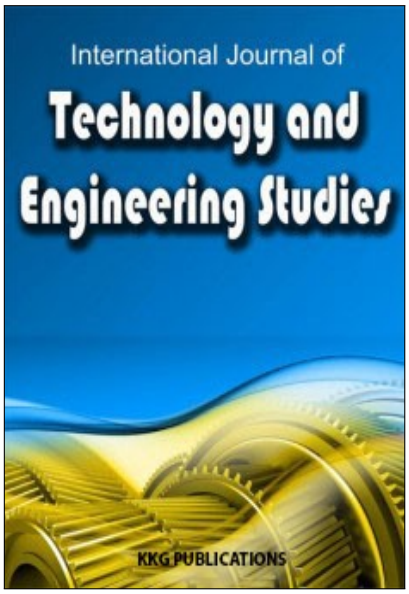

Key Knowledge Generation
Publication details, including instructions for author and
subscription information:
http://kkgpublications.com/technology/

Research on Single-Board Computers Clustering the Computing Performance

CHAO HSI HUANG [1], MIN HAO CHANG [2], I HSUAN LIN [3]

[1, 2, 3] Department of Computer Science and Information Engineering, National Yilan University, Yilan, Taiwan

Published online: 24 October 2016

To cite this article: C. H. Huang, M. H. Chang and I. H. Lin, "Research on single-board computers clustering the computing performance," *International Journal of Technology and Engineering Studies,* vol. 2, no. 5, pp. 125-133, 2016.
DOI: https://dx.doi.org/10.20469/ijtes.2.40001-5

To link to this article: http://kkgpublications.com/wp-content/uploads/2016/2/Volume2/IJTES-40001-5.pdf

KKG PUBLICATIONS

International Journal of Technology and Engineering Studies
vol, 2, no. 5, pp. 125-133, 2016

IJTES

RESEARCH ON SINGLE-BOARD COMPUTERS CLUSTERING THE COMPUTING PERFORMANCE

CHAO HSI HUANG [1], MIN HAO CHANG [2*], I HSUAN LIN [3]

[1, 2, 3] Department of Computer Science and Information Engineering,
National Yilan University, Yilan, Taiwan

Keywords:
SBC(Single-Board Computer)
Raspberry Pi
Clusters
Hadoop
Map/Reduce

Received: 27 July 2016
Accepted: 03 September 2016
Published: 24 October 2016

Abstract. This research checks effectiveness of test of a single computer (CPU I5-4440 3.1 GHz and 8G memory). And tests the use of Raspberry Pi3 erected six groups of Hadoop clusters. This study puts forward setup clusters called Hadoop on the new SBC (Single-Board Computer) Raspberry Pi. This Hadoop cluster provides computing and storage services. Pattern recognition was run on the SBC clusters and the single computer; they are similar in price; their performance was observed. It is revealed the SBC clusters were better than single computers in performance that were increased by about 20% and increased number of SBC can improve cluster processing speed. In this research, experimental data were apparent, and Hadoop was used by Hadoop Distributed File System (HDFS) in the data storage; it has better security than a single computer. This research can be used as Hadoop infrastructure in the future.

INTRODUCTION

In recent years, vigorous development of computer science and technology has taken place. Single-Board Computer (SBC) [2] development is very rapid, smart phone and tablet are from SBC group, originally used the PC, a lot of change in use of the SBC, because PC may be substituted, for example like the Raspberry Pi.

Raspberry Pi is single-board computer of credit card size that can install many OS and coding, visit the website, file processing like PC, etc. The Raspberry Pi Foundation was laid in 2012 with its first launch. Raspberry Pi is Linux based operating system kernel, and the new Raspberry Pi can install Ubuntu operating system that only installed on the PC before.

If Raspberry Pi achieves better computing performance, it must be built as Raspberry Pi Cluster [6] [7], and through a cluster of distributed computing can become outstanding computing unit. Very fast evolution of technology products, but also emerged out of carbon reduction issues, SBC computer is small in size and low power consumption, improves many hardware and has far exceeded the former SBC computer.

LITERATURE REVIEW

In order to more clearly present research, we will introduce the SBC Cluster and Hadoop [13].

