Broadband Wireless Networking in the Era of Big Data

Dr. Tamer Omar, East Carolina University

Tamer Omar is an Assistant professor with the department of Technology systems at East Carolina University. Dr. Omar earned his Ph.D. from the Electrical Engineering department at Iowa State University, USA and his MBA with emphasis on MIS from the Arab Academy for Science and Technology, Egypt and his B.S. degree in Electrical Engineering from Ain Shams University, Egypt. Dr. Omar research interests include wireless networks architecture, resources allocation in wireless networks, heterogeneous networks, self-organized networks, big data implementation and analysis, RDBMS and decision support systems. Dr. Omar has 6 years of experience in academia and more than 10 years of industrial experience in different ICT positions.

Ms. Sirena A. Hardy, East Carolina University

Sirena Hardy thrives on the ever-changing world of information technology and the various ways technology has advanced our society. She has acquired over 10 years of information technology experience in the areas of software consulting and implementation; software training and application support. She gained valuable insight and knowledge during her time traveling around the country providing software training as well as assisting various colleges with the implementation of an enterprise resource planning system. Currently she is providing human resource management system software training to the public school districts of North Carolina and assisting with the statewide implementation of a new applicant tracking solution. She holds a MS in Information Science from North Carolina Central University and is currently pursuing a MS in Networking Technology at East Carolina University.

Broadband Wireless Networking in the Era of Big Data

Abstract

Organizations accumulate huge amounts of data from various systems but more often than not the data is stored but not organized or analyzed by the organizations. When certain characteristics define this data such as volume refers to a large quantity of data received and stored; velocity refers to a high speed of receiving data from different data streams; variety involves the ever-changing data formats from new services, and new data types that are being captured; and finally that this data is valuable. Any data characterized by the aforementioned characteristics is articulated as big data and the systems managing such data is referred to as Big Data Systems (BDSs). Mobile service providers (MSPs) in their efforts to provide more efficient heterogeneous networks (HetNets) deal daily with data characterized by the same features. The successful implementation of a BDS involves having the required infrastructure in place to process the data. There are three key areas involved with a big data infrastructure which includes data acquisition, data organization, and data analysis. Since big data involves higher velocity, volume, and variety an organization must have the ability to capture this data. MSPs need to employ a system to actually extract and analyze network utilization big data to determine if it brings value to them and their customers. This work discusses the design, implementation and utilization aspects of a Hadoop system that can help MSPs to delve deep into their big data stores to analyze the potential of adding value to the organization. A Hadoop system would allow an entity to organize and process their big data. A system architecture for the BDS supporting the HetNet operations will be proposed together with the recommendations of an analytics framework. The BDS architecture together with the analytics framework aims at helping the MSPs in forecasting the network traffic. The results of the traffic big data analytics and the network load forecasting can be used to adjust different network operating parameters. These adjustments can definitely enhance the HetNet performance. The proposed big data architecture and the analytics framework proposed in this study will be used as a decision support system component in an educational and research pilot project that aims at introducing the role of big data analytics in guiding the self-healing process used in cellular self-organized networks.

Keywords: Big Data, Hadoop, HetNets, Self-Organized Networks (SON)

Introduction

Big data is becoming an increasing topic discussed by organizations because we as a society are generating data at an exponential rate due to all of the devices we use on a daily basis. Big data can be classified as either structured or unstructured data. Structured data involves data obtained from relational databases such as customer relation management systems and enterprise resource management systems. Unstructured data involves data which does not have a pre-defined data model or data that is not organized such as log files, metadata, audio and video files. Unstructured data also includes the enormous amount of data generated by wireless networks. Now is the time for organizations to determine if implementing a big data solution will bring value to their organization. There are three key areas involved with a big data infrastructure which includes data acquisition, data organization, and data analysis.

