HADOOP INTRANET CLUSTER SETUP AND DATA ACCESS USING WIRELESS NETWORKS

Mr.S.S.Aravinth / AP / CSE / Knowledge Institute of Technology, Salem
Mr.D.Saikumar / IV / Year / CSE   Knowledge Institute of Technology, Salem
Ms.S.Habipriya / IV / Year / CSE   Knowledge Institute of Technology, Salem
Ms.S.Sowmya / III / Year / CSE   Knowledge Institute of Technology, Salem
D.R.Sudharsan /III / Year / CSE   Knowledge Institute of Technology, Salem

Abstract:

This modern world relies on information transmission and data processing which is always challenging even for the best data managers in the world. This crisis must be overcome with some solution like **hadoop system architecture.** The data needs in various organization can be fulfilled greatly using this system and it is a beneficial way of big data approach. The Hadoop system is a part of Big Data which is used to handle large quantity of data with short time and great accuracy. The system can analyse, process, store huge data within short span of time. The entire system consists of a **Name node, secondary name node,** and **slave nodes.** The name node acts as a central server which is directly connected to the client systems. The slave nodes are processing node which is again connected to the name node. The name node has the meta-data of the data in the secondary node and the slave nodes will have the data stored in HDFS file system. All these are interconnected through networks. The client and the name node are connected through wireless network so that the data can be access through a local Wi-Fi access point. Hadoop system architecture helps improve the data analytics and processing of large amount of data and making the system more reliable by data redundancy in multiple nodes with least risk of data loss in the system. As the data is split and processed by multiple nodes, the possibility of data in the queue is reduced and the processing system's efficiency is improved. The name node can serve large queries compared to other data systems. So this system is better than many data processing systems that are currently used. Hadoop system can satisfy the future data needs and it will be the system for change.

748

KEYWORDS

PHP: Personal Home Page, **SQL**: Structured Query Language **FTP:** File transfer protocol, Authentication, Map Reduce.

I) OVERVIEW

Data requirement and analysis is the major part in the current IT world. The data processing methods that are currently implemented are neither fast nor reliable. As this requirement grows day by day new systems must be implemented to overcome the problem of lack of performance and reliability. Considering this, Hadoop distributed system has found a place to improve the efficiency of the database system. Unlike a normal distributed system the hadoop architecture is a robust and it is capable of mining, processing, and storing the data within a reduced span of time.

Using this feature, an intranet can be set up in the local domain and the data can be managed all across the domain. This is done using wireless networks and any web browser can access the architecture to fetch process and store the data. The data from the database can be accessed using FTP protocol. The clients or peers are connected to the wireless access points and these access points are in turn connected with the database. Providing security modules the data in the database are isolated from one another as they cannot be accessed without the proper authentication. This makes data in the database more secure as only the authenticated user has the privileges to fetch the data. The front end will be PHP and the backend will be SQL. Hadoop system serves the user with considerably greater performance and the data in these systems are more reliable than a traditional database system. It also provides enhanced security to the data compared to the existing system. Implementing this system will make the data accessible without bottle-necking the server hence this system will serve multiple demands simultaneously. This makes the data portable within the local domain securely.

II) DBMS- INTRODUCTION

A database management system (DBMS) is a collection of programs that enables the user to store the data , modify the data , and also extract the information from database .types of DBMSs are ranging from small systems that may run on personal computers to huge systems ,and also run on mainframes. The following are examples of database applications:

- computerized library systems

- automated teller machines

- flight reservation systems

- computerized parts inventory systems

III) DATA WAREHOUSE

A data warehouse could be a electronic information service that is designed for question and analysis instead of for group action process. it contains historical information derived from group action information, however it will embrace information from alternative sources. It separates analysis work from group action work and permits a corporation to consolidate information from many sources.

IV) APACHE HADOOP

Apache Hadoop is 100 % open supply to all the users , and also is pioneered essentially a new approach of storing and process knowledge. rather than wishing on expensive ,proprietary hardwares and completely different systems to store and method knowledge, Hadoop allows distributed data processing of giant amounts of informations across the cheap, and also the industry-standard servers may store and method the information, and may scale while not limits. With Hadoop, no knowledge is simply too huge. And in today's hyper-connected world wherever lot of knowledge is being created on a daily basis, Hadoop's breakthrough blessings mean that companies and organizations will currently realize price in knowledge that was recently thought-about useless.

HADOOP CLUSTER MANAGEMENT

Cluster management is that the method of managing the cluster of slave nodes that ar connected to the name node. Organizing the clusters properly create the whole system additional reliable with improved performance. This includes the choice of setting of of HDFS on the nodes and running the map scale back programs on the clusters within the system. a correct and economical system ought to have well organized cluster to run cleanly.

