International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

Digital Transformation of Data Center using Redfish, Django and Hadoop technologies

Shreekara C S[1], Nagaraj Bhat[2]

[1]Student, Department of Electronics and Communication Engineering, RV College of Engineering, Bengaluru -59, Karnataka, India
[2]Assistant Professor, Department of Electronics and communication Engineering, RV College of Engineering, Bengaluru – 59, Karnataka

---

Abstract - *Data center is a space which accommodates the different storage, telecommunication and computer systems. Data center has many redundant infrastructure for the power, data communication connections, and security devices. This results in the increase of operational expenditure and reduces the efficiency of the data centers. The main issues with the present data center management systems are that they don't have, a data driven decision making systems, a smart data storage and data collectors, and have poor user experience. This project is trying to, build a smart data collector system, to enhance the present asset management system by automating it, to create a data lake to store all the different types of data to make it more efficient for faster access. Smart data collectors are developed using Nmap and Redfish. Automation of the asset management system is done using Django and MySQL. Data lake is created using Hadoop clusters, and the data is ingested to the data lake using Nifi. This system of smart data collectors, Asset management, and Data Lake, will improve the efficiency of the data center.*

***Key Words*: Nmap, Redfish, Django, MySQL, Nifi, Asset Management, Data Lake.**

1. INTRODUCTION

Wide ranges of methods can be employed to improve the efficiency of the data center, like, a standard for the data center like redfish [1], data analysis to understand the usage of the components, pipeline model to improve the data flow and different ways to ingest the data into the data lake in the data center, thus improving the efficiency of the data center [2-5], Data lake to store the data in an efficient manner using sql [6], and using the Hadoop technologies to create the data lake [7-9]. And also understanding the different issues regarding the data center and suitable technologies to solve those issues [10]. This paper focuses on development of a data center management system using the new technologies and automating certain aspects of the data center. Developing a smart data collection system, automated asset management system, building the data lake to store all the data about the data center and using these data for data analysis to improve the efficiency of the data center.

2. SYSTEM DEVELOPMENT

2.1 Methodology Adopted

Existing tools for data collection, asset management, and Data Lake were studied for identifying improvements [11-13]. A smart data collector is developed to collect all the possible data from the various components of the data center using red fish, which involves information about the chassis, managers, IP addresses, etc., and this obtained data in Java Script Object Notation format is given to the asset management system. Asset management system was created to account for all the components present in the data center, and a web application using Django for an easy access to all the information and maintenance of the devices was built. Data modelling system to process the output data from the data collectors and was stored in the data lake which was developed using a Hadoop based server. And the data ingestion to the data lake was done using NiFi.

2.2 Experimental Details

The block diagram, shown in the Fig -1, depicts the major components and sequence of work flow for the data center management system.

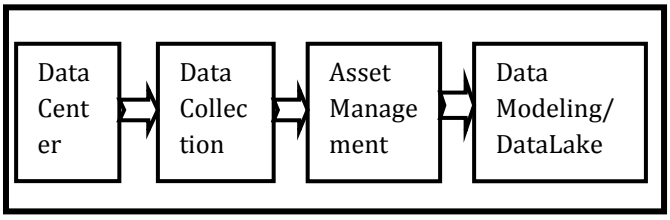

Fig -1. Block Diagram of Data Center Management System

Data Center has various components like storage units, processors, routers, switches, and other networking components. First step is collecting all the possible data from the various components of the data center using red fish, which involves information about the chassis, managers, IP addresses, etc., this obtained data in Java Script Object Notation format is given to the asset management system. Asset management involves creating a data base to account for all the components present in the data center, and also building a web application using Django for an easy access to all the information and maintenance of the devices. Data

modelling involves processing the JSON data and storing it as a raw data in a repository and also making models of the data on the basis of use, storage, etc. and storing them in various other repositories in the data lake which is a Hadoop based server, using Nifi (a special software built for data ingestion).

