HADOOP CONFIGURATION IMPLEMENTATION AND DEPLOYMENT MODEL IN VIRTUAL CLOUD ENVIRONMENT

[1]Nasrullah, [2] Akanksha Bana

[1] M.Tech (SE) School of Engineering & Technology, Noida International University

Plot 1, Yamuna Expy, Sector 17A, Uttar Pradesh 203201

[2] Assistant Professor, School of Engineering & Technology, Noida International University

Plot 1, Yamuna Expy, Sector 17A, Uttar Pradesh 203201

Abstract Recognizing the value of big data in any organization an in age which reach to huge amount of data. That cannot be handled by the traditional computing techniques to sort, collect and analyze their business intelligence, by which can't properly recognized the need to add big data capabilities to their own efforts. The complex data volumes to store, process and analyzed by Hadoop in virtual cloud environment, will be faster than pervious computing technologies. This growth of demands for new strategies big data processing and analyzing information based on Hadoop, MapReduce, Spark and HDFS has become a powerful computation Models which addresses these problems in the cloud environment. Our Research's mainly effort to evaluate Apache Hadoop in different distribution environment to provide a model-driven approach for all organizations, easy implementing and Hadoop based core components but its ever changing demand for deployment Hadoop with a great challenges, difficulty and complexity. We have to propose a model-driven for deployment and transformation of Hadoop Model in this paper which have generic implementation options. There for we mainly focus to deliver an approach, which has bidirectional synchronization between running system and Hadoop deployment model which will be efficient, reliable and low-cost to support any organization needs in any virtual cloud environment.

Keywords: *Cloud Computing, Big Data, Hadoop, Hadoop on the Cloud, Related Work and Background, Overview on Different Model-Driven Approaches, Future Work*

1. INTRODUCTION

Now a days about 2.5 trillion bytes of huge amount of data is generated every day, and cloud computing has been driven fundamentally creating suitable environment by the need to process an exploding quantity of huge data in terms of terabyte to petabytes data, to store and process according to different organization needs. If we look to the data which is generated from different sources and we are surrounded the exponential growth to present the challenge, risk, and data lock-in situation in the cloud environment to this huge problem cutting–edge business such as Facebook, Yahoo, Amazon, Microsoft, Twitter, Google etc. These companies generated huge amount of data which cannot be handled by the traditional computing techniques to process and analysed within a specific time limit and it's quite difficult these huge amount of datasets to store, process, analysed and make a proper report within specific environment. There for to implement Hadoop in virtual cloud environment is the best solution which is used to define a bright future of ITs. The traditional database management system has a size cap as how much data it can process. If this is the case, then it is absolutely impossible to overcome the problems faced by the big data which grows exponentially. Moreover, some database supports infinite data but the data accessing time and processing time is extremely large which is not feasible. Data has become the raw material of production, a new source for immense economic and social values. Advances in data mining and analytics the massive increase in the computing power and data storage capacity have expanded by order of magnitude the scope of information available businesses and government. Over the past few year the data collected by private and government organization has exploded. But now the data is generated from online transactions, emails, videos, images, clickstream, logs, searches queries, health records and social networking interactions. This and much more contribute to good amount of heap of data. There for Doug Cutting and Mike Cafarella invented Hadoop and the name became Hadoop after his son's toy elephant. The Nutch search engine project was first supported by this Hadoop. Later on it was acquired by Apache and was made open source framework. Whenever a company want to process huge amount of data like an in age of big data there for the company has two choice in the first choice they have to buy a bigger machine which have more Disk Space, RAM, CPU Power etc. this is known as scaling vertically. But this should not be a proper solution for the company there is a limit how bigger machine you have to buy for it this option is not work if we are talking and thinking about internet scalability so this solution is not working properly in this environment. The other solution would be horizontally that's means connecting some database vender to bring the beigest solution for the current problems these solution is not cheap as needed this usually mean contacting some significant investment and need high cost ratio. There for the Cloud is used to provide on demand self-services, unlimited data storage, on demand computation power, from world wide access resources by using internet with limited bandwidth to process these huge dataset with minimal effort. Typically cloud is hosted on server from which provide fully hardware and software resources they user only pay for those resources which they are used in their daily life these resources can be increased or decreased base on the service level agreement between the user and the cloud service provider here the user only care about the contents how the provider take care of the services, maintenance, security issues and networking, the complete service is usually virtualized which is not required to include in a cloud base environment. Virtualization is a complete and the most important pillar in the cloud computing. Are virtualization is a logically representation using computer software which enables for the provider to