Mobile service providers (MSPs) are particularly faced with big data issues because of the proliferation of wireless devices and mobile applications. Most of these devices access the Internet using cellular networks. Cellular networks are becoming the primary method for accessing the Internet [1]. MSPs must contend with the challenge of managing and planning their network resources to address the continuous increase in traffic demands. Due to the increased traffic from mobile applications, MSPs are generating huge amounts of data. A big data solution such as Apache Hadoop will provide MSPs with an opportunity to capture and analyze this data in a way that was not available with earlier technologies[2].

Big data has come to the forefront of information technology because of the huge accumulation of data generated by web logs, social media, email, photos as well as many other sources. Hadoop Distributed File System (HDFS) was modeled after Google's file system (GFS) paper and Hadoop's MapReduce framework was developed using Google's MapReduce paper. Today, the open-source Apache Hadoop is used by organizations to manage big data[3,4,5].

Facebook is a huge Hadoop and big data supporter, Facebook claims to operate the largest single Hadoop cluster[6]. Facebook initially used relational database management systems (RDBMS) to store its data. However, due to the increasingly large volume of data generated on a daily basis and the popularity of the social media application, it became almost impossible and simply impractical to maintain their RDBMS infrastructure[7,8,9].

It is projected that by 2017 mobile device traffic will reach 1.1GB a month compared to the 2012 level of 350MB a month[10]. By 2020, it is estimated that we will generate 50 times more data than we did in 2011[11]. It may be an assumption that big data automatically brings value to an organization. However, it requires the ability to develop an advanced system to analyze the data. Organizations accumulate huge amounts of data but do not necessary know the value unless they implement a system to process this large volume of data. According to Li, the big data life cycle consists of generation, collection, aggregation processing and application delivery[12]. As mentioned previously big data is classified based on volume, variety, and velocity. However, it does not mention value because it is yet to be determined if analyzing the big data, in fact, brings value to the organization. The future networks must have computational, data, and effectiveness intelligence to handle the network complexities as well as the required network data analytics needed to manage these networks[13].

We are researching the different approaches for developing and implementing a big data solution for MSPs. The collected information in our survey will be used to specify the best approach based on our assumptions. We will determine the feasibility as well as the difficulties of implementing a big data solution using Apache Hadoop in an educational testbed environment. This research is significant because it is important to determine whether or not accumulating, storing, organizing and analyzing big data using a Hadoop system will bring value to MSPs. Wireless networks generate huge amounts of data, it is important to determine if MSPs can make use of their big data analysis to improve future network functionality and prepare for the ever-increasing number of mobile user subscribers. Our research will contribute on determining whether a big data solution will help MSPs improve their capability of handling this exponential increase in network traffic and produced data in future cellular networks.

The rest of this paper is organized as follows. Section two presents the project description and the testbed that includes both the wireless 4G heterogeneous network (HetNet) and the big data system. Section three discusses the current operations and testing for the implementation and finally section four draws our conclusions.

Project Description

The purpose of this project is to implement and run an academic lab for studying a 4G HetNet that depends on a big data infrastructure in order to collect, organize and analyze the network performance. The analysis results will be used to tune the network parameters in order to enhance the self-organized network (SON) self-healing functions. This lab is currently used to perform research tasks. The students in the lab will be able to work extensively with designing, implementing and operating both a wireless 4G Heterogeneous Network (HetNet) and a Big Data System (BDS) in a testbed platform.

The future plan is to extend the project scope to create two courses. The first course will concentrate on the information technology aspects of managing, administrating and operating wireless HetNets. The scale of this networks is definitely moving towards distributed administration by end users. These requirements create a need for qualified IT candidates capable of managing such type of networks in different industries and businesses that is expected to highly depend on wireless access for their daily business transactions. The second course will include BDSs administration and management concepts needed to implement and operate BDSs. The adoption of most companies to BDSs has increased the demand for qualified candidates that can implement, operate and maintain these systems.

Feasibility Study

A feasibility study is conducted to determine the best approaches for implementing both the simulated 4G Small Cells (SC) access network together with the core network and the infrastructure for the BDS.

