	7	126	13	24	8	34	10	273	231	134	180	124	164	245	273
59	436	7	97	13	13	12	54	9	285	230	133	117	115	167	247	298
60	457	7	126	15	52	11	37	10	277	244	185	117	117	197	224	274
61	398	7	120	14	38	11	37	9	263	223	133	171	118	168	130	273
62	384	8	97	14	26	8	36	9	155	223	133	115	119	195	245	289
63	290	8	122	13	49	12	53	9	249	229	131	116	119	163	247	263
64	520	8	124	13	53	11	57	9	283	240	175	119	121	175	243	266

8.2 Metadata Evaluation

HDFS was evaluated on 4, 16 and 64 nodes for metadata operations. Below is the data for the same.

4 Nodes:

0 (0 KB)	
10000	
Write	**Read**
73	16
74	15
73	16
75	16

Instances #	**node-1**	**node-2**	**node-3**	**node-4**
1	6.664	-	-	-
2	6.662	6.662	-	-
4	7.434	7.438	7.182	10.336

16 Nodes:

0 (0 KB)	
10000	
Write	**Read**
103	22
102	21
104	21
96	21

103	21
101	22
103	21
99	21
96	21
105	21
104	22
101	21
96	20
100	22
82	22
102	21

64 Nodes:

0 (0 KB)	
10000	
Write	**Read**
344	36
350	36
350	39
351	37
349	39
341	35
349	35
343	37
343	37
351	39
343	37
341	38
311	39
343	38
348	39
351	39

337	38
345	39
343	37
342	36
338	39
344	35
351	37
351	39
351	35
340	37
335	38
336	38
341	39
337	36
349	37
342	38
348	39
350	36
342	35
352	37
343	38
349	39
343	39
349	39
344	37
339	38
338	36
348	38
335	37
349	37
345	39
342	36
344	38

333	36
345	35
351	37
343	39
339	38
349	39
350	37
345	36
344	38
340	39
348	38
350	39
348	39
350	36
341	38

9. FusionFS Evaluation

9.1 Read / Write Evaluation

FusionFS was evaluated on 4, 16 and 64 nodes for Read and Write operations. Below is the data for the same.

4 Nodes:

Size (byte)	1kb		10kb		100kb		1mb		10mb		100mb		1gb		10gb	
Count	10000		1000		1000		1000		1000		100		10		1	
Nodes	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read
1.0	52.7	26.3	4.591	2.555	6.584	3.312	26.521	12.393	228.256	172.416	216.395	171.528	211.332	193.424	220.553	193.594
2.0	53.8	26.5	4.401	2.533	6.772	3.268	24.685	16.89	221.758	172.471	216.395	168.365	211.161	193.26	218.674	193.88
3.0	51.1	26.1	4.644	2.672	6.462	3.223	24.371	14.171	223.367	172.762	220.933	172.729	213.76	193.339	221.286	192.12
4.0	51.0	26.2	4.601	2.713	6.611	3.354	26.389	11.515	229.09	172.516	224.631	173.303	215.537	193.592	220.899	193.96

16 Nodes:

Size (byte)	1kb		10kb		100kb		1mb		10mb		100mb		1gb		10gb	
Count	10000.0		1000.0		1000.0		1000.0		1000.0		100.0		10.0		1.0	
Nodes	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read
1.0	51.3	26.3	4.8	2.6	6.7	3.1	25.6	11.7	223.7	173.2	218.4	179.3	227.2	199.3	223.4	194.1
2.0	53.0	27.8	4.8	2.6	6.6	3.2	25.6	12.1	232.1	172.2	220.0	177.9	237.9	197.1	225.7	194.4
3.0	51.3	25.9	5.0	2.6	6.9	3.2	26.0	12.5	226.0	173.7	223.0	173.0	233.4	199.4	220.3	193.1
4.0	51.7	26.0	4.7	2.6	6.6	3.2	25.7	11.5	225.7	170.6	219.9	173.2	233.1	194.5	221.8	193.4
5.0	52.7	26.3	5.3	2.7	7.0	3.2	26.0	13.4	238.0	170.6	229.1	165.2	229.2	197.3	222.3	192.7
6.0	53.8	26.5	4.9	2.7	6.6	3.1	25.4	13.6	225.0	172.0	222.4	178.3	234.9	197.9	217.2	192.4
7.0	51.1	26.1	5.0	2.6	6.8	3.1	25.4	11.8	225.8	172.0	218.4	178.3	235.1	193.6	221.6	194.2
8.0	51.0	26.2	4.8	2.6	7.0	3.1	25.4	11.8	224.1	174.2	221.7	178.3	234.8	197.7	227.6	193.9
9.0	53.6	26.8	5.2	2.6	6.9	3.2	25.8	11.9	224.4	169.5	218.2	177.2	220.2	194.2	220.2	194.4
10.0	50.9	25.0	4.9	2.6	6.8	3.2	25.5	11.5	225.9	173.6	221.5	177.4	230.7	199.6	219.4	194.5
11.0	52.9	26.4	5.0	2.9	6.9	3.3	24.9	13.7	225.9	170.6	222.6	178.4	227.5	195.8	220.9	191.7
12.0	48.8	26.0	4.8	2.6	6.7	3.2	25.3	11.8	226.8	170.6	219.8	177.8	228.6	197.8	220.8	196.8
13.0	54.9	25.6	5.3	2.7	7.2	3.5	25.7	11.5	222.7	173.3	223.3	179.8	232.4	202.3	218.3	193.0
14.0	52.9	26.8	4.9	2.6	6.8	3.2	26.0	12.0	222.7	175.4	221.6	178.8	232.2	195.3	225.8	193.8
15.0	54.5	24.3	5.2	2.7	7.3	3.3	25.9	11.8	238.3	172.4	225.5	178.6	228.1	194.3	219.1	193.0
16.0	54.3	26.9	5.2	2.6	7.3	3.1	25.4	12.3	226.8	174.4	225.9	175.7	227.7	195.6	220.7	193.3

64 Nodes:

Size	1kb		10kb		100kb		1mb		10mb		100mb		1gb		10gb	
Count	10000.0		1000.0		1000.0		1000.0		1000.0		100.0		10.0		1.0	
Nodes	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read	Write	Read
1.0	49.3	25.1	5.7	2.8	6.9	3.2	23.8	13.6	238.1	174.6	221.4	172.1	221.1	197.7	221.5	193.0
2.0	51.4	25.7	5.5	2.7	7.1	3.3	26.7	11.5	216.9	162.9	220.4	171.4	222.6	198.5	223.4	190.9
3.0	48.2	25.7	5.2	2.6	6.7	3.3	25.7	13.5	217.7	162.2	206.1	160.5	217.2	198.1	219.3	191.2
4.0	50.9	26.0	5.6	2.7	7.0	3.7	25.5	13.6	212.0	161.5	220.2	171.6	197.2	177.5	218.3	188.0
5.0	48.6	26.1	5.1	2.6	6.8	3.2	25.3	11.5	217.5	165.5	223.2	173.1	219.2	198.4	220.3	193.4
6.0	48.7	26.2	5.1	2.6	6.8	3.2	26.6	11.6	213.5	162.9	221.9	168.6	221.0	198.6	219.4	193.2
7.0	49.3	26.6	5.7	2.6	6.9	3.2	25.4	13.6	226.6	173.5	219.8	172.9	218.8	197.3	219.2	193.5

