

Cluster Computing Using Raspberry PI

Anuj Shinde¹, Shashank Pal², Vedant Singhvi³, Manish Rana⁴

¹(BE Student, Computer Engineering, TCET, Mumbai, India)

²(BE Student, Computer Engineering, TCET, Mumbai, India)

³(BE Student, Computer Engineering, TCET, Mumbai, India)

⁴(Associate Professor, Computer Engineering, TCET, Mumbai, India)

Abstract: This project aims on creating a Raspberry Pi based cluster computer which will be used for data analytics. People are becoming dependent on computer and their computational power more and more every day. Increasing amount of data is coming into existence each day and power to process this data calls for cluster computer which are a set of computer that operate together, only to be viewed as a single computer, to process and compute data. In this project, we are creating a miniature model of cluster computer which are used on servers such as windows server 2003 or Beowulf cluster using Raspberry Pi which will give us an idea about the working and power of cluster computers. Our main idea is to create a cluster computer and provide the facility of data analysis to demonstrate the power of cluster computing and the uses of data analytics. This project will also provide a cheaper alternative to teacher who wants to teach their class about working of a cluster and student who wants to build their own small cluster or want experience how a cluster works or any industry personnel who wants to use the power of cluster computer.

Keywords - Beowulf cluster, Cluster computer, Raspberry Pi, Data analytics.

I. Introduction

With the appearance of huge data, organizations need to formulate and compute statistics in an efficient manner with a purpose to reach logical conclusions as mentioned in [5]. Those outcomes need to be accurate and be calculated fast. For this, cluster computing is used. Cluster computing may be used to create redundancy in a computer network to ensure that it will usually be to be had and that it's going to not fail. Some other software is in huge tasks that require excessive performance computing. A few computations are extraordinarily complex, and they require using multiple computers that may speak quick with every different, as adjustments in you'll be able to exchange the complete gadget for example, the simulations used to test theories in meteorology are regularly run on computing clusters. Without a cluster, the calculation is probably impossible to do, or it would take a very long term to technique.

Raspberry Pi have actually taken the embedded Linux community by typhoon. For those surprising, but, a Raspberry Pi (RPI) is a small (credit card sized), cheaper single-board pc this is capable of going for walks Linux and other lightweight working systems which run on ARM processors. Figure[1] suggests some information on the RPi abilities.

As the part of this project we will first start by installing Raspbian and Hadoop on single raspberry pi and checking its performance. Then we will start integrating one node (means a raspberry pi) at a time and perform various test to check the performance of the cluster. This will continue till all the four nodes (as we are creating a 4 node cluster) are attached to each other and a cluster is formed the performance of will be tested as a whole.

II. Background

As noted in [1], The choice to get extra computing capacity and enhanced stability by using orchestrating some of low-fee commercial computers has resulted in to a diffusion of architectures and configurations. The clustering approach commonly connects a number of computing nodes through a quick local area network (Lan). The sports of the computing nodes are coordinated by way of "clustering middleware", a software program layer that is above the nodes and lets in the customers to treat the cluster as one big cohesive computing node, pc clustering is predicated on a centralized control method which makes the nodes to behave as coordinated shared servers. It's different from different strategies such as peer to peer or grid computing which additionally use multiple nodes, however with a greater distributed nature. A cluster may be a easy -node device which simply connects multiple personal computer systems or may be a faster supercomputer. A fundamental method to constructing a cluster is that of a Beowulf cluster which can be constructed with some personal computer systems to provide a cheaper alternative to conventional high performance computing.

As mentioned in [7], Hadoop is an open-supply software programming architecture used for distributed storage and processing of very big data sets. It includes clusters built from commodity hardware. All the modules in Hadoop are designed with a basic assumption that hardware disasters are common in occurrences and must be automatically handled by the system. The heart of Hadoop comprises of a storage element, called Hadoop distributed report machine, and a processing part that's a mapreduce programming version. Hadoop splits documents into huge blocks and distributes them throughout nodes in a cluster as represented in Figure[2]. It then transfers packaged code into nodes to process the statistics in parallel. This technique takes benefit of information locality – nodes manipulating the facts they've get entry to to – to allow the dataset to be processed faster and more effectively than it would be in a extra traditional supercomputer structure that relies on a parallel record system wherein computation and statistics are disbursed thru high-pace networking.