maximized raw hardware and gives the cloud elasticity, ability for the user to scale the instances required. In a physical computer system there is run one are more application by using the virtualization concept a software that run on a single physical computer system that consist single physical resources that running in to different virtual machines. The most important advantages of the virtualization one machine is not affect the other machine resources and effectively used of the hardware which is, sharing the resources to other virtual machines billions of dollars have been invested on the research to control view and analyzed the big data to use the less number of servers and machines and heat dissipation in data centers. Hence virtualization allow fewer physical hardware and reduce fully cost saving, maintenance, less electricity and backup is easy to make by using the snapshot mechanism. Now days there are a lot of open service computing solution for providing infrastructure environment in hybrid, public and private clouds such as open Nebula, Cloud Stack, and Eucalyptus etc. These open source solutions and easily deploy on the private/public and hybrid infrastructure which is cloud based in (IaaS) model that do not need any homogenous configuration but it is also possible different hypervisors on different GUI/Linux based virtual machine clusters. In this paper we meanly focus to reduce the hardware and software resources in virtual cloud environment but it is a ever-changing demand for deployment Hadoop which is face to challenges, difficulty and complexities we also conclude most of the present researchers lie in the configuration and implementation fully parameter setting up Hadoop Deployment model in virtual cloud environment. They don't take any specific diversion of infrastructure and scalability issue in the cloud by in order to expanded Hadoop services deployment according to their own demand but here we firstly proposed Hadoop deployment and demand model are presented and secondly a specific transpiration method form one model to another model is proposed, in the last there will be a bidirectional synchronization between the Hadoop demand model and running systems based on the runtime generate platform and will be completed model for any clouds.

2. Cloud Computing

Cloud computing refers to both which provide the services and applications delivered as service and provide the environment for development and deployment services over the internet and also provide physical resources such as a hardware and data centers that provide those services over the internet, these resource can be access and denied with minimal request. The serves themselves the cloud service level which provide as Software as a Service (SaaS) a some vendors use terms such as IaaS (Infrastructure as a Service) and (Platform as a Service) PaaS to describe their products for different venders, but we gathered these accepted definitions for them, and still used very widely. The line between "low-level" Infrastructure and a higher-level infrastructure "platform" model is not crisp in different environment. We believe the two are more alike than different, and we consider them together for different cloud environment. Similarly, the related term "grid computing" from the high-performance issues and challenges computing community have suggests different tools and protocols to offer shared computation and storage over long distances in the cloud environment, but those software and protocols did not lead to software environment that grew beyond its community. From a hardware and some of resource provisioning and pricing point of view from different venders, three aspects are new in cloud computing base on the service level agreement. The appearance of infinite computing resources available on demand self-services, quickly and efficiently enough to follow load balancing, thereby eliminating the need for cloud computing as the users to plan for ahead provisioning the services. The elimination of an up-front in each layer commitment by cloud users, thereby allowing companies to start small business and increase hardware resources but this is not a good method only when there is an increase in their needs which might be to invest high cost. The ability of the cloud computing to pay for use of computing resources on a short-term basis as needed (for example, RAM, STORAGE, CPU by the hour and storage by the day) and release them as needed, thereby rewarding conservation by letting different machines and storage go when they are no longer useful from the venders.