8.0	52.9	27.4	5.3	2.7	7.2	3.3	27.2	12.5	222.5	173.9	223.7	173.3	216.7	197.0	217.8	194.0
9.0	50.6	26.1	5.1	2.7	6.8	3.6	25.4	12.5	223.3	173.8	221.8	170.7	219.6	197.5	223.7	188.4
10.0	45.8	25.8	4.9	2.6	6.7	3.2	25.5	11.8	223.0	174.5	220.7	170.5	215.8	197.8	217.9	193.2
11.0	51.1	26.7	5.2	2.7	7.0	3.3	26.3	11.5	214.7	157.8	220.6	170.2	221.2	197.0	219.6	197.7
12.0	51.1	25.5	5.8	2.7	7.2	3.4	26.6	12.3	211.6	161.6	217.8	169.2	218.9	197.7	217.9	194.1
13.0	50.5	26.1	5.1	2.7	7.0	3.2	25.7	13.7	214.3	163.3	221.8	172.8	217.0	198.4	222.1	193.9
14.0	51.3	26.4	5.5	2.6	7.0	3.2	26.3	11.7	216.1	162.8	219.0	171.7	217.7	198.5	217.7	193.1
15.0	51.7	26.6	5.2	2.6	6.8	3.2	26.0	11.5	211.9	163.2	222.1	171.7	216.1	194.7	218.8	192.8
16.0	50.5	26.5	5.7	2.7	6.8	3.3	26.1	11.8	222.9	172.4	200.4	158.4	217.9	192.6	221.9	194.4
17.0	47.0	27.3	5.2	2.6	6.8	3.1	26.1	11.8	216.0	169.1	208.3	160.8	195.4	169.7	217.7	193.0
18.0	52.0	26.6	4.9	2.6	6.6	3.1	25.4	13.5	216.1	163.7	220.4	170.6	196.5	175.9	220.5	194.5
19.0	47.5	26.0	5.3	2.7	6.7	3.2	25.9	11.5	210.7	160.5	202.3	157.1	216.2	196.3	222.9	193.0
20.0	48.0	26.3	5.3	2.6	7.2	3.3	26.0	12.5	218.4	166.6	219.3	172.3	196.8	177.0	218.5	192.7
21.0	47.5	26.5	4.9	2.5	6.7	3.1	25.4	13.6	222.7	170.6	219.7	173.0	217.7	198.3	244.0	193.7
22.0	53.8	27.1	4.8	2.6	6.7	3.1	24.8	14.1	222.1	169.6	219.5	172.2	218.7	204.3	219.0	193.5
23.0	47.8	25.8	5.4	2.7	6.6	3.2	25.4	12.6	212.6	163.3	219.1	167.9	216.9	199.9	224.2	193.4
24.0	49.4	26.6	5.2	2.6	7.3	3.4	26.1	13.0	221.7	172.2	224.5	172.7	221.3	198.2	221.8	192.8
25.0	49.8	25.9	5.0	2.6	6.6	3.1	26.2	11.7	226.5	173.4	223.9	172.9	224.3	198.2	218.0	191.4
26.0	48.6	26.3	5.3	2.6	6.8	3.3	25.6	13.5	213.4	163.1	217.7	167.2	218.9	198.2	282.5	193.3
27.0	49.7	26.3	5.2	2.7	6.7	3.3	26.6	11.8	222.4	173.9	218.7	170.9	219.1	198.2	217.7	194.1
28.0	50.2	25.7	5.1	2.6	6.7	3.3	25.4	12.5	212.9	162.5	202.9	179.8	217.8	196.6	218.7	192.7
29.0	47.5	25.7	5.0	2.6	6.7	3.3	26.0	13.5	103.5	177.8	211.4	161.7	223.4	179.6	257.3	193.5
30.0	53.7	26.1	5.0	2.6	6.8	3.2	24.7	13.5	217.3	169.3	222.3	172.4	222.6	201.7	219.3	193.9
31.0	51.4	27.7	5.6	2.7	7.9	3.9	25.7	11.8	215.2	163.3	221.3	170.5	219.5	195.2	219.0	192.9
32.0	54.2	26.9	4.8	2.6	7.5	3.4	26.1	12.6	222.6	174.5	220.4	173.1	219.1	198.4	221.1	193.3
33.0	47.9	25.8	5.4	2.7	6.8	3.2	26.0	11.8	222.6	174.2	203.1	158.8	197.4	176.5	217.1	194.4
34.0	48.9	26.4	5.1	2.6	7.3	3.2	25.9	11.8	221.5	168.5	221.9	170.8	245.0	187.7	218.3	191.4
35.0	47.3	25.7	4.9	2.7	6.8	3.4	26.0	11.8	222.7	170.7	218.5	170.5	216.4	198.3	219.2	192.7
36.0	48.3	26.1	5.1	2.5	6.6	3.2	25.7	11.8	212.4	162.6	219.8	171.8	218.6	196.3	225.2	193.5
37.0	47.5	25.4	4.9	2.6	6.6	3.2	24.9	12.2	222.4	171.5	220.2	171.2	221.9	204.2	218.4	194.1
38.0	53.0	26.9	5.3	2.5	6.7	3.3	25.8	11.8	213.1	163.0	222.1	173.2	218.3	198.5	216.8	193.2
39.0	48.7	26.8	5.5	2.7	6.7	3.2	25.3	13.5	224.9	173.9	222.3	169.3	218.8	196.6	219.0	193.2
40.0	46.6	25.9	5.0	2.6	7.3	3.4	26.1	11.8	216.6	164.3	202.5	160.4	195.6	175.0	219.5	192.6

41.0	49.1	26.2	5.1	2.6	6.6	3.3	25.7	11.5	217.0	165.6	221.6	173.4	218.9	198.4	223.4	189.1
42.0	52.2	27.1	5.2	2.5	6.7	3.1	25.7	11.8	222.4	173.6	224.1	170.3	225.0	196.6	217.0	193.3
43.0	49.5	26.3	5.3	2.7	6.9	3.2	26.5	12.6	221.2	161.3	219.3	172.7	219.7	195.2	221.3	193.1
44.0	51.8	26.1	5.4	2.6	7.1	3.3	26.3	13.6	222.5	166.3	208.4	159.6	197.6	173.7	220.8	192.9
45.0	55.2	26.9	5.6	2.7	6.9	3.2	25.9	13.2	219.4	168.9	221.7	172.5	219.4	197.7	219.3	193.5
46.0	50.9	27.3	5.7	2.7	7.1	3.3	26.5	11.7	223.6	174.2	217.6	170.1	218.9	196.0	281.7	194.2
47.0	47.5	26.2	5.2	2.9	7.5	3.3	25.6	13.6	211.4	161.4	214.0	167.2	222.9	181.3	223.5	193.6
48.0	52.5	26.9	5.3	2.7	7.1	4.6	26.0	18.4	215.2	159.2	221.9	170.2	220.9	199.3	221.6	195.0
49.0	51.3	25.8	5.7	2.7	6.8	3.3	26.0	11.8	223.7	173.6	221.7	171.7	221.2	198.5	220.4	192.6
50.0	52.7	26.5	5.3	2.7	7.2	3.3	25.9	12.5	213.9	162.8	221.8	172.4	217.9	200.7	219.5	168.8
51.0	48.7	26.1	5.4	2.6	7.0	3.2	25.8	12.5	218.2	168.0	220.0	171.7	219.6	197.9	224.1	195.4
52.0	51.1	26.9	5.1	2.7	7.0	3.3	25.9	11.8	211.2	163.1	205.7	160.6	197.3	176.0	223.1	188.1
53.0	48.7	25.7	5.3	2.7	7.0	3.3	25.7	13.5	217.2	166.1	223.1	171.6	220.0	197.7	282.4	194.3
54.0	48.1	26.5	5.1	2.6	7.0	3.2	25.7	11.5	214.0	161.9	219.5	169.2	217.7	198.5	220.4	193.4
55.0	47.0	26.4	5.1	2.6	6.9	3.2	25.5	12.6	216.9	161.6	208.3	162.0	195.5	176.9	221.2	189.3
56.0	51.1	25.7	5.1	2.6	6.8	3.2	25.2	11.5	215.3	164.8	217.1	172.2	217.9	197.8	222.9	194.5
57.0	51.8	26.1	5.4	2.7	6.6	3.1	25.9	12.5	215.5	163.4	217.7	172.8	216.1	195.8	227.3	193.5
58.0	53.0	26.4	5.3	2.6	6.9	3.6	25.4	14.2	213.7	163.9	203.9	158.9	219.5	198.4	220.1	188.9
59.0	47.0	25.7	5.6	2.7	7.0	3.2	26.0	13.5	220.6	167.0	206.5	161.3	195.9	176.9	221.9	192.1
60.0	47.2	25.8	5.1	2.6	7.3	3.2	25.3	13.5	217.3	166.1	217.8	169.5	225.9	179.1	219.6	194.0
61.0	51.0	26.4	5.2	2.6	6.7	3.2	25.5	18.4	218.4	166.6	224.7	173.0	219.2	198.9	221.3	193.4
62.0	52.5	26.5	5.3	2.6	6.7	3.3	25.9	12.5	214.9	162.3	220.2	171.7	222.2	197.7	269.8	192.9
63.0	47.4	25.8	5.5	2.6	6.7	3.3	26.1	17.6	220.6	165.5	223.2	165.2	217.5	199.8	220.7	190.8
64.0	48.8	25.1	5.4	2.6	7.0	3.2	25.3	15.6	213.9	163.4	217.7	167.1	222.9	190.4	218.9	188.0

9.2 Metadata Evaluation

FusionFS was evaluated on 4, 16 and 64 nodes for metadata operations. Below is the data for the same.

