# Power Efficiency of a SBC Based Hadoop Cluster

Basit Qureshi[1(✉)] and Anis Koubaa[1,2,3]

[1] Prince Sultan University, Riyadh, Saudi Arabia
{qureshi,akoubaa}@psu.edu.sa
[2] Gaitech Robotics, Hong kong, China
[3] CISTER, INESC-TEC, ISEP, Polytechnic Institute of Porto, Porto, Portugal

**Abstract.** Security, reliability and energy efficiency are key requirements in any smart cities applications. In this paper, we investigate the use of a power efficient cloud cluster composed of Single Board Computers (SBC). A Low-Cost Cloud Computing Cluster (LoC4) is designed using Raspberry Pi and Odroid Xu4 Single Board Computers. The LoC4 cluster is thoroughly tested for performance and power efficiency using TeraGen and TeraSort benchmarks that are used primarily for stress testing the Hadoop cluster. Results show that Odroid Xu4 computers consume more energy compared to RPi computers, however the overall energy consumption in the cluster for completion of a task was 91% and 72% better for Odroid Xu4 with workloads of 800 MB and 400 MB respectively.

**Keywords:** Hybrid cloud computing · Low cost cluster · Power efficiency

## 1 Introduction

In recent times, Cloud computing technology is widely being adopted by businesses and organization. A major motivation for this change is the decrease in maintenance of infrastructure, deployment and management overheads as well as overall reduced operating costs. Energy consumption in data centers is a major concern for green computing research. NRDA [1] estimated in 2013, in US alone, the data centers consumed 91 billion kilowatts hours (kWh) of energy, which accounts to 5% of total energy consumed per year. It is estimated to increase by 141 billion kWh every year until 2020, costing businesses $13 billion annually in electricity bills and emitting nearly 100 million metric tons of carbon pollution per year. Rightly so, the environmental impact of maintaining large computational infrastructure and data centers is a big concern, prompting the need for research in "greener" technologies for data centers.

Apache Hadoop framework [2] is a popular platform commonly used for analysis of data intensive operations and is widely used for research in Big Data analysis where large volumes of data cannot be analyzed using traditional technologies. Hadoop's Map/Reduce [3] has become a benchmark tool for comparing performance of various architectures for compute, network, storage and IO operations [7, 8]. Recent works have provided an opportunity for further investigating efficiency of Map/Reduce workloads in a Hadoop clusters. Authors in [4, 5] and [6] outline the need for understanding the


R. Mehmood et al. (Eds.): SCITA 2017, LNICST 224, pp. 52–60, 2018.
https://doi.org/10.1007/978-3-319-94180-6_7

potential for energy saving in MapReduce Jobs in the context of CPU-bound, IO-bound or network-bound workloads. Tiwari et al. in [7] argue that varying MapReduce parameters have a significant impact on computation performance and energy consumption for typical MapReduce workloads.

On the other hand, deploying private cloud requires large investment in addition to costs incurred due to energy usage, maintenance and software upgrades etc. For small healthcare organizations and universities, it is cost abhorrent to establish expensive datacenters and maintain these. Recently, researchers in academia have used low cost and low power cloud clusters composed of Single Board Computers (SBC) [9, 10]. Baun [9] developed a Hadoop based cluster using 8 Raspberry Pi Model 2B Computers. Qureshi et al. in [10] utilize a Hadoop cluster composed of 20 Raspberry Pi 2B and 20 Odroid Xu-4 computers for image processing applications.

In this paper, we investigate power consumption behavior of Hadoop Map/Reduce jobs utilizing the LoC4 cluster. LoC4 is an affordable and low-cost cloud computing cluster for research in universities and academic institutions. The contributions of this work are in two folds. We first describe the design and implementation details for the LoC4 cluster as well as deployment settings and parameters. Second, we extensively test this cluster and measure power consumption for medium and heavy intensity workloads. We also study the impact of heavy workloads on the two types of SBCs used in this cluster in terms of power efficiency and, completion time of various stages of applied workloads.

