BIG DATA BASED SECURITY SOLUTION FOR PROTECTING VIRTUALIZED ENVIRONMENT IN CLOUD COMPUTING

HarshiniUrs[*1], Tanusri C Jain[*2], Swathi R[*3], Sai Swaroop[*4], Radhika T V[*5]

[*1,2,3,4] *Student, Department of Information Science and Engineering, Dayananda Sagar College of Engineering, Bengaluru, Karnataka, India.*

[*5] Assistant Professor, Department of Information Science and Engineering, Dayananda Sagar College of Engineering, Bengaluru, Karnataka, India.

Abstract: Usage of cloud computing to store enormous amount of data is increasing every day. Almost all organizations store their data on a virtualized platform like the cloud. Virtualized infrastructure which is used in cloud computing has become a platform for hackers to extract data easily from a single place. To detect such attacks on huge amount of data, a novel big data based security approach is used.

A large cricket based recent dataset is collected and categorized the data into different classes along with attributes. Through the usage of AWS web services and EC2 instances a space for storing data on virtualized platforms like the cloud and deploying the OS on the instances and accessing through communication protocols on a local computer. Security service is enabled to the data stored in the cloud by enabling the SSH, HTTP and other such ports and by remote desktop accessing, it is able to access the cloud PC where the dataset is stored, updated and analyzed. An elastic IP address is created to host the front end which contains the analyzed part of the data which can be run on any local systems.

Index Terms: Data security, application security, security key pairs.

I.INTRODUCTION

Big Data is defined as large amount of data which can be analyzed by bringing out information systematically out of it. The data can be dealt with easily unlike other traditional application software. Big data analysis is the process of performing statistical functions on the data which will have computing capabilities in it, where data can be captured, stored and analyzed. Data can also be shared, transferred, updated and can perform queries on it.

Cloud Computing is defined as a group of virtualized computers which can be used to perform operations like a normal computer but it is not physically present and it is only virtually created.

To protect the virtualized environment from any malicious activity or any other host admin to try accessing is the main security solution to be given. It is based on AWS web services and instances. The method used helps to develop a virtual environment where the deploying of Operating system on the instances and accessing it on the local computer over communication protocol. In the above process of operation, the cloud OS is installed utilizing the EC2 instance, and the OS it is completely user friendly system, and in the same instance which are running an application by accessing the instance remotely, where it is completely visible in the local system.

The trust in the current cloud computing system is a major problem. The user once deploys his data in the cloud, they have to trust the cloud provider totally and unconditionally. The circumstances will increase in worries about the data privacy in cloud computing. The main aim is to understand the security threats and identify the appropriate security techniques used to mitigate them in Cloud Computing

II.   SYSTEM ARCHITECTURE

System design satisfies specified requirements of the system by defining the modules, components, architecture, interfaces and data. Systems design infers an efficient systematic approach to plan a framework. It might use a top-down approach or bottom-up approach , but either way the process is said to be systematic in which it takes under consideration all associated variables of the system that desires to be created—from the architecture, to the desired hardware and software, and data.

System design is one of the most important phases of software development process. The purpose of the design is to plan the solution of a problem specified by the requirement documentation. In other words the first step in the solution to the problem is the design of the project. System design can be seen as the application of systems for product development. Object-oriented analysis and design methods are becoming the most widely used methods for computer systems design

System architecture diagram:

A system architecture is a conceptual version that defines the structure, behavior, and extra perspectives of a device. In Fig. 1, the structural description is a formal description and illustration of the device, arranged in manner that helps in reasoning approximately the systems structure and behaviors of the system. A system architecture can comprise system components that will work together to implement the overall system.
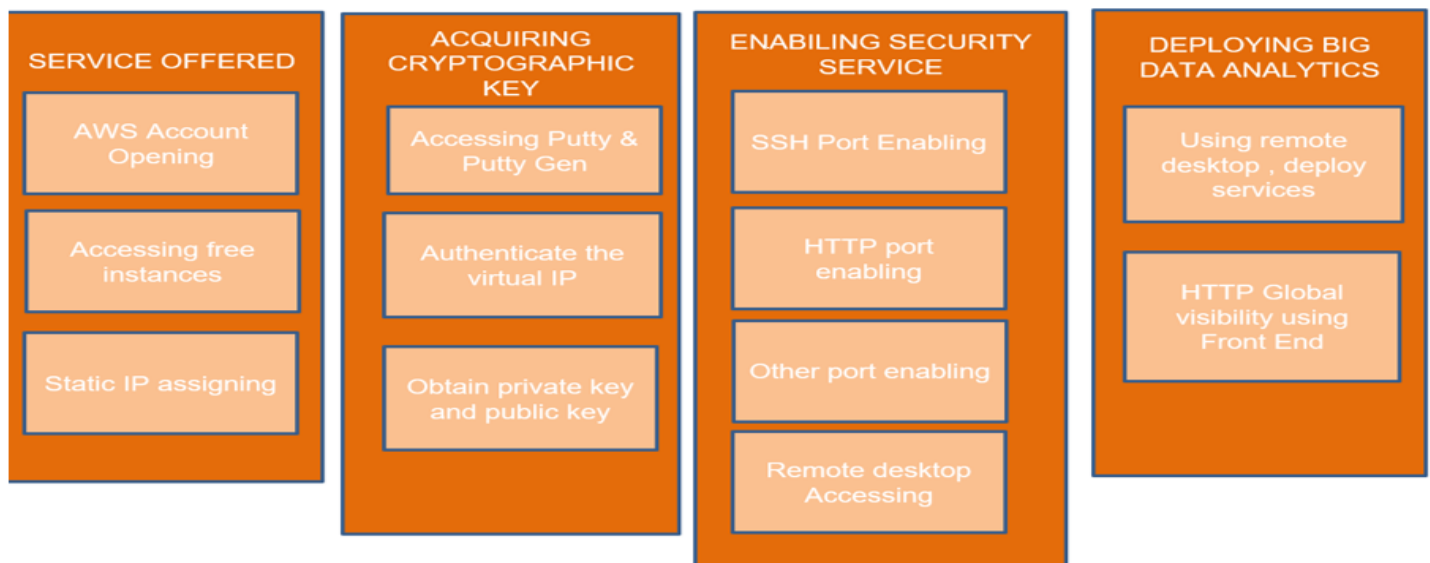

Fig 1: System Architecture

Putty:

Putty is a free and open-source terminal emulator, serial console and network file transfer application. It supports several network protocols, including SCP, SSH, Telnet, rlogin, and raw socket connection. The virtual IP address obtained from instance creation can be used in the software to bring about security. Putty software will bring about authentication and provide a public key and private key require to bring security in the system.

SSH Port Enabling:

Secure Shell (SSH) is a network protocol used for a secure connection between a client and a server. Each interaction between the server and a client is encrypted.

Amazon EC2 instance enable port forwarding through created SSH session. To get access in SSH session in the Amazon EC2 instance there should be a SSH key pair for secure connection, then the key pair is generated through Putty.

Remote Desktop Accessing:

Remote Desktop Application is used to connect to a remote PC or to virtual apps and desktops made available by the operation systems admin.

The Remote Desktop Connection available on windows is used to connect any other system terminal that is also running windows.

Aws account opening:

Amazon web services is a container of various resources. The user has to create an account to access AWS resources. The user will be provided with admirative capabilities for access and billing.

t2.micro instance :

Elastic Computing Cloud (EC2) is a virtual server which is used by the user for running applications on AWS. EC2 provides a wide selection of instance types optimized to fit different use cases. T2.micro comes under the instance types.

Static IP:

A static IP address is an IP address that was manually configured for a device instead of one that was assigned by a DHCP server. The value assigned for static address does not change.