4 Nodes:

metadata	0kb	
0 byte	10000	
Touch	Write	Read

8.45	50.769	24.774
8.376	49.834	25.227
8.452	49.725	24.962
8.314	49.292	25.22

16 Nodes:

metadata	0kb	
0 byte	10000	
touch	Write	Read
7.964	49.057	25.032
8.073	49.122	25.507
8.084	49.842	26.161
7.937	50.087	26.519
7.902	48.739	24.913
8.093	47.715	24.94
8.118	51.075	25.659
8.536	47.702	24.982
7.951	50.252	25.086
7.996	46.635	24.918
7.913	47.892	24.793
7.967	49.548	25.662
7.983	48.176	26.329
7.949	47.485	25.756
8.486	49.224	26.754
8.011	48.62	25.102

64 Nodes:

metadata	0kb	
0 byte	10000	
Touch	Write	Read
7.981	50.868	25.75
8.002	46.883	24.707

8.129	49.02	24.736
7.988	47.852	24.54
8.048	47.006	25.308
7.92	47.841	24.977
7.965	50.974	25.695
7.952	49.525	25.056
8.045	45.777	24.483
8.039	48.346	25.174
8.218	49.294	24.925
8.164	48.747	24.869
7.941	50.329	24.679
8.049	48.922	24.385
7.938	50.21	25.227
7.884	48.61	24.869
7.978	45.771	24.1
8.029	46.867	24.715
7.961	50.962	25.167
8.088	47.144	24.727
8.073	46.734	24.885
8.263	46.377	24.138
8.064	51.728	25.509
7.901	47.619	24.177
7.951	48.91	24.964
8.072	46.963	24.731
8.225	48.911	24.748
7.903	48.288	25.193
8.016	49.127	18.443
8.084	45.356	24.49
7.947	53.65	25.11
7.849	50.035	26.275
7.916	46.286	25.25
7.923	46.659	24.802
7.943	45.359	24.402

7.221	47.498	24.229
8.133	46.703	23.914
7.931	47.136	24.835
7.955	46.365	24.553
8.669	46.782	25.152
7.901	52.137	25.73
8.005	48.68	24.534
7.964	49.564	24.783
8.004	53.074	25.185
8.125	48.815	24.223
8.076	45.809	25.164
8	51.833	25.18
8.003	49.642	24.783
7.974	51.709	25.161
8.034	47.957	24.379
7.95	49.223	24.391
7.962	46.965	25.717
7.98	45.725	24.577
8.123	51.481	25.043
7.949	49.852	24.597
8.013	51.767	25.909
7.968	46.19	24.108
8.018	47.285	24.18
8.157	48.47	24.877
7.963	51.174	24.893
7.927	45.903	24.343
7.679	48.798	25.171
7.887	47.632	25.329
8.039	49.211	26.157

10. Analysis & Conclusion

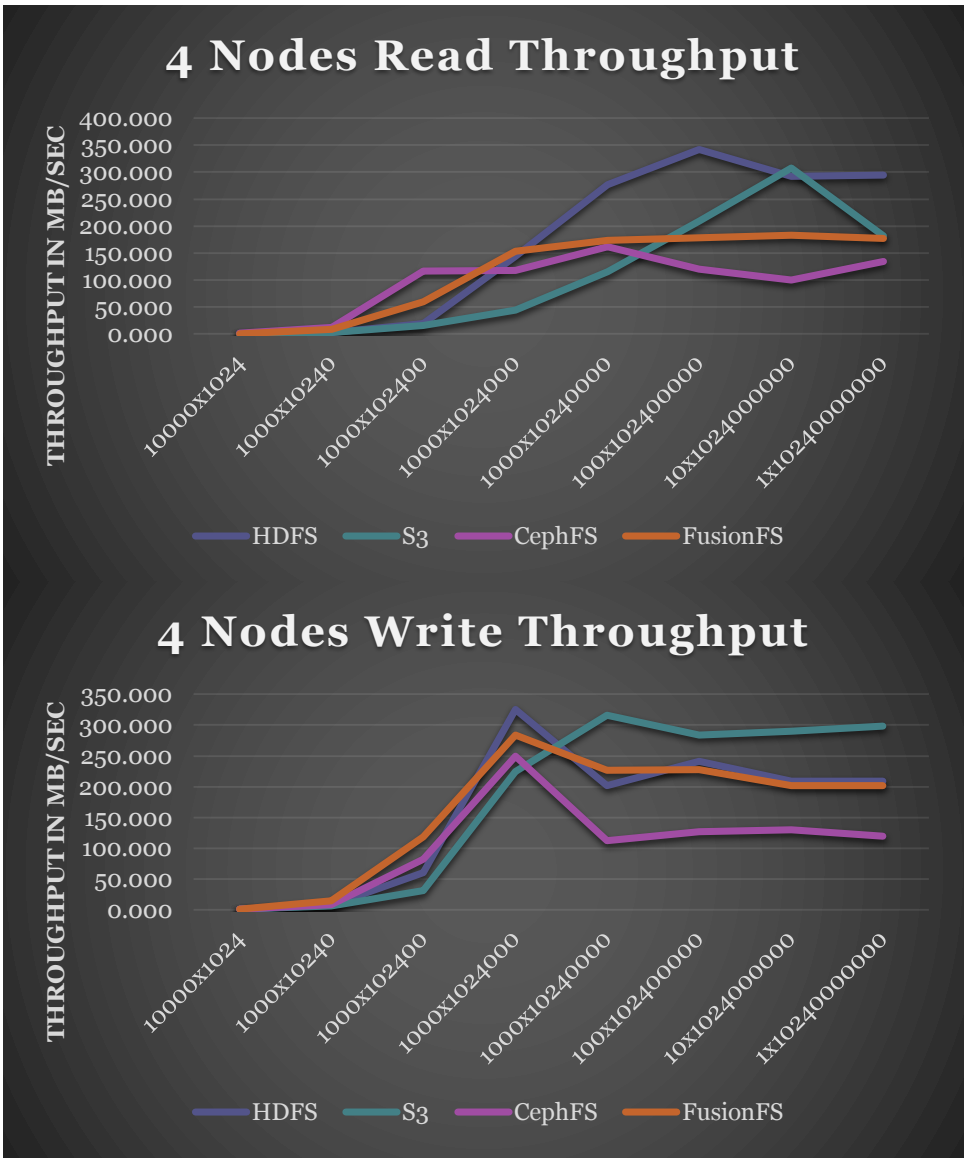
Read Write Evaluation

In this section we compare and analyze the Read and Write throughput of the different systems at 4, 16 and 64 nodes.