The rest of the paper is organized as follows. Section 2 describes the LoC4 cluster. Section 3 presents the power efficiency results of the cluster followed by conclusions and discussion in Sect. 4.

## 2 LoC4: SBC Based Cluster

A single-board computer (SBC) is a complete computer built on a single circuit board [9, 10]. A SBC incorporates microprocessor(s), memory, I/O as well as host of other features required by a functional computer. In this work, we build a cluster using Raspberry Pi Model 3B [12] and Hard Kernel Odroid Xu-4 [11] SBCs.

Figure 1 shows the LoC4 cluster. The LoC4 cluster is composed of 11 SBCs of while five RPi Model 3B and six Odroid Xu-4 computers are interconnected with power supplies, network cables, Storage modules, connectors and cases. All the Raspberry Pi computers are equipped with 16 GB Class-10 SD cards for primary bootable storage. The Odroid Xu-4 devices are equipped with 32 GB eMMCv5.0 modules. The Odroid Xu-4 s are housed in a compact layout racks using M2/M3 spacers, nuts and screws. The Raspberry Pi 3B's are housed in cases and connected to the rest of the cluster. Currently each Raspberry Pi computer is individually supplied by 2.5 Amp power supply; each Odroid Xu-4 computer is supplied by a 4.0 Amp power supply that provides ample power for running each node.

Each SBC's network interface relates to a Cat6e Ethernet cable through the RJ-45 Ethernet connector. All Ethernet cables connect to a 16-port Cisco switch which in turn is connected to the university network equipment. The size of the cluster can be easily scaled by introducing a core switch that connects to the 16-port switch. Additional

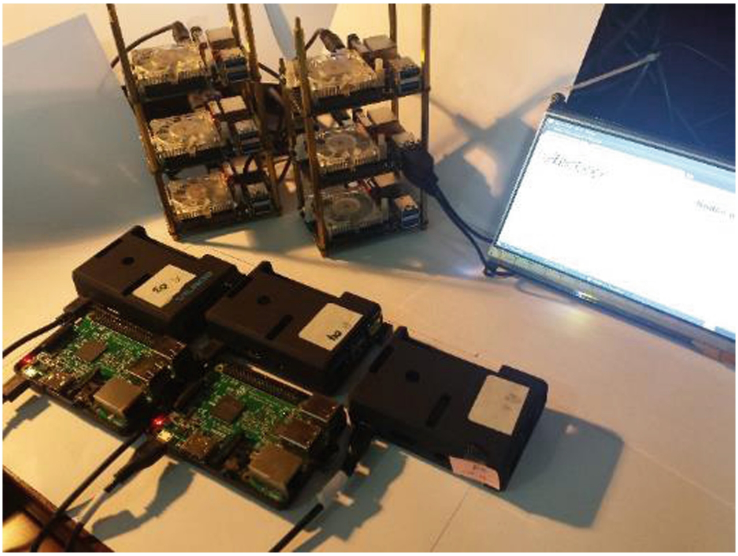

**Fig. 1.** The LoC4 cluster with 6 Xu4 (Rack mounted) and 5 RPi 3B in cases.

nodes can be included in the network configuration scaling the size of the cluster. For Hadoop deployment, one of the Xu-4 SBC is used as the master node whereas the rest of the nodes server as slaves' nodes. We compare the performance of Xu-4 sub cluster (composed of five Xu-4 nodes) with RPi sub cluster also composed of five RPi 3B nodes. The purchase cost for all equipment for the LoC4 Cluster was as low as $1179.

