Benchmark of a Cubieboard cluster

Journal of Physics: Conference Series **664** (2015) 052032 doi:10.1088/1742-6596/664/5/052032

Benchmark of a Cubieboard cluster

M J Schnepf, D Gudu, B Rische, M Fischer, C Jung and M Hardt

Steinbuch Centre for Computing, Karlsruhe Institute of Technology, Karlsruhe, Germany

E-mail: `matthias.schnepf@student.kit.edu`, `diana.gudu@kit.edu`, `bernd.rische@kit.edu`, `max.fischer@kit.edu`, `christopher.jung@kit.edu`, `marcus.hardt@kit.edu`

Abstract.
We built a cluster of ARM-based Cubieboards2 which has a SATA interface to connect a harddrive. This cluster was set up as a storage system using Ceph and as a compute cluster for high energy physics analyses. To study the performance in these applications, we ran two benchmarks on this cluster. We also checked the energy efficiency of the cluster using the preseted benchmarks. Performance and energy efficency of our cluster were compared with a network-attached storage (NAS), and with a desktop PC.

1. Introduction

In this paper we present the results of two benchmarks of a cluster of Cubieboards. In the first benchmark, we analyzed the performance and energy efficiency of the Cubieboard cluster as a storage system using the object storage system Ceph. In the second benchmark, we looked at an I/O-intensive high energy physics analysis and compared the energy consumption and processing rate of the Cubieboard cluster with those of a desktop PC.

2. Cubieboard Cluster Set-up

The Cubieboard is a single-board computer produced by Cubietech, that can run Android and several Linux distributions. The specifications of the version used in our setup, Cubieboard2[1], are listed in Table 1.

Table 1. Specification of Cubieboard2

CPU	AllWinnerTech SoC A20 @1.0 GHz, ARM©Cortex^{TM}-A7 Dual-Core
GPU	ARM©Mali400 MP2 Complies with OpenGL ES 2.0/1.1
RAM	1 GB DDR3
Network	1x 10/100 Mb/s ethernet
Storage	3.4 GB internal NAND flash, up to 32 GB on SD slot and 2.5 SATA up to 2 TB
Power	5VDC input 2A od USB OTG input

The cluster comprises 15 Cubieboards, each with an 8 GB SD card for the operating system and programs. The operating system used on the boards is the Debian based

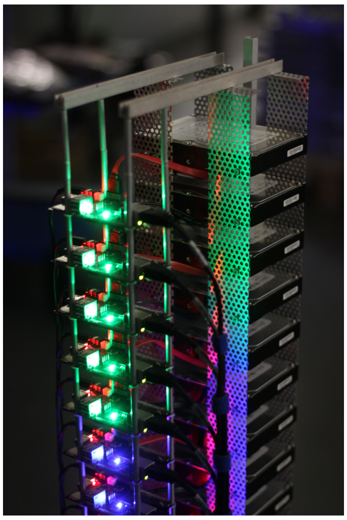

Figure 1. Picture of the Cubieboard cluster

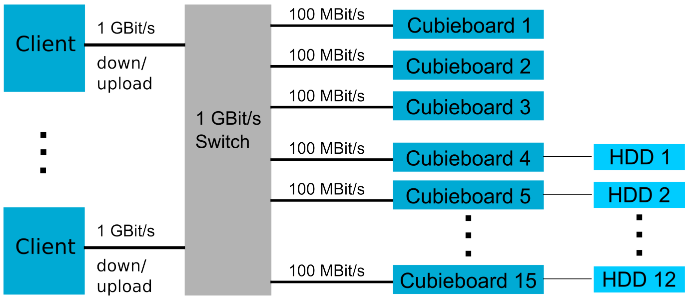

Figure 2. Hardware set-up of the Cubieboard cluster

Cubian. Twelve Cubieboards also have a 3.5-inch Seagate Baracuda ES hard disk drive (HDD) with a capacity of 750 GB, connected via SATA. The Cubieboards are connected to a 48-Port 1 Gb/s switch—Cisco Catalyst 2960G, see Figure 2. The cluster is powered by two 400 W power supplies: the first power supply is used by the 15 Cubieboards and the second one by the 12 HDDs. Each power supply is connected via a socket plug from Plugwise[2] to measure the used electric power every second.