2.1 Cloud Computing in Today Environment

Today more and more organizations are based on cloud computing. Some of them are large companies such as Google, Cloudera, Facebook, Amazon and Twitter some of them are small, but provide useful tools such as Dropbox, Open Nebula, Go Giddy, Ali Baba etc. In today environment online storage people can use very often in all over the world. Currently the cloud computing market for SMB is expanding includes more and more companies to use the services, each of them developing the business as much as possible in different area. The main determiner is the increasing popularity of this technology is used to provide complete virtualization. One of the pioneer, Amazon supplies many types of cloud computing services over the internet. The most popular is considered as Amazon EC2, which allows users to access virtual servers through a web interface. The payment process is per-hour according to the number and size of the installed virtual machines with any types of commodity hardware's and an extra fee for data transfer rate. Other cloud services offered by Amazon includes Amazon Simple Storage Service and simple processing Model, in cloud front is so simple DB and Elastic MapReduce. The partner list for Amazon Web Services which includes names like: Facebook, Red Hat, Intel, and IBM.

2.2 Future of Virtual Cloud Computing Based on Hadoop

Cloud computing technology implies a larger storage and processing capacity than private PCs and servers in a virtual cloud and on the other hand, the dependency for IT personals for maintaining the system lower cost. The main revenue then this technology promises is cost reduction as compare the pervious technologies, through the payment made step by step as the service request, thus avoiding significant investments in infrastructure as service, which depreciates quite fast. Still easy to implement Hadoop in the virtual cloud environment to provide normal solution as platform for big data there are some difference opinion in which concerns the real cost of the supplied services base on Hadoop because it appears that some new packages, this technology can come with hidden costs and open source implementations, concerning, conversion of the current company system based on the given platform to easy implement Hadoop core components, libbers in Integration, Implementation, and Training the staff and redesigning the architecture processes through life cycle will be benefiter form small to huge organizations. By which though on the short terms, are at the start of the business these costs are low to handled big data and easy report generated, on the medium and long term schedule based, but these costs can be unpredictable in different virtual cloud environment.

Fig 1: Big data and cloud computing from A-whizlabs

3. Big Data

Big data is fully identically and modernistic sizzling topic in to today environment big data is an enormous volume amount of data which conventional databases does not have the ability to capture, process, store, manage and analyze these complex datasets due to structured, unstructured, and semi-structured data. Everyday about 2.5 trillion [3] bytes of data is generated. These huge amount of data is generated from different source like social media, sensors, smart phones, emails, blogs, etc. Big data and virtual cloud goes hand and hand Hadoop is very hot growing technology for deferent organization's needs. The two breakthroughs have helped to adopt complete solution for handling big data first availability for cloud base solutions. Secondly distribution of data over many servers, this amount of big data implementation using Hadoop and Hadoop is fully demand now a days in virtual cloud environment.

Fig 2. Four V's of big data

4. Hadoop

Hadoop is an open source distributed computing framework which is design and develop by Apache foundation organization. For the solution of big data using Hadoop low-price equipment to build large computing resource pool to improved speed and efficiency of massive data analyzing. Nowadays Hadoop is widely used in the virtual clouds computing as a low cost of the

cloud computing solutions. Hadoop works in the cloud computing in different ways but mostly used in the SaaS layer in computing to process massive data in a scalable, reliable, analyzed data efficient manner, provide data storage, processing and simple computing services.

