Firefly Erasure Coding CPU Tests

Mark Nelson
mnelson@rehdat.com
2/21/2015

Table of Contents

SETUP AND NOTES

These tests were performed in the spring of 2014 during the final stages of Ceph Firefly development using a pre-release version of Ceph Firefly. The results here may not be 100% representative of new releases of Ceph. Several tests were performed comparing performance and CPU utilization of different erasure coding and replication configurations. The system was configured as follows:

HARDWARE

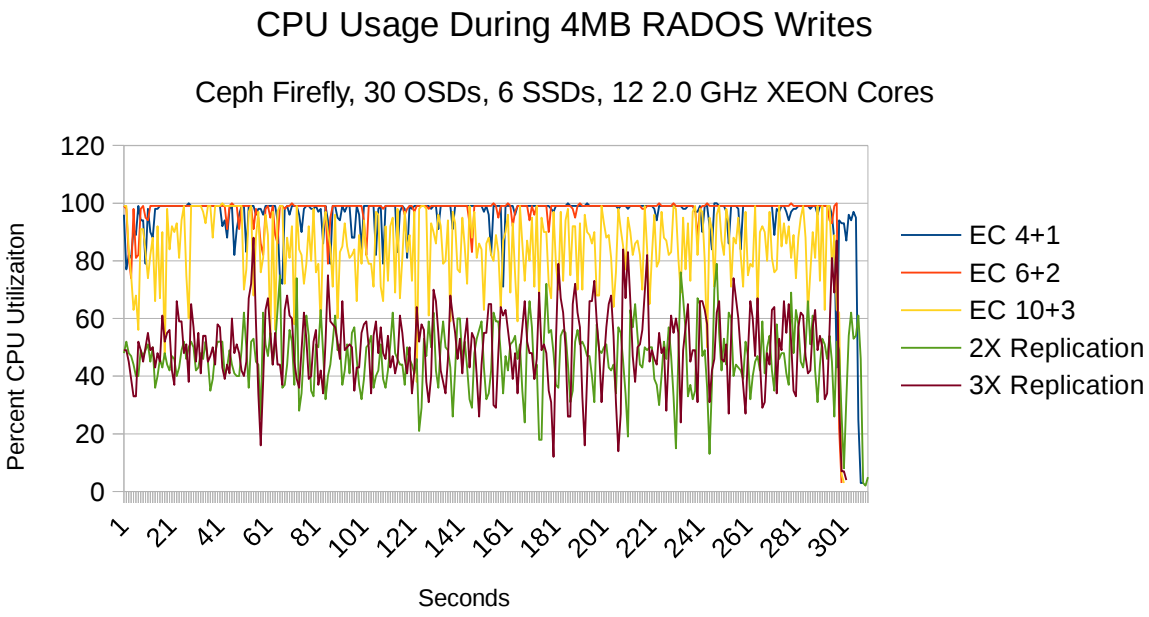
Device	Model
Chassis	Supermicro SC847A
Motherboard	Supermicro X9DRH-7F
CPUS	2 X Intel XEON E5-2630L (2.0GHz, 6-core)
RAM	8 X 4GB Supermicro ECC Registered DDR 1333
NIC	Intel X520-DA2 10GbE (bonded configuration)
Spinning Disks	30 X 7200RPM Seagate Constellation ES.2 (1 OSD per Disk)
SSDs	6 X Intel 520 SSDs (5 Journals per SSD)