Hadoop version 2.6.2 was installed due to availability of YARN daemon which improves the performance of the map-reduce jobs in the cluster. To optimize the performance of these Clusters, `yarn-site.xml` and `Mapred-site.xml` were configured with 512 MB of resource size allocation. The primary reason for this setting is the limited amount of RAM available in the RPi Model 3B. The Raspbian operating system as well as the shared GPU memory bus consume over 200 MB or RAM out of a total of 1 GB. The default container size on the Hadoop Distributed File System (HDFS) is 128 MB. Each SBC node was assigned a static IPv4 address based on the configuration and all slave nodes were registered in the Master node. For YARN Resource manager, we allocated up to 4 cores which means up to 4 containers can execute per node (one container per core). The replication factor for HDFS is 2 which means only two copies of each block would be kept on the file system.

## 3 Power Efficiency of LoC4

As mentioned earlier in our cluster setup, we use a single SBC to run the master node of the cluster, which executes the namenode as well as the YARN ResourceManager Hadoop applications. The slave nodes execute the datanodes as well as the YARN NodeManager tasks. In order to avoid influence of the namenode, we attach the power measurement equipment to the clusters slave nodes only and collect power consumption data. The WattsUp Pro.net[1] meter is capable of recording power consumption in terms of watts, each reading is collected every second and is logged in the meter's onboard memory. The meters are initialized 10 s before each TeraGen and TeraSort job is initiated and stops reading 10 s after the job is completed. In addition to power consumption readings, we also periodically measured (every minute) the CPU temperature (Celsius) for both RPi's as well as Odroid Xu-4 boards in the cluster. Average Power Consumption in terms of kilo watts per hour $W_t$, can be given by the following equation:

$$W_t = \Delta t \,.\, \frac{E}{kWh} \tag{1}$$

Where is $\Delta t = t_f - t_s$ the interval time from start of the task to its completion, $E$ is the Energy consumed in terms of Watts as recorded using the Wattsup meter.

We study the power consumption on LoC4 cluster using two types of applications widely used in benchmarking of Hadoop clusters:

- The TeraGen is light to middle weight workload benchmark that stresses CPU and IO (Network, Storage etc.) in the cluster
- TeraSort benchmark gives intensive workload stressing CPU and IO (heavy workload).

In this experimental setup, we isolate RPi devices from Xu4 devices and study the performance of the two using medium workload (TeraGen) and heavy workload (TeraSort) benchmarks with 100 MB, 200 MB, 400 MB and 800 MB datasets. The experiments are repeated 10 times for each run to obtain a weighted average.

Figure 2(a) shows the comparison of power consumption as well as CPU Temperature for both clusters for the TeraGen using 800 MB datasets. It can be noted, the power consumption for Xu4 devices peaks at 18.2 watts whereas RPi devices consumes at most 12.1 watts. The temperature on RPi SBC mostly stays within the range 29–32C, on the other hand Odroid Xu-4 SBC's are equipped with a cooling fan. At 45 °C the fan turns on due to the built-in hardware settings yielding in increased power consumption on the Odroid Xu-4. Since TeraGen is IO bound job, initially mappers start executing and writing to the HDFS, as the progress continues some of the mappers complete the tasks assigned. Consequently, we observe reduction in the overall power consumption of the cluster, this effect can be clearly observed in Fig. 2(a) with 800 MB Data size for both clusters.

Figure 2(b) shows the power consumption for both clusters when TeraSort is used. We observe that TeraSort requires more time for completion. Initially mappers read through the input files generated by TeraGen and stored in HDFS, as the TeraSort shuffle process for keys and values initiates, we observe increased power consumption which continues until the mappers as well as majority of reduce jobs complete. As the mappers continue to complete the tasks, the incoming results start processing in the

[1] Wattsup: https://www.wattsupmeters.com/secure/products.php?pn=0.