3. Ceph

Ceph[3] is an open source object storage and file system. The main component of Ceph is RADOS (Reliable Autonomic Distributed Object Storage), which supports replication, failure detection and recovery. Ceph consists of two types of daemons: Ceph Monitor (MON) and Ceph Object Storage Daemon (OSD). The MON maintains the cluster map, while the OSD handles the I/O operations on the storage disk.

The location of an object in a Ceph cluster can be determined with the CRUSH-algorithm[4] by any member of the cluster, without the use of a central directory. This decentraliesed approch promises an excellent scalability. CRUSH is a pseudo-random hashing algorithm which maps a placement group (PG) (an aggregation of objects stored on the same OSD) to the OSD or OSDs where it will be stored. Objects are assigned to PGs based on their name. Every OSD supports a specified number of PGs.

3.1. Ceph Benchmark Set-up

The Cubieboard cluster uses Ceph version 0.80.8. We used three MONs, each running on an individual Cubieboard. The minimum recommended memory configuration for a Ceph Monitor is 1 GB of RAM, which coincides with the available memory on a Cubieboard2. Consequently, no other daemons (such as OSDs) can be co-located with the monitoring daemons on the same board. The other 12 Cubieboards run OSDs, which store data on local HDDs. The total amount of storage space available on the Cubieboard cluster with Ceph is 8.2 TB.

The maximum transfer rate of this setup is limited by the Cubieboard's network card to 100 Mb/s per node. Hence, for the entire Ceph cluster with 12 OSDs, the theoretical limit is $12 \times 100\,\mathrm{Mb/s} = 150\,\mathrm{MB/s}$.

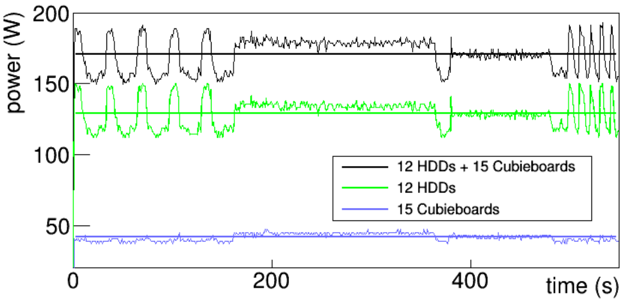

Figure 3. Average power consumption of HDDs and Cubieboards over time for the Ceph benchmark with 5 clients

For benchmarking the Cubieboard cluster as a storage system, 5 external client nodes were connected via 1 Gb/s Ethernet to the same switch as the Cubieboard cluster. To test the read and write speed over the network, the following benchmark procedure was employed: at first, one by one, the clients create their individual object pools. Every client then reads or writes with 4 parallel threads for a fixed duration, using the benchmark tool *rados bench*. The size of the objects written to the cluster is 4 MB. The benchmark duration is approximately 200 s for the write operations and 100 s for read operations. After the write and read phases, the pools are deleted. The time between the creation of two pools is 30 s, and the time between the deletion of two pools is 20 s. This time includes the creation of the pool, as well as sleep time necessary to allow Ceph to create all the PGs in a pool and bring the cluster into a healthy state. We also measured the power consumption, which include the entire benchmark procedure described above: the creation of pools, read and write operations, and deletion of pools. The power was averaged over the entire time of the benchmark. We performed the power measurements 5 times. The systematic uncertainties for the power measurement are 1 W on power per power supply and 3 s on duration.

3.2. Results of the Ceph Benchmark

The power consumption of the Cubieboard cluster over time is shown in Figure 3. Approximately 75% of the power is needed by the 12 HDDs, while the Cubieboards live up to the promise of low power consumption: the 15 boards use less than 50 W in total. The five peaks observed at the beginning of the time series correspond to the pool creation phase of the benchmark. The two long plateaus correspond to the write and read phases. The last five peaks depict the deletion of the five pools.