Single-Board Computer (SBC)

Now, we can see the products of smart phone, tablet or Raspberry Pi, Arduino on market. Those are called SBCs. The computer will be composed of part on IC by SBC, including Microprocessor, Memory, IO interface, RAM, ROM and so on. In addition [8], something about external device, like Keyboard, Slots.

You can connect other external components, sensors and so on. SBC is better than SCM in terms of features. It is suitable for production process control. Also, it can be operated on Breadboard, and SBC can be applied to teach.

Raspberry Pi

Raspberry Pi is SBC and it is developed by Raspberry Pi Foundation. Raspberry Pi has USB, HDMI, Ethernet, and RCA audio terminal interface which is the concept given in the United States ATMEL semiconductor plant in 2006. Prior to release of the final version manufactured in 2011, the first batch of circuit boards were to be officially released in March 2012, Raspberry Pi released the second generation product in February 2015 and released the third generation in March 2016, enhanced CPU performance and increased hardware device. Raspberry Pi can be installed through the System Manager (NOOBS), also available on the official website to download and install other operating systems. Add Wi-Fi module, Bluetooth module on the third-generation hardware for improving the user's convenience.

*Corresponding author: Min Hao Chang
†Email:spurs20406@gmail.com

TABLE 1
RASPBERRY PI

	Pi 1	Pi 2	Pi 3
CPU	ARM 1176JZF-S 700MHz	ARM Cortex-A7 900MHz	A 1.2GHz quad-core ARMv8 CPU
GPU	Broadcom Video Core IV 1080p 30 h.264/MPEG-4 AVC		
RAM	512MB	1024MB	1024MB

Arduino

Arduino is a microcontroller board which can be connected to a variety of electronic devices on board, if you need.

Arduino can be connected to LED lights, speakers, motors, switches, Ultrasound devices, water level sensors, CO_2 sensors, temperature and humidity sensors, infrared transmitter and receiver, LCD monitor and Ethernet, Wi-Fi, XBee, Bluetooth, RFID, GPS other communication module, and then written the automatic control program by the development platform, you can use Arduino to do a variety of automatic controls. For example, you can control the lamp to be bright and dark, and adjust air conditioning by temperature sensors or control the fan.

Arduino and Raspberry Pi are from SBC group, however Arduino is different from Raspberry, Arduino is used for automatic control, Pi and Arduino both differ greatly in hardware, Arduino can't be installed in Linux OS, so this research can't be used for Arduino setup cluster and experiment. And the mobile, smart phone and tablet belong to SBCs. Computer, phone and tablet aren't Open Source which OS and hardware by the manufacturers decided. They can't install other OS. So, we choose Raspberry Pi as research equipment, because it can be freely installed on OS and performance is good. The Raspberry Pi can be connected to many sensors on board which are developed and applied in Internet of Things (IoT) [3].

SBC Hardware Comparison

TABLE 2
SBC HARDWARE

	Arduino	Raspberry Pi 3	tablet
CPU Clock	16MHz	1.2GHz	
GPU	X	Broadcom Video Core IV 1080p 30 h.264/MPEG-4 AVC	
RAM	2 KB	1024 MB	Manufacturers decided
Vol	5V	5V	
Storage	32KB	SD	card Size
OS	X	Linux Base	
OpenSource	V	V	X

Cluster

First introduce the openmpi and the mpich. These are software in cluster, both the cluster software are implemented by the MPI (Message Passing Interface) standard [19] [20] [21], however, the standard MPI cluster software in the development program is closer to stand-alone system, different from Hadoop Map/Reduce.

MPI of running the program is a one-time task scheduling which is the advantage, if the system applies for many processes. These are processes that will run together and be dead. These are processes that will be synchronized with each other and can send data. Through the MPI packaging, the operation and sending data will become more convenient.

The MPI standard with respect to the Hadoop has two drawbacks. The first, fault tolerance is very poor, because the MPI cluster has one process dead, the mission will be stopped on MPI standard. But the Hadoop cluster has the task assigned which has Management System for monitoring the implementation process of all nodes; if the single node dies (task failed), the mission will not crash; the Hadoop has high fault tolerance [18]. The second, MPI standard cluster is smaller than Hadoop cluster in scale, the Hadoop node is increased to ten thousands which can handle any problem, but MPI standard cluster can't.