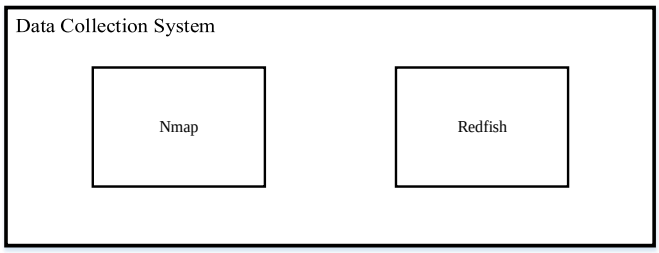

Fig -2. Components of a data collection system

Data collection system comprises of two main components, Nmap for network discovery and Redfish for the actual collection of the data as shown in the Fig -2. Nmap first pings all the IPs present in the data center and obtains the status of these devices. And it sends the list of all the IPs that are active to Redfish system for the collection of the required information about the devices.

Fig -3. Shows the model of Asset Management System which has MySQL as the database and Django as the web application framework. MySQL is a relational databases, which has a lot of tables, to accommodate different types of devices that are present in the Data center. Django is used to represent all these data about the devices and to provide the users with different functionalities necessary to improve the ease of managing the Assets present in the data center. As shown in the model diagram below, all the different types of devices that are present are divided into two categories, storage devices and server devices, which is owned by a User. Django displays the devices that belongs to the user in an Account page and a User also has the option to assign these devices to the subordinates. This is how the asset management system works. There are many other functionalities that are present in the web application along with the view like, options to update information about the device and decommission the device from a user, and more. They are categorized into the tables related to the user, and the tables related to the device, and the tables related to the device pool. There are over 30 different types of devices that have over 25 attributes. Django does the integration between the web application development and translating these tables and manipulating these tables using python. A login page is developed where in the user can view his account, and the account page has the details of all the device that a use owns, and he can assign these devices to his subordinates, and when the work is completed, he can take away the device. This helps in keeping track of all the devices that are not being used at a point in time. If there is a central repository where the information about all these devices are stored, it will be easier to manage the devices.

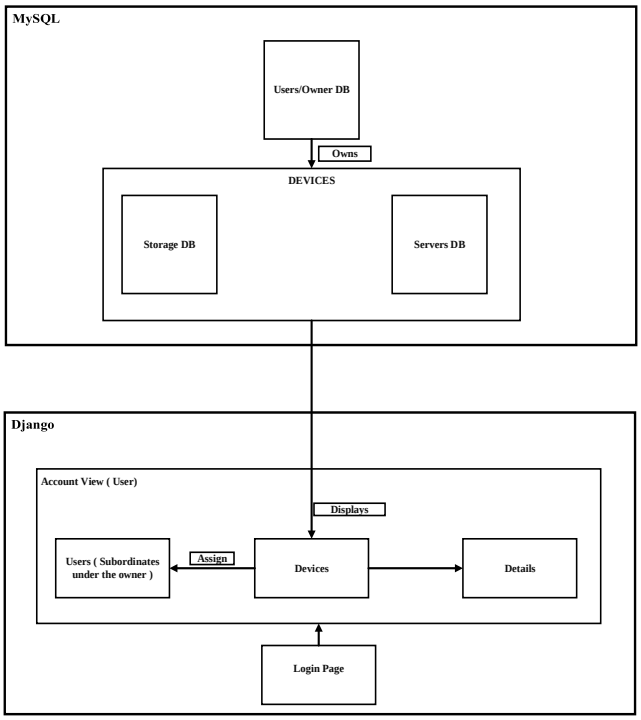

Fig -3. Model of the Asset Management System

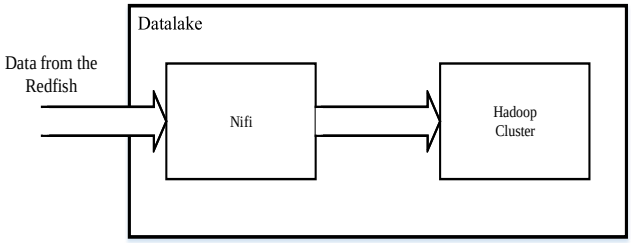

Fig -4. Components of the data lake

Data Lake is the place where all the data about the data center is stored in the raw form. As shown in the Fig -4. Data Lake is made up of Nifi for data ingestion and Hadoop cluster for storage of the data. Several copies of the data is stored in the data lake, there is a central repository where are the structured and unstructured data collected using the Redfish is stored. And several other repositories where the copies of the same data is stored, but it will be grouped based on the types, etc., which enables for a faster access of the data in the common used cases. Data Lake is made up of Hadoop clusters, which stores all these data collected from all the active IP's. Nifi is the software used to ingest the data into the data lake. Nifi also has some inbuilt features which can be used in the project. It has the option to filter different types of files, and more. This is also the part of data analytics.