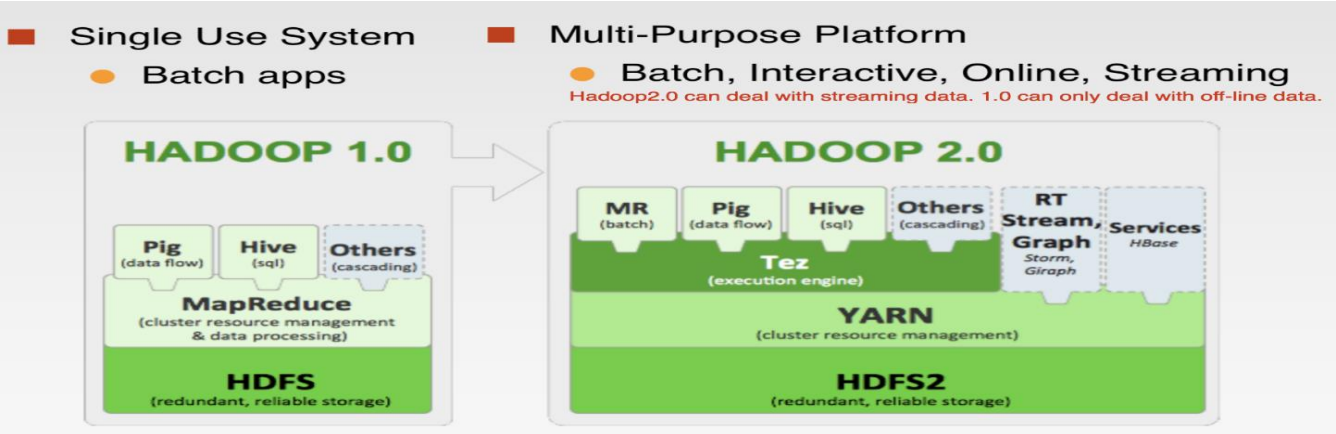

Fig: 3 Compression of Hadoop 1.0 with Hadoop 2.0 from Jingxuan's blog jingxuan.li

5. Hadoop on the Cloud

Hadoop as service in the public cloud, which provide by different organizations like amazon's EMR (Elastic MapReduce), MapR, Cloudera, Intel etc. The MapReduce can join easily and quickly run in any environment without installing Hadoop clusters. The advantages is easy setup instant recovery and low cost latency. The distribution of Hadoop in Virtual cloud Environment like (Cloudera, IBM, Hortonwork, HDF, and CDH) launched and run Hadoop as a service on public cloud here is everyone has to configure and install the Hadoop cluster in the virtual cloud on his own design environment, but it is still unknown what will be complete generic configuration and implementation Hadoop in cloud environment.

5.1 Top Challenges and Deploying Hadoop in Virtual Cloud Environment

Hadoop was design to work highly secure, on promise data centers and hence the important design features are based on the following assumptions**:**

- The cluster totally topology is faxed usually a fat-tree structure.
- In the data center consist of large number of homogeneous servers within NON-RAID disk. The whole cluster is dedicated the network to the trusted users.
- Failures are due to disk, and memory overheating fashion some other case top of the Rack Switch might be fail.
- In some case might be failures are independent.

Hadoop stores data on local disk of datanode and the replicates data is stores across the nodes and each rack for fully fault-tolerance in different nodes and fully "rack-awareness" the code which is running in different clusters. However in the virtual cloud environment of a cloud platform, Hadoop datacenters do not always hold platform for the following reasons.

- Physical nodes in a cluster might be heterogeneous and a physical node might be hosting in multiple machines to compute each other for Storage and Bandwidth for different resources.
- Replication for fault tolerance is not reliable as the data now written in to transient, virtual drives.
- The Network Bandwidth would be slow due to share number of virtual machine is running in the Infrastructure. And the cluster topology is unknown as VMs can be allocated anywhere in the infrastructure.
- The Infrastructure can be share across many organizations hence securing data is a major concern in hadoop implementation in virtual cloud environment.

