

## Multi-Cloud Services

**Aniruddha Vaze**

*CSE Department*

*Rajarambapu Institute of Technology, Rajaramnagar  
Sangli, India*

**Rushikesh Suryanwanshi**

*CSE Department*

*Rajarambapu Institute of Technology, Rajaramnagar  
Sangli, India*

**Digvijay Gore**

*CSE Department*

*Rajarambapu Institute of Technology, Rajaramnagar  
Sangli, India*

**Pruthvi Belgaonkar**

*CSE Department*

*Rajarambapu Institute of Technology, Rajaramnagar  
Sangli, India*

**Abstract—** Data management has a major and important role in the systematic working of an organization and institute. There are a lot of problems when it comes to data management and storage. Many organizations don't have a proper way to store, organize and handle data, and also can't afford to have that storage facility in the early stage of a startup, but now with help of new technology like cloud computing, it is possible to store, handle, organize and retrieve easily at any place and any time with an additional benefit of paying off what we use and how much we use, But this also has a drawback, many news organizations don't know how to use this facility a do whole big time consuming manual process, But this problem can also be solved by doing this whole process of data storage automatically, This can be done with the new side of technology such as Ansible, Terraform etc.

This paper presents the design and implementation of an automation framework for data management and storage on a cloud platform with help of automation tools like Ansible. This system will be useful for everyone who wants to store and handle their data easily and autonomously in just one click without any long procedure.

**Keywords -** Hadoop, Ansible, Cloud Computing, Terraform

### I. INTRODUCTION

In this, today's rapidly growing world data management and storage play the most important role in one's life. Speaking of data storage, management and handling when it comes to organizations, institutes and startups it plays a very crucial role. Organizations face problems with data storage and management without having a proper way or platform to store their data. When it comes to data storage on cloud platforms it is a very long and difficult process. To overcome these we need a smart and efficient way to which we can solve the problem of data storage and management autonomously.

Hadoop is a Data Distribution tool which is generally used in many industries because it is majorly used for managing large amounts of data and also configuring higher computational power with MapReduce Cluster.

This Clustering process is very complicated in nature and also very difficult to Handle so Ansible Automation Tool comes into place to configure Hadoop Cluster automatically also it can be managed. Ansible is used to launch multiple processes very quickly and efficiently also main use cases of Ansible are provisioning, application deployment, software management, continuous deployment of applications, automation etc. Ansible is used to provide the underlying hosts and network devices also hypervisors, and computer hosts. It can install services, and add computer hosts, services and applications inside any environment.

### II. LITERATURE REVIEW

*Iqbaldeep Kaur, et al. [1]*, According to author, "Big Data" refers to methods and tools for quickly storing, distributing, managing, and analysing massive datasets. Big data can be structured, unstructured, or semi-structured, making it impossible for traditional data management techniques to handle it. Hadoop is the main platform for organising Big data, which also addresses the issue of how to make it useful for analytics. With a relatively high level of fault tolerance, Hadoop is an open-source software project that enables the distributed processing of enormous data collections.

*Mansaf Alam and Kashish Ara Shakil, et al. [2]*, Describes the Big Data in this category in terms of its quantity, worth, variety, and speed. On the other hand, bulk data is consumed and possibly produced by long-running analytical and decision support queries employing Hadoop-based systems.

*Harshawardhan S. Bhosale, et al. [3]*, According to the possibility for faster scientific discipline advancements due to the analysis of massive amounts of data, these technical hurdles must be overcome for effective and quick processing of Big Data. at all phases of the analysis pipeline, from data gathering through

interpretation, heterogeneity, a lack of structure, error-handling, privacy, timeliness, provenance, and visualisation. Since these technical difficulties are prevalent across a wide range of application domains, it would not be cost-effective to address them in the context of a single domain. The article discusses Hadoop, an open-source programme used to process Big Data.

**Pranav T P, et al. [4]**, The usage of Ansible, an open-source automation tool, in server management within a DevOps framework is covered in the article. The authors then describe the features and capabilities of Ansible, including its use of declarative language, its ability to manage multiple servers simultaneously, and its support for a wide range of platforms and technologies. The article also discusses the benefits of using ansible in server management, including automating complex tasks, reducing errors and downtime, and improving efficiency and scalability.