4 Nodes:

Read	4 Nodes			
	HDFS	S3	CephFS	FusionFS
10000x1024	0.290	0.185	1.084	0.749
1000x10240	2.264	2.065	11.685	8.568
1000x102400	19.778	15.112	117.111	59.121
1000x1024000	143.349	44.691	118.293	153.237
1000x10240000	276.549	115.897	162.246	173.136
100x102400000	341.904	209.053	120.720	177.890
10x1024000000	292.603	308.283	99.708	183.437
1x10240000000	294.811	181.700	133.944	177.272

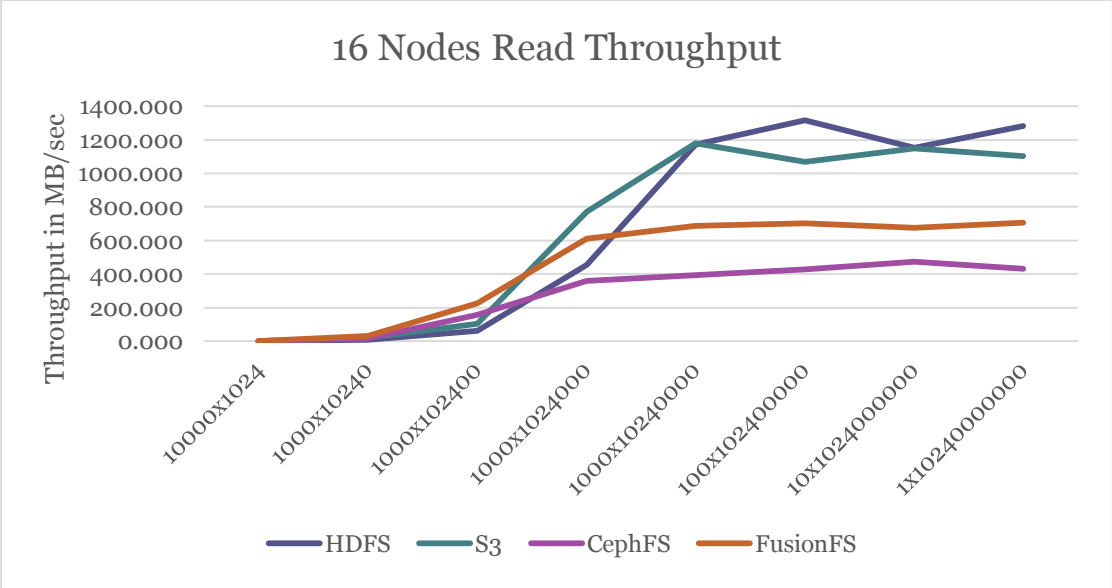
Write	4 Nodes			
	HDFS	S3	CephFS	FusionFS
10000x1024	1.756	0.749	0.883	1.487
1000x10240	7.102	6.323	8.377	14.919
1000x102400	60.096	31.718	82.380	118.758
1000x1024000	325.521	225.021	249.095	284.251
1000x10240000	201.873	316.049	112.811	226.395
100x102400000	241.499	283.492	127.338	227.795
10x1024000000	208.611	290.234	129.899	201.974
1x10240000000	209.450	298.790	119.338	201.990

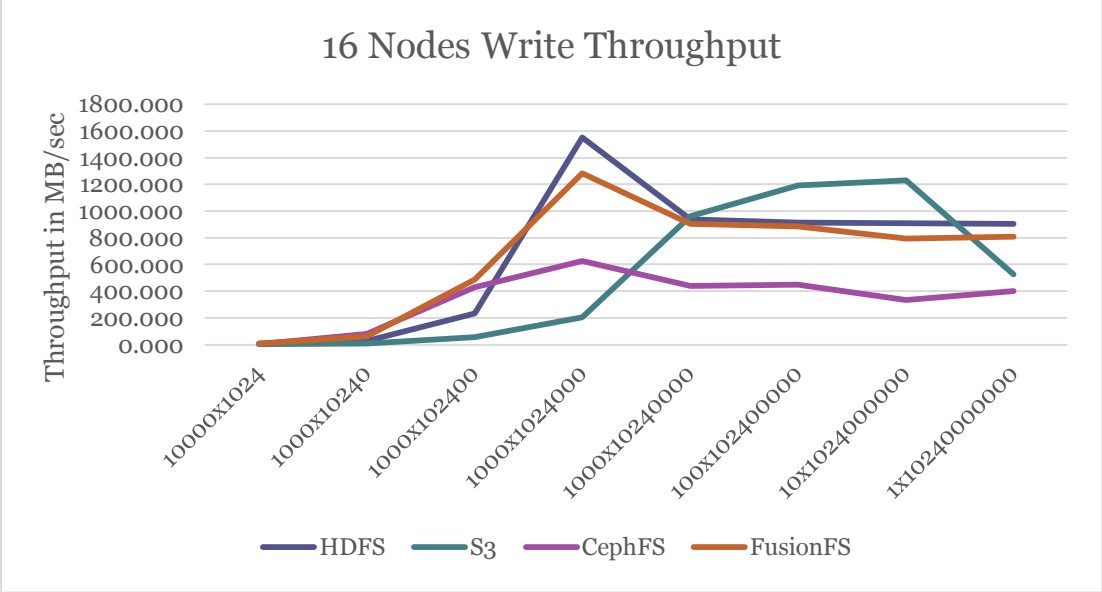


16 Nodes:

Read	16 Nodes			
	HDFS	S3	CephFS	FusionFS
10000x1024	0.457	2.5	1.309	2.981
1000x10240	7.082	19.8	17.847	31.245
1000x102400	61.576	105.8	159.118	227.169
1000x1024000	456.204	769.4	361.275	610.258
1000x10240000	1172.608	1178.9	392.968	687.973
100x102400000	1316.482	1067.2	427.630	703.955
10x1024000000	1150.483	1150.3	474.500	676.958
1x10240000000	1282.051	1104.7	431.729	705.219

Write	16 Nodes			
	HDFS	S3	CephFS	FusionFS
10000x1024	5.708	0.7	3.156	5.969
1000x10240	25.000	7	79.509	59.029
1000x102400	233.645	53.4	430.968	488.005
1000x1024000	1552.795	206.4	627.950	1281.716
1000x10240000	936.330	959.6	436.848	906.389
100x102400000	913.743	1193.5	449.185	884.295
10x1024000000	911.079	1231.7	333.232	793.252
1x10240000000	905.469	527.6	401.300	806.802