V) MODULES IN THIS PROPOSED WORK

Module-1: SET UP CENTOS SERVER

- ✓ Installing centos in the server.
- ✓ Configuring the server.
- ✓ Installing the dependencies.

The cent os is a linux distribution which runs on the server system and manages all the clusters connected to the server (Name node). This operating system is used because it has all the basic components that are required for the server setup is present during installation. This operating system is installed and configured in the server and then the Hadoop package is deployed.


Module-2: CONFIGURING THE CLUSTERS

- ✓ Installing UBUNTU on the clusters.

751

✓ Configuring and installing dependencies.

✓   Allocating data nodes.

✓ Connecting the clusters to the domain.

The data nodes will be installed with the Ubuntu operating system as this has the dependencies pre-installed for map reduce and other operations in the Hadoop system. The slave nodes and the Name nodes are connected through wired network. A wired network is used in this case to reduce the packet loss and reduce the latency in reaching the host. The dependencies like SSH and java packages are installed in the slave nodes. These packages are then configured to connect to the primary node. After configuring the slave nodes, the path for the domain in the server is set and a SSH connection is created to the server. This enables all the nodes in the cluster to fetch the server through SSH and the server will maintain a track all the clusters connected to it. After connecting start the Job   tracker   and   task   tracker   services   in   the   nodes.
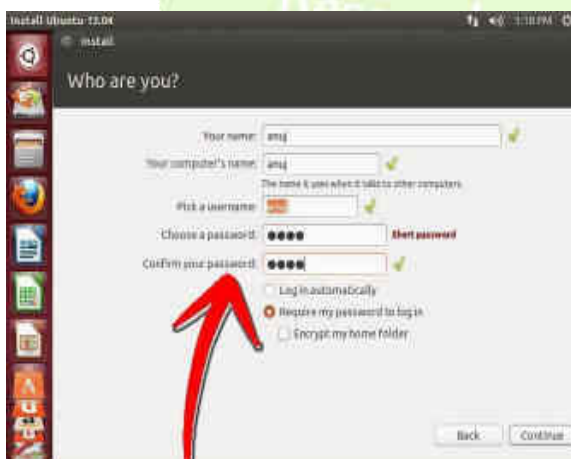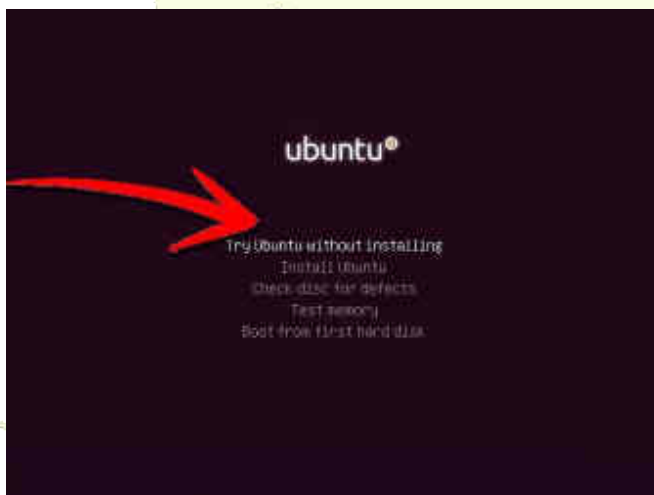
Module-3: Creating Front End Application

✓ Adding login page.

✓ Accessing the database for authentication.

✓ After authentication the data in hadoop cluster is accessed.
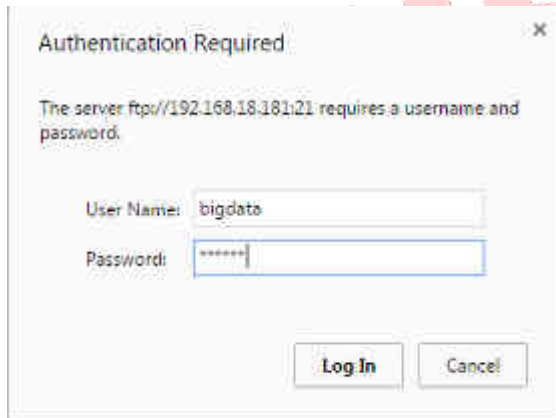
The login page has to be created in order to login securely into the server. This allows only authenticated users to have access to the server. By this the security of the system is improved and the data loss and the data theft can be eradicated. For the login page and other interface, a web page is built which connects the client system to the server system. These interfaces are designed using HTML and XML languages. This acts as an interface to connect with the server.

Module-4: Implementing Wireless Connectivity

✓ Setup wireless access point.