Overall, this article provides a comprehensive overview of the use of ansible for automating server management tasks in the context of DevOps practices. It highlights the key features and capabilities of the tool, as well as its potential benefits and challenges, and offers insights into its potential for future development and adoption

**Pranay Dutta, Prashant Dutta, et al. [5]** The writers of this article compare and contrast the cloud services provided by Google Cloud Platform, Microsoft Azure, and Amazon Web Services (AWS) (GCP). The authors start off by giving a general introduction to cloud computing and outlining the salient characteristics and advantages of the three platforms. After that, they evaluate the different services provided by each platform, such as infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS), and they explore the main contrasts and similarities between the platforms.

The authors also compare the pricing and billing models of the three platforms and discuss the factors that influence how much it costs to use cloud services, including the type and amount using resources used, as well as the location and duration of The. The authors conclude by discussing the strengths and weaknesses of each platform and offering some recommendations for organizations considering the adoption of cloud services.

**Beakta R, et al. [6]** In this article, the authors offer a review of Hadoop's use in the context of big data, an framework which is open-source for distributed storage and processing of huge data sets. The authors start out by defining big data and outlining the opportunities and difficulties it poses. After that, they give a general review of Hadoop and its main elements,

such as the MapReduce programming methodology and the Hadoop Distributed File System (HDFS).

The authors continue by discussing how to utilise Hadoop to store, process, and analyse massive data sets and go on to cover the several tools and technologies that are used in combination with Hadoop, including Pig, Hive, and Spark. Additionally, they go over the advantages and drawbacks of utilising Hadoop for large data, including scalability, efficiency, and security. The authors wrap up by examining the current state and anticipated future developments in Hadoop and big data, including the advent of new tools and technologies like artificial intelligence and machine learning, as well as their prospective effects on the industry.

**Kaushik, Prakarsh, et al. [7]**, According to research, businesses are moving their software from on-premise data centres to the cloud in an effort to innovate, cut costs, and boost agility, which is driving uptake of cloud computing. The three most well-known public cloud providers are Google Cloud Platform, which is for those looking for alternatives to Azure and AWS with lower costs, and Amazon Web Services (AWS), which is preferred by most businesses due to its abundance of tools and services. Microsoft Azure also offers a fully compatible platform where all of your apps can use enhanced and new features almost immediately.

**Howard, Michael, et al. [8]** The author discussed the features and capabilities of HashiCorp Terraform, including its use of declarative language, its ability to manage multiple infrastructure components simultaneously, and its support for a wide range of platforms and technologies. The article also discusses the benefits of using Terraform in IaaS, including automating complex tasks, reducing errors and downtime, and improving efficiency and scalability.

The authors then present a case study of a company that implemented Terraform in its IaaS processes, highlighting the challenges and successes of the implementation process.

### III. METHODOLOGY

There are certain steps involved in the Proposed Methodology, like having to search for requirements then designing and also with Development and testing in it.

#### 1) Requirement Analysis

In the requirement analysis, we are working on Big Data Hadoop Cluster, the name given as the requirement of this project, we will use Apache Hadoop Tool for the cluster which will build on top of AWS Cloud. But for automation purposes, Ansible does the job in this project. This is how we decided to

implement these tools and create a Big Data Cluster for users.

## 2) System Design

In the System Design part, we will discuss what are the different technologies that can be used for this whole Big Data Hadoop cluster, also we will get some ideas about how this technology works independently.

