ISSN: 2278 – 909X
International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 7 Issue 5, May 2018

Analytics of Big data for Hybrid Cloud Layered Architecture

Archna Aggarwal, Ms. Nisha Pandey

***Abstract*— Cloud computing, is the paradigm of computing on the fly. Technically said, Cloud computing is the virtualization and central management of data centre resources as software-defined pools. Big Data represents content and Cloud Computing is infrastructure. For analytics of these larger sized information new techniques, technologies, tools and architectures are required. This big data can be analyzed for better decisions and strategic business moves for insights through cloud computing. A number of Cloud computing architectures are there for analytics of big dat. Cloud can be public cloud or public cloud. Hybrid cloud is combination of both private and public cloud. New architectures and analytics methods are needed to manage and extract value and hidden knowledge from big data because as it is complex, vast and diverse. This paper presents new hybrid cloud architecture HCEA for analytics of big data.**

***Index Terms*—Cloud Computing, Private Cloud, Public Cloud, Hybrid Cloud, Hadoop, Data Centers, Big Data Analytics**

I. INTRODUCTION

Big Data[2] is obtaining data out of "Velocity, Variety and Volume" from the data resources accessible, while cloud centres around on-request, versatile, adaptable, pay-per utilize self-service models. In basic words, the cloud takes information from different assets to process that information and move that outcome into the internet which is presently accessible for the entire world to get to through this cloud. Some of the major clouds computing service providers are Amazon Web Services, Google Cloud, Microsoft Azure, and more. Cloud[3,4] providers usually offer three basic services: Infrastructure as a Service (IaaS) that delivers infrastructure which means storage, processing power, and virtual machines and satisfies the needs of the client by virtualizing resources according to the service level agreements (SLAs), Platform as a Service (PaaS) is a level where big data DBMS are implemented and is built atop of IaaS and allows users to deploy cloud applications created using the programming and run-time environments supported by the provider, Software as a Service (SaaS) one of the most known cloud systems and consists of applications that are running in the cloud provider directly.

Cloud computing[6] is a technology used to store data and information on a remote server rather than on a physical hard drive. Internet which acts as an infrastructure as a service is referred as cloud. Cloud is a new technology for computing of resources. Big data can exist without cloud computing while Cloud requires big data for computing resources.

A Cloud computing system [7] based on infrastructure location could be implemented as Private, Public or Hybrid cloud. The private model is a local implementation of cloud computing system. In this model, hardware is located in local data centers and uses cloud software applications to provide service to local users. This model is the best option for consumers who needs cloud computing capabilities with low-risk in IT departments because this model allows an IT department to migrate from the traditional model to the cloud computing system and does not require data to be migrated to another location (such as cloud vendor location). This model is implemented for local trusted users. This model still allows scalability, on-demanded self-service, and elastic service. However, this model requires high investment in maintenance, recovery, disaster control, security control, and monitoring.

In addition, the private cloud computing model enables an IT department to handle a local organization's big data by its own infrastructure, such as the storage of big data and computing big data. This model provides a flexible resource assignment and could enhance the resource availability.

The public model is a regular model of cloud computing system. This model is provided by cloud vendor who supports billing and a subscription system for public users. This model, unlike a private model, does not require high investment, because consumers could pay on pay-per-use basis for cloud storage or cloud computing services on demand.

The hybrid model [17] composes private and public clouds. This model could connect a private cloud to public cloud through network connection, such as the Internet. This model has several advantages, which are listed below:

- Collaboration between cloud computing systems: Often collaboration between two clouds led to emergence of hybrid cloud model. An organization could keep their own cloud security and maintenance, and simultaneously have collaboration with other clouds. This collaboration could be permanent or temporary.
- Scalability: This model also is useful for extending the scalability of a private cloud computing system, because in case of limited resources at a peak time, a cluster of new resources could be added temporary from another cloud.

II. RELATED WORK

Chaowei Yanga & Qunying Huang et.al. states that Big Data has emerged as a new paradigm providing. A number of opportunities to improve decision-support

applications including business, sciences and engineering. But presents some challenges also to store, transport, process, mine and serve the data. Cloud computing helps to handle these challenges with shared computing resources that includes computing, storage, networking and analytical software. There paper surveys the two frontiers – Big Data and cloud computing and the utility of cloud computing to tackling Big Data in the relevant domains. Various observations of their research are that cloud computing and Big Data enable science discoveries and application developments, provide solutions for Big Data. [7]