The static IP address is useful for hosting a website from home, or to use remote access program, etc.

Deploying Big Data:

Deploying big data is done in three steps: Data Ingestion, Data Storage and Data Processing.

First extract the data from different sources. The data can be a log file, social media files, documents, etc. The next step is data storage, the extracted data is stored in HDFS. The final step is to process big data in data processing through different framework such as Mapreduce, Pig, etc.

III.FLOWDIAGRAM

Flow Diagram of a project depicts the flow in which a project is carried out. It emphasizes the sequence of movements involved in the complex system. The purpose of the flow diagram is to reveal the underlying structure of the elements and their interactions.
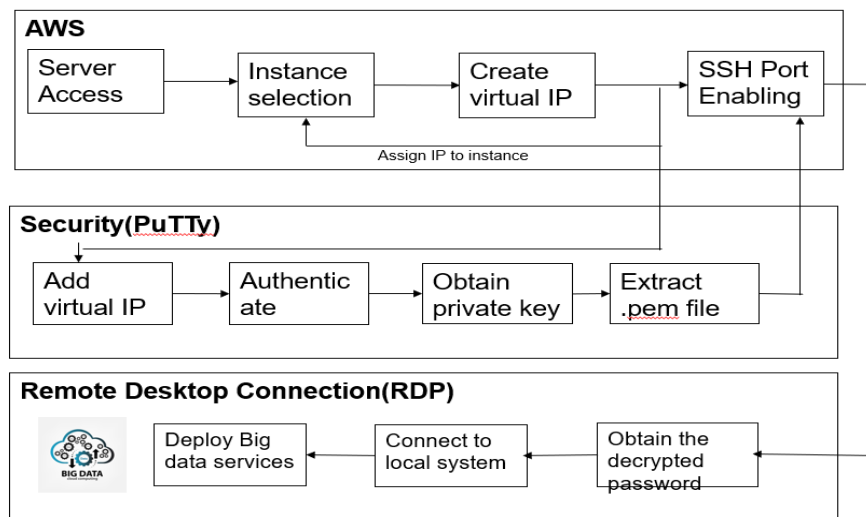