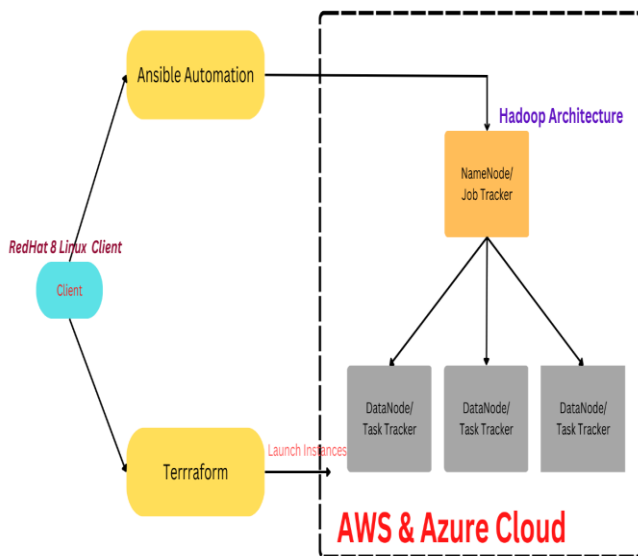


Fig 1 Overall Architecture

### i. Big Data Hadoop: -

Apache's Hadoop is a widely used open-source programme that may be implemented on a single processing node or cluster. When handling massive amounts of data, Hadoop and MapReduce programmes are utilised. Hadoop is useful for processing and storing huge data in applications including bioinformatics research, report generating, file analysis, and data mining.

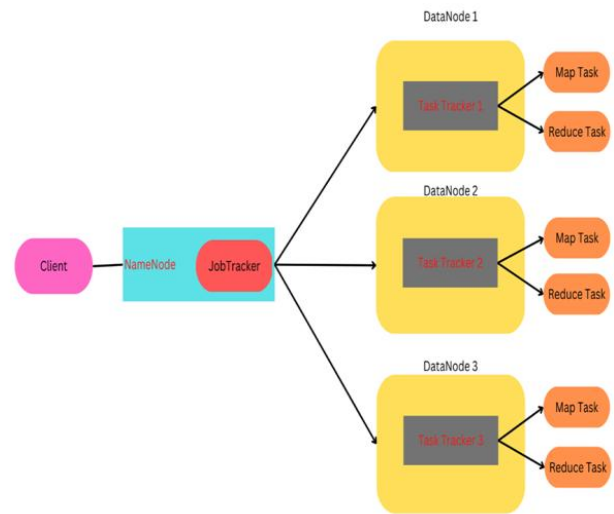


Fig 2 Hadoop Architecture

### ii. AWS Cloud

There are many uses of Cloud Computing Storage Purpose, Big Data, Deployment etc. those are below

#### a) Big Data Analytics:-

Big data is a disruptive movement that is upending the business sector when it comes to gathering more data. Big Data Powerhouses like Amazon and Facebook gather data on consumer purchasing patterns, preferences, and likes in order to forecast future purchases and expand their businesses. All businesses today work to gather and comprehend large data in order to make decisions about sales, marketing, R&D, and other things. The cloud is an extremely effective tool for storing, managing, and analysing this data.

#### b) Virtual Desktops or Desktop as a Service (DaaS):-

Employees are increasingly bringing their own gadgets to work since a mobile workforce is becoming more common. IT companies can now unify security and content access across devices thanks to virtual desktops and DaaS. Because they are housed in the cloud and are simple to access from any device, VDI and DaaS can lessen the effects of a disaster.

#### c) Email:-

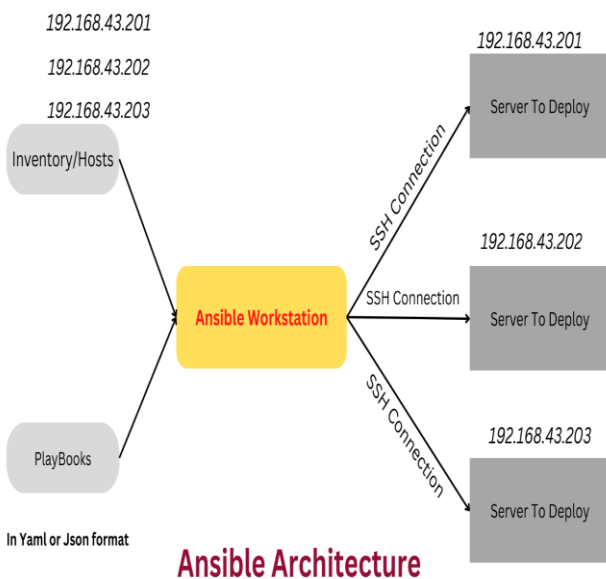
Email is a technology that has been around for a while in the SaaS category. Regular customers and others can be reached online and are integrated into key company procedures. Email has it, whether it's

for marketing, sales, or IT. There are several use cases for it in every industry, and cloud accessibility is crucial.