Mehdi Bahrami & Mukesh Singhal et.al. discussed the definition of big data, its importance and major challenges and issues related to big data. They specify the importance of cloud computing technology to handle big data as a solution for computing and storage purposes. They examine the capabilities like resource scalability, resource shrink-ability, resource pool sharing, on-demanded servicing, elastic servicing of cloud computing systems which are important for big data, and also its collaboration with other cloud computing systems. They defined cloud architecture service layers and role of each service layer to handle big data.[8]

Christine Ouyang & Marcio Moura et.al. defines hybrid cloud as a connection between private cloud and one or more public cloud as shown in fig. 1. and provides the best feature of each environment. It provides the flexibility to locate data and services based on the need of business organization. It helps their clients to provide freedom to choose and change environments, data and services as required which allows cloud applications and services to be quickly made using the related data and insights available, while retaining visibility, integrated control, and security. The facility or openness of the hybrid astronaut allows the data then the associated analytical workload after keep placed where such makes the most experience within terms concerning commercial enterprise needs. The statistics privateness yet security is managed or controlled persistently across all the systems over the hybrid bird environments. Hybrid cloud's openness and flexibility allows the data and its analytics to be placed where it is most required in business. The privacy and security of information is controlled and managed by Hybrid cloud across all systems in its environment. [9]

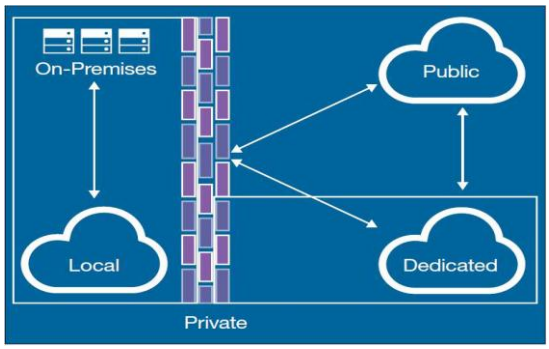

Fig. 1: The Hybrid Cloud

Cloud computing allows businesses of small to medium size to execute big data technology with a promise of less use of company resources. The efficiency of big data model could give new perceptions of knowledge to the business relating to execution change, basic leadership support, and development in plans of action, items, and administrations. The advantages of executing big data technology via cloud computing are profitable in terms of cost in hardware and processing, additionally the capacity to test big data technology before making a generous use of organization resources. Many models of cloud computing services are available for businesses, where each model has business trade-off between the cost savings and data security and loss of control.[10]

Dinh Thai Hoang, Chonho Lee, Dusit Niyato & Ping Wang et.al. states that mobile cloud computing will become one of the future trends technology as it combines the features ans advantages of both cloud computing and mobile computing and providing optimal services to users. They provide the overview and definition of Mobile cloud computing, its architecture and advantages.[11]

George Lackermair et.al. provides a comprehensive introduction to Cloud Computing. It mainly discusses scenarios for a hybrid integration of net-based infrastructure, platform, software and collaboration from a retailers' point of view. The integration of cloud-based services in various domains is already quite common-payment services and user tracking are widely used in online commerce. Besides that, a retailer needs to evaluate the economic stability of the supplier. Crowd sourcing on top of the Cloud Computing stack is relatively new to online retailers. There are various different forms of integration of consumers in the value creation, whereas it is still unclear how the Social Web can be used for collaboration with consumers. In his opinion is the human layer a promising field of future E-commerce research [12].

III. Proposed Hybrid Cloud Architecture- HCEA

Architecture describes [8,9] the basic structure of a system with its elements, the relations between these elements and the relationships of the system to the environment. In addition, principles for the design, development and use of the system should be described are. Architecture of other models: (1) differs by a holistic view of a system in terms of the width of the elements under consideration, and (2) by a coarsened consideration.

The data architecture represents, specific part architecture of the information system architecture, and describes the data structure of an information system. In the data architecture[11] such entity types, which we use in our daily data modelling and their relationships to each other are shown. In terms of a coarsened consideration it may be necessary to map only selected entity in the data architecture or summarize the data architecture entity types.

For Big Data Architecture[12], there is no specific way. However, semi-formal notations can be used by practitioners. Big Data Architectures can be referred to as enterprise data architecture to the basic data structures of an entire company or refer as the data architecture of an application system on a section of the company. However Big Data Architectures are to be distinguished from the enterprise data models, which have the aim of documenting the history of a company. This is an approach, which is being utilized from the 90s, however, is uptight with difficulties, since an immense care expense is created by the constant change of a company.

In this archive, we will outline what hybrid cloud is, clarify why it is essential with regards to enormous

information and investigation, and examine execution contemplations and providing a new hybrid cloud Architecture for Big Data and Analytics to give direction on the sending of enormous data and investigation arrangements in hybrid cloud.