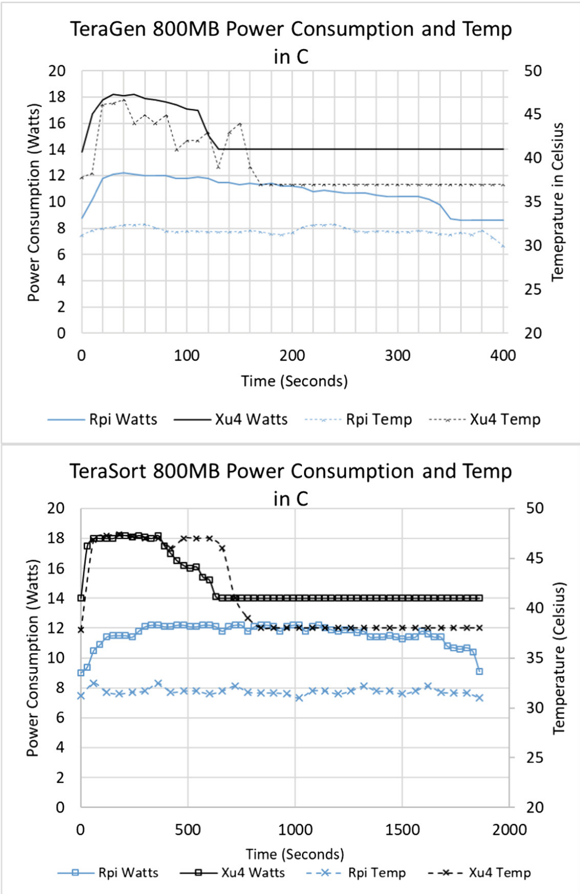


**Fig. 2.** Power consumption and temperature against execution time for (a) TeraGen with 800 MB dataset (b) TeraSort with 800 MB dataset.

reduce jobs. Before the completion of all map functions, the reduce functions initiate sorting and summarizing process requiring CPU as well as IO resources towards completion of the tasks.

We plot the percentage of map and reduce completion against the power consumption for RPi devices with and 800 MB data size in Fig. 3(a) and for Xu4 devices in Fig. 3(b). As can be seen, the percentage of maps and reduces completed correlates with the power consumption. In particular, when the map and reduce complete, the power consumption decreases therefore highlighting underutilized nodes in the clusters. Both TeraGen and TeraSort exhibit different power consumption. TeraSort on both sets of SBCs has a relatively long phase of higher power consumption from initialization of map jobs until about 80% of map jobs completion indicating high CPU utilization. Afterwards, the power consumption decreases slightly fluctuating while both map and reduce jobs are executing in parallel. Finally, the power consumption steadies with minor tails and peaks in the plot towards reduce jobs completion. We observe that the trends for power consumption relevant to task completion are similar for larger data sizes used in this study. Table 1 shows the comparison of completion times for various TeraGen and TeraSort jobs. It also shows the average power consumption for TeraSort Tasks in terms of kWh.

It is difficult to monitor and normalize the energy consumption for every test run over a period of time. We observed that the MapReduce jobs in particular, tend to consume more energy initially while map tasks are created and distributed across the cluster, while a reduction in power consumption is observed towards the end of the job. For the computation of power consumption, we assumed max power utilization (stress mode) for each job, during a test run.

## 4 Related Works

A leading incentive for adoption of cloud computing is the reduction of installation and operational cost for small businesses and enterprises. Apache Hadoop has proven to be an effective platform using the MapReduce framework allowing quick and parallel processing of data. Setting up cloud infrastructure in universities could be a very costly endeavor. Hadoop Clusters built using Single Board Computers have been presented in recent research work.

In 2016, Baun in [9] presented the design of a cluster geared towards academic research and student scientific projects building a 8-node Raspberry Pi Model 2B cluster. They develop this cluster and conduct a thorough study of quality of service parameters such as CPU execution time, Memory and IO performance etc. Qureshi et al. in [10] develop a 40 node cluster using Odroid Xu4 devices and Raspberry 2B SBCs. They utilize this cluster for image analysis using popular image processing libraries. Earlier in Tso et al. [13], researchers built a small-scale data center consisting of 56 RPi Model B boards. The Glasgow Raspberry Pi Cloud offers a cloud computing testbed including virtualization management tools. In 2013, Whitehorn [14] presented the first ever implementation of a Hadoop cluster using five Raspberry Pi Model B nodes.