On Hadoop cluster system the task will be assigned by Management System. The distributed computing system has good fault tolerance.

We need stable performance operation for a long time and high fault tolerance. So, we choose the Hadoop as research cluster system.

Hadoop Map/Reduce

The Hadoop cluster software has been developed by Apache company, it is reliable Open Source [1] and provides Distributed computing [5] [12]. Hadoop architecture is composed of two. One of the architectures is HDFS (Hadoop Distributed File System) [9] [14]. User can store data to Hadoop cluster. Hadoop will disperse data and store it in several hundred or several thousands of computers. Another architecture is the Map/Reduce, it can analyze a big program into many small programs. And then, each node will do the processing.

Map/Reduce is a Hadoop architecture, it can easily compose, and on cluster can parallel process big data [10] [11] [15] [16] [17]. The Map/Reduce will divide the mission into small tasks, Map tasks to a parallel computing on nodes, this architecture will sort the Map output to the relay result, usually this job input and output are stored in a file system, the Map / Reduce structure is responsible for scheduling tasks, monitoring, and re-executing the failed task.

The computing and the storage are usually on same node, that is representative of Map/Reduce and HDFS at the same time running on nodes, this configuration is allowed, and can efficiently schedule tasks. The data are stored on the nodes that can have improved high storage space on cluster.

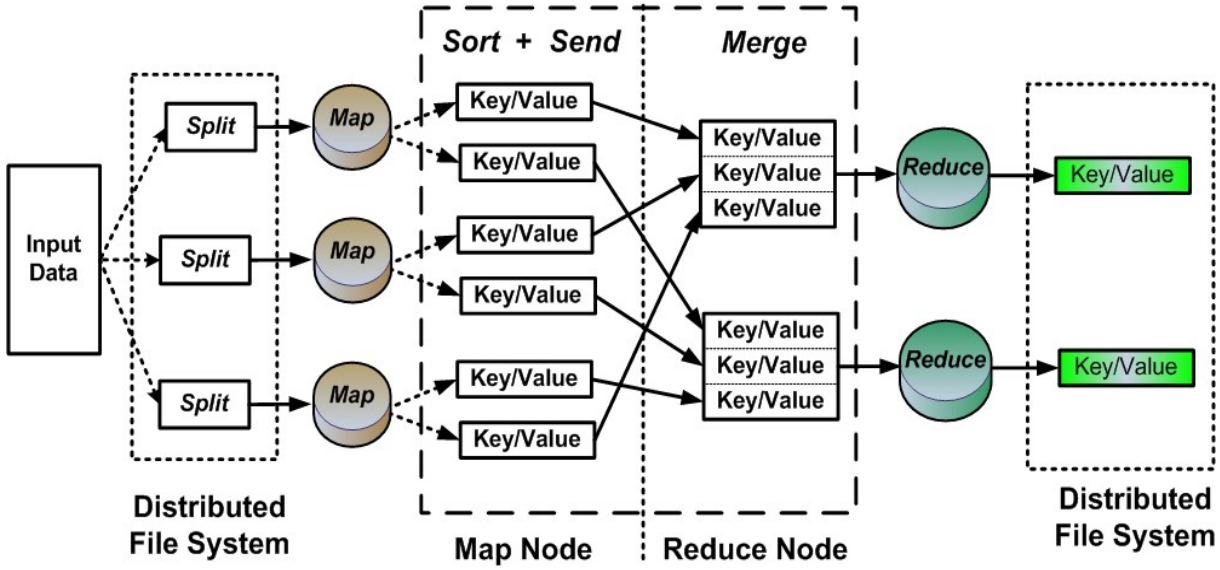

Fig. 1 . Map/reduce work

METHOD AND MATERIALS

In this section we discuss design and implementation of the Pi Cluster and Map/Reduce code, highlighting the rationale as well as the operational and design details of the individual components. Hadoop has the four main setup files, core-site.xml, hdfs-site.xml, mapred-site.xml and yarn-site.xml. We emphasize especially on mapred-site.xml and yarn-site.xml setting the value.