The minimal power needed by the Cubieboard cluster is (148.7 ± 1.4) W and the maximal power is (195.4 ± 1.4) W. The average power consumption of Cubieboards and HDDs increases with the number of clients doing I/O operations in parallel to the cluster. The average power with 5 clients is $(170.8 \pm 0.2 \pm 1.4)$ W. Three Cubieboards are only used for monitoring but comprise approximately 20% of the Cubieboard power consumption (approximately 8 W). By adding more Cubieboards that run OSD daemons, the percentage of power used by the monitors will decrease, while the cluster's total amount of storage and the I/O througput will increase.

A comparable NAS from QNAP with 12 HDDs (TS-1270U-RP) has an average power consumption (in operation mode) of 188.7 W, according to the vendor specification[5].

As seen in Figure 4 and Table 3, the maximal read bandwidth for one client is (84.6 ± 1.1) MB/s and the maximal write bandwidth is (73.7 ± 0.3) MB/s. The read and write bandwidth increase with the number of clients. The maximal bandwidth of the Cubieboard cluster was achieved in our Ceph benchmarks during sequential read tests with 5 parallel clients: (119.9 ± 1.3) MB/s, which is 79.3% of the theoretical maximum. The maximal write bandwidth was (115.7 ± 0.8) MB/s or 77.1% of the theoretical maximum.

Table 2. Power consumption of 15 Cubieboards, 12 HDDs and complete Cubieboard cluster with average, maximum and minimum values

Number of Clients: 1	avg. ± stat. ± sys	min. ± sys	max. ± sys
15 Cubieboards :	$(41.1 \pm 0.1 \pm 1.0)$ W	36.6 ± 1.0	45.1 ± 1.0
12 HDDs:	$(125.7 \pm 0.2 \pm 1.0)$ W	112.2 ± 1.0	152.3 ± 1.0
15 Cubieboards + 12 HDDs:	$(166.8 \pm 0.3 \pm 1.4)$ W	148.7 ± 1.4	193.2 ± 1.4
Number of Clients: 2	avg. ± stat. ± sys	min. ± sys	max. ± sys
15 Cubieboards :	$(41.6 \pm 0.1 \pm 1.0)$ W	36.6 ± 1.0	47.3 ± 1.0
12 HDDs:	$(127.9 \pm 0.1 \pm 1.0)$ W	112.2 ± 1.0	150.2 ± 1.0
15 Cubieboards + 12 HDDs:	$(169.4 \pm 0.2 \pm 1.4)$ W	148.7 ± 1.4	191.1 ± 1.4
Number of Clients: 3	avg. ± stat. ± sys	min. ± sys	max. ± sys
15 Cubieboards :	$(41.7 \pm 0.1 \pm 1.0)$ W	36.6 ± 1.0	47.1 ± 1.0
12 HDDs:	$(128.8 \pm 0.3 \pm 1.0)$ W	112.2 ± 1.0	152.3 ± 1.0
15 Cubieboards + 12 HDDs:	$(170.5 \pm 0.4 \pm 1.4)$ W	148.7 ± 1.4	195.3 ± 1.4
Number of Clients: 4	avg. ± stat. ± sys	min. ± sys	max. ± sys
15 Cubieboards :	$(41.8 \pm 0.1 \pm 1.0)$ W	36.6 ± 1.0	47.3 ± 1.0
12 HDDs:	$(129.0 \pm 0.3 \pm 1.0)$ W	110.0 ± 1.0	152.3 ± 1.0
15 Cubieboards + 12 HDDs:	$(166.8 \pm 0.3 \pm 1.4)$ W	146.6 ± 1.4	195.4 ± 1.4
Number of Clients: 5	avg. ± stat. ± sys	min. ± sys	max. ± sys
15 Cubieboards :	$(41.7 \pm 0.0 \pm 1.0)$ W	36.6 ± 1.0	47.3 ± 1.0
12 HDDs:	$(129.1 \pm 0.2 \pm 1.0)$ W	112.2 ± 1.0	154.4 ± 1.0
15 Cubieboards + 12 HDDs:	$(170.8 \pm 0.2 \pm 1.4)$ W	148.7 ± 1.4	193.2 ± 1.4