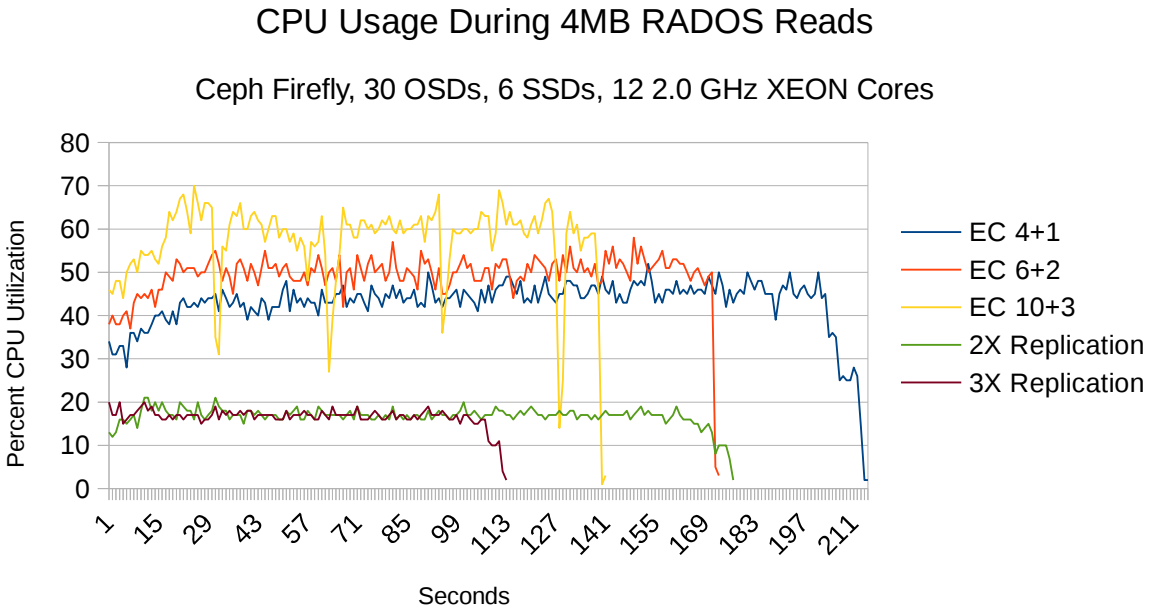
SOFTWARE

Software	Version
Ubuntu	13.04
Kernel	3.8.0
Ceph	Firefly Pre-release (0.78-385-gfb20823-1saucy)
Collectl	V3.6.7-1

Tests were run at 4MB, 128KB and 4KB IO sizes. CPU Utilization statistics were recorded for each test using the collectl utility.