4G small cells access network

Three approaches have been studied to implement the needed access network. The first approach is to create a production testbed that consists of several SCs, a controller for the SCs, and a SON management system. The SCs are planned to be distributed over one of the university buildings. The network is designed to use the university Local Area Network (LAN) to backhaul the system traffic to the MSP. The advantages of this approach are the ability to collect real data samples from the network that can map a real-time production network, exposing the students to the operations of commercial equipment, and performing the different analytics and self-healing testing on a real-time production system. However, the disadvantages of this approach are the interference caused by the testbed with current existing production distributed antenna system (DAS) owned by the university, security concerns regarding data collection privacy, and SCs access network traffic and operations overhead on the university resources.

The second approach considered is to use network emulators utilized by broadband network implementers in industry to design and simulate the access network. A survey is conducted to specify the key players in the area of SC access network emulators. Some network emulators are specified as a qualified candidate to replace the production network. The emulator consists of HW and SW components that can perform the needed functions to generate the network traffic and integrate with external products to perform the required analysis and the SON functions.

The advantages of this approach are overcoming interference concerns with current wireless DAS, ease of implementing and operating the emulator, and prevents any MSP or data collection security issues and concerns. While the disadvantages of this approach are the complexity of the network and parameter configurations for the Evolved Packet Core (EPC) network emulator, and the loss of student exposure and hand on experience with production networks equipment.

The third approach investigated is using SW simulators to simulate the operations of the access network and create the needed traffic to manipulate the BDS. After analyzing this data for optimal network operation conditions, the optimized network parameters are fed back to the simulator to test its effect on the network. Two simulators are investigated to determine the one that best match the research purpose.

The advantages of this approach are the availability and affordability of educational license agreements targeted for educational purposes, the open source code for some simulators and the interfacing features that can help in integrating the simulators with external resources. However, the disadvantage of this approach is the complexity of the input/output operations between the simulators and the BDSs.

A decision was made to implement the network using a combination of the three approaches. The access network will be simulated using a traffic load simulator in addition to implementing one real production SC in a controlled environment (e.g. unlicensed frequency spectrum). The EPC will be emulated to provide the functionalities needed from the core network. This prototype will help avoid some of the disadvantages accompanied by the three approaches (e.g. avoid interference with production systems) and offers the students an opportunity to work with a real production system while performing and maintaining the network functionalities through SW simulation and the HW emulators.

Big Data System

There are several Hadoop vendor distributions that can be implemented to help MSPs organize and analyze their network data. The standard Apache Hadoop distribution is comprised of three components including MapReduce which provides the capability to run computations in parallel; HDFS based on Google's File System (GFS); and Hadoop Common which includes the utilities needed by various Hadoop modules[14]. HDFS is a distributed file system which allows the capability of storing and organizing big data into managing data blocks. HDFS also provides a mechanism to create replicas of the data blocks for redundancy purposes. The Hadoop framework is designed to accommodate the processing of big data using multiple commodity computers organized into clusters. Apache's Hadoop distribution also includes Yet Another Resource Negotiator (YARN), a component used for data processing. HDFS and YARN work together to allow storage and processing handling within the same system[15].

There are several approaches that can be used to implement a Hadoop system to organize and analyze big data. There are several questions MSPs should address to determine the feasibility of developing and implementing a big data solution. MSPs need to assess what insight they are trying to achieve, will they be using different data sources than existing data sources, how large of a cluster will they need to implement if the pilot is successful, will they be able to easily add more equipment if necessary and another consideration revolves around the integration of the BDS with their other applications[2,16]. For the purposes of this research we will focus on implementing a pilot consisting of a simulated HetNet, an Application Programmable Interface (API) to acquire the data from the access network and importing it into the Hadoop system, and then using statistical analysis software such as IBM's Watson Analytics or SPSS and an optimization tool such IBM's CPLEX Optimizer to analyze the cellular network traffic.