✓ Configuring the access point.
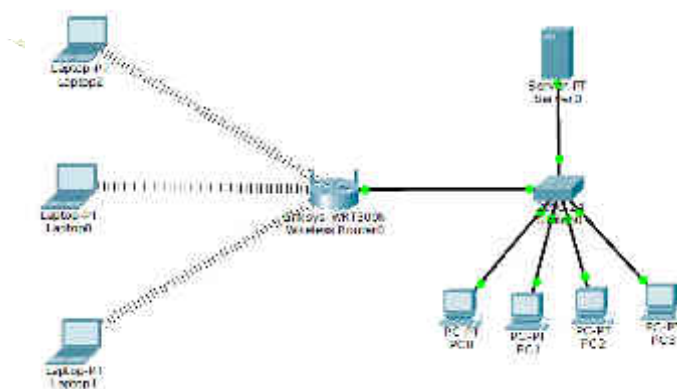
The client system is connected to the server through the wireless networks. This enables the portability of the machine without messing up with the wires. This also provides robust support that the server can be fetched through the web interface and this doesn't require any particular tool to make it platform dependent. The name node and the slave

753

nodes are connected through the switch in which a wireless router will be connected so that the client systems will be connected to that router. The router is configured with the ip address to route the server directly and the data is fetched.

Module-5: Configuring FTP and data access

✓ Configuring the FTP in the server.

✓ Linking login page.

✓ Accessing the data from hadoop cluster using web browser.
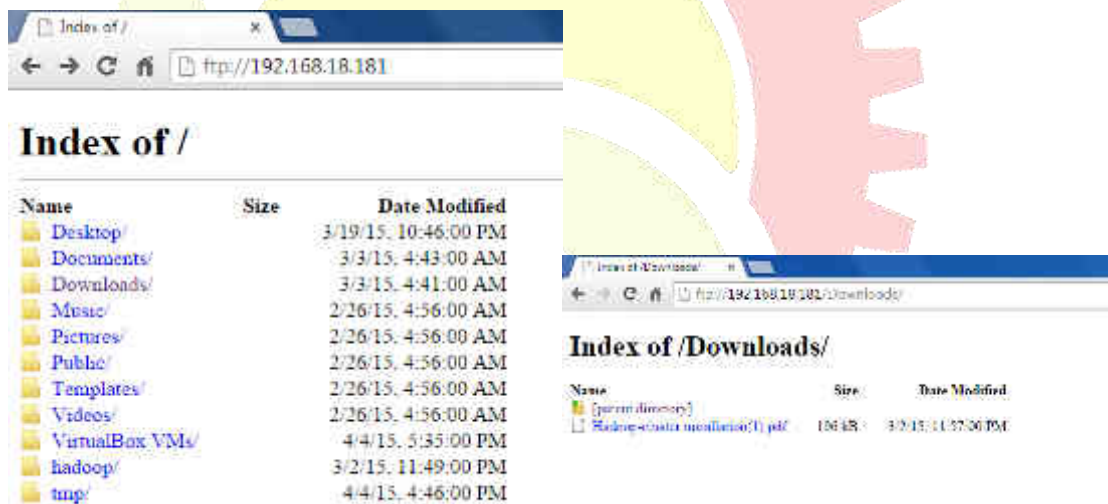
The ftp server is setup and then configured to the requirement. The domain name and the IP address is set to connect the server. The local path is set where all the files will be stored in the server. The path can be accessed remotely through the ftp service that runs in the server. The surever intron stores the data in the slave node and takes a metadata of all the files that are being stored. The FTP service is run on the web browser and the credentials are given for the authenticated access. After that, the files in the server will be displayed and they can be accessed through the web browser.

CONCLUSION:

The Hadoop Intranet setup has been played a major in an organization which wants to access multi tasks on Hadoop cluster setup. Without going to the physical master node

754

setup , the rest nodes can be accessed in the system that anywhere else in the same premises.

REFERENCES:

[1] Michael Minelli, Michelle Chambers, and Ambiga Dhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses", Wiley,2013.

[2] P. J. Sadalage and M. Fowler, "NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence", Addison-Wesley Professional, 2012.

[3] Tom White, "Hadoop: The Definitive Guide", Third Edition, O'Reilley, 2012.

[4] Eric Sammer, "Hadoop Operations", O'Reilley, 2012.

[5] E. Capriolo, D. Wampler, and J. Rutherglen, "Programming Hive", O'Reilley, 2012.

[6] Lars George, "HBase: The Definitive Guide", O'Reilley, 2011.

[7] Eben Hewitt, "Cassandra: The Definitive Guide", O'Reilley, 2010.

[8] Alan Gates, "Programming Pig", O'Reilley, 2011.

[9]1.http://tecadmin.net/setup-hadoop-2-4-single-node-cluster-on-linux/