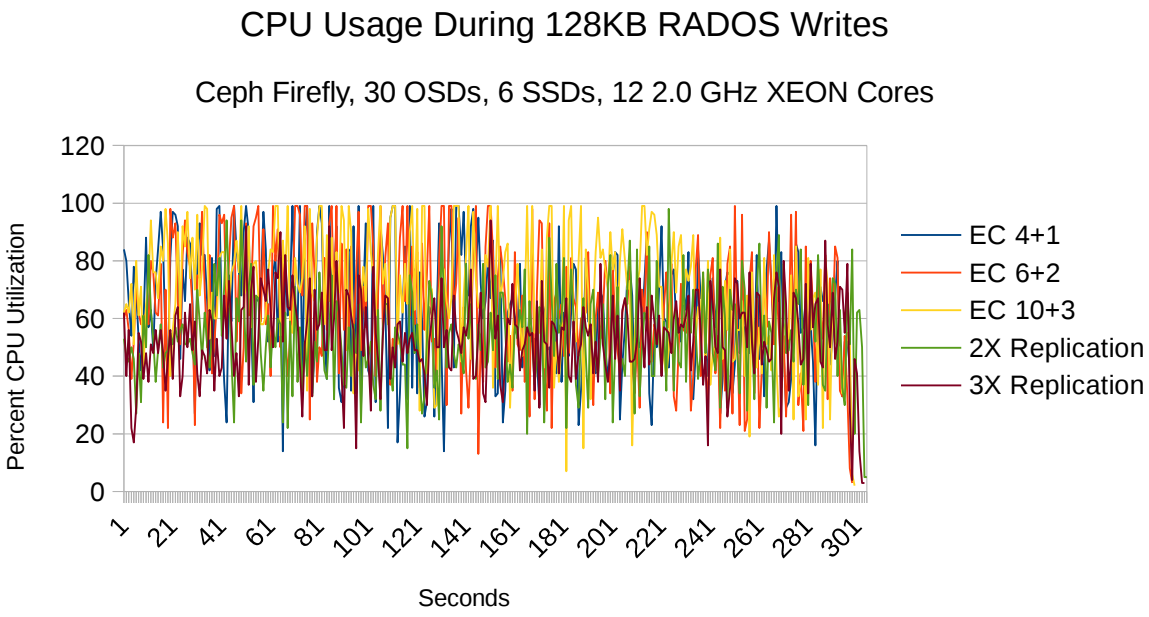
CPU USAGE DURING 4MB CLIENT IO

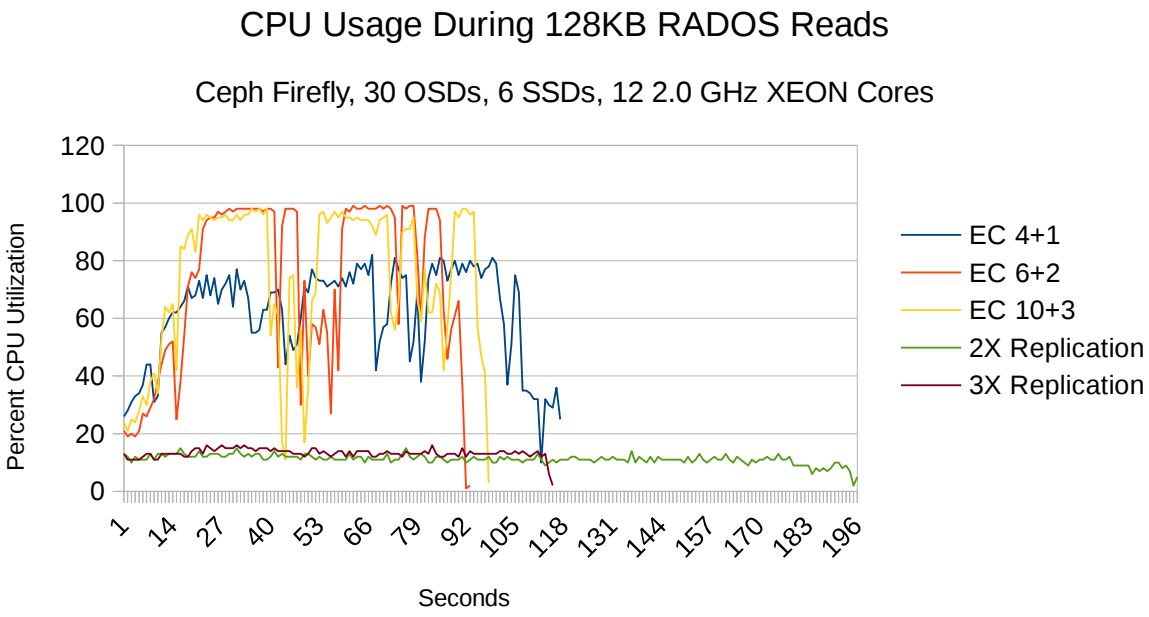




Average Throughput (MB/s)					
Test	EC 4+1	EC 6+2	EC 10+3	2X Replication	3X Replication
4MB Writes	1063	750	494	1171	755
4MB Reads	1546	1309	1061	2111	2131