5.2 Research Issues Deploying Hadoop Model in Virtual Cloud Environment

The major challenges in the storage as Hadoop relies three-way replication on the local drives data nodes for providing fully fault-tolerance in virtual cloud. The physical topology is abstracted and the concept of rack awareness is become continues and some time that two virtual machine is storing a replica reside on the same server so there for virtual machine is on important issue while deploying Hadoop on virtual cloud environment its most important to compare different virtual machine while deploying Hadoop in the cloud let's compare Amazon (EC2) and Amazon (EMR) which is used by the Hibench Hadoop Benchmarks. (EC2) is suitable for those application which less data intensive and size is less than 150 GB, and (EMR) is suitable for huge data processing it also help the resource, cluster managing and security group management. But there is also some other technologies is used for virtualization Hadoop in virtual cloud environment is VMware and Docker is also

available Docker is a light width which is the best in memory, libraries and kernel management, whereas VMware used for workstation hypervisor to manage VMs. But finally we observe deployment strategies difference only until the cluster is fully loaded, this would be a standard practice in the cloud data centers, can prove fully beneficial in virtual Hadoop clusters in different virtual cloud environment.

5.3 Related Work and Background

In [1] The author propose the approaches and deployment Hadoop on the cloud platforms also define analysis with their research issues, challenges and compared different development strategies. Big data highly impact on the technologies the author suggested a new paradigm platform on Hadoop in the cloud environment Hadoop cluster which is fully driven by the cloud and highly flexible, scalable, elastic service were explored. In [2] Introduces fully propose model on Hadoop open source platform model on Eucalyptus cloud infrastructure which arrange the service both public and private clouds. In [3] the highly identify economic cloud storage based on Hadoop the author fully introduces the pervious and current storage system in virtual cloud and design a prototype system of the cloud storage which is efficient for all cloud service providers. [4] As declared the author elastic deploying Hadoop and various application in the cloud environment on which a cloud create and destroy VM nodes automatically on demand for (N) number of users. [5] Author says that we fully address the problems on delegated access for Hadoop clusters in virtual cloud environment and how to integrated data confidentially and massive processing in the big data scenario, and provided widely accepted solutions.[6] The author proposed based on the related work processing of big educational data using Apache Hadoop in the cloud which focus on more detail analysis of user interaction and finalized report form big data in efficient manner.[7] As mention that big data processing with Hadoop-MapReduce in virtual cloud system. In the cloud via data uploaded to the cloud venders who have implemented Hadoop/Hbase. Here HDFS provide high-throughput access to these information. There for in order to manage hardware and software resources effectively and handle to big data properly we highly proposed a model-driven approach for Hadoop deployment validate, functionality, highly feasibility in virtual cloud environment. The proposed work will be implementation in near future based on diversified scenario configuration for all clouds platform must be supported.