iii. Ansible:-

Ansible is a component of the Red Hat-owned Fedora Linux distribution and is also accessible through Extra Packages for Enterprise Linux (EPEL) for Red Hat Linux Enterprises, CentOS, OpenSUSE, SUSE Linux Enterprises, Debian, Ubuntu, Scientific Linux, and Oracle Linux.

Ansible manages several machines, or operating systems, simultaneously by allowing users to choose components from the Ansible inventory that are kept in plain ASCII text files. The target OS inventory can be defined and received dynamically or via cloud sources in a variety of forms, including YAML and INI.



**Implementation and Development**

In this step, we will do all the coding and development parts, also develop terraform script and the connectivity of this script with our Hadoop Cluster. We will also do the Ansible development part because this whole part mainly depends on Ansible Automation. Ansible is the one who creates and destroys Hadoop clusters and also manages cloud services like EC2 instances, VPC, security groups etc.

3) Testing

In the testing part, front-end testing, back-end testing and system testing will take place. During the

backend testing we check the terraform script is connected to both Azure, and AWS cloud properly, then run the Ansible playbooks and check the formation of the Hadoop Cluster by connecting the name node URL which has port number 50070, from this URL we get information about the connection of data node to name node.

**IV. RESULT AND DISCUSSION**

The proposed system meets all the specifications and provides the required functionality to automatically create a storage space according to the needs of the user. The system is tested to perform in real-time and provide real-time automation on the cloud platform. This system enhances the current manual cloud storage facility selection procedure

**V. CONCLUSION**

The proposed system is successfully designed and implemented. It is tested for reliability and accuracy. This system helps the user autonomously store their data on a cloud platform. It reduces the time of users by providing easy and reliable cloud storage accessibility. Using this system users can store and handle data anywhere. This system represents a prototype model which visualizes the status of data storage wirelessly.

**VI. FUTURE WORK**

This system can be modified and developed in future. This system has a wide scope. The system further can be integrated with the front end to allow users to access more features. This system can be extended and developed using web development and application development by giving users a complete platform by providing various features which can automatically perform the tasks.

**VII. REFERENCES**

- [1] Kaur, Iqbaldeep. "Navneet Kaur, Amandeep Ummat, Jaspreet Kaur, Navjot Kaur, "Research Paper on Big Data and Hadoop"." *International Journal of Computer Science and Technology* 7940 (2016).
- [2] Alam, Mansaf, and Kashish Ara Shakil. "Big data analytics in cloud environment using Hadoop." *arXiv preprint arXiv:1610.04572* (2016)
- [3] Bhosale, H. S., & Gadekar, D. P. (2014). A review paper on big data and Hadoop. *International Journal of Scientific and Research Publications*, 4(10), 1-7.

- [4] Pranav, T. P., S. Charan, and M. R. Darshan. "DevOps Methods for Automation of Server Management using Ansible." *International Journal of Advanced Scientific Innovation* 1.2 (2021): 7-13.
  
- [5] Dutta, Pranay, and Prashant Dutta. "Comparative study of cloud services offered by Amazon, Microsoft & Google." *International Journal of Trend in Scientific Research and Development* 3.3 (2019): 981-985.
  
- [6] Beakta, Rahul. "Big data and hadoop: A review paper." *International Journal of Computer Science & Information Technology* 2.2 (2015): 13-15.
  
- [7] Kaushik, P., Rao, A. M., Singh, D. P., Vashisht, S., & Gupta, S. (2021, November). Cloud Computing and Comparison based on Service and Performance between Amazon AWS, Microsoft Azure, and Google Cloud. In 2021 International Conference on Technological Advancements and Innovations (ICTAI) (pp. 268-273). IEEE.
  
- [8] Howard, Michael. "Terraform--Automating Infrastructure as a Service." *arXiv preprint arXiv:2205.10676* (2022).