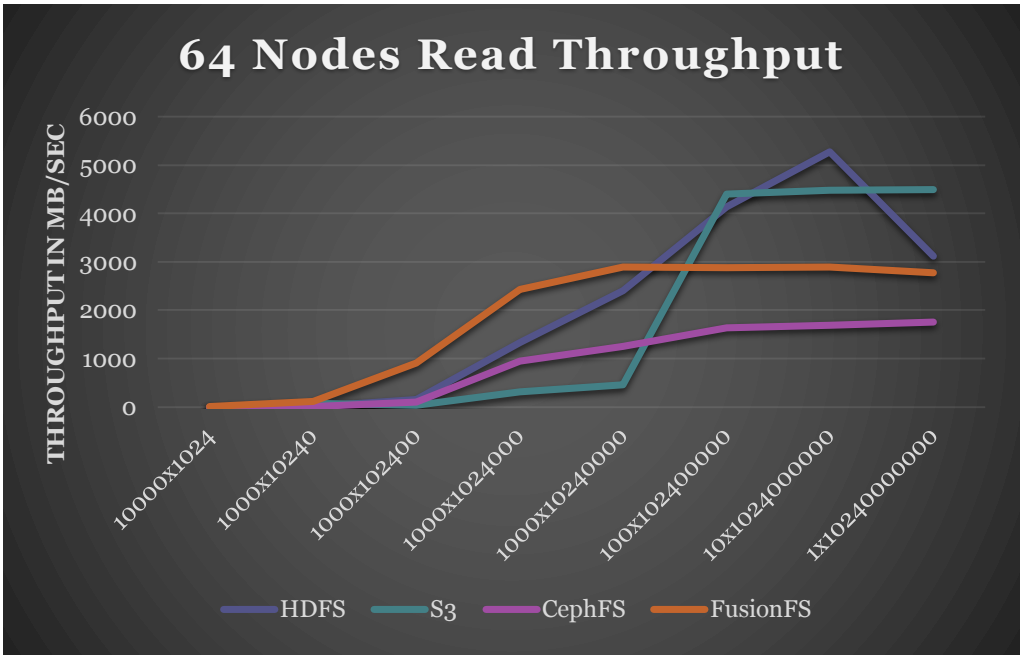


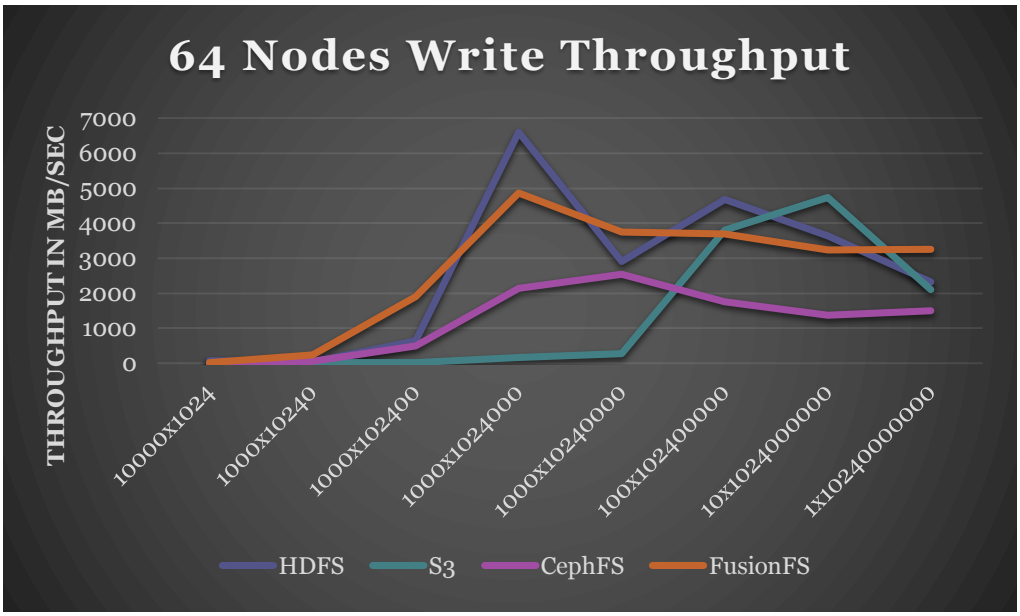


64 Nodes:

Read	64 Nodes			
File Sizes	HDFS	S3	CephFS	FusionFS
10000x1024	1.59	0.009	0.86	12.5
1000x10240	5.86	75.4	12.77	118.8
1000x102400	156.43	38.2	101.66	903.5
1000x1024000	1336.45	312.2	952.56	2423.5
1000x10240000	2400.10	454.2	1257.86	2887.9
100x102400000	4131.38	4402.4	1630.69	2873.5
10x1024000000	5270.09	4475.3	1682.50	2889.4
1x10240000000	3120.37	4486.6	1752.02	2778.5

Write	64 Nodes			
	HDFS	S3	CephFS	FusionFS
10000x1024	81.13590264	0.002	2.43876	23.8
1000x10240	45.45454545	22.5	53.16038	236.8
1000x102400	643.0868167	23.5	485.57243	1898.2
1000x1024000	6600.660066	169.9	2145.23452	4873.6
1000x10240000	2904.654709	276.6	2541.01096	3745.8
100x102400000	4686.035614	3808.7	1759.79835	3694.5
10x1024000000	3646.308113	4730.3	1376.25667	3232.1
1x10240000000	2329.509056	2107	1497.48096	3247.8





Based on the above data, the following analysis and conclusions can be drawn.

1. HDFS has very good throughput for read operations involving large files.
2. S3 performs well for writing large files compared to other systems.
3. Fusion performs very well for smaller files and both read and write operations.

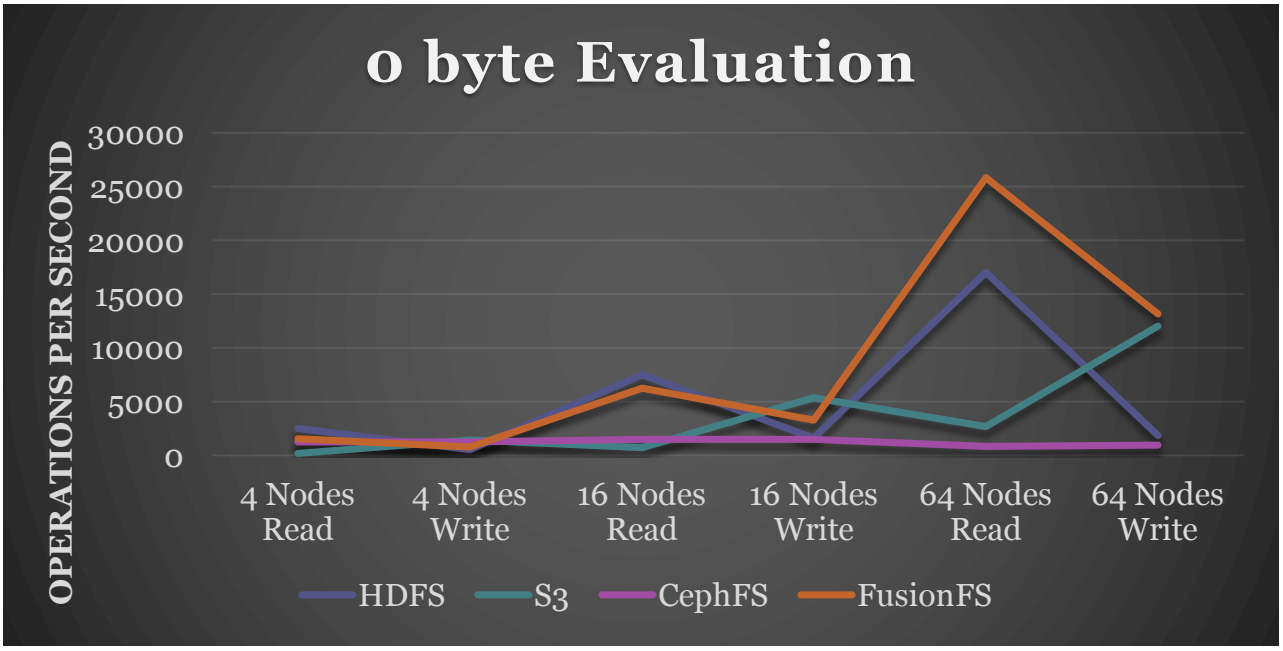
Metadata Evaluation

In this section we compare and analyze the metadata handling in terms of operations per second on the different systems at 4, 16 and 64 nodes scales. We evaluate by using touch and read / write of empty / blank files.

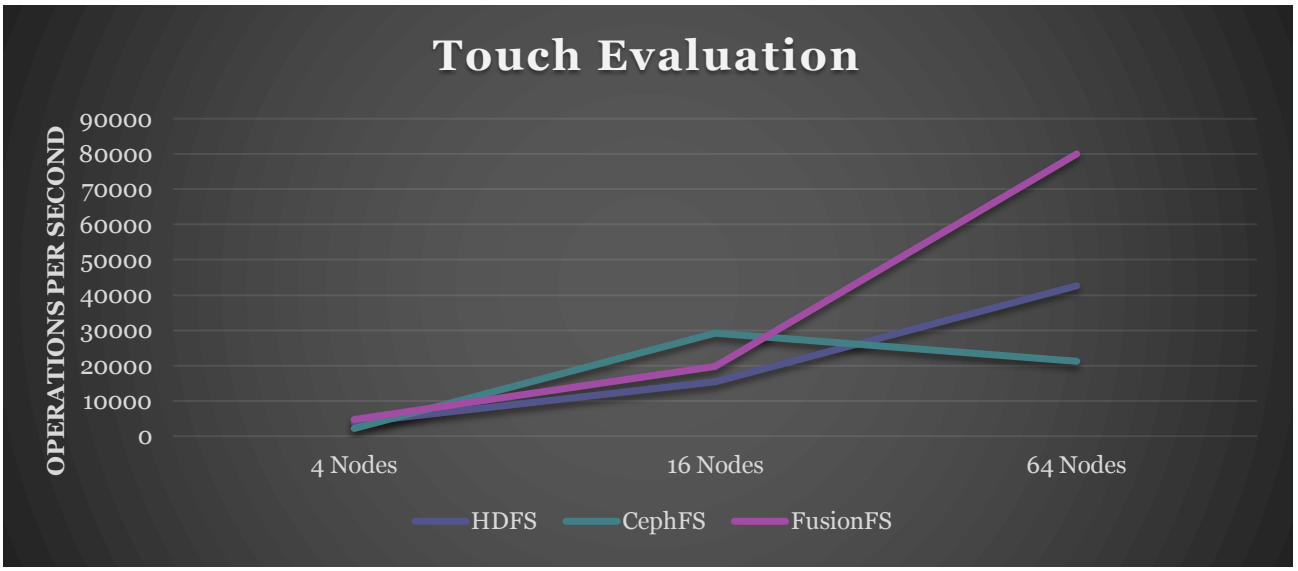
0 byte Read Evaluation in operations per second				
Metadata	HDFS	S3	CephFS	FusionFS
4 Nodes	2539.68254	207.4400984	1250.22465	1597.077348
16 Nodes	7529.411765	759.6754761	1520.246898	6272.772492
64 Nodes	17066.66667	2722.811899	848.5564027	25821.04428

