

A Comparative Study of Andrew File System and Hadoop Distributed File System Framework to Manage Big Data

Mr. Rajesh R Savaliya, Dr. Akash Saxena

Abstract-Big-Data is the storing and retrieving is the current and first challenge in the field of IT. The large amount of data generated by different information system .Those generated data in the forms of structured, unstructured or semi-structured .now a day it become big-challenge to store the Big-Data. This paper introduce study and the comparison of two distributed file system managing and handling different parameters with respect to frameworks of both distributed file system which is used to handle and maintain storage and access of the large amount of generated big data in the Andrew distributed file system and Hadoop distributed file system. This study and analyses will more useful for understanding the insight into framework of Andrew and HDFS and highlight the features those are common and difference between both distributed file systems.

Keywords: HDFS, ANDREW, DFS, BIG-DATA, HADOOP, VICE, VENUS.

I. INTRODUCTION

Big-Data is the one useful word which is used to manage and maintain the large size of data, that is generated by different digital communication or e-transaction and different social media used to communication in the whole world now a day. Andrew Distributed File System and Hadoop distributed File System implement to manage and handle large size of data. To store as well as retrieve the Big-data challenges in the field of IT. Differences and similarities in the both distributed file system that design based on different parameters, its levels and different criteria to handle the large amount of data file [11]. The main objective of HDFS and Andrew ware developed to maintain and handle the large size of data files arriving from different terminals in a variety of formats and large amount of data. Both HDFS and Andrew distributed file system are developing to handle big-data. In this paper analysis and comparatively study done with respect to different parameter and features of both Andrew and HDFS. Andrew distributed file system was first distributed file system to handle the large amount of the big data. Hadoop is the latest distributed file system to handle the large amount of the big data.

II. ANDREW FILE SYSTEM (AFS)

Andrew file system (AFS) was developed by Carnegie Mellon University in the 1980's Project led by Prof. M. Satyanarayanan. The main objective of this project was to develop a scalable Distributed file system to support so many clients as possible [2]. Andrew File System (AFS) is a distributed file system used for Enterprises use an AFS to facilitate stored and access between AFS client machines located in different location on the network [1]. Andrew Distributed file system was designed for location independent file system using local cache to increase the performance in the distributed computing framework and also reduce work load for the system. The first request comes for data from the any terminal is satisfied by the server and then this request put into the local cache. When second request comes for same data that is satisfied from the cache [3]. The designing and implementation went to through different reversion: AFS1 in 1983, AFS1 in 1987, AFS1 in 1988, AFS1 in 1989, AFS1 in 1993, AFS1 in 1998, AFS1 in 2000, AFS1 in 2001, AFS1 in 2005, AFS1 in 2008, AFS1 in 2010[5].

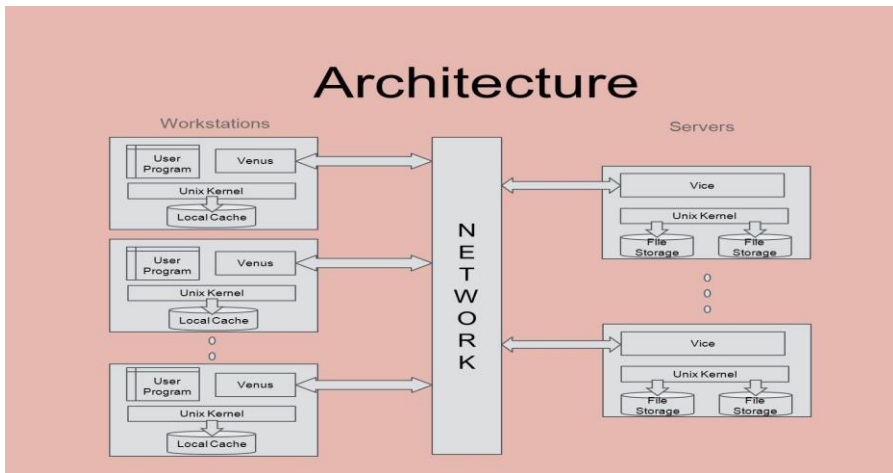


Figure 1: Andrew Architecture [4].

Andrew file system (AFS) has two important software section with the two UNIX processes Called Voce and Venus. **Vice:** Vice is the Name of the Server Software that will run as a User-level UNIX process in each server Computer. **Venus:** Venus is the user level process that runs in each client computer. Files are available in two

different formats either in local or shared form. Local files are stored on the Terminal's disk and those files are available only to local user processes. Shared files are stored on the Servers and copies of those files are cached on the local disk of Terminal [9].

III. HADOOP DISTRIBUTED FILE SYSTEM (HDFS)

Hadoop Distributed File system which is the framework to manage and maintain large size data file. Hadoop Distributed File system is an open source and design by Apache. So many network based application implemented using the concepts of hadoop such as social media like Whatup, Facebook, and Amazon. Hadoop Distributed File system which is used to store large size file dataset. HDFS again divided into mainly two components such as NameNode and DataNode [7].

