Hadoop environment management App based on mobile cloud computing

Feng-Qi Cheng, I-Ching Hsu*

Department of Computer Science and Information Engineering

National Formosa University

64, Wenhua Rd., Huwei Township, Yunlin County 632, Taiwan

hsuic@nfu.edu.tw*

Abstract—**Cloud computing has become an important subject in both the academia and the industry. Apache Hadoop is an open source software project that enables the data-intensive distributed applications across scalable clusters of computer servers. Although Hadoop environment provides a web interface for information display, it does not support management using mobile device. With the current popularity of 3G and Wi-Fi networks, mobile devices can access the internet anytime, anywhere. This work develops a Hadoop environment management App for android phone, which enables administrators to manage Hadoop clusters via 3G or Wi-Fi without the time and location restrictions.**

Index Terms—**Hadoop、Android、Mapreduce**

I. INTRODUCTION

Hadoop [1] is a widely applied cloud computing platform, which provides two main functions: MapReduce [2] and the Hadoop Distributed File System (HDFS) [3]. The former performs the function of parallel computing, whereas the latter enables distributed file backups. Research on cloud computing is present within various domains. A study in Norway established algorithms on Hadoop to store and analyze historical data concerning petroleum and gas explorations, thereby enhancing the performance [4], and another research used the HDFS of Hadoop to save school data [5]. Many other studies have also integrated cloud computing to achieve data mining [6], implement algorithms on Hadoop [7], and execute computational biology. These studies focus on using Hadoop to solve and enhance performance issues. There are only a few studies have focused on Hadoop management. [8].

It is very inconvenient for manager using the Secure Shell (SSH) [9] protocol to control and management Hadopp cluster. Although the Hadoop cluster information is displayed in a web-based interface, there are no management functions. The Hadoop manager need to link to the Namenode with an SSH connection. The original SSH connection information display in a mobile phone is restricted and incomplete. This work develops a Hadoop environment management App for android phone, which enables administrators to manage Hadoop clusters via 3G or Wi-Fi without the time and location restriction.

This paper is organized as follows. The next section introduces some related technologies. Section 3 presents and develops an android-based mobile for Hadoop environment management. Section 4 presents an applicable demonstration and experimental results are presented. Finally, summary and concluding remarks are included.

II. BACKGROUND

A. Hadoop

Hadoop originated from the Apache Nutch [10] project, an open source web-search engine. In 2003 and 2004, Google published papers on GFS [11] and MapReduce [12], resolving issues encountered in Nutch. Derived from these studies. Hadoop have became a top-level project of Apache, which contains of different modules such as Common, Avro, MapRaduce, HDFS, Pig, Thrift, Hive, HBase, Zookeeper, and Sqoop. Any user can use Hadoop to set up their own cloud computing application and environment. Hadoop primarily provides PaaS and IaaS services, which were written in Java. These services can be executed under multiple operating systems and provide distributed computing for exploiting large datasets. Hadoop-based cloud services are integrated with MapReduce, HDFS, and Hbase.

B. HDFS

HDFS is the fundamental file system of Hadoop. The large datasets are required to store within HDFS, which adopts distributed storage technologies to increase storage space and fault tolerance. It utilizes the advantages of parallel extension in Hadoop and low hardware requirements to improve space limitations. HDFS consists of Namenode and Datanode; the former is responsible for access control and Datanode management, whereas Datanode is responsible for data storage. Namenode divides the data into multiple segments, each of which is 64 MB. Based on the set number of copies, backups are created for each segment and then stored on Datanode. In this manner, fault tolerance is enhanced. As shown in the figure 1, a 192 MB file is divided by Namenode into three segments. According to the settings, two duplicates are made and stored. Segment 1 is stored on Datanode1 and Datanode2; Segment 2 is stored on Datanode2 and Datanode3, and

Segment 3 is stored on Datanode3 and Datanode4.
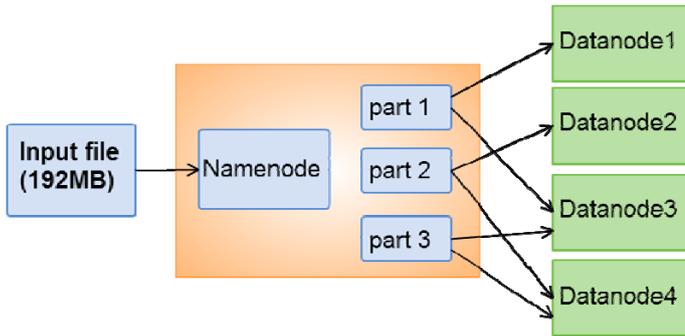
Figure 1.An example of HDFS framework