0 byte Write Evaluation in operations per second				
Metadata	HDFS	S3	CephFS	FusionFS
4 Nodes	542.3728814	1434.553002	1244.09057	801.5228935

16 Nodes	1603.005636	5383.308379	1488.480267	3277.131384
64 Nodes	1859.620449	12030.44266	979.9439729	13185.97831



Touch Evaluation			
Metadata	HDFS	CephFS	FusionFS
4 Nodes	3869.96904	2285.714286	4763.038819
16 Nodes	15479.87616	29239.09816	19850.65484
64 Nodes	42666.66667	21272.75031	79990.78231



Based on the above data, the following analysis and conclusions can be drawn.

1. FusionFS is the best at metadata management at all scales.
2. HDFS struggles at metadata management for write operations.
3. CephFS does not show any variations even at large scales.

11. Source Code Repository

We have open sourced all the code (scripts & java) that we wrote for setup and benchmarking HDFS, S3, S3FS, Ceph and FusionFS. Source code is available at github.

https://github.com/akashmahakode/benchmarking_file_systems

12. References

1. High-Performance Storage Support for Scientific Applications on the Cloud http://datasys.cs.iit.edu/publications/2015_ScienceCloud15-cloud-storage.pdf
2. FusionFS: Toward Supporting Data-Intensive Scientific Applications on Extreme-Scale High-Performance Computing Systems -- http://datasys.cs.iit.edu/publications/2014_BigData14_FusionFS.pdf
3. FusionFS -- http://datasys.cs.iit.edu/projects/FusionFS/
4. Google Protocol Buffers -- https://developers.google.com/protocol-buffers/
5. FUSE -- https://github.com/libfuse/libfuse
6. Amazon EC2 -- http://aws.amazon.com/ec2
7. EC2 m3.large -- https://aws.amazon.com/ec2/instance-types/
8. GitHub Source code Repository -- https://github.com/akashmahakode/benchmarking_file_systems
9. Hadoop users Guide -- https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-hdfs/HdfsUserGuide.html#Purpose
10. Hadoop setup -- http://insightdataengineering.com/blog/hadoopdevops/
11. Ceph Setup -- http://docs.ceph.com/docs/master/start/

12. Setting up Ceph -- http://blog.programster.org/ubuntu-14-04-deploy-a-ceph-cluster-part-1/, http://blog.programster.org/ubuntu-14-04-deploy-a-ceph-cluster-part-2/

13. Appendix

a. CephFS Setup

CREATE A CLUSTER

If at any point you run into trouble and you want to start over, execute the following to purge the configuration:

```
ceph-deploy purgedata {ceph-node} [{ceph-node}]

ceph-deploy forgetkeys
```

To purge the Ceph packages too, you may also execute:

```
ceph-deploy purge {ceph-node} [{ceph-node}]
```

If you execute purge, you must re-install Ceph.

On your admin node from the directory you created for holding your configuration details, perform the following steps using ceph-deploy.

1. Create the cluster.

   ```
   ceph-deploy new {initial-monitor-node(s)}
   ```

   For example:

```
ceph-deploy new node1
```

Check the output of ceph-deploy with ls and cat in the current directory. You should see a Ceph configuration file, a monitor secret keyring, and a log file for the new cluster. See ceph-deploy new -h for additional details.

2. Change the default number of replicas in the Ceph configuration file from 3 to 2 so that Ceph can achieve an active + clean state with just two Ceph OSDs. Add the following line under the [global] section:

```
osd pool default size = 2
```

3. If you have more than one network interface, add the public network setting under the [global] section of your Ceph configuration file. See the Network Configuration Reference for details.

```
public network = {ip-address}/{netmask}
```

4. Install Ceph.

```
ceph-deploy install {ceph-node}[{ceph-node} ...]
```

For example:

```
ceph-deploy install admin-node node1 node2 node3
```

The ceph-deploy utility will install Ceph on each node. **NOTE**: If you use ceph-deploy purge, you must re-execute this step to re-install Ceph.

5. Add the initial monitor(s) and gather the keys:

```
ceph-deploy mon create-initial
```

Once you complete the process, your local directory should have the following keyrings:

- {cluster-name}.client.admin.keyring
- {cluster-name}.bootstrap-osd.keyring
- {cluster-name}.bootstrap-mds.keyring
- {cluster-name}.bootstrap-rgw.keyring

1. Add two OSDs. For fast setup, this quick start uses a directory rather than an entire disk per Ceph OSD Daemon. See ceph-deploy osd for details on using separate disks/partitions for OSDs and journals. Login to the Ceph Nodes and create a directory for the Ceph OSD Daemon.

```
2. ssh node2

3. sudo mkdir /var/local/osd0

4. exit

5.

6. ssh node3

7. sudo mkdir /var/local/osd1
```

```
exit
```

Then, from your admin node, use ceph-deploy to prepare the OSDs.

```
ceph-deploy osd prepare {ceph-node}:/path/to/directory
```

For example:

```
ceph-deploy osd prepare node2:/var/local/osd0
node3:/var/local/osd1
```

Finally, activate the OSDs.

```
ceph-deploy osd activate {ceph-node}:/path/to/directory
```

For example:

```
ceph-deploy osd activate node2:/var/local/osd0
node3:/var/local/osd1
```

8. Use ceph-deploy to copy the configuration file and admin key to your admin node and your Ceph Nodes so that you can use the ceph CLI without having to specify the monitor address and ceph.client.admin.keyring each time you execute a command.

```
ceph-deploy admin {admin-node} {ceph-node}
```

For example:

```
ceph-deploy admin admin-node node1 node2 node3
```

When ceph-deploy is talking to the local admin host (admin-node), it must be reachable by its hostname. If necessary, modify /etc/hosts to add the name of the admin host.

9. Ensure that you have the correct permissions for the ceph.client.admin.keyring.

```
sudo chmod +r /etc/ceph/ceph.client.admin.keyring
```