Fig 2 . Flow diagram
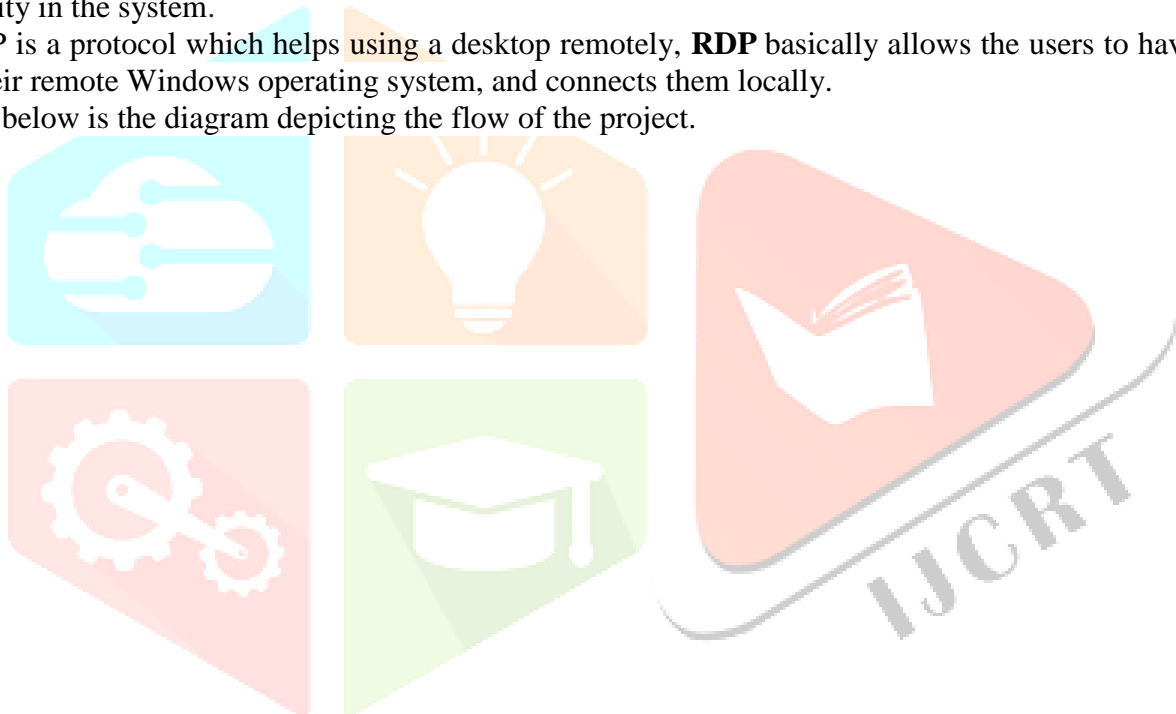
IV.IMPLEMENTATION

The task is huge information and security dependent. The project is executed by utilizing python language. The models are created utilizing numerous libraries accessible in python language. One of the model is AI based and is carried out with the assistance of python libraries, for example, SciKit learn, numpy and pandas and the other model is a major information based model and is executed with the assistance of guide reducer capacities. The prepared models are conveyed on a web worker which is carried out utilizing html structures and libraries. For the execution of backend, python structures are utilized and for execution of frontend, html and CSS are utilized.

Amazon Web Services (AWS) is also a secure cloud services platform, which helps in database storage, provides content transmission to customer. The AWS Management Console is a program based on GUI for Amazon Web Services (AWS). Through the support, a client can deal with their distributed computing, distributed storage and different assets running on the Amazon Web Services foundation.

Putty is a software or a open-source terminal emulator, serial console and network file transfer application. It has supported several network protocols, including SCP, SSH, Telnet, rlogin, and raw socket connection. This virtual IP address obtained from instance creation can be used in the software to bring about security. Putty software will bring about authentication and provide a public key and private key require to bring security in the system.

RDP is a protocol which helps using a desktop remotely, **RDP** basically allows the users to have a control of their remote Windows operating system, and connects them locally.