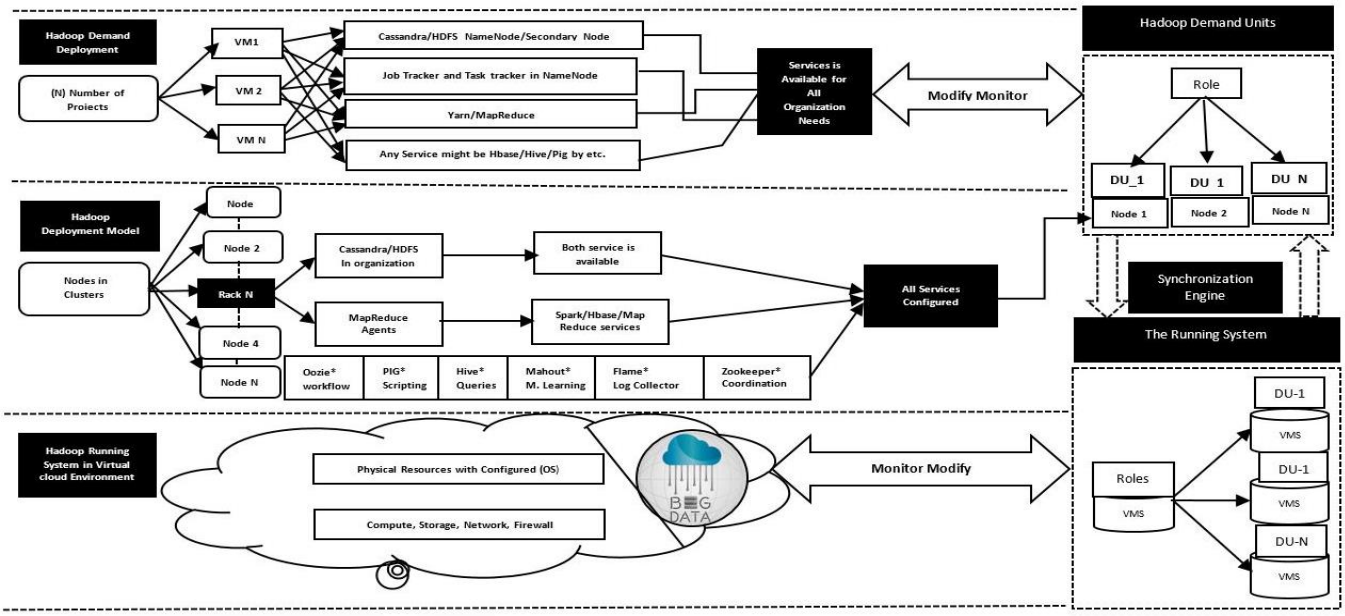

Fig 4: Hadoop Deployment Model in Virtual Cloud Environment

6. Overview on Hadoop Model-Driven Approaches

Overview on the model-driven approach to Hadoop deployment in virtual cloud environment as a shown in Fig: 4 the total work is divided into three parts. Firstly we proposed Hadoop demand model services for different types of projects for big data solutions which consist of several number of virtual machines with all configurable resources based on the Deployment model of the system, these resource platform is highly available for big data processing, storage, analytics operations and proper report generated from the Deployment Model. And secondly the Deployment Model is management view which is fully oriented the Deployment Model of the system based on Hadoop Running system which is the middle ware to provide all libraries and services to upper layer to handle big data based on transformation from the demand model to deployment model it also proved fully bidirectional synchronization Hadoop deployment model and Hadoop Running system layer and finally in virtual cloud the physical infrastructure resources is fully mange, control, and secure based on the running Hadoop with configurable resources with available managements tools.

Model Description and Implementation Hadoop in the Cloud

7. Hadoop Demand Model

In Hadoop Demand model which consists of (N) number of cluster nodes and cluster services here we divide each cluster two section infrastructure and services in infrastructure consist Nodetypes includes all nodes configuration, images is the list of nodes and images in each cluster sections, in a Nodetype like +id, +name, +ram, +disk, +cups +eid same as in images nodes +id, +name, +size, +status, + mean disk +mini ram in services consist like HDFS voice +id, +name, +version, +stats

of HDFS in MapReduce Service +id, +name, +version, +Hbase Services +id, +name, +version, +stats, these three model also have some +agent +nodeld in the demand model a part form Hbase we can configure Pig*, Sqoop*, Oozie*, Mahout* Ganglia* Nag iOS* etc. in different clusters. In cluster service section which represent the total list of services, including elements such as MapReduce, Spark, HDFS, and Hbase Services and so on. In each agent contains fully contents and access location of node where the model resides.

7.1 Hadoop Deployment Model

In Hadoop deployment model also consists the important clusters nodes and cluster services, in the cluster node section for all type huge datasets project will be represent and fully supports, a project containing virtual machine types and images, elements in a virtual machine types is list of virtual machines configuration and VMs is the list of virtual machine and images is the list of VM image. And the service section consist list of service model with the mean elements service units within the health unit, and attributes stats of deployment unit and proper location of VM where the software module is resides.