III. Proposed System

The main focus of this project is to create a small and compact 4-node Hadoop cluster which will work on the same principle as the one on the server works and understands how it works. This project also provides any teacher, student or industry personnel who their own cluster and provides with the step to do so.

This project aims to create a model of cluster computer more specifically a model of Hadoop cluster which are used for big data analysis as refered in [6]. It will be using Raspberry Pi's to run and make the system. The cluster we will be creating will be a 4-node cluster.

As the part of this project we will first start by installing Raspbian and Hadoop on single raspberry pi and checking its performance. Then we will start integrating one node (means a raspberry pi) at a time and perform various test to check the performance of the cluster. This will continue till all the four nodes (as we are creating a 4 node cluster) are attached to each other and a cluster is formed the performance of will be tested as a whole. Upon constructing the required cluster computer and installing the required softwares, the processing and implementation of data will be divided into 3 parts.

1. The input data or file will be provided to the master Rpi.
2. The master pi will distribute the file to other slave nodes.
3. The slave node along with master will process the input file using MapReduce software installed on Rpi and give the result to user.

The input for the system will be the input file on which the processing is needed to be done. For the output, the data will be analyzed and return in structured format according to the need required and useless data is segregate.

IV. Figures and Tables

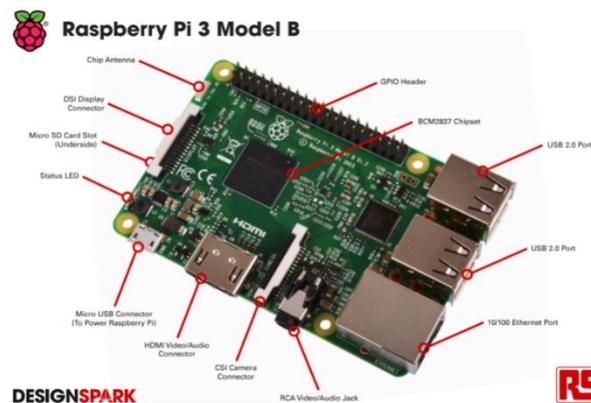


Figure 1: Raspberry Pi 3 Model B and its Features.

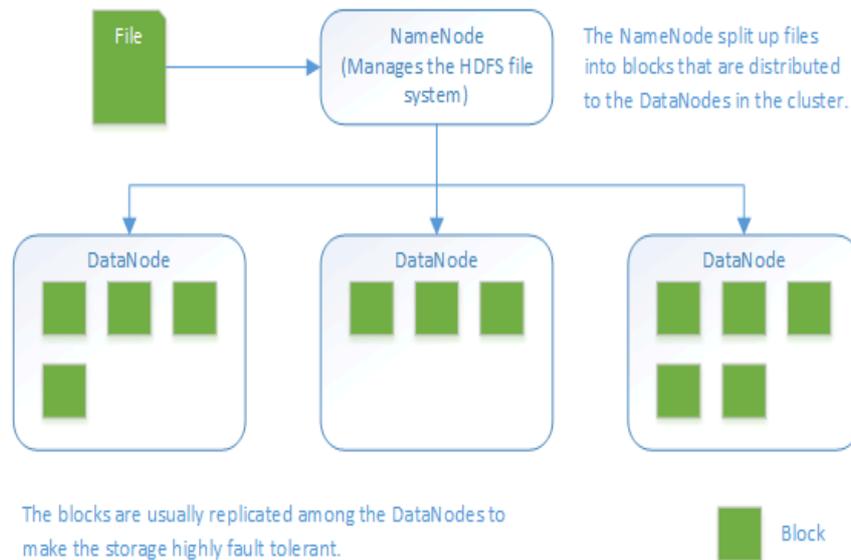


Figure 2: Hadoop Distributed File System.