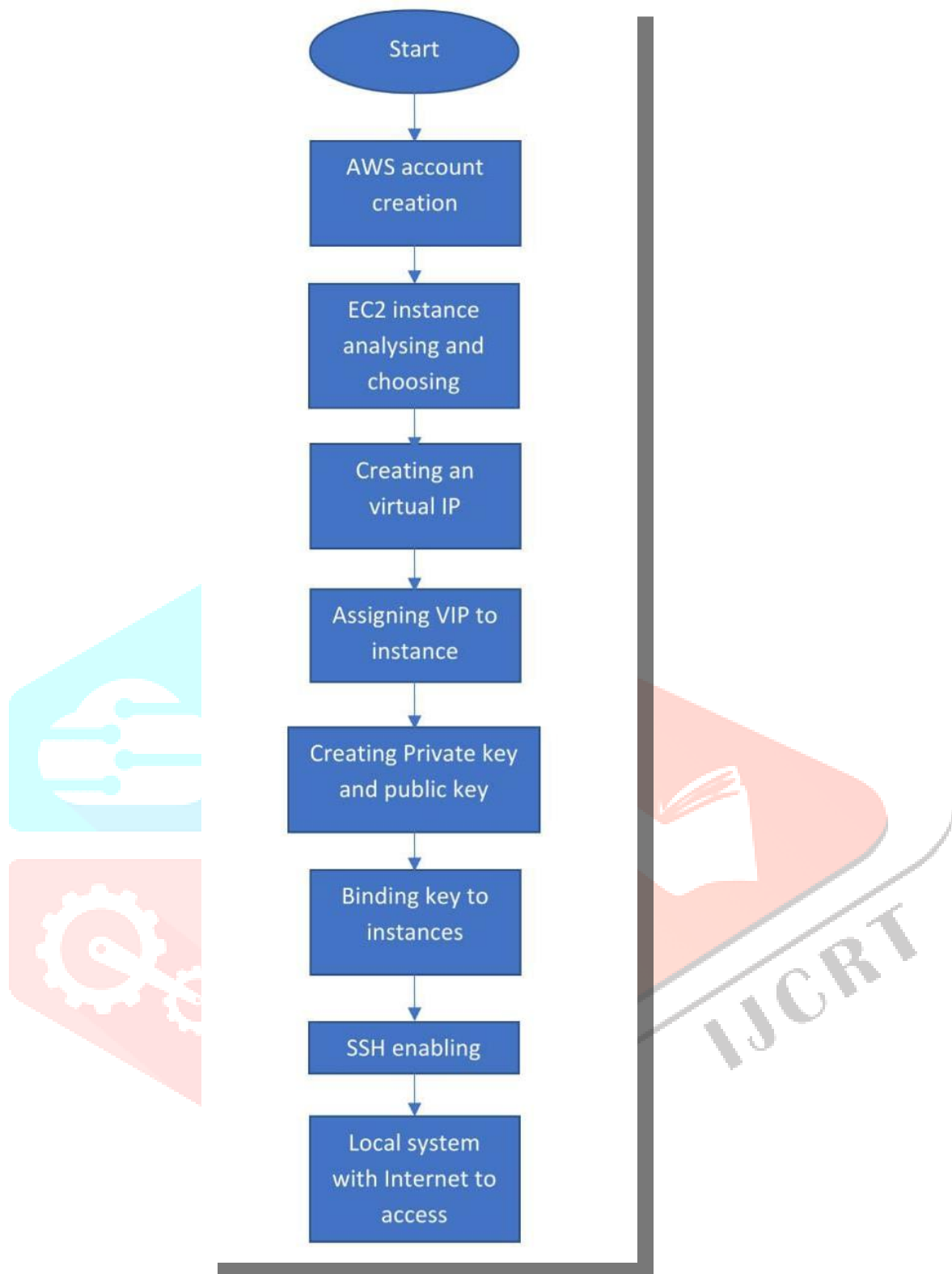The below is the diagram depicting the flow of the project.

```
                    ┌─────────────┐
                    │    Start    │
                    └─────────────┘
                           │
                           ▼
                    ┌─────────────┐
                    │ AWS account │
                    │  creation   │
                    └─────────────┘
                           │
                           ▼
                    ┌─────────────┐
                    │EC2 instance │
                    │analysing and│
                    │  choosing   │
                    └─────────────┘
                           │
                           ▼
                    ┌─────────────┐
                    │ Creating an │
                    │ virtual IP  │
                    └─────────────┘
                           │
                           ▼
                    ┌─────────────┐
                    │Assigning VIP│