3. CONCLUSIONS

Data center transformation using new digital technologies is essential so as to increase the efficiency of the data center and also to handle the increase in quantity of data, as the data center expands. To know which device is active in the data center is very important, as these devices consume a lot of power. So recognizing which device is active is very

important to reduce the power and to reduce expenditure on the power. Smart data collectors do this work of recognizing which device is active and collecting the data from them. Asset management system will have a database of all the devices present in the data center, automating it to continuously update without needing the manual work will reduce the time of data retrieval and also reduce the operational expenditure by removing the need for manual monitoring of the asset management system. Data Lake stores all the data about all the devices, and this data is used for data analytics to draw inference about the condition of the data center.

Asset Management system is automated using Django and MySQL data base. Django is a python web application framework, using the many tools present in the python, and by linking it to the MySQL database where all the asset data is stored, automation is done. Data lake is created using Hadoop clusters which stores the data in the formation of Hadoop File System, which can store all forms of data and will reduce the time of retrieval, as it has an efficient way to store the data. Nifi is a data ingestion tool used for ingesting data into the data lake. Smart data collectors are developed using the Nmap(a network discovery device) and Redfish(a tool for collecting the data from the devices present in the data center), and this data is fed into the Asset management system and the Data lake. The smart data collection system, Asset Management system and the Data Lake reduced the time for data retrieval by almost 50%.

REFERENCES

[1] G. Gonçalves et al., "A Standard to Rule Them All: Redfish," in IEEE Communications Standards Magazine, vol. 3, no. 2, pp. 36-43, June 2019.

[2] P. Minet, É. Renault, I. Khoufi and S. Boumerdassi, "Data Analysis of a Google Data Center," 2018 18th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID), Washington, DC, 2018, pp. 342-343.

[3] H. Isah and F. Zulkernine, "A Scalable and Robust Framework for Data Stream Ingestion," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 2900-2905.

[4] F. Lei, L. Yu, B. Shao, F. Teng and B. Zhou, "Large Scale Data Centers Simulation Based on Baseline Test Model," 2018 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), Vancouver, BC, 2018, pp. 57-68.

[5] M. Levy and J. O. Hallstrom, "A new approach to data center infrastructure monitoring and management (DCIMM)," 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, 2017, pp. 1-6.

[6] Dholakia et al., "Designing a high performance cluster for large-scale SQL-on-hadoop analytics," 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, 2017, pp. 1701-1703.

[7] Yamashita et al., "Hadoop triggered opt/electrical data-center orchestration architecture for reducing power consumption," 2017 19th International Conference on Transparent Optical Networks (ICTON), Girona, 2017, pp. 1-4.

[8] H. Kodama, H. Endo, S. Suzuki and H. Fukuda, "High Efficiency Cloud Data Center Management System Using Live Migration," 2017 IEEE 10th International Conference on Cloud Computing (CLOUD), Honolulu, CA, 2017, pp. 733-738.

[9] Verma, A. H. Mansuri and N. Jain, "Big data management processing with Hadoop MapReduce and spark technology: A comparison," 2016 Symposium on Colossal Data Analysis and Networking (CDAN), Indore, 2016, pp. 1-4.

[10] H. I. Bahari and S. S. M. Shariff, "Review on data center issues and challenges: Towards the Green Data Center," 2016 6th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), Batu Ferringhi, 2016, pp. 129-134.

[11] E. Salbaroli, E. Benetti, A. Fiocchi and G. Mazzini, "Database as a service in a public administration datacenter," 2016 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), Varna, 2016, pp. 1-5.

[12] S. G. Manikandan and S. Ravi, "Big Data Analysis Using Apache Hadoop," 2014 International Conference on IT Convergence and Security (ICITCS), Beijing, 2014, pp. 1-4.

[13] M. Pawlish, A. S. Varde, S. A. Robila, C. Alvarez, C. Fleischl and G. Serviano, "GreenDSS tool for data center management," 2014 5th International Conference on Information and Communication Systems (ICICS), Irbid, 2014, pp. 1-6.