C. MapReduce

The MapReduce of Hadoop is a parallel computation method to simplify the process of parallel programming. It is particularly effective for big data, as MapReduce divides the problem into a number of smaller problems and assigns them to each individual Mapper for processing. Subsequently, Reduce integrates the outputs of each Mapper to form the final result. Therefore, MapReduce is suitable for big data computing. It uses multiple nodes to increase computing capabilities and reduce computation time. MapReduce also provides another feature that solves the issue of data transfer from multiple nodes by sending the operational program to the data end for computation rather than sending the data to be processed in the computing system. This greatly reduces the time that have been required to transfer big data. Mapper uses key/value pairs to divide problems into many smaller problems. Figure 2 shows that the input data comprises the grades of thousands of students, and MapReduce is used to find the student with the highest score. MapReduce distributes 100 students to each segment, which is assigned to an individual Mapper for computation. Reduce then performs the final executions and output.
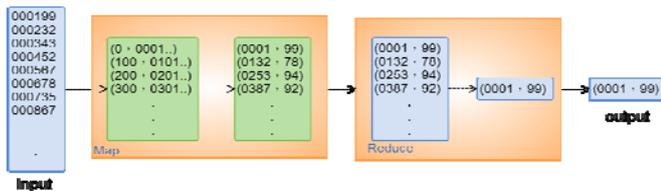

Figure 2.An example of MapReduce framework

D. Android

After developing the Android system [13], Andy Rubin sold it to Google to function as the base of the Open Handset Alliance. Google simplified the development procedures of Android and provided free access as an open source operating system. Google further provided the Google Play platform for programmers to develop applications on, and users can purchase and download them from the platform.

Android is a Linux-based operating system for mobile devices. It inherits the mobility of Linux and is compatible with many mobile devices other than smartphones and tablet computers. Android can also be used in televisions sets, MP4 players, and handheld game consoles. Android has became increasingly popular, particularly in smartphones and tablet computers.

III. SYSTEM FRAMEWORK

This work developed a android-based management App to manage Hadoop cloud computing clusters via 3G or Wi-Fi networks. The main objective was to achieve the function of real-time management. Generally, the management issues of Hadoop cluster arise in Linux. This work used the SSH protocol to enable the synchronous management of Hadoop clusters and the Linux operating system. The Hadoop manager can handle quickly emergency events to reduce losses. As Hadoop is generally applied to big data, computing time may take several days. The App software developed in this study can immediately enable the Hadoop operation and management to meet the requirement of mobile cloud computing.

The structure of the Hadoop environment management App developed in this study is displayed in the figure 3. The Android mobile device connects to multiple servers via a 3G or Wi-Fi wireless communication, whereby the servers to control their respective Hadoop clusters. As a result, our App can manage multiple Hadoop clusters at the same time.
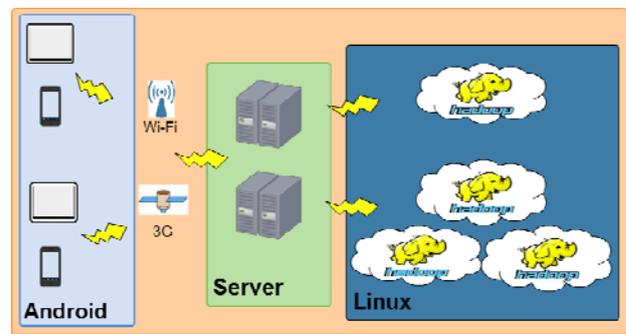

Figure 3.Hadoop management using Android phone

A. Android App

The design concept of the Android App was to serve as a communication bridge between users and Hadoop, thereby achieving real-time management and simplifying the management of Hadoop. The App provides connection settings and command sending. The connection settings is responsible for storing and setting information associated with Hadoop clusters. Managers must provide the Namenode IP address, SSH port, user name, user password, server IP address, and path of the Hadoop clusters. The information is saved on the mobile device for making connections and sending commands. Our Android App will also share the connection information with the Java server for command sending. The command sending function also enables users to manually input commands and receives information from the remote Hadoop.