7.2 Model Transpiration form Cloud platform Services to Upper-Models

The mean method of due to transpiration from Hadoop demand model to Hadoop deployment model which consist to parts of the work-area model transformation of cluster node and model transformation of cluster services. A model transpiration of cluster node which consist of deployment model under different scenario. It is necessary to establish the element and images section mapping relation between deployment model and node section of demand model to implement the model transformation the mapping relation is defined in the following two ways.

- Helper: relation of elements
- FSCH Image Configuration in the NameNode
- Mapper: describe each mapper relations attributes.

The total management of each node is achieved through the model operations which content five mean types, Set, Get, List, Add, and Remove. Form lower services running Hadoop in virtual cloud environment to different deployment model and the model transpiration of cluster services which consist of computing and storage framework with some specific methods of deployment and configuration, dependent and constraint relationships within them, which realized automatic transformation of cluster service through different algorithms. In proper implementation in future that how to archived and proposed automatic transformation form Hadoop running system to Deployment model to upper stage demand model. Which would be automatically, effectively, deploy a customized Hadoop cluster in virtual cloud environment but we spent much more time to ensure correctness of cluster deployments, configurations and finally we come with presented model-driven approach for all cloud providers and today newly market demands.

8. Future Work

For Big data and business solutions how to design Hadoop a generic platform with front end tools. In this paper we present a complete design driven-model of Hadoop deployment in virtual cloud environment but it is still unclear after fully implementation what will be the new platform for big data and business solutions in all clouds supported environment. The generated platform will be provide with the correct Hadoop configuration steps and all tools need in any virtual cloud. But this platform and framework must prove in the future comprehensive solution for deploying, securing, managing in private and hybrid clouds environment. This will be a future proof design with integration different upcoming tools for next generation's visualization, analytics, and hardware less consumption solutions. In this Apache Hadoop platform elements to simple deployment pre-integrated and management need as compare to pervious design platforms for big data and business environment, that platform will be help to minimize training, security issued, vender lock-in stations, and financial investments, as well as enable faster time to market new upcoming demands for big data and business solutions based on Hadoop.

9. Conclusion

In this paper we present a model driven approach to store, process and analyzed effectively big data in virtual cloud environment using Apache Hadoop deployment model in order to reduce difficulties. Deploying Hadoop cluster form various resources of hardware, software and demand of different services for the solutions of big data using Hadoop and it core components in the cloud. But if we look to the cloud environment there are certain number of cloud service provider to provide number of platform for big data processing but it is still unclear what is the correct and generic configuration must be to implement Hadoop in the cloud. The size of huge dataset my be required new type of data storage and processing system, which is highly available in our model-driven approach if any cloud service provider will implement this model in near future. We are sure it must provide the best solution for big data and fully physical resource management in virtual cloud environment it's found that virtual Hadoop clusters are driven by the cloud services for running MapReduce and Spark jobs. In our research and driven-model challenges associated with converting Hadoop into flexible, scale and elastic service were explored. Hadoop implementation by using in physical clusters which can be provide additional management and controls for "Hadoop on the physical cloud management" and it is very important for all organizations to consider the three major aspects while selecting our model-driven approach in any cloud service architecture, workload utilization, demand operation, usability and generic prancing structure with generic solutions. We plan to conduct some farther research to applying Hadoop Model-driven Platform into realistic cloud environment to add some more and advance management features and functions by employing the new model technology to reduce configuration and implementation difficulty of Hadoop service and management platform in virtual cloud environment.