Hadoop Setting

When we set up Hadoop cluster, we must set all the nodes. Some of the parameters must be same. But each node can be different about Resource allocation and CPU and RAM of settings. Each node is independent individual, then node will be managed by Master. Each node will return resources and setting, then it is assigned workload by Master.

Mapred-Site.XML

Usually Map and Reduce startup time is not same, when Map execution reached 5%, it will start the Reduce, it's not any program or any algorithms using this parameter very well, sometimes the algorithms take a long time in Map, but the Reduce has extremely short time, if Map completion rate to 5% started the Reduce, Reduce has been waiting for Map that will take up resources, so we must adjust the Map and Reduce startup time. And we should set Map/Reduce memory to maximum.

TABLE 3
MAPSITE SET

Parameter ID	Remark
mapred.reduce.slowstart.completed.maps	Default value 0.05(5%), max 1(100%) ,We set 0.98
mapreduce.map.memory.mb	Value set 682MB
mapreduce.reduce.memory.mb	Value set 1364MB

Yarn-Site.XML

Adjust number of virtual cores used on node which by default is 8, but Raspberry Pi CPU clock is not high, so adjusted number of virtual cores is same as the physical cores number. And adjusting the run time can use the total memory size on node, Hadoop has used the virtual memory, because physical memory size is 1024MB which is insufficient.

And we set Job that can be used as max/min memory, The yarn can analyze the set Job memory minimum and total memory (virtual and physical), determine the number of containers. We test more memory values, the 682mb is better.

TABLE 4
YARN-SITE SET

Parameter ID	Remark
yarn.nodemanager.resource.cpu-vcores	Virtual CPU amount, default 8, we set 4.
yarn.nodemanager.resource.memory-mb	Total memory, we set 2G.
yarn.scheduler.minimum-allocation-mb	Job can used min memory, default 1G, we set 682mb.
yarn.scheduler.maximum-allocation-mb	Job can used max memory, default 8G, we set 2048mb.
yarn.nodemanager.vmem-pmem-ratio	Virtualization memory used ratio, default ratio 2.1, we set 4.

Experimental Program Algorithm

We went through Open CV which offers SURF algorithms, when image is input, we will acquire the image feature.

Each feature is composed of sixteen blocks, and each block has four values. Individual expression is sum of horizontal direction value, sum of the absolute value of the horizontal direction, sum of Vertical value and sum of the absolute value of the vertical direction. So, each feature is 16*4=64 vector. We want to compare the feature and other features are similar, as long as calculate the dimensional spaces for features.

First, we do calculate the dimensional space for features, find the minimum dimensional spaces for features, then by the threshold to determine two features that are similar in algorithm. If the feature dimensional spaces can't be within the threshold value, then can't find the feature with a similar feature.

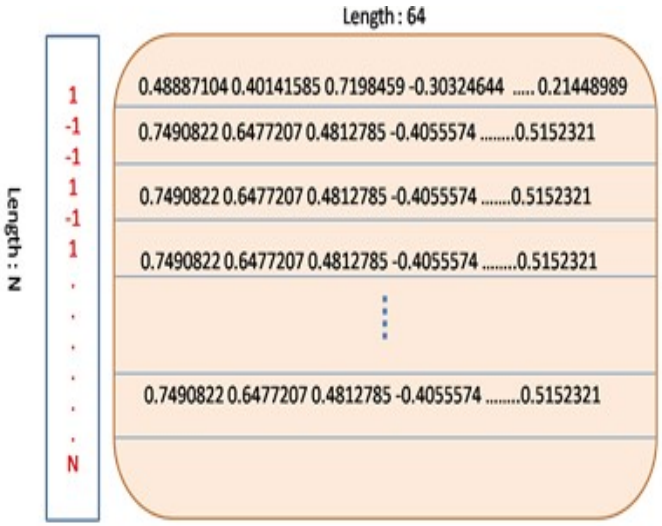

Fig. 2 . SURF feature

TABLE 5
ALG VARIABLE

M	The number of point sources feature
N	The number of data sample feature points
Fi	Source sample feature vector
Kj	Data sample feature vector
D	Dimensional spaces Distance
count	Counter