The hybrid cloud is called a combination of local cloud resources and integrated on-premises with one or more dedicated cloud and one or more public clouds as shown in fig. 2. Combined with on-premises and dedicated cloud of local cloud known as "personal environment". These public clouds and private environments are structured in such a way that they can operate independently using technologies that provide portability of applications and data movement, but can also communicate with each other through a secure connection on private and / or on the public network.

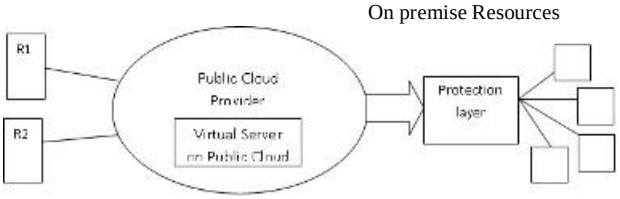

Fig. 2: The Hybrid Cloud

A hybrid cloud enables enterprises to incorporate information from the undertaking frameworks on the private domain with applications running on public cloud, while utilizing the public cloud's computational assets and capacity. For instance, associations can produce significant insights by combining the data from Systems of Record (private condition) with Systems of Engagement in an open cloud or by applying edge-investigation on the tools in public cloud.

Additionally, where the private environment doesn't provide adequate computational power, hybrid cloud increases scalability by allowing organizations to use public cloud resources for situations. Moreover, to expand the portability of workloads between private environments and public cloud, containers can be used. Hybrid Cloud Data allows better management of sovereignty and compliance, which is the best fit for global distribution of applications and data

Present cloud architecture has many challenges such as bandwidth, latency, energy consumption. In keeping the future demands in mind and to maintain the system efficiency, those challenges always remain for the big data for analytics. Large organizations are often required to interconnect their big data centres from remote locations, which creates demand to move servers not only inside the big data center but also between them. For this reason, we need to provide high throughput for the connections among servers in the big data center. Due to the large number of operations created by Big Data analytics, the traffic at the server level will increase substantially. The traditional Ethernet link is not sufficient to provide such high throughput requirements. Hence, we propose to connect public cloud with private cloud servers directly to make a hybrid cloud that could fulfil the requirement.

Fig.3. Shows the proposed HCEA, an hybrid architecture for big data analytics with cloud while keeping security as a vital requirement. The architecture of HCEA, is segmented into three layers- Adaptation Layer, Middle Layer and Service Layer. We discuss each layer below in brief.

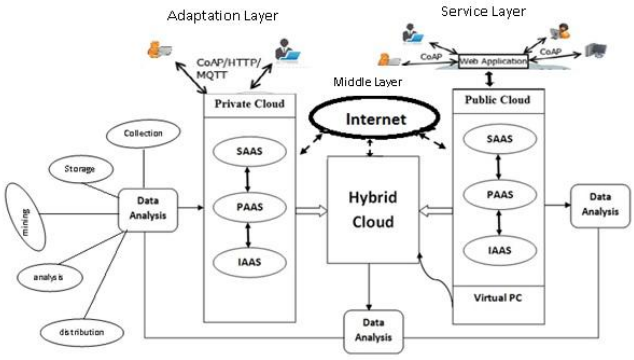

Fig. 3: The proposed hybrid cloud architecture for big data analytics (HCEA).

a. Adaptation Layer

Having public cloud for data access and services, security of information may be compromised by the outsiders. Private clouds have certain policies described by specific organization for data access to restrict the users from getting all type of services. To ensure security of data within a single domain, data gathered from the Device Layer is first stored to a private cloud through gateways. This layer provides data storage in private cloud within an organization where clouds offer services like SaaS, Paas and IaaS to its intra-domain users. Web services over cloud use RESTful and SOAP architecture to offer light weight services. REST adopts CRUD (Create, Read, Update, and Delete) operations provided by the HTTP protocol such as GET, POST, PUT, and DELETE. SOAP manages the sensor's data and their geographical working location. User can use web application through HTTP, CoAP or MQTT protocols. Constraint Application Protocol (i.e. CoAP) is a RESTful application layer web transfer protocol used between low powers embedded wireless networks using HTTP interfaces. CoAP incurs less parsing complexity, low overhead, and runs on top of User Datagram Protocol (UDP). Hyper-text Transfer Protocol (i.e. HTTP) is a simple text based protocol that supports many libraries. Message Queuing Telemetry Transport Protocol (i.e. MQTT) is specially designed to support lossy networks. It is used to publish and subscribe data transport with efficient bandwidth but it doesn't provide device-to-device transfer or multicast. The integration of WSN and the web using REST/CoAP with web applications (REST/HTTP) through CoAP/HTTP proxy helps to visualize the measurements of WSN on the HTTP web browser based on CoAP. Security of the data is maintained by the Datagram Transport Layer Security (DTLS) protocol. Use of private cloud at each domain restricts information access to the outside users, which in turn ensures the security of data within a specific domain.