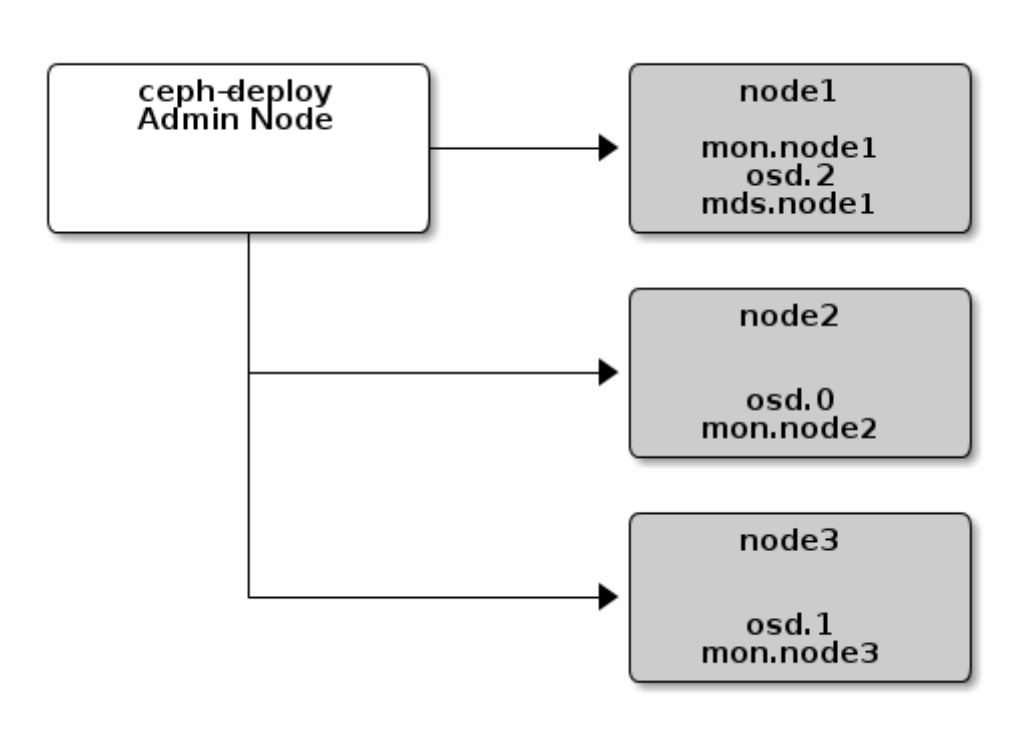
10. Check your cluster's health.

```
ceph health
```

Your cluster should return an active + clean state when it has finished peering.

EXPANDING YOUR CLUSTER

Once you have a basic cluster up and running, the next step is to expand cluster. Add a Ceph OSD Daemon and a Ceph Metadata Server to node1. Then add a Ceph Monitor to node2 and node3 to establish a quorum of Ceph Monitors.



ADDING AN OSD

Since you are running a 3-node cluster for demonstration purposes, add the OSD to the monitor node.

```
ssh node1

sudo mkdir /var/local/osd2

exit
```

Then, from your ceph-deploy node, prepare the OSD.

```
ceph-deploy osd prepare {ceph-node}:/path/to/directory
```

For example:

```
ceph-deploy osd prepare node1:/var/local/osd2
```

Finally, activate the OSDs.

```
ceph-deploy osd activate {ceph-node}:/path/to/directory
```

For example:

```
ceph-deploy osd activate node1:/var/local/osd2
```

Once you have added your new OSD, Ceph will begin rebalancing the cluster by migrating placement groups to your new OSD. You can observe this process with theceph CLI.

```
ceph -w
```

You should see the placement group states change from active+clean to active with some degraded objects, and finally active+clean when migration completes. (Control-c to exit.)

ADD A METADATA SERVER

To use CephFS, you need at least one metadata server. Execute the following to create a metadata server:

```
ceph-deploy mds create {ceph-node}
```

For example:

```
ceph-deploy mds create node1
```

ADDING MONITORS

A Ceph Storage Cluster requires at least one Ceph Monitor to run. For high availability, Ceph Storage Clusters typically run multiple Ceph Monitors so that the failure of a single Ceph Monitor will not bring down the Ceph Storage Cluster. Ceph uses the Paxos algorithm, which requires a majority of monitors (i.e., 1, 2:3, 3:4, 3:5, 4:6, etc.) to form a quorum.

Add two Ceph Monitors to your cluster.

```
ceph-deploy mon add {ceph-node}
```

For example:

```
ceph-deploy mon add node2 node3
```

Once you have added your new Ceph Monitors, Ceph will begin synchronizing the monitors and form a quorum. You can check the quorum status by executing the following:

```
ceph quorum_status --format json-pretty
```

All the installation steps for Ceph have been automated into scripts. The code is committed to GitHub, below steps detail how to use them.

1. Checkout all scripts and the archives from GitHub
2. Execute init_script.sh which runs setup_ceph.sh to setup a ceph cluster
3. Execute benchmarking.sh to run evaluations

b. FusionFS Setup

How to install fusionfs:

- Make sure all dependent libraries are specified, e.g. echo $LD_LIBRARY_PATH :/usr/local/lib:/home/dongfang/fusionFS/src/ffsnet/:/home/ubuntu/fusionFS/src/udt4/src
- Install FUSE for your Linux distribution
- Install Google Protocol Buffer
- Go to ./src/zht/ and run Makefile
- Go to ./src/udt/ and run Makefile
- Go to ./src/ffsnet/, run Makefile
- Go to ./src, run Makefile

How to use fusionfs on clusters:

- Specify list of nodes on /src/zht/neighbor
- ./cleanAll to `make clean` everything
- ./compileAll to `make` everything
- ./start_service_all to start services on all nodes
- ./start_all to start fusionfs on all nodes
- ./stop_all to stop all fusionfs on all nodes
- ./stop_service_all to stop all services on all nodes

All the above steps have been automated into scripts. The code is committed to GitHub, below steps detail how to use them.

4. Checkout all scripts and the archives from GitHub

5. Execute init_script.sh to copy call contents to the nodes
6. Run setup_fusion.sh on each node to setup fusion
7. Execute benchmarking.sh to run evaluations

c. HDFS Setup

It is expected that you should have all the required permissions and authorization key-pair available with you which allows running ec2 instances.

1. Setup passwordless ssh for all ec2 instances. Clone the github repository -
   > git clone https://github.com/akashmahakode/benchmarking_file_systems.git
   > cd benchmarking_file_systems/hdfs
   > Run setPasswordless.sh as
   setPasswordless.sh someKeyPair.pem

2. You need to change the permissions of the pem key that you downloaded from AWS earlier. Do this with the command otherwise SSH'ing will complain that the key is too open (bolded code would be different for your setup):
   > sudo chmod 600 ~/.ssh/**pem_key_filename**

3. Export AWS access key and secret key
   export AWS_ACCESS_KEY_ID=your_access_key
   export AWS_SECRET_ACCESS_KEY=your_secret_access_key
4. To enable PSSH (parallel ssh), we need all the public dns names of the ec2 instances. Run the following command
   >ec2-describe-instances --aws-access-key $AWS_ACCESS_KEY_ID --aws-secret-key $AWS_SECRET_ACCESS_KEY | grep INSTANCE | grep running | grep akash | awk -F" " '{print $4}' > /tmp/dns.txt

5. Install JDK on all nodes
   allnodes$ sudo apt-get update
   allnodes$ sudo apt-get install openjdk-7-jdk
OR install jdk on one of the node and create an AMI. Use this AMI to launch multiple instances

6. Download and extract Hadoop -

> pssh -h /tmp/dns.txt -p 64 -l ubuntu -t 100000000 -x "-oStrictHostKeyChecking=no -i akash_mahakode.pem " -o /tmp/del_out 'wget http://apache.mirrors.tds.net/hadoop/common/hadoop-2.7.1/hadoop-2.7.1.tar.gz -P /home/ubuntu; sudo tar zxvf hadoop-* -C /usr/local; sudo mv /usr/local/hadoop-* /usr/local/hadoop'

7. Export environment variables -

export JAVA_HOME=/usr
export PATH=$PATH:$JAVA_HOME/bin
export HADOOP_HOME=/usr/local/hadoop
export PATH=$PATH:$HADOOP_HOME/bin
export HADOOP_CONF_DIR=/usr/local/hadoop/etc/hadoop

> pssh -h /tmp/dns.txt -p 64 -l ubuntu -t 100000000 -x "-oStrictHostKeyChecking=no -i akash_mahakode.pem " -o /tmp/del_out 'echo export JAVA_HOME=/usr >> ~/.profile; echo export PATH=$PATH:$JAVA_HOME/bin >> ~/.profile; echo export HADOOP_HOME=/usr/local/hadoop >> ~/.profile; echo export PATH=$PATH:$HADOOP_HOME/bin >> ~/.profile; echo export HADOOP_CONF_DIR=/usr/local/hadoop/etc/hadoop >> ~/.profile; . ~/.profile'

8. For a basic setup of Hadoop, we'll be changing a few of the configurations in the Hadoop directory defined now by HADOOP_CONF_DIR environment variable. All the current configuration changes will be applied to the NameNode and all the DataNodes. After these changes, we will apply configurations specific to the NameNode and DataNodes.