Table 3. Average write and sequential read speed from external clients on the Cubieboard cluster

Number of Clients	write (avg. ± stat.) MB/s	seq. read (avg. ± stat.) MB/s
1	73.74 ± 0.29	84.58 ± 1.08
2	94.18 ± 1.33	100.49 ± 2.44
3	105.83 ± 1.34	111.00 ± 1.91
4	111.92 ± 0.85	117.10 ± 1.22
5	115.67 ± 0.76	118.95 ± 1.34

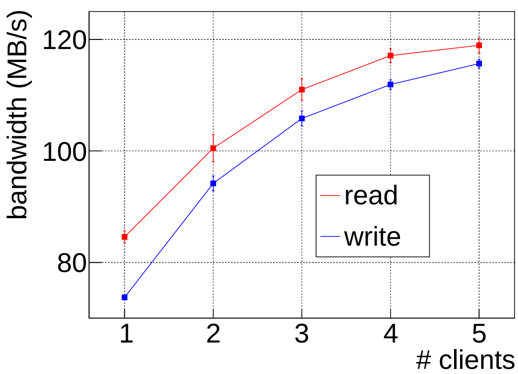

Figure 4. Write and sequential read bandwidth with 1 to 5 clients

4. High Energy Physics Benchmark

To evaluate the energy efficiency of the Cubieboard cluster for high energy physics analysis, we ran an I/O-intensive HEP analysis. We then compared the results with the ones obtained from running the analysis on a desktop computer. The analysis used for benchmarking was the study of muon events and calculation of invariant di-muon mass. The data sample includes 133 Mio. CMS events (simulated and observed). From these events, only the information about muons is saved in 644 files. Both systems had ROOT version 5.34.20 installed and read 191.5 GB from the I/O buffer. The total size of the files is 259.9 GB uncompressed and 107.5 GB compressed.

4.1. HEP Benchmark Set-up

The desktop PC used for comparison had an Intel Core i7 processor 3770K with 4 cores @3.5 GHz with hyper threading, 8 GB DDR3 RAM, 4 Seagate Barracuda ES 750 GB HDDs and operating system Arch Linux. The desktop PC analyzed compressed files, which is the default setting in ROOT. On the desktop PC the jobs run as coordinated threads, with two threads per HDD. All four HDDs have copies of all the files for the analysis. The measurement time is from the start of the first job to end of the last job.

Each Cubieboard runs the analysis program with files from the local HDD, so that the network does not become a bottleneck due to processing of non-local data.

The input files for the HEP analysis were uploaded and distributed on the Cubieboard cluster with Ceph.

To distribute the files better and increase the number of files per Cubieboard, each file was replicated once in the Ceph cluster. The HEP benchmark on the Cubieboard cluster runs analysis programs which read uncompressed files, because the analysis programs which read compressed files brought the CPU usage to 100% on all the Cubieboards.

We wrote a rudimentary batch sysytem to coordinate the run of HEP jobs on the Cubieboard cluster. The workflow of this system is shown in Figure 5. The master of the batch system runs on the same Cubieboard as one of the Ceph MONs, while the analysis runs on the Ceph OSD nodes. The communication between the master and the analysis nodes is done over SSH. At the beginning of the benchmark, the master broadcasts a list of all the files which are needed for the analysis. Then, the master sends requests for the lists of files located on each Cubieboard. A local program runs on each Cubieboard, which uses the CRUSH-algorithm to get the list of files and their paths on the local HDD. Every node then sends this list to the master. In the next step the master checks which files on the list have not yet been analysed by other nodes and then starts an analysis job for the file on the respective Cubieboard. Two jobs run on each Cubieboard in parallel, summing up to a total of 24 jobs on the Cubieboard cluster. The local result file of the analysis programs were uploaded in the Ceph cluster, to be available for external clients.