We researched various approaches and recommendations on getting value from Hadoop and determined the most feasible implementation method to address our research needs. Hadoop distributions implementation is divided into various classifications ranging from using a company that uses the Apache Hadoop code but adds its own proprietary code, these companies are classified as completers. Another method is to use a distribution from the second classification known as builders which are companies that focus on building functionality within Hadoop but it is shared within the open-source community. Next, there is the third distribution classification developed by companies known as embedders which can help an organization use Hadoop by incorporating it into other systems. And finally, organizations may choose to follow the fourth classification known as customizers by developing their own Hadoop code to serve its purposes[17].

Based on market research there are three widely used distributions utilized by organizations when implementing Hadoop. While there are many Hadoop distributions available; Cloudera, Hortonworks, and MapR Hadoop distributions are the most commonly implemented solutions[14]. Cloudera has the majority of the market share of Hadoop distributions. Hortonworks and MapR control a smaller share of the market of Hadoop distributions[18]. Cloudera's and Hortonworks' distributions are based on open source Apache Hadoop while MapR has incorporated proprietary modules within its distribution. Although Cloudera uses the core aspects of Hadoop it has also developed proprietary software to make the Hadoop implementation process easier[18].

Hortonworks' Hadoop distribution stays true to the open source code. Their distribution allows companies to organize, store and manage big data using the open source Apache Hadoop without the need for proprietary software. This allows organizations the ability to make changes without being locked in compared to the case if they were using proprietary software. Hortonworks can be downloaded free and installed directly from the Hortonworks' website[18].

MapR has a Hadoop distribution but it has replaced the HDFS component and developed its own MapR file system (MapRFS). MapR's file system is more robust and provides enterprise level functionality within Hadoop and giving organizations the ability to manage data efficiently, it provides the reliability needed and it is easy to use. MapR does provide a free version but it does not include the proprietary software included with the enterprise version[18]. Vendor distributions can provide the MSPs with better functionality, greater ease of use and better capability in addressing their big data needs. Of the three distributions, Hortonworks is the vendor that uses the open source Apache Hadoop without proprietary software[18].

Hadoop vendor distributions are ever changing to meet the demands of organizations as they manage big data. Recent trends that affect Hadoop distributions include third-party integration, performance enhancements, and data security. There is also an increasing need to develop new capabilities in the open source Apache Hadoop code[14]. Based on the market research and the purposes of this research Hortonworks Data Platform (HDP) 2.3.2 Hadoop distribution will be used to store, organize and analyze the network data.

Project Groups

Three groups of students will be formed to implement, operate, and maintain the testbed systems.

Group 1 – Access Network Administrators

Students in this group are responsible for implementing the access network using the simulator and the sampler real production SC to generate the traffic needed for simulating a real operating access network. Another responsibility for this group is the integration of the EPC emulator with the access network to complete the structure of the 4G network. The third responsibility is to maintain the core switching and routing equipment to connect the different networks to each other and to the internet.

Group 2 – Big Data System Administrators

Students in this group are responsible for implementing the Hadoop system, maintain, and administer the system HDFS. The BDS will be implemented using virtualization, this group second responsibility is to maintain the virtual platform used and ensure the reliability of the system. Their last responsibility is to use the MapReduce in order to produce the data sets needed for analysis by the analytic tools.

Group 3 – Developers

Students in this group are responsible for importing the datasets created by group two into the analytic tools. They are also responsible for conducting the analysis using different algorithms and optimization techniques. Finally, they are responsible for creating the scripts needed for collecting the empirical data from the access network and exporting it to the BDS and importing the optimized network parameters recommended by the SON module to the access network to enhance the network performance.

Testbed Design

The small cells access network together with the BDS is designed to imitate the real life production systems. This design will generate the traffic needed by the small cells access network from which the needed data will be collected on the distributed BDS. Using data analytic tools, the collected data will be analyzed to predict optimal operating conditions in the form of network parameters optimization and traffic load analysis. The analysis results will be used by the SON framework to control both the network parameters for optimal operations and the self-healing parameters for failover and fault tolerance purposes. Figure 1 shows a complete diagram for the proposed architecture of the testbed. The four layers of the testbed proposed architectures shown in Figure 1 are:

- Small Cells Access Network Simulator
- EPC Network Emulator.
- Big Data System.
- Data Analytics System.