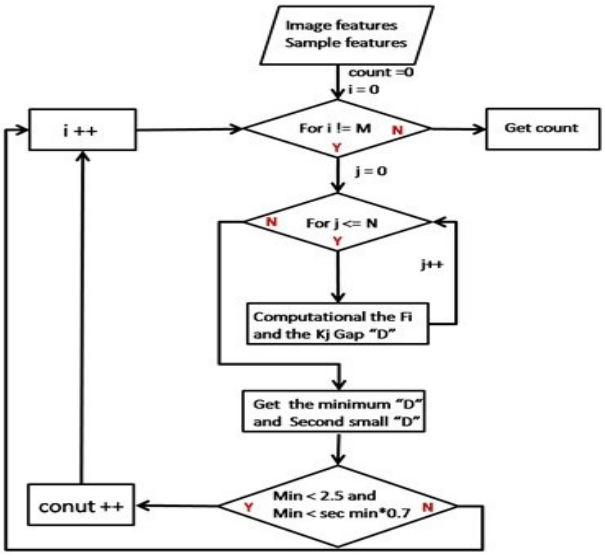

Fig. 3 . SURF feature

Computing Efficiency

All image features are obtained which are stored as text by Open CV. These data store into Hadoop HDFS as sample. This research writing Map/Reduce program in the Map doing the "dimensional spaces algorithm," will be calculating samples and the source image of the dimensional spaces, and then these data are handed over to Reduce final operation. Finally, information is returned at the sample ID and the amount of similarities of features.

RESULTS

We have completed the program, the program output jar file, and put it in the Hadoop. The program is running image feature comparison, to test the effectiveness of the overall operation. The following is the amount of data in HDFS, each group picture features data size.

TABLE 6
HDFS SAMPLE GROUP

Group	Data size	Piece of data
1	251MB	1,000
2	504.1MB	2,000
3	1GB	4,000
4	2.01 GB	8,000
5	4.03GB	16,000
6	8.07GB	32,000
7	16.13GB	64,000
8	32.26GB	128,000

This research checks effectiveness of test of a single computer (CPU I5-4440 3.1 GHz and 8G memory). And tests the use of Raspberry Pi3 erected six groups of Hadoop clusters, the number of nodes in each group were 5, 6, 7, 8, 9, 10.

The statistical group operation time consuming experiments were conducted, and recorded the computing 100 pieces of data of each group experiment of elapsed time. And observed increase in data on the computing impact, and Hadoop nodes increase or decrease the impact.

Single PC

The experimental information in this research, all experimental groups as single computer, computing time of each is quite close at all tests, the data decrease or increase which does not affect the computing efficiency.

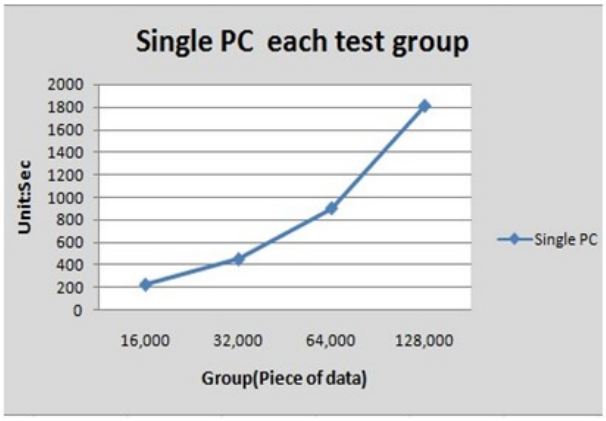

Fig. 4 . Single PC total test

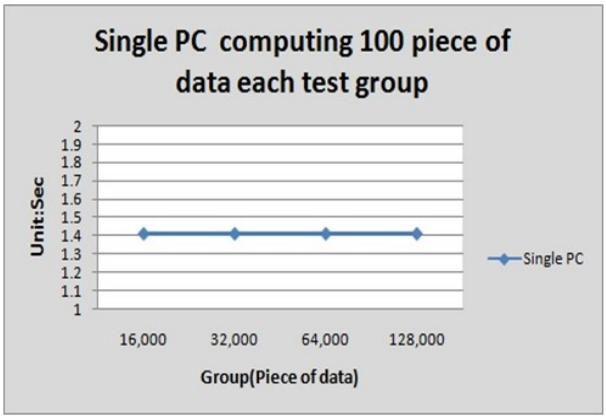

Fig. 5 . Single PC 100 pieces of data

TABLE 7
PC 100 PIECE OF DATA

Group (computing 100 piece of data)	Consuming(sec)
16,000	1.4140.009
32,000	1.4110.007
64,000	1.4100.004
128,000	1.4110.005

Raspberry Pi3 Cluster

The Raspberry Pi3 clusters are 6 groups with node amount respectively as 5, 6, 7, 8, 9, 10.