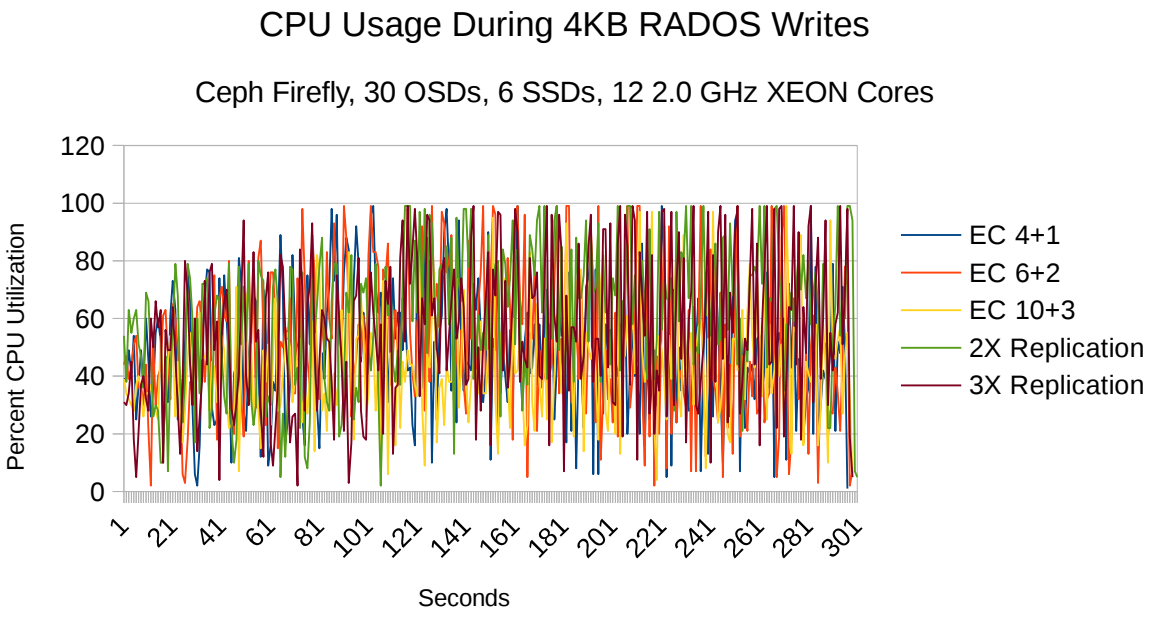
CPU USAGE DURING 128KB CLIENT IO

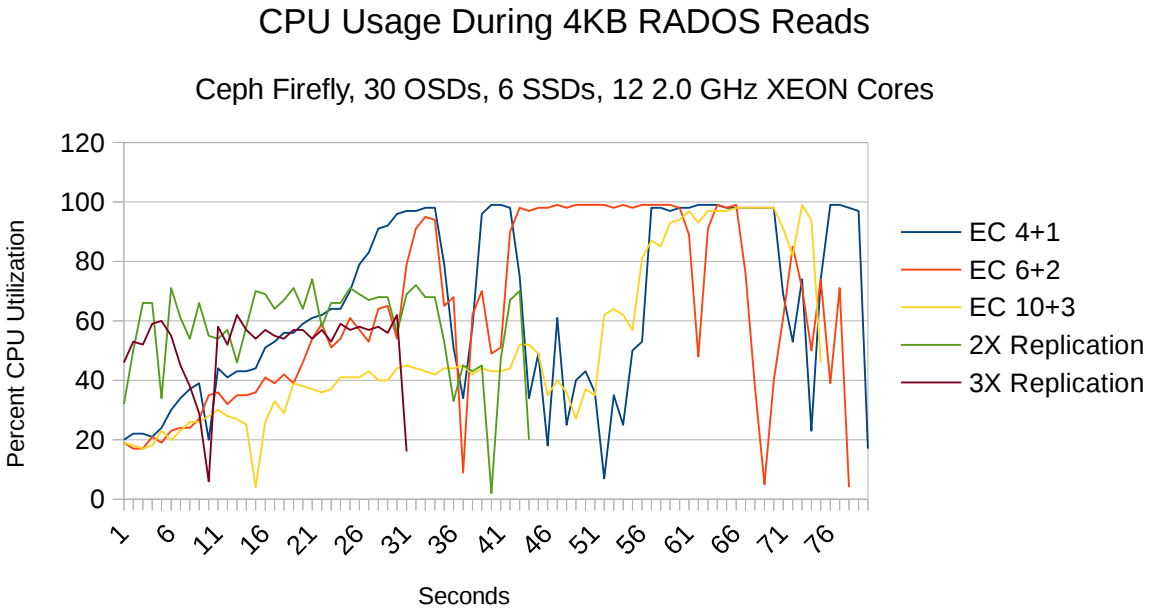




Average Throughput (MB/s)					
Test	EC 4+1	EC 6+2	EC 10+3	2X Replication	3X Replication
128KB Writes	100	60	43	228	154
128KB Reads	276	196	134	370	407