Here are the following files to focus on:

- $HADOOP_CONF_DIR/hadoop-env.sh
- $HADOOP_CONF_DIR/core-site.xml
- $HADOOP_CONF_DIR/yarn-site.xml
- $HADOOP_CONF_DIR/mapred-site.xml (This file currently does not exist in the default Hadoop installation, but a template is available. We'll make a copy of the template and rename it to mapred-site.xml)

9. Common Hadoop Configurations on all Nodes

a. Let's start with $HADOOP_CONF_DIR/hadoop-env.sh. Currently only root users can edit files in the Hadoop directory, but we'll change this after all configurations have been applied.

The only thing that needs changing is the location of JAVA_HOME in the file. Simply replace${JAVA_HOME} with /usr which is where Java was just previously installed.

> pssh -h /tmp/dns.txt -p 64 -l ubuntu -t 100000000 -x "-oStrictHostKeyChecking=no -i akash_mahakode.pem " -o /tmp/del_out 'sudo sh -c "echo export JAVA_HOME=/usr >> /usr/local/hadoop/etc/hadoop/hadoop-env.sh"'

b. Edit $HADOOP_CONF_DIR/core-site.xml:

```
<configuration>
 <property>
  <name>fs.defaultFS</name>
  <value>hdfs://namenode_public_dns:9000</value>
 </property>
</configuration>
```

c. Edit $HADOOP_CONF_DIR/yarn-site.xml:

```
<configuration>

<!-- Site specific YARN configuration properties -->

 <property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
 </property>
 <property>
  <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
  <value>org.apache.hadoop.mapred.ShuffleHandler</value>
 </property>
 <property>
  <name>yarn.resourcemanager.hostname</name>
  <value>namenode_public_dns</value>
 </property>
</configuration>
```

d. Edit $HADOOP_CONF_DIR/mapred-site.xml

> sudo cp $HADOOP_CONF_DIR/mapred-site.xml.template $HADOOP_CONF_DIR/mapred-site.xml

```
<configuration>
 <property>
  <name>mapreduce.jobtracker.address</name>
  <value>namenode_public_dns:54311</value>
 </property>
 <property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
 </property>
</configuration>
```

10. **NameNode Specific Configurations**

Now that all the common configurations are complete, we'll finish up the NameNode specific configurations. On the NameNode, all that remains are the following:

- adding hosts to /etc/hosts
- modifying the configurations in $HADOOP_CONF_DIR/hdfs-site.xml
- defining the Hadoop master in $HADOOP_CONF_DIR/masters
- defining the Hadoop slaves in $HADOOP_CONF_DIR/slaves

a. Let's start with adding to the hosts file located under /etc/hosts. We will need to add each node's public DNS and hostname to the list. The hostname can be found with the following

> echo $(hostname)

or by taking the first part of the private DNS (e.g. ***ip-172-31-35-242***.us-west-2.compute.internal)

By default, 127.0.0.1 localhost is present, so we can add under it to look like the following (ignoring the IPv6 settings):

/etc/hosts should look like:

```
127.0.0.1 localhost
namenode_public_dns namenode_hostname
datanode1_public_dns datanode1_hostname
datanode2_public_dns datanode2_hostname
datanode3_public_dns datanode3_hostname
# The following lines are desirable for IPv6 capable hosts
::1 ip6-localhost ip6-loopback
fe00::0 ip6-localnet
ff00::0 ip6-mcastprefix
ff02::1 ip6-allnodes
ff02::2 ip6-allrouters
ff02::3 ip6-allhosts
```

b. We can now modify the $HADOOP_CONF_DIR/hdfs-site.xml file to specify the replication factor along with where the NameNode data will reside. For this setup, we will specify a replication factor of 1 for each data block in HDFS. Scroll down in the xml file to find the configurations tag and be sure to change the file to look like the following

$HADOOP_CONF_DIR/hdfs-site.xml:

```
<configuration>
 <property>
  <name>dfs.replication</name>
  <value>1</value>
 </property>
 <property>
```

```
  <name>dfs.namenode.name.dir</name>
  <value>file:///usr/local/hadoop/hadoop_data/hdfs/namenode</value>
 </property>
</configuration>
```

The current path where data on the NameNode will reside does not exist, so we'll need to make this before starting HDFS.

namenode$ sudo mkdir -p $HADOOP_HOME/hadoop_data/hdfs/namenode

c. Next we'll need to add a masters file to the $HADOOP_CONF_DIR directory

namenode$ sudo touch $HADOOP_CONF_DIR/masters

then insert the NameNode's hostname in that file

d. We will also need to modify the slaves file in the $HADOOP_CONF_DIR directory to the following. By default, localhost is present, but we can remove this.

$HADOOP_CONF_DIR/slaves

```
datanode1_hostname

datanode2_hostname

datanode3_hostname
```

e. Now that all configurations are set on the NameNode, we will change the ownership of the $HADOOP_HOME directory to the user ubuntu

namenode$ sudo chown -R ubuntu $HADOOP_HOME

11. **DataNode Specific Configurations**

a. Let's now move onto the final configurations for the DataNodes. We will need to first SSH into each DataNode and only configure the $HADOOP_CONF_DIR/hdfs-site.xml file

Scroll down in the xml file to find the configurations tag and be sure to change the file to look like the following.

$HADOOP_CONF_DIR/hdfs-site.xml:

```
<configuration>
 <property>
  <name>dfs.replication</name>
  <value>3</value>
 </property>
 <property>
  <name>dfs.datanode.data.dir</name>
  <value>file:///usr/local/hadoop/hadoop_data/hdfs/datanode</value>
 </property>
</configuration>
```

b. Just like on the NameNode, we will need to create the directory specified in the$HADOOP_CONF_DIR/hdfs-site.xml file.

datanodes$ sudo mkdir -p $HADOOP_HOME/hadoop_data/hdfs/datanode

c. Now that all configurations are set on the DataNode, we will change the ownership of the$HADOOP_HOME directory to the ubuntu user

datanodes$ sudo chown -R ubuntu $HADOOP_HOME

12. **Start Hadoop Cluster**

We can now start up HDFS from the Namenode by first formatting it and then starting HDFS. An important thing to notes is that every time the NameNode is formatted, all of the data previously on it is lost.

namenode$ hdfs namenode -format

namenode$ $HADOOP_HOME/sbin/start-dfs.sh

You can go to **namenode_public_dns**:50070 in your browser to check if all datanodes are online. If the webUI does not display, check to make sure your EC2 instances have security group settings that include All Traffic and not just SSH.

d. S3 Setup

1. Start the number of required EC2 instances
2. Copy the s3_java.jar to all instances. You can use pscp to copy all jars in parallel as follows-

pscp -h /tmp/ips.txt -l ubuntu -t 100000000 -x "-oStrictHostKeyChecking=no -i <<keypair.pem>>" s3_java.jar /home/ubuntu/s3_eval

This will scp s3_java.jar to all ec2 instances under /home/ubuntu/s3_eval directory

3. Export following variables AWS_ACCESS_KEY_ID, AWS_SECRET_ACCESS_KEY, BUCKET_NAME and run the jar by passing file size and number of files for evaluation. This can be done using **pssh** as follows

pssh -h /tmp/ips.txt -l ubuntu -t 100000000 -x "-oStrictHostKeyChecking=no -i <<keypair.pem>>" -o /tmp/del_out 'export AWS_ACCESS_KEY_ID= <<KEY>>; export AWS_SECRET_ACCESS_KEY=<<SECRET_KEY>>; export BUCKET_NAME=bucket_name;java -jar /home/ubuntu/s3_eval/s3_java.jar 0 10000; java -jar /home/ubuntu/s3_eval/s3_java.jar 1024 10000'

4. In the above pssh command we have exported the required variables and specified file size for 0 bytes and 1 KB for evaluation. Similarly, we can specify variable file sizes

5. s3_java.jar creates the specified files locally on the ec2 instance and then puts in the S3 bucket