The measurement time of the benchmark starts when the master sends file list requests to all the nodes and ends when all jobs finished.

4.2. Results of HEP Benchmark

The maximal processing rate of the Cubieboard cluster is approximately 5 times lower than the processing rate of the desktop PC. The energy consumption is higher for the Cubieboard cluster than for the desktop PC (see Table 4 and Figure 6).

The processing rate of the Cubieboard cluster is lower than the data rate of a single HDD. Therefore the bottleneck of the Cubieboard Cluster is the CPU.

For the desktop computer, two parallel jobs read from one HDD, which causes an increase in random access and hence disk seek. With more parallel jobs per HDD, this effect will increase. The HDDs work up to 60% of their maximal throughput (~ 70 MB/s). The CPU usage is approximately 25%. Running our HEP benchmark with uncompressed files on the desktop PC is therefore slower than with compressed files; for this reason, we use the results of the analysis of compressed files on the desktop PC for the comparison with our Cubieboard cluster. The desktop PC is not at the maximum of its performance,

Journal of Physics: Conference Series **664** (2015) 052032 doi:10.1088/1742-6596/664/5/052032

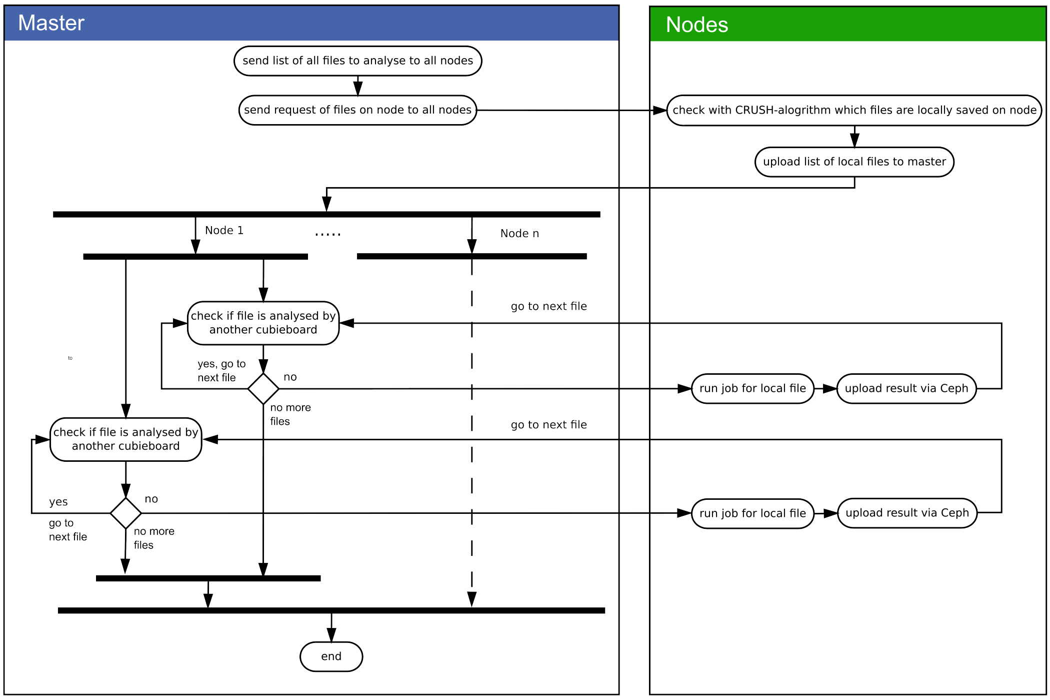

Figure 5. Workflow of batch-system for HEP benchmark on Cubieboard cluster

Table 4. Average energy consumption and processing rate from our Cubieboard cluster and the desktop PC for our HEP benchmark

System	Energy consumption (avg±stat±sys) Wh	Processing rate (avg±stat±sys) MB/s
Cubieboard cluster (15 Boards)	$74.4 \pm 0.4 \pm 1.6$	$33.7 \pm 28.3 \pm 3.0$
Cubieboard cluster (15 Boards + 12 HDDS)	$272.0 \pm 1.5 \pm 2.3$	$33.7 \pm 28.3 \pm 3.0$
desktop PC	$39.6 \pm 1.3 \pm 0.3$	$169.0 \pm 4.0 \pm 0.4$