B. Middleware Server

The middleware server is a bridge between Android App and Hadoop. It receives and recognizes commands made by the Andorid App, and then sends commands to the remote Hadoop. Finally, the output message of Hadoop is passed to Android App through middleware server (see figure 4). Command recognition determines whether the command contains information for connection settings or is a general command. When the middleware server receives a command with connection setting information, it will retrieve and save the Hadoop cluster information for future

use.
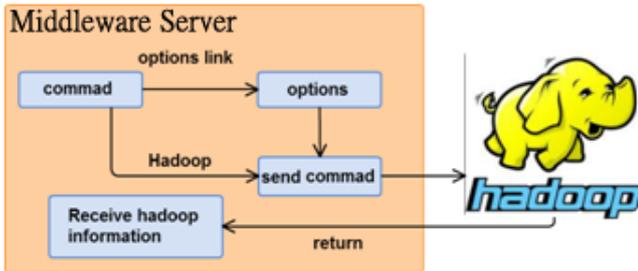
Figure 4. The interaction between middleware server and Hadoop

IV. ANDROID APP DEMONSTRATION

The Options menu in our Android App provides three functions, including add, update, remove, and settings, as shown in figure 5. When a new Hadoop cluster is added, the Namenode IP address, SSH port, user name, user password, and Hadoop installation path must be input (see figure 6). When data input is completed, the "save" function will store the information in the mobile device for future use. After adding the new Hadoop cluster, the system will automatically capture and display the Hadoop information, as shown in figure 7. Before using any function, the user must confirm that the Hadoop has been added.
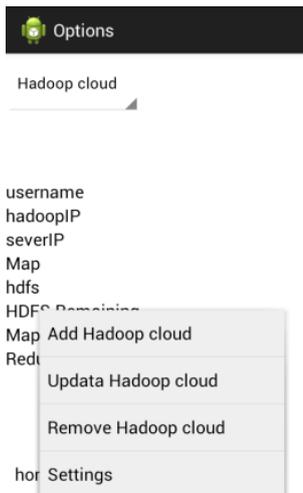
Figure 5.Options menu
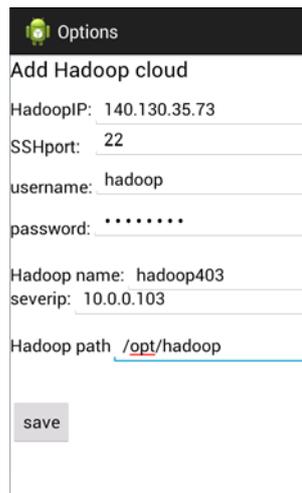
Figure 6. input information
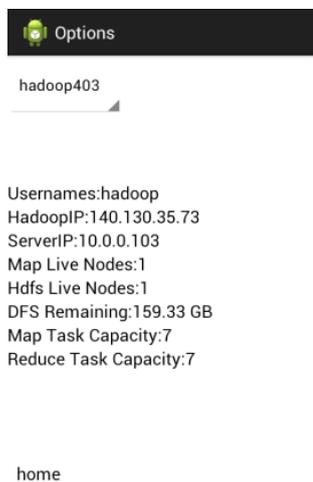
Figure 7.Hadoop cluster information

The operation of starting and stopping Hadoop are similar. We give an example of stopping Hadoop to explain how the Android App is used to stop Hadoop. First, a new Hadoop cluster must be added in the options so that the Control function can be used to manage the Hadoop cluster. After the stop-hadoop button is clicked , the App will automatically send the stop command. If the Hadoop system has been modified, the command can be manually altered, as shown in Fig. 8.When the send button is clicked, the system will send the command to the middleware server, which will send the command to Hadoop cluster. When the middleware server received the returned information from Hadoop cluster, it will first process the warnings or error messages. Manager clicks "more" button to display the complete message, as shown in figure 9. Once the Hadoop cluster has shutdown, the online light in the upper right corner will be off in the refreshed image, thereby indicating that the shutdown was successful, as shown in figure 10.
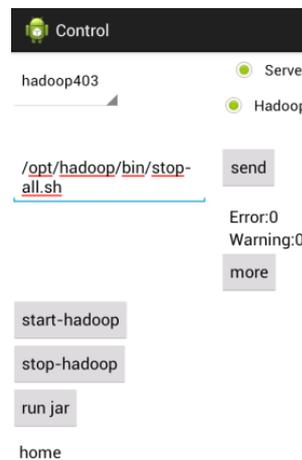
Figure 8.Stop Hadoop