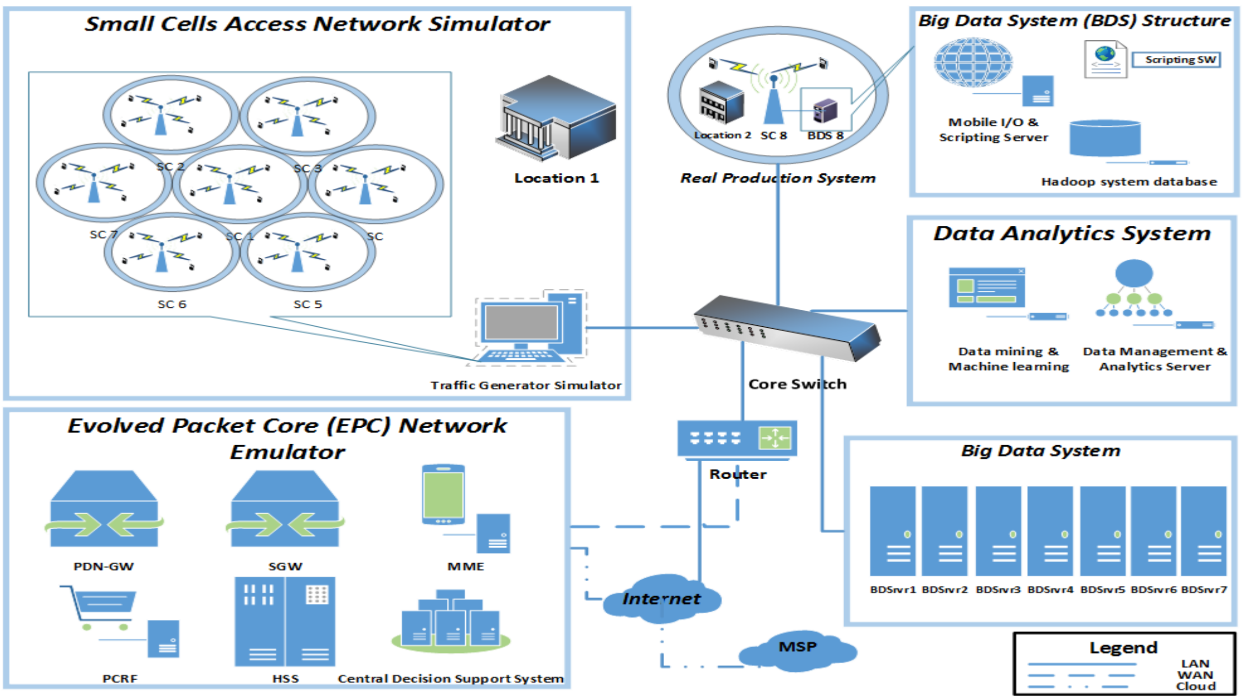

Figure 1: Network Architecture

Testbed Implementation

Phase 1 – Individual systems

During this phase, each system will be implemented individually to ensure the smooth operability and reliability of each standalone system.

Small Cells Access Network Simulator

Primarily seven SCs will be created using the simulator. Each cell will be serving from 10 to 20 user equipment (UE). A simulated traffic flow will be generated to produce the variable traffic load across the seven SCs. The traffic produced will be stored in PCAP files and together with the network load and different SC operating parameters will be acquired from the simulator by the BDS mobile I/O and scripting server.