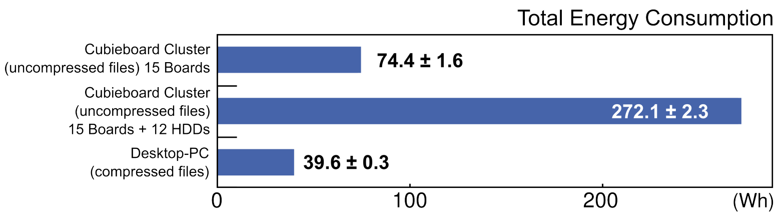

Figure 6. Total energy consumption of the Cubieboard cluster (with and without HDD) and the desktop PC for our HEP benchmark

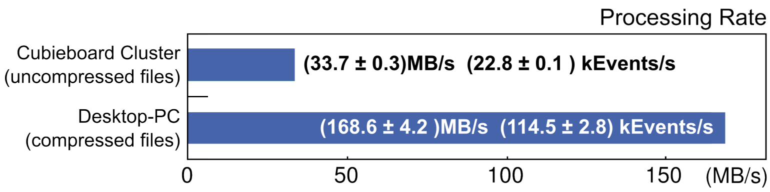

Figure 7. Processing rate of our Cubieboard cluster and the Desktop-PC for our HEP benchmark

but it reaches a saturation point, because jobs interfere with each other. The same problem can be observed with a parallel cluster with calculation nodes which read and write files over network: jobs interfere with each other with random disk access, or they reach the limit of network or I/O bandwidth of fileservers.

5. Conclusion

Ceph on a Cubieboard cluster is a good storage solution in terms of I/O throughput and energy consumption. Our setup uses approximately 80% of the maximal theoretical network speed and has a power consumption similar to a NAS solution with comparable hardware configuration. The limiting factor is the 100 Mb/s network card of the Cubieboards.

Nevertheless, an ARM-based cluster could be an option in future. The biggest limitation for the storage system use case (the network bandwidth) could be alleviated by using ARM boards with a 1 Gb/s network card, such as the Cubieboard3[6].

For HEP analysis, the setup is not power efficient. Most of the energy is used by the HDDs. Even considering only the Cubieboards' energy consumption, the Cubieboard cluster is not more efficient than a desktop PC. The problem of the Cubieboard cluster is not that jobs interfere with each other, but rather the low performance of the CPU and the energy efficiency.

To improve the energy efficiency in the HEP analysis use case, it is important to use a fast and more energy efficient storage medium, such as SD cards. That would provide the bandwidth needed by the analysis at a lower energy consumption.

Another possible improvement is to increase the processing power for the same electric power consumption. There exist already ARM boards with quad-core CPUs, such as the Raspberry Pi 2[7]. Furthermore, new ARM CPUs with higher clock rates are being developed, yielding better performance and offering 64-bit support[8]. In the near future it could be that ARM boards, such as the Cubieboard2, are powerful and energy efficient enough to be used for high energy physics analysis.

[1] http://docs.cubieboard.org/products/start from 30.3.2015.
[2] https://www.plugwise.com/ from 2.4.2015.
[3] http://ceph.com/ from 2.4.2015.
[4] Weil S, Brandt S A, Miller E L and Maltzahn C 2006 *SC '06 Proceedings of the 2006 ACM/IEEE conference on Supercomputing* Article No. **122**.
[5] https://www.qnap.com/i/de/product/model.php?II=38&event=2 from 30.4.2015.
[6] http://docs.cubieboard.org/tutorials/cubietruck/start from 8.4.2015.
[7] http://www.raspberrypi.org/products/raspberry-pi-2-model-b/ from 8.4.2015.
[8] http://www.arm.com/about/newsroom/arm-launches-cortex-a50-series-the-worlds-most-energy-efficient-64-bit-processors.php from 9.4.2015.