Gomez in [15] use Raspberry Pi Computers as sensors in an industrial environment. They were able connect various SBCs over the network to collect sensory information. In 2016, Fernandes et al. [16] used Odroid Xu4 devices for an image processing application for matching patterns in different sketches. In both of these works, researchers demonstrate the benefit of using SBC's to accomplish application specific tasks.

To the best of our knowledge, there is a lack of literature in extensive study for evaluation of power consumption in SBC based Hadoop clusters. In this paper we present a Hadoop cluster using two kinds of SBCs, Raspberry Pi 3B and Odroid Xu4 devices. We extensively study the power consumption for data and compute intensive TeraSort and contract the average energy consumption for both sets of devices.

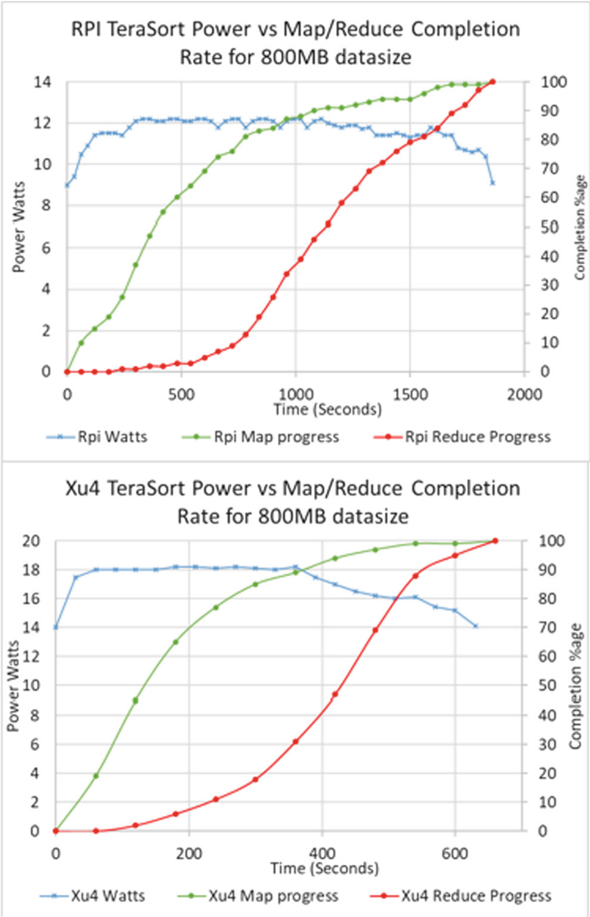


**Fig. 3.** TeraSort power consumption with 800 MB dataset size for (a) RPi (b) Xu4.

**Table 1.** Task completion time and average power consumption against various workloads for SBC devices in LoC4

| | Workload (MB) | Task completion time (seconds) | | Average power consumption for TeraSort task (kWh) |
|---|---|---|---|---|
| | | TeraGen | TeraSort | |
| RPi devices | 100 | 52.8 | 210.1 | 0.68 E-3 |
| | 400 | 122.3 | 681.3 | 2.23 E-3 |
| | 800 | 343.9 | 1873.2 | 6.13 E-3 |
| Xu4 devices | 100 | 13.2 | 59.2 | 0.29 E-3 |
| | 400 | 58.1 | 325.9 | 1.61 E-3 |
| | 800 | 124.3 | 692.7 | 3.42 E-3 |

## 5 Conclusions

In this paper, we investigate the use of SBC in a low-cost cloud computing cluster. We consider two kinds of popular platforms Raspberry Pi 3B and Odroid Xu-4 using ARM Cortex Processors. The LoC4 cluster was deployed comprising of 11 SBCs interconnected in a network topology over a gigabit Ethernet. Hadoop was installed on the cluster with configuration to suit the SBC's memory and storage requirements. Results from these studies show that while SBC based clusters are energy efficient overall, the operation cost to performance ratio can vary based on the workload. The Odroid Xu4 computers consume more energy compared to RPi computers, however the overall energy consumption for completion of a task was 91% and 72% better for workloads of 800 MB and 400 MB respectively, in favor of Xu4 Computers. We conclude that, for light to middle workload applications, SBC clusters are a very cost effective and power efficient solution for deployment in hybrid cloud environments.