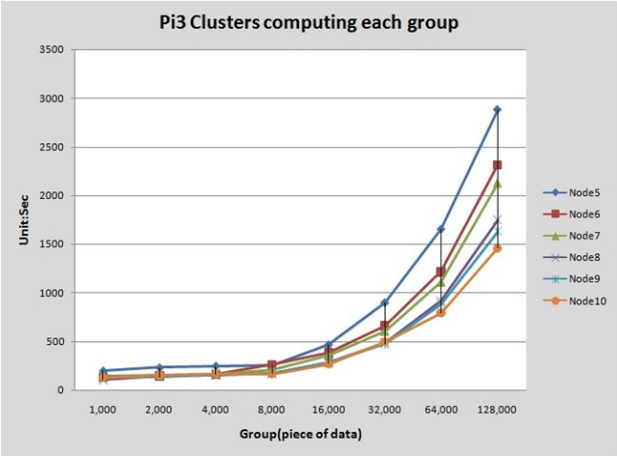

Fig. 6 . Pi3 Cluster all data

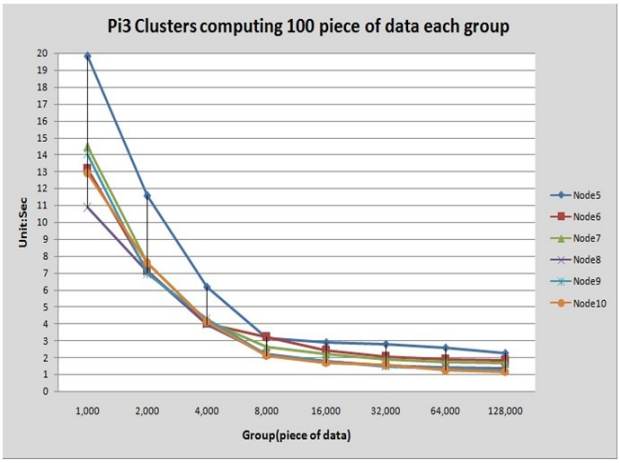

Fig. 7 . Pi3 Cluster 100 piece of data

The experimental information in this research, Raspberry Pi cluster nodes 10 better Raspberry Pi cluster nodes 5 which approximately 90% higher performance. Experimental information show that each additional node which computing time is relatively reduced, proved nodes increase can effectively improve computing performance.

For example the 128,000 piece of data group , for each additional Pi cluster node will be improve 10% performance than the previous nodes amount. If the data is too small, the cluster doesn't significantly enhance computing efficiency .

TABLE 8
PI CLUSTER GROUP 128,000

Node amount (computing 100 piece of data)	Consuming(sec)
5	2.250.11
6	1.800.08
7	1.660.03
8	1.360.03
9	1.270.03
10	1.130.05

Pi3 Cluster and Single PC

The experimental information in this research, the single PC and the Pi cluster were compared which was for 128,000 pieces of example. The single PC takes about 1.41 seconds of time to compute. The Pi cluster nodes 10 take about 1.13 seconds of time to compute. Pi cluster is better than single PC which is 23% higher in computing performance.

The experimental information in this research, the single PC and the Pi cluster are compared which was for 16,000 pieces of example. The single PC takes about 1.41 seconds of time to compute. The Pi cluster nodes 10 take about 1.6 seconds of time to compute. Pi cluster shows 23% decrease in computing performance.