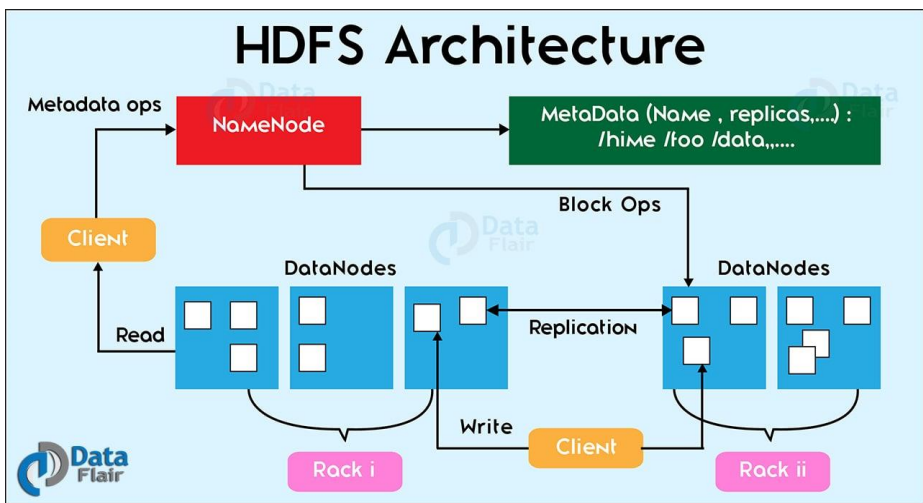


Figure 2: Andrew Architecture [6].

➤ NameNode

The NameNode is used to handle and store the namespace of large size of data file stored in the DataNode. NameNode is also used to maintain heart beat between NameNode and DataNode. NameNode handle the access control of data file access by the clients.

➤ DataNode

The DataNode is used to manage the storage of large size data file in the available different data block in the cluster of HDFS. DataNode is used to store replica in different block around the available cluster of HDFS.

IV. A COMPARATIVELY STUDIES AND ANALYSIS OF ANDREW AND HDFS DISTRIBUTED FILES SYSTEMS

| Operational Keywords | Hadoop Distributed Files System | Andrew Distributed Files System |
|--------------------------|--|--|
| Goal to Design | To design for handle the large size files in the HDFS file system. | To design for handle the large size files in the Andrew file system. |
| Languages used to Design | Java Programming | UNIX |
| Supported OS | Linux and Windows are supported, but BDS, Mac OS/X, open solaris are known to work [10]. | Linux and Windows ,BDS,HP-UX, Mac OS/X, open solaris [10] |
| File Management | HDFS supports a traditional hierarchical directories data structure. HDFS also supports third party different file system such as Cloud Storage, and Amazon storage services [8]. | Location-transparent UNIX file Name space to clients. AFS combin files into Volumes and distribute files across many machine [8]. |
| Processes Management | NameNode, DataNode. | Vice (Server) and Venus (client). |
| Default Size | The default size is 128 MB and it can be changed by the user. | It has no default size. |
| Cache Management | HDFS uses Distributed cache. This facility provided by MapReduce Framework. | Cache on the local disk. This facility provided by UNIX kernel in the main memory. |
| Replication | Replication done automatically. | Volume has frequently access data can be read-only replicated on the server. |
| Currently implemented by | Yahoo, Facebook | OpenAFS |
| Snapshot | Snapshot facility is available | Snapshot facility is not available |

V. CONCLUSION

From the comparatively analysis, it is concluded that this paper analyze an insight of comparatively studies towards Firstly develop Andrew Distributed File System and currently most used Hadoop Distributed File System powerful distributed big-data processing Framework. This studies was performed to observe the parameters for both Andrew and HDFS big-data transactions such as storing, managing, maintain, implementation, replication, cache management as well as retrieving, etc. of large scale data file. HDFS provide the better Snapshot, Replication and file management as compare as Andrew Distributed File System.

VI. REFERENCES

- [1] <https://www.techopedia.com/definition/11029/andrew-file-systemafs>. [Accessed: December. 1, 2018]
- [2] <http://www.ece.eng.wayne.edu/~sjiang/ECE7650-winter-15/topic-S.pdf> [Accessed: December. 3, 2018]
- [3] <https://searchstorage.techtarget.com/definition/Andrew-file-system> [Accessed: December. 3, 2018]
- [4] <https://slideplayer.com/slide/4842653/> [Accessed: December. 4, 2018]
- [5] <https://www.slideshare.net/ManfredFuruholmen/afs-introduction> [Accessed: December. 6, 2018]
- [6] <https://data-flair.training/blogs/hadoop-hdfs-architecture/> [Accessed: December. 8, 2018]
- [7] http://en.wikipedia.org/wiki/Big_data .[Accessed: December. 10, 2018]
- [8] R.Vijayakumari, R.Kirankumar and K.Gangadhara Rao, “Comparative analysis of Google File System and Hadoop Distributed File System”, International Journal of Advanced Trends in Computer Science and Engineering, Vol.1, PP. 553–558, 24-25 February 2014.
- [9] Monali Mavani, ” Comparative analysis of Andrew File System and Hadoop Distributed File System”, Lecture Notes on software engineering, Vol. 1 No. 2, PP. 122-125, May 2013.
- [10] Akram Elomari, Larbi Hassoni, Abderrahim Maizate, “The main characteristics of five distributed file system required for big data: A comparative study”, Advance in science technologies and engineering system, Vol. 2 No. 4, PP. 78-91, 26 June 2017.

- [11] Gemayel. N, “Analyzing Google File System and Hadoop Distributed File System”, Research Journal of Information Technology, PP. 67-74, 15 September 2016.

AUTHOR’S PROFILE



Mr. Rajeshkumar Rameshbhai Savaliya from Ambaba Commerce College, MIBM & DICA Sabargam and master degree in Master of science and Information Technologies(M.Sc-IT) from Veer Narmad South Gujarat University. Rajesh R Savaliya has teaching as well programming experience and PHD Pursuing from RAI University.

Dr. Akash Saxena PhD professor and guide from Rai University.