Figure 9.Reture message

Figure 10. successful shutdown

The Hadoop management App simplifies the management of Hadoop platforms via the 3G or Wi-Fi mobile networks. It enables the execution of MapReduce so that each MapReduce process can be quickly performed. A prerequisite of the Android App sending commands to the Hadoop clusters via middleware servers is that the Hadoop connection settings must first be completed before MapReduce can be executed. When the "run jar" button is selected, the system will automatically input simple commands to execute MapReduce. Managers can modify them based on their needs (see figure 11). When "send"

button is clicked, the middleware server will receive the command and send it to the Hadoop cluster for execution.
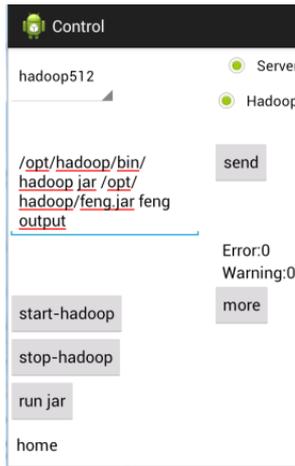
Figure 11. Automatically generate commands

V. CONCLUSION

The constant promotion of Hadoop cloud computing platforms has increased the importance of Hadoop management. However, there are no Android App available for Hadoop management. This work develops a Hadoop environment management App for android phone, which enables administrators to manage Hadoop clusters via 3G or Wi-Fi without restriction to time and place. The Android App enables managers to analyze or test data without the restriction of time or place. We plan to provide more functions in the future and offer additional information on Hadoop clusters so that managers can manage Hadoop clusters more easily. Another future direction of development is to adopt MapReduce operations to mobile devices for mobile cloud computing applications [14-16].

REFERENCES

[1]Hadoop-Apache Hadoop, Available from:http://hadoop.apache.org/.

[2] Gunarathne, T.,2010, "MapReduce in the Clouds for Science", Cloud Computing Technology and Science, 2010 IEEE Second International Conference on,pp.565-572,Dec. 2010.

[3] Oriani, A.,2012, "From Backup to Hot Standby: High Availability for HDFS,Reliable Distributed Systems", 2012 IEEE 31st Symposium on,pp.131-140,Oct. 2012.

[4] Baodong Jia、Wlodarczyk, T.W.、Chunming Rong, 2010, "Performance IEEE Second International Conference on, pp. 5 Considerations of Data Acquisition in Hadoop System", Cloud Computing Technology and Science (CloudCom), 201045-549 , 30 2010-Dec.

[5] Patrick Donnelly、Peter Bui., 2010, "Attaching Cloud Storage to a Campus Grid Using Parrot Chirp and Hadoop", Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference on, pp.488-495 , 30 2010-Dec.

[6] Yang Lai, Shi ZhongZhi, 2010, "An Efficient Data Mining Framework on Hadoop using Java Persistence API", Computer and Information Technology (CIT), 2010 IEEE 10th International Conference on,pp.203-209 , 29 2010-July.

[7] ChenHao 、 QiaoYing, 2011, "Research of Cloud Computing based on the Hadoop platform", Computational and Information Sciences (ICCIS), 2011 International Conference on ,pp.181-184, 21 2011-Oct.

[8] Guanghui Xu, Feng Xu, Hongxu Ma, 2012 , "Deploying and Researching Hadoop in Virtual Machines", Automation and Logistics (ICAL), 2012 IEEE International Conference on, pp.395-399, 15 Aug. 2012.

[9] Diatchenko L, Lau Y F, Campbell A P, Chenchik A, Moqadam

F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov

E D, Siebert P D. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specificc DNA probes and libraries. Proc Natl Acad Sci USA, pp. 6025–6030, 1996.

[10] Nutch-Apache Nutch, 2012; Available from:http://nutch.apache.org/.

[11] Sanjay Ghemawat, Howard Gobioff, Shun-Tak Leung, 2003, "The Google File System" ,pp.1-15 ,October 2003.

[12] Jeffrey Dean, Sanjay Ghemawat , "MapReduce-Simplified Data Processing on Large Clusters" ,pp.1-13, 2004.

[13] Butler, M.,2011,Android: Changing the Mobile Landscape,Pervasive Computing, IEEE,10 , Issue: 1 ,pp.4-7,IEEE,Jan.-March 2011

[14] Niroshinie1 Fernando, Seng W. Loke, Wenny Rahayu, Mobile cloud computing: A survey, Future Generation Computer Systems, 2013, 29(1), pp.84-106

[15] M. Reza Rahimi, Jian Ren, Chi Harold Liu, Athanasios V. Vasilakos, Nalini Venkatasubramanian, Mobile cloud computing: A survey, state of art and future directions, Mobile Networks and Applications, 2014, 19(2), pp. 133-143

[16] Yating Wang, Ing-Ray Chen, Ding-Chau, A Survey of Mobile Cloud Computing Applications: Perspectives and Challenges, Wireless Personal Communications, 2015, 80(4), pp. 1607-1623