**Acknowledgements.** This work is partially funded by the Robotics and Internet of Things Unit (RIoTU) at Prince Sultan University.

## References

1. Data Center Efficiency Assessment: Scaling up energy efficiency across the data center industry, Natural Resources Defense Council, August 2014. https://www.nrdc.org/sites/default/files/data-center-efficiency-assessment-IP.pdf
2. The Apache Hadoop Project, 2017. http://www.hadoop.org
3. Feller, E., Ramakrishnan, L., Morin, C.: Performance and energy efficiency of big data applications in cloud environments: a Hadoop case study. J. Parallel Distrib. Comput. **79–80**, 80–89 (2015)
4. Conejero, J., Rana, O., Burnap, P., Morgan, J., Caminero, B., Carrión, C.: Analyzing hadoop power consumption and impact on application QoS. Future Gener. Comput. Syst. **55**, 213–223 (2016)
5. Krish, K.R., Iqbal, M.S., Rafique, M.M., Butt, A.R.: Towards energy awareness in hadoop. In: 4th International Workshop on Network-Aware Data Management, New Orleans, LA, pp. 16–22 (2014)
6. Tiwari, N., Bellur, U., Sarkar, S. Indrawan, M.: Identification of critical parameters for MapReduce energy efficiency using statistical Design of Experiments. In: 2016 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), Chicago, IL, pp. 1170–1179 (2016)
7. Bilal, K., et al.: A taxonomy and survey on green data center networks. Future Gener. Comput. Syst. **36**, 189–208 (2014)
8. Deng, X., Wu, D., Shen, J., He, J.: Eco-Aware Online Power Management and Load Scheduling for Green Cloud Datacenters. IEEE Syst. J. **10**(1), 78–87 (2016)
9. Baun, C.: Mobile Clusters of Single Board Computers: An Option for Providing Resources to Student Projects and Researchers. SpringerPlus PMC. Web. **5,** 360 (2016)
10. Qureshi, B., Javed, Y., Koubâa, A., Sriti, M.-F., Alajlan, M.: Performance of a low-cost hadoop cluster for image analysis in cloud robotics environment. Procedia Comput. Sci. **82**, 90–98 (2016)

11. Odroid Xu-4 detailed specifications. http://www.hardkernel.com/main/products/prdt_info.php?g_code=G143452239825. Accessed 10 Oct 2016
12. The Raspberry Foundation. https://www.raspberrypi.org. Accessed 20 Nov 2015
13. Tso, F.P., White, D.R., Jouet, S., Singer, J., Pezaros, D.P.: The Glasgow raspberry pi cloud: a scale model for cloud computing infrastructures. In: IEEE 33rd International Conference on Distributed Computing Systems Workshops (ICDCSW). IEEE, Philadelphia, PA, USA, pp 108–112 (2013)
14. Whitehorn, J.: Raspberry flavored Hadoop. http://www.idatasci.com/uploads/1/4/6/6/14661274/jamiewhitehorn_raspberryflavouredhadoop_annotated.pdf
15. Gómez, A., Cuiñas, D., Catalá, P., Xin, L., Li, W., Conway, S., Lack, D.: Use of single board computers as smart sensors in the manufacturing industry. Procedia Eng. **132**, 153–159 (2015)
16. Fernandes, S.L., Bala, G.J.: ODROID XU4 based implementation of decision level fusion approach for matching computer generated sketches. J. Comput. Sci. **16**, 217–224 (2016)