EPC Network Emulator

The emulator contains five main components that are required to emulate the EPC and a proposed component for central decision making:

- The Packet Data Network Gateway (PDN-GW) responsible for IP addresses allocation to UE.
- The Service Gateway (S-GW) responsible for transferring all the IP traffic in the access network.
- The Mobility Management Entity (MME) responsible for the signaling between the UE and the EPC through the SCs.
- The Policy Control and Charging Rules Function (PCRF) responsible for policy control and Quality of Service (QoS) parameters.
- The Home Subscriber Server (HSS) responsible for user's subscription and roaming between different SCs.

A central Decision Support System (DSS) is proposed to aid in supporting the SON modules. The system uses the MapReduce tools to aggregate data from the distributed BDS and create global data sets. These data sets are then used by the analytics to calculate the network Key Performance Indicators (KPIs) and resources allocation decisions[19].

Big Data Servers

The BDS consists of Big Data Servers (BDSrvrs) distributed through the access network with each server supporting one or multiple small cells according to the network geographical distribution and the cell loads. Each BDSrvr acts as a node of the HDFS. Each BDSrvr is used to host a Mobile I/O & Scripting application agent and Hadoop database.

Data Analytics System

This system consist of two main servers, the first one contain data mining and machine learning tools such as IBM's Watson Analytics or SPSS. Data sets for long term (days or weeks) operations are exported from the BDSrvrs according to different criteria (e.g. daily load threshold, signal strength variations) and imported to this server in order to create learning patterns and mine the data for useful information that can help optimize the network performance or perform fault tolerance for faulty cells. The Data Management and Analytics server are responsible for short-term (hours) operations in order to optimize the resources allocation in the access network using optimization techniques and IBM's CPLEX Optimizer.

Phase 2 – Compatibility Testing

Once the implementation of the individual system passes the local operations testing successfully. The integration of all the systems in the testbed will be performed through the core switch. Part of network traffic will be routed through the internet to test the cloud-based networking performance for the traffic generated by the simulated access network and the EPC. The real production network controlled sampler will be operated by Qualcomm proposed equipment working in the unlicensed frequency spectrum to show the students an analogy between the real access network configuration and the simulated/emulated lab testbed.

Phase 3 – Full Implementation

Finally, a full implementation that includes all the systems will go into production to start generating the daily traffic and perform the needed analysis for the research purposes. A lab manual will be developed to document the network scenarios and create the different lab procedures.

Current Operations and Testing

It was determined that Hortonworks Data Platform (HDP) 2.3.2 would be used for testing purposes. Hortonworks provides a sandbox environment to help with learning and developing within the Hadoop environments. We initially opted to install HDP 2.3.2 on a personal machine using the virtual box method. However, using the sandbox proved to be challenging due to the resources available on the computer. Since the use of the virtual box was unsuccessful due on the personal computer, we installed the sandbox using Windows Server 2012 on a virtual machine with the necessary resources. Once configured the sandbox environment was used to access the Hadoop tutorials and learn how to develop within the Hadoop environment.

There are several processes that should be started before performing data analysis. After logging into Ambari which is the administrator interface that makes it easier to manage your Hadoop cluster, you need to start all services. It is recommended specifically to start Apache Spark because it helps to speed up MapReduce programs. MapReduce jobs are created and run for analysis of big data but sometimes the jobs may run longer than expected. MapReduce was designed for a certain set of use cases. However, the need for a greater framework to address other use cases existed. Apache Spark was developed to advance the MapReduce framework and achieve much faster processing times compared to MapReduce[15].