CPU USAGE DURING 4KB CLIENT IO

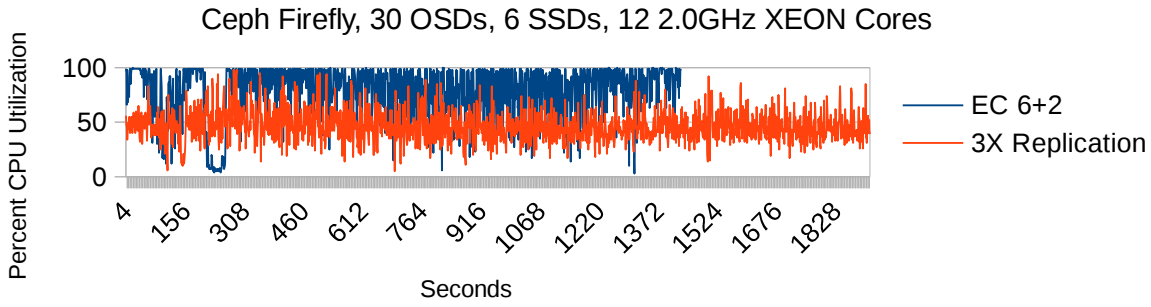




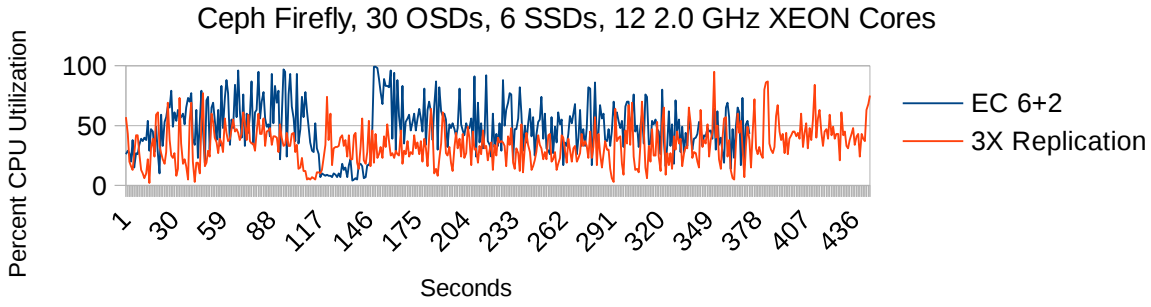
Average Throughput (Obj/s)					
Test	EC 4+1	EC 6+2	EC 10+3	2X Replication	3X Replication
4KB Writes	639 Obj/s	397 Obj/s	208 Obj/s	2067 Obj/s	1289 Obj/s
4KB Reads	2418 Obj/s	1532 Obj/s	825 Obj/s	13663 Obj/s	11967 Obj/s

CPU USAGE DURING RECOVERY AND CLIENT IO

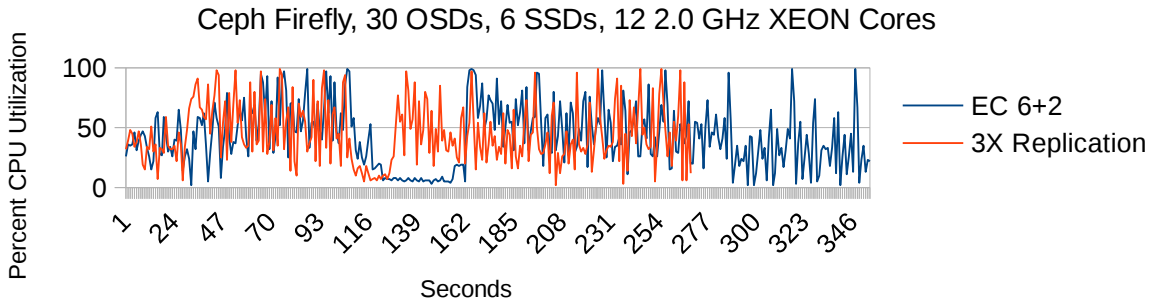
CPU Usage During OSD Recovery, and Simultaneous 4MB RADOS Writes



CPU Usage During OSD Recovery, and Simultaneous 128KB RADOS Writes



CPU Usage During OSD Recovery, and Simultaneous 4KB RADOS Writes



Client Performance During Recovery		
Test	EC 6+2	3X Replication
4MB Writes	553.3 MB/s	643.1 MB/s
128KB Writes	47.3 MB/s	102.7 MB/s
4KB Writes	366 Obj/s	1367 Obj/s

CONCLUSION

In many of the tests, erasure coding has a significant CPU overhead over replication relative to the client IO performance that is obtained. In some cases CPU load is higher at the same level of performance, while in other cases CPU load is roughly the same while performance lower. A good system architecture involving erasure coding with Ceph likely will involve utilizing large objects and fast CPU cores to help offset the increase in the CPU utilization. Fast CPUs may especially be important during recovery and also potentially during scrub and deep scrub operations. In these tests, 12 2.0GHz XEON cores were used with 30 OSDs. During large object writes, the CPU was often pegged at 100% CPU indicating that greater write performance may have been achieved with faster CPUs. At the same time, reads and writes did not consistently show a dramatic increase in CPU utilization, but did show a dramatic decrease in performance versus replication.