                    │ to instance │
                    └─────────────┘
                           │
                           ▼
                    ┌──────────────┐
                    │Creating Private│
                    │key and public │
                    │     key       │
                    └──────────────┘
                           │
                           ▼
                    ┌─────────────┐
                    │ Binding key │
                    │to instances │
                    └─────────────┘
                           │
                           ▼
                    ┌─────────────┐
                    │ SSH enabling│
                    └─────────────┘
                           │
                           ▼
                    ┌─────────────┐
                    │Local system │
                    │with Internet│
                    │  to access  │
                    └─────────────┘
```
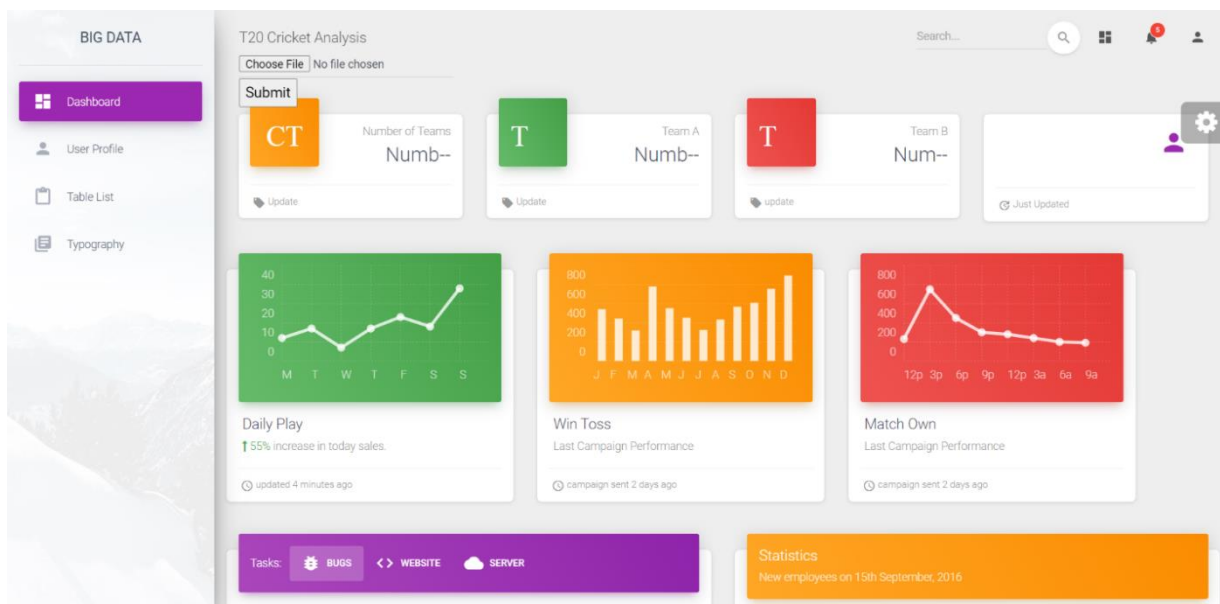
Fig 3:Flow of the project