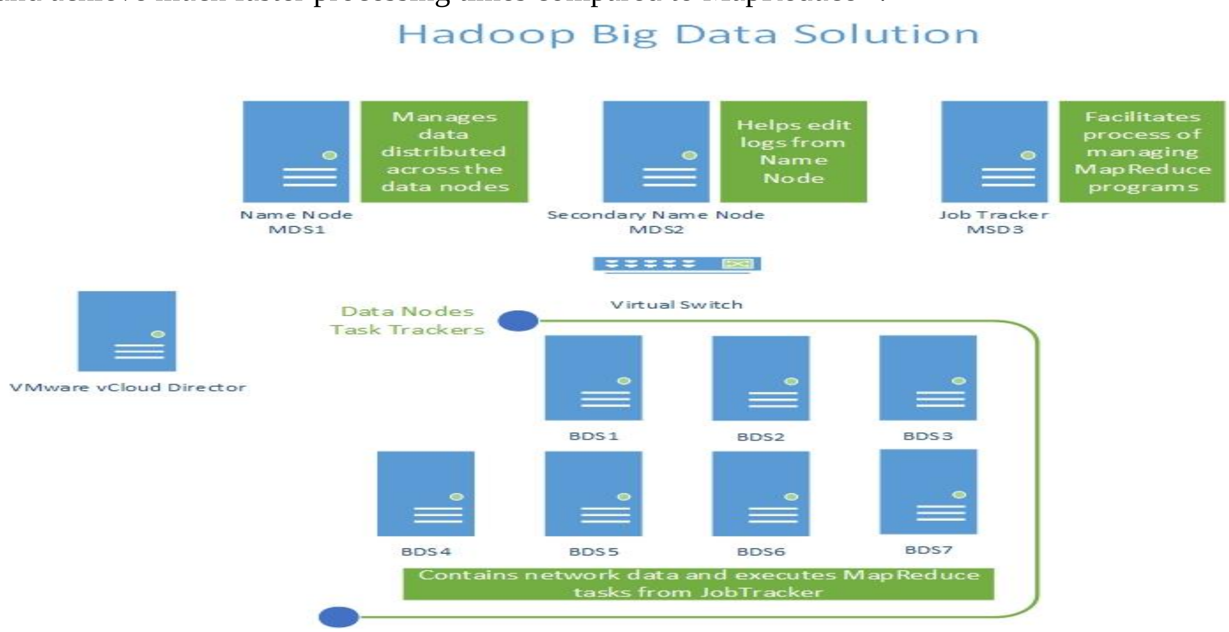

Figure 2: Hadoop Big Data Solution

All servers used in our topology are created as virtual machines and connected through a virtual switch as shown in Figure 2. VCloud Director is used as the hypervisor for this BDSrvrs. Figure 2 shows the model for our Hadoop testbed that includes:

- Master Data Servers (MDSs)
- VMware Vcloud Director
- Virtual Switch
- Big Data Server (BDSrvrs)

The seven Big Data Servers (BDSrvrs 1-7) nodes will act as the data nodes for the system that contain the data collected from each cell in our seven small cells cellular network model along with three Master Data Servers (MDS 1-3) nodes including the NameNode, JobTracker, and Secondary NameNode. The NameNode is overseeing and managing the data distributed across the seven data nodes. JobTracker is the service that facilitates the process involved with managing MapReduce programs as they are submitted by the client application.

Job Tracker works in conjunction with the NameNode to determine file locations on the data nodes in order to process the MapReduce jobs. Task Tracker runs on each data node to execute Map and Reduce tasks received from the Job Tracker. The Secondary Name Node provides a checkpoint for the Name Node which contains the metadata for the files stored on the data nodes. It is not a backup of the Name Node. The purpose of the Secondary NameNode is to help merge edit logs created by the Name Node. The Secondary Name Node assumes some of the responsibility from the NameNode. Other applications associated with Hortonworks Hadoop distribution such as Hive and Pig will be used to store and analyze the data. Pig is a scripting language used for creating MapReduce Programs while Hive is a part of the Hortonworks Data Platform that provides the ability to create tables and serves as a database query interface for Hadoop. Hive provides structure to the data stored in HDFS[4]. These services work together for easy administration of the Hadoop environment.