In HCEA architecture, users can also share data, resources, and functionalities with other private cloud by using internet on the conceptual, logical, and procedural level and this technology is known as Cloud Federation. One organization can extend their working area in future to

one more geographical location with new technologies and protocols. To share data and resources between different locations, there is no need to upgrade old setup. Adaptation Layer not only ensures security using private cloud but also consider the issues of scalability and interoperability in heterogeneous big data analytics environment. Having distinct private cloud to process heterogeneous data collected from different devices reduces complications involved in operating heterogeneous data on the same cloud. Bringing enormous data on the same platform and accessing it using standard protocols solves interoperability issues. Though, introducing adaptation layer in Cloud-big data environment derives significant advantages, issue of extra data overhead might be present in HCEA.

Private cloud represents a particular type of cloud computing where only authorized customers can operate. Private cloud computing keep all characteristics of cloud computing, including the virtualization of resources. The benefits consists in more privacy and control for the organization which owns it. It is difficult to identify the techniques through which the services are provided. The exclusive use for one customer and the network security are ensured by the ring fetching of the cloud.

Different from a public cloud, the private cloud services drag the resources from different linked computers, which could be held internally or externally. The resources are dragged through private rented connections or encoded connections via public networks. The most important characteristics of private cloud are - higher security and privacy, higher level of control, efficiency of costs and energy, improved reliability and cloud bursting.

The most used type of cloud computing is the public cloud which provide the services through virtualization, accessed using a public network (for example, the internet). In contrast with the private cloud, in the public cloud more customers have access to the resources, sharing the same infrastructure.

b. Middle Layer

Middle Layer provides global communication for information exchange between different private clouds. Data from these private clouds can then be processed and uploaded to a public cloud, so that it can be accessed globally by the users. This layer also communicates resources, and functionalities between private-private or private-public clouds using Internet. To facilitate large number of tools in big data analytics, all the protocols and hardware used in this layer must support IPv6 for internetworking.

Hybrid cloud represents [17,18,19] a cloud computing environment in which different on-premises are used, private cloud and third-party, public cloud services with an important orchestration between the two platforms. By having the possibility to allow workload to move between private and public clouds as the needs of computing and according to the changing costs, the hybrid cloud will offer to business a high flexibility and many options for data deployment.

The private architecture was built to work with a unique cluster, and it's not capable to coordinate activities in multiple domains environment like hybrid cloud. One of the common strategies used in the hybrid cloud is to use the paid resources from the public cloud only when the private cloud resources weren't enough to achieve the desired performance. However, this is a hard task for users. The private cloud is dynamic and its resources are allocated and released as the users' need. Thus, the user would have to evaluate the situation at the moment and decide how many public cloud's resources would be necessary to the desired process. To address these questions we propose architecture to do the application orchestration in hybrid clouds. Our architecture incorporates the IaaS service characteristics of private cloud and public cloud to the complex hybrid cloud environment.

To move easily between multiple domains, the architecture defines a model for configuration of domains that can be implemented in files of the repository in the cloud storage managed by OE. In this model each domain must be identified by attributes and classes. The domain's attributes are: the name, the type (public cloud, private cloud, etc), the business model (PaaS, IaaS, SaaS), the provider, the server gateway, etc. The domain classes identify classes of hosts (type and capacity of CPU, memory, number of cores, etc.), describe the images and identify the flavours of VMs, that is, must contain all the information to create appropriate VMs.

c. Service Layer

Users access services or data across different organizations through a public cloud. The public cloud provides SaaS and access to shared data globally. CoAP/HTTP/MQTT protocols, integrated with RESTful and SOAP architecture, can be used to access various web services which helps visualize the data acquired by WSN.

Public cloud has the following main characteristics - ultimate scalability, efficient costs, high reliability, elasticity, and location independence.

The flexibility and scalability of public clouds will take out from the discussion the needs for a company to create a massive capital costs outgoing in order to accommodate in short-term spikes in demand. The provider for public cloud will give the necessary supplies for resource computing, and further, the company will only pay for the resources that he consumed. Thus, the administrators of such infrastructures can to confirm (through Android systems, messages, and/or different channels) the fact that some resources which are newly integrated in the infrastructure are legitimate. This thesis proposes a solution represented by novel architecture for big data analytics in hybrid clouds.