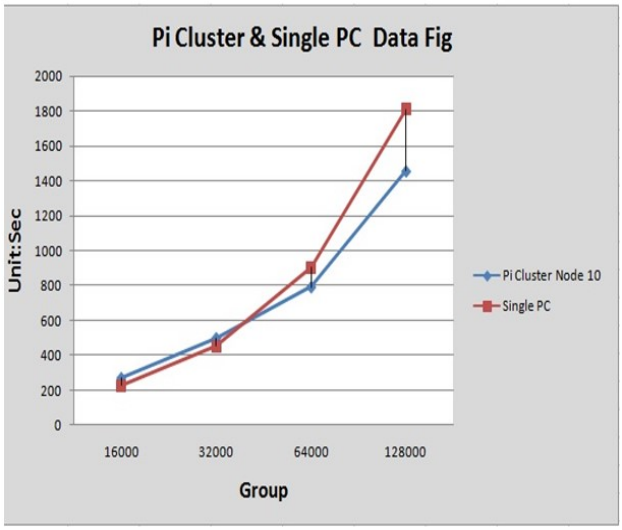

Fig. 8 . Pi3 cluster & PC data

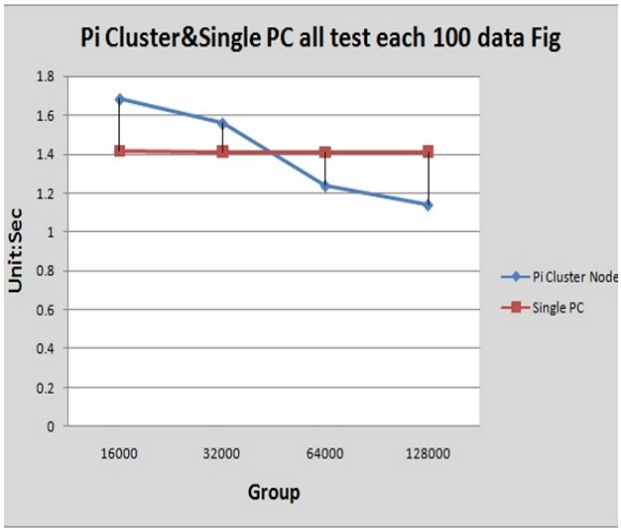

Fig. 9 . Pi3 cluster & PC 100 pieces of data

Pi More Easily Increases Node

When Raspberry Pi of systematic SD card is copied to the new SD card, it can be inserted to another Raspberry Pi and can be started to use. Not only fast but also you don't need to install any driver to hardware devices. However, as long as that adjusts the IP and ID, it can be expanded to the Hadoop architecture, and lets Hadoop cluster become more convenient to expand node when Raspberry Pi is setting up.

Beside, Hadoop cluster can adjust for number of nodes which is contrary to amount of data.

When the amount of data is increasing, we have to increase the number of nodes for improving Hadoop cluster efficiency. while the space is limited, the cluster nodes increasing of PC will be limited. But, Raspberry Pi Hadoop clusters compared to PC, you can add more nodes.

TABLE 9
PRICE COMPARE

	Item	Price
	Raspberry Pi	$35
	CPU I5-4440	$200
PC	8G RAM(4G*2)	$45
	Low end power	$40
	Hard Disk 500G	$60

CONCLUSION AND RECOMMENDATIONS

When there is a large amount of data, Raspberry Pi cluster is better than single PC on distributed computing dominant. But, in case of a small amount of data, single PC is better than Raspberry Pi clusters.

While the space is limited, and takes long time for analysis for large amounts of data, using Raspberry Pi clusters is nice choice. Because Hadoop architecture in the operation of the process failure node or interruption contact, Hadoop system can still complete the work. So, it is suitable for taking long time to compute.

Hadoop cluster is easier to expand nodes, while the Raspberry Pi cluster can expand nodes easier than computer which is the Raspberry Pi's advantage on setting up the Hadoop.

This research can also be used as Hadoop infrastructure in the future, research SBC and computers are to set up Hadoop cluster number of the node ratio, which can be reached at low cost and high efficiency [4].

In the Big Data generation, the improved operational efficiency and energy conservation, they are worth exploring topics.

Declaration of Conflicting Interests

No conflicts if interests.

REFERENCES

[1] K. R. Lakhani and E. Von Hippel, "How open source software works:'free' user-to-user assistance," *Research Policy,* vol. 32, no. 6, pp. 923-943, 2003.