Conclusion

This project aims at providing insight into the significance of using Hadoop big data solution to organize and analyze cellular networks traffic. A test bed is studied to simulate the process of data acquiring, organization and analysis from a mobile service provider (MSP) network. This research is designed to determine the value that can be added to MSPs by using a Big Data System (BDS) in optimizing their networks performance. There is an increasing trend of heavy network utilization due to the growth in the number of mobile devices accessing the Internet through cellular networks which have a negative impact on user experience as well as a significant impact on application performance. We aim that by proposing our architecture and implementing a Hadoop BDS along with a statistical analysis methodology that MSPs will benefit from analyzing the network traffic and enhance the network utilization. Our testbed could be a model for MSPs to assess the value of their big data and help them plan for better network performance and more accurate network traffic forecasting. This will mainly improve the user's experience on the network as well as boost the performance of applications that will certainly increase the MSPs revenues.

Bibliography

1. **Djicks, Jean-Pierre.** *Oracle: Big Data for the Enterprise.* Redwood Shores : Oracle Corporation, 2014. White Paper.

2. **Juniper Networks.** *Introduction to Big Data: Infrastructure and Networking Considerations.* s.l. : Juniper Networks, 2012. White Paper.

3. **Woods, Dan.** A Quick Guide To Choosing The Right Way To Use Hadoop. *Forbes.* [Online] July 16, 2013. [Cited: November 8, 2015.]

4. **White, Tom.** *Hadoop: The Definitive Guide.* Sebastopol : O'Reilly, 2015.

5. **Rijmenam, Mark van.** https://datafloq.com/read/big-data-history/239. *A Short History of Big Data.* [Online] January 7, 2015. [Cited: November 8, 2015.]

6. **Rathbone, Matthew.** A Beginners Guide to Hadoop. *Matthew Rathbone Blog.* [Online] April 17, 2013. [Cited: November 8, 2015.]

7. **Hortonworks.** http://hortonworks.com/big-data-insights/how-facebook-uses-hadoop-and-hive/. *How Facebook uses Hadoop and Hive.* [Online] March 25, 2013.

8. **Borovick, Lucinda and Richard L. Villars.** *The Critical Role of the Network in Big Data Applications.* Framingham : IDC, 2012. White Paper.

9. *Hive: a warehousing solution over a Map-Reduce framework.* **Thusoo, Ashish, Joydeep Sen Sarma, Namit Jain, Zheng Shao, Prasad Chakka, Suresh Anthony, Hao Liu, Peter Wyckoff and Raghotham Murthy.** 2009. Proc VLDB Endow. pp. 1626-9.

10. *Building a Network Highway for Big Data: Architecture and Challenges.* **Yi, Xiaomeng, Fangming Liu, Jianchuan Liu and Hai Jin.** 2014, IEEE Network, pp. 5-13.

11. *Monitoring and Analyzing Big Traffic Data of a Large-Scale Cellular Network with Hadoop.* **Liu, Jun, Feng Liu and Nirwan Ansari.** 2014, IEEE Network, pp. 32-39.

12. *Big Data: Transforming the Design Philosophy of Future Internet.* **Yin, Hao, Yong Jiang, Chuang Lin, Yan Luo and Yunjie Liu.** 2014, IEEE Network, pp. 14-19.

13. *A comprehensive view of Hadoop research - A systematic literature review.* **Polato, Ivanilton, Reginaldo Re, Alredo Goldman, and Fabio Kon.** 2014, Journal of Network and Computer Applications, Vol. Volume 46, pp. 1-25.

14. **Matti, Mona and Tor Kvernvik.** Applying big-data technologies to network architecture. *Ericsson Review.* 2012.

15. **Gantz, John and David Reinsel.** Extracting Value From Chaos. *IDC.* June 2011.

16. **Grigorchuk, Kirill.** Comparing the top Hadoop distributions. *Network World.* [Online] June 26, 2014. [Cited: November 17, 2015.]

17. **Experfy.** Cloudera vs Hortonworks vs MapR: Comparing Hadoop Distributions. *Experfy.* [Online] September 2014, 2014. [Cited: November 17, 2015.]

18. **Cisco.** *Top 10 Considerations for a Successful 4G LTE.* s.l. : Cisco Systems, Inc., 2013. White Paper.

19. **Hortonworks.** Understanding Hadoop Ecosystem. [Online]