IV. IMPLEMENTATION

To cater the need of more flexible and scalable cloud environment in large businesses, I have implemented the hybrid cloud environment through an application which earlier when made over private cloud network, proves to be more expensive and dedicated to single organization but now by increased scalability and flexibility it proved to be a low cost and more friendly to access shared resources than private cloud. Basically it serves the purpose of both private as well as public network. It is a small application that only demonstrates the working environment of private as well as hybrid cloud network. The application is a small customer

management application that enables the user to keep a record of maintaining customer detail. One can also filter the customer list as per requirement. It also facilitates user to add customer, view customer and detail of orders each customer place. The deletion option is there along with necessary authentication required through login screen. Earlier when the application resides over the private cloud network whose server resides on the private network, it was not feasible to share it with remote users. After deploying the application on virtual server resides over the hybrid cloud network and keeping the data on private cloud end of hybrid network, more secure environment with greater flexibility provided to the application owning organization. The framework used for application is MVC framework for Microsoft Visual Studio. MVC framework consists of angular framework that supports the proposed application. The backend support for application is MongoDB. It is an open source database that uses a document-oriented data model. MongoDB [reference] is built on architecture of collection and documents rather than using tables as in relational databases. In MongoDB documents comprise pairs of key values as basic units. This collection contains the set of documents and function as those in RDBMS. MongoDB supports dynamic schema design just like other NoSQL databases allowing the documents in a collection to have different fields and structures. As per the article posted on tricerat.com MongoDB was created by Dwight Merriman and Eliot Horowitz, who had encountered scalability and development issues with traditional relational database approaches while building Web applications at DoubleClick, an Internet advertising company that is now owned by Google Inc. According to Merriman, the name of the database was derived from the word humongous to represent the idea of supporting large amounts of data. Merriman and Horowitz helped form 10Gen Inc. in 2007 to commercialize MongoDB and related software. The company was renamed MongoDB Inc. in 2013. The database was released to open source in 2009 and is available under the terms of the Free Software Foundation's GNU AGPL Version 3.0 commercial license.

V. Conclusion

We have proposed HCEA, a secure hybrid private-public cloud architecture which provides security to the heterogeneous network data in big data environment, while ensuring scalability and interoperability. We have discussed various integration issues while bringing these two all-encompassing technologies to talk to each other and derive mutual benefit from respective strengths and provision great services to users. We have added an adaptation layer in proposed architecture which implements private cloud to provide secure and seamless communication between the network and sensors. We have made use of Cloud Federation to address the issue of security and communication between distinct private clouds. Some of the research challenges encountered with proposed architecture are also discussed.In this architecture we present a solution in order to offer to the clients a high level of security for a private computational infrastructure which is divided between more public and private cloud environments. Starting from on-demand scalability (turning on the virtual machines in different clouds), it is very important to integrate the automation of resource managing with the confirmation of identities by the operators.

Acknowledgment

We are thankful to our guide Ms. Nisha Pandey, our HOD Dr. Dinesh Chaudhary who provided expertise that greatly assisted the research. I would also like to thank the director of our college, Dr. S. K. Garg for his valuable support.

References

[1] G. Pallis, “Cloud Computing: The New Frontier of Internet Computing”, IEEE Internet Computing, vol. 14, issue 5, pp. 70 – 73, 2010, ISSN 1089-7801
[2] V. V. Mayer, and K. Cukier, “Big Data: A Revolution That Will Transform How We Live, Work and Think”, John MurrayPress: UK, 2013.
[3] R. Buyya, C. S. Yeo, S. Venugopal, J. Brobergand and I. Brandic, ”Cloud computing and emerging IT platforms: vision, hype,and reality for delivering computing as the 5th utility”, Future Generation Computer Systems 2009; 25(6): 599–616.
[4] P. Shukla, B. Radadiya, and K, Atkotiya,“An Emerging Trend of Big Data for High Volume and Varieties of Data to Search of Agricultural Data”, Orient. J. Comp. Sci. and Technol;8(2).
[5] P. Neves, B. Schmer, J. Bernardino, and J. Camara, “Big Data in Cloud Computing: features and issues”, 3030-290 Coimbra, Portugal.
[6] S. Mythreyee, P. Purohit, D. R. Apoorva, R. Harshitha, P. V. Lathashree, "Big Data in Cloud Computing", International Journal of Advance Research, Ideas and Innovations