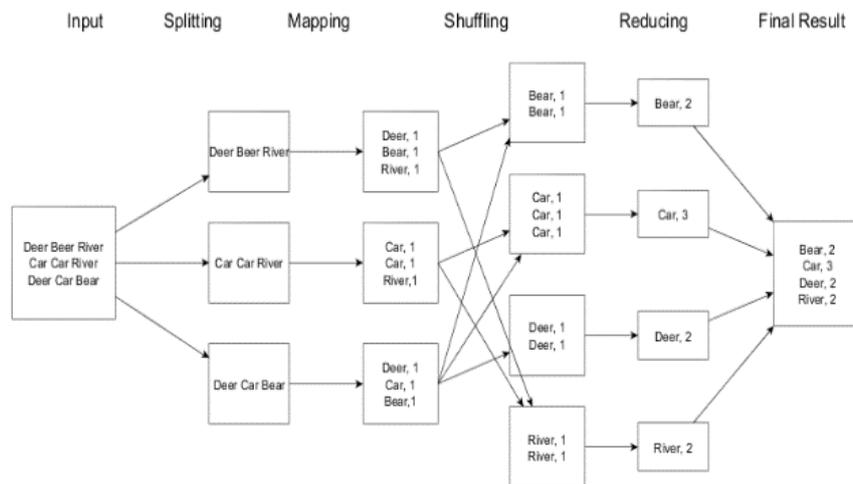


Figure 3: MapReduce flow overview.

V. Conclusion

In the design proposed above we have implemented a cluster computing environment using raspberry pi which is very cheap and feasible to use. As in today’s world there is much emphasis placed on, using the commodity-based hardware and software components to achieve high performances and scalability and at the same time keep the ratio of price versus performance low. This project tries to recognize the importance and uses of Cluster Computing in the IT industry. This project is a small scale project, simply aimed to demonstrate the prowess of cluster computing and data analytics. It can be used as a learning ground for institutions and be used in small scaled industries and projects as well.

Acknowledgement

We sincerely thank our guide Mr. Manish Rana for his guidance and constant support and also for the stick to our backs. We also thank the project coordinators for arranging the necessary facilities to carry out the project work.

We thank the HOD, Dr.SheetalRathi, Dean Academic Dr. R.R Sedamkar, the Principal, Dr. B. K. Mishra and the college management for their support.

References

- [1] High-Performance Computing: Clusters, Constellations, MPPs, and Future Directions Jack Dongarra, Thomas Sterling, Horst Simon, and Eric Strohmaier, IEEE Computing Volume 7 No. 2, March/April 2005
- [2] BEOWULF: A PARALLEL WORKSTATION FOR SCIENTIFIC COMPUTATION Donald J. Becker, Thomas Sterling, Daniel Savarese, John E. Dorband, Udaya A. Ranawak, Charles V. Packer, Proceedings, International Conference on Parallel Processing, 2009
- [3] " Creating a Raspberry Pi-Based Beowulf Cluster ", Joshua Kiepert, UC Berkeley Technical Report, December, 2013.
- [4] CUSTOMER DATA CLUSTERING USING DATA MINING TECHNIQUE,Dr.SankarRajagopal, Second Extreme Linux Topics Workshop,Jun2005.
- [5] Cluster Computing and Applications, Mark Baker,Amy Apon,RajkumarBuyya,Hai Jin, Second Extreme Linux Topics Workshop Jun, Monterey CA,2000.
- [6] A Review Paper on Big Data and Hadoop, Harshawardhan S. Bhosale, Prof. Devendra P. Gadekar, Department of Computer Engineering, JSPM's Imperial College of Engineering & Research, Wagholi, Pune.
- [7] The Hadoop Distributed File System, Konstantin Shvachko, HairongKuang, Sanjay Radia, Robert Chansler, Yahoo! Sunnyvale, California USA.
- [8] Big Data Analytics using Hadoop, BijeshDhyani, Anurag Barthwal, International Journal of Computer Applications (0975 – 8887) Volume 108 – No 12, December 2014.
- [9] Research on Reliability of Hadoop Distributed File System, Daming Hu, Deyun Chen, Shuhui Lou and Shujun Pei, College of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China.
- [10] High Performance Cluster Computing: Systems and Architectures, R. Buyya (Ed.), Vol. 1, 1st Edition, Prentice-Hall, Englewood Cliffs, NJ, 1999.