V. Result

Fig 4: Frontend output

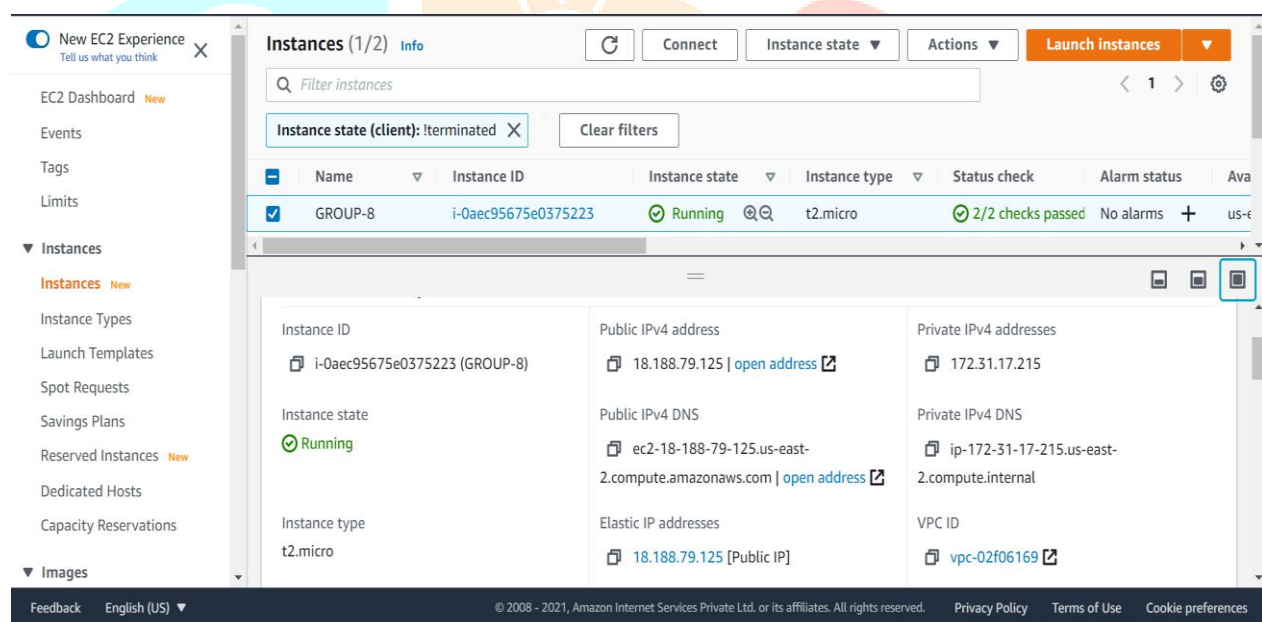

Fig 5: Instance Created for the application

VI. CONCLUSION AND FUTURE SCOPE

When dealing with big data we have two major concerns Data security and privacy .Control security becomes a primary concern when application is hosted on a Cloud platform due to its open environment and limited user .On contrary, Big data uses a lot of third-party services and infrastructure as it is an open source application . The systems are integrated in Private Cloud Solution that is Elastic and Scalable.

In the proposed model, implementation is done by amazon web services which is fully integrated portfolio of cloud computing services and helps to build, secure, and also deploy big data applications as proposed model is a cost effective approach . The raw data is collected and stored in the secure remote desktop using big data technique. By performing more advanced functions the data is converted from its raw state into a consumable format.

Therefore, the actionable insight of the big data approach will be visualized through the data assets. With enhancement of Cloud technology, big data analysis has become more advanced which has led to effective results. Hence, the companies opt to comply big data analysis in the Cloud. These two Big data and Cloud technology are reducing the cost burden for enterprise purposes which helps in financial affairs and also in bringing value to the enterprise.

VII. REFERENCES

[1.]   I. Hababeh, A. Gharaibeh, S.Nofal, I. Khalil, "An Integrated Methodology for Big Data Classifications and Security for Improved Cloud Systems Data Mobility", 2169-3536(c) 2018 IEEE.

[2.]   Durumeric, J. Kasten, D. Adrian, J. A. Halderman, M. Bailey, F. Li, N. Weaver, J. Amann, J. Beekman, M. Payeret al., "The matter of heartbleed," in Proceedings of the 2014 Conference on Internet Measurement Conference. Vancouver, BC, Canada: ACM, 2014, pp. 475–488.

[3.]   K. Cabaj, K. Grochowski, and P. Gawkowski, "Practical problems of internet threats analyses," in Theory and Engineering of Complex Systems and Dependability. Springer, 2015, pp. 87–96.

[4.]   J. Oberheide, E. Cooke, and F. Jahanian, "Cloudav: N-version antivirus in the network cloud." in USENIX Security Symposium, San Jose, California, USA, 2008,pp. 91–106.

[5.]   P. K. Chouhan, M. Hagan, G. McWilliams, and S. Sezer, "Network based malware detection within virtualized environments," in Euro-Par 2014: Parallel Processing Workshops. Porto, Portugal: Springer, 2014, pp. 335–346.

[6.]   J. V. Gautam, H. B. Prajapati, V. K. Dabhi, and S. Chaudhary, "A survey on job   scheduling algorithms in big data processing," in Electrical, Computer and Communication Technologies (ICECCT), 2015 IEEE International Conference on. IEEE, 2015, pp. 1–11.

[7.]   A. Nasridinov and Y.-H. Park, "Visual analytics for big data using r," in Cloud and Green Computing (CGC), 2013 Third International Conference on. IEEE, 2013, pp. 564–565.

[8.]   7. A. Katal, M. Wazid, and R. Goudar, "Big data: issues, challenges, tools and good   practices," in Contemporary Computing (IC3), 2013 Sixth International Conference on. IEEE, 2013, pp. 404–409.