[2] W. P. Birmingham and D. P. Siewiorek, "MICON: A knowledge based single board computer designer," in *Design Automation, 21st Conference on IEEE,* 1984, pp. 565-571.

[3] R. Morabito. (2016). *A performance evaluation of container technologies on internet of things devices* [online.] Available: https://goo.gl/9KHvS0

[4] S. J. Cox, J. T. Cox, R. P., Boardman, S. J., Johnston, M., Scott and N. S. O'brien, "Iridis-pi: A low-cost, compact demonstration cluster," *Cluster Computing,* vol. 17, no. 2, pp. 349-358, 2014.

[5] F. P. Tso, D. R. White, S. Jouet, J. Singer and D. P. Pezaros, "The glasgow raspberry pi cloud: A scale model for cloud computing infrastructures," in *IEEE 33rd International Conference on Distributed Computing Systems Workshops*, 2013, pp. 108-112.

[6] P. Abrahamsson, S. Helmer, N. Phaphoom, L. Nicolodi, N. Preda, L. Miori, ... and .S. Bugoloni, "Affordable and energy efficient cloud computing clusters: The bolzano raspberry pi cloud cluster experiment," in *Cloud Computing Technology and Science (CloudCom), IEEE 5th International Conference on*, 2013, vol. 2, pp. 170-175.

[7] A. Anwar, K. R. Krish and A. R. Butt, "On the use of microservers in supporting hadoop applications," in *IEEE Inter national Conference on Cluster Computing (CLUSTER)*, 2014, pp. 66-74.

[8] W. Rosch, *Hardware Bible,* 5th ed. Indianapolis, IN: Que Publishing, 1999, pp. 50-51.

[9] A. K. Karun and K. Chitharanjan, "A review on hadoopHDFS infrastructure extensions," in *Information & Communication Technologies (ICT), IEEE Conference on,* 2013, pp. 132-137.

[10] G. Yang, "The application of mapreduce in the cloud computing," in *Intelligence Information Processing and Trusted Computing (IPTC), 2nd International Symposium on,* 2011, pp. 154-156.

[11] M. B. Mollah, K. R. Islam and S. S. Islam, "Next generation of computing through cloud computing technology," in *Electrical & Computer Engineering (CCECE), 25th IEEE Canadian Conference on,* 2012, pp. 1-6.

[12] L. Kulik, "Mobile computing systems programming: A graduate distributed computing course," *IEEE Distributed Systems Online,* vol. 8, no. 5, pp. 4-4, 2007.

[13] J. Nandimath, E. Banerjee, A. Patil, P. Kakade, S. Vaidya and D. Chaturvedi, "Big data analysis using Apache Hadoop," in *Information Reuse and Integration (IRI), IEEE 14th International Conference on,* 2013, pp. 700-703.

[14] K. Shvachko, H. Kuang, S. Radia and R. Chansler, "The hadoop distributed file system," in *IEEE 26th symposium on mass storage systems and technologies (MSST),* 2010, pp. 1-10.

[15] A. Pal and S. Agrawal, "An experimental approach towards big data for analyzing memory utilization on a hadoop cluster using HDFS and mapreduce," in *Networks & Soft Computing (ICNSC), First International Conference on,* 2014, pp. 442-447.

[16] S. G. Manikandan and S. Ravi, "Big data analysis using apache hadoop," in *IT Convergence and Security (ICITCS), International Conference on,* 2014, pp. 1-4.

[17] K. Danielson. (2008). *Distinguishing cloud computing from utility computing* [online]. https://goo.gl/8b9U6l

[18] M. Snir, *MPI–The Complete Reference: The MPI Core,* vol. 1, Cambridge, MA: MIT Press, 1998.

[19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision,* vol. 60, no. 2, pp. 91-110, 2004.

[20] H. Bay, A. Ess, T. Tuytelaars and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding,* vol. 110, no. 3, pp. 346-359, 2008.

[21] E. Uma, A. Kannan, "Self-aware message validating algorithm for preventing XML-based injection attacks," *International Journal of Technology and Engineering Studies*, vol. 2, no. 3, pp. 60-69, 2016.

— This article does not have any appendix. —