10. REFERENCES

[1] Jayalakshmi, DS, Syeda Rabiya Alam, and R Srinivasan "Approaches to Deployment of Hadoop on Cloud Platforms: Analysis and Research Issues" IEEE International Conference on Recent Trends in Electronics Information & Communication Technology (RRTEICT), May 19-20-2017 India

[2 Vahid Amiry,Mohammad Kazaem Abari, Shayan Zamani Rad, and Morteza " Implementing Hadoop Platform on Eucalyptus Cloud Infrastructure " Seventh International Conference on P2P, Parallel, Grid Cloud and Internet computing 2012

[3] Lin Weiwei, Liang Chen and LiuBo "Hadoop-Based Efficient Economic Cloud Storage System" Project, https://ieeexplore.ieee.org/abstract/document/5990265, July 2011

[4] Hong Mao, Zhenzong Zhang, Bin Zaho, limin Xiao and Ruan "Towards Deploying Elastic Hadoop in the Cloud" International Conference on Cyber-Enabled Computing and Knowledge Discovery, 2011

[5] David Nunez, Isaac Agudo, Javaier Lopez "IEEE 6th Delegated Access for Hadoop Clusters in the Cloud" International Conference on Cloud Computing Technology and Science, 2014

[6] Renata Machova, Jitka Komarkova and Martin Lnenicka "Processing of Big data Educational Data in the Cloud Using Apache Hadoop" International Conference on Information Society, (i-society 2016)

[7] Rabi Prasad Padhy "Big Data Processing with Hadoop-MapReduce in the Cloud System" International Journa Of Cloud Computing and Services Science (IJ-CLOSER), Vol.2, No.1, February, pp.16~27 2013

[8] Vijay Malav, Dr.Amit Sharma "Effect and benefits of deploying Hadoop in private cloud" National Journal of Multidisciplinary Research and Development" Vloume3; Issue 1; page No.1057 1062; January 2018

[9] Antonio Celesti, Maria Fazio, Antonio Puliafito, and Massimo Villari " Automating the Hadoop Configuration For Easy Setup in Resilient Cloud systems" IEEE Symposium on Computers and Communications (ISCC), June 2014

[10] Avanish Singh, P. Gouthaman, Shivankit Bagla and Abhishek Dey "Comparative Study of Hadoop over Containers and Hadoop Over Virutal Machine" Interantial Journal of Applied Engineering Research, Volume 13,PP.4373-4378, Number 6 2018

[11] Pedro Roger Magallhaes Vasconcelos, Gisele Azevendo De Araujo Freitas "Evaluating Virtualization for Hadoop MapReduce on an OPenNebula Cloud" International Journal Multimedia and Image processing (IJMIP), Volume 4, Issued ¾, September/December 2014

[12] Nikhil Gupta, Komal Saxena "Cloud computing Techniques for Big Data and Hadoop Implementation",International Journal of Engineering Research & Technology (IJERT), vol. 3, (4), April-2014

[13] Huang Lu, Chen Hai-Shan Hu Ting-ting "Research on Hadoop Cloud Computing Model and its Applications" Third International Journal Conference on Networking and Distributed Computing, Published in: https://ieeexplore.ieee.org/abstract/document/6386653,2012

[14] Zheyi Chen, Tao Xiang Xing Chen, " Model-Driven Approach to Hadoop Deployment in Cloud" 5th IEEE International conference on Mobile Cloud Computing and Engineering" in 2017

[15] Mythreyee S, Poornima Purohit, Apoorva D.R "A Sudy on Use of Big Data in Cloud Computing Environment" International Journal of Advance Research, Ideas and Innovations in Tehcnoloy. Volume 3, Issues 3 pp 1312-1318, 2017

[16] Dr. Vinay Kumar, Ms. Arpana Chaturvedi "Challenges and Security Issues in Implementation of Hadoop Technology in the Current Digital Era" International Journal of Scientific & Research Vol. 8
Issue 4, April 2017

[17] Mohammad Faraz Hyder, Mohammad Ali Ismail, Hameeza Ahamd "Performance Compression of Hadoop Clusters Configured on Virtual Machines and as a Cloud Service" IEEE, 2014

[18] Chaitanya P. Garware, Prof. Bharat A. Tidke "A Security Framework for Big Data omputing through Distributed Cloud Data Centers in G-Hadoop" International Journal of Computer Science and Mobile Computing Vol.5 Issue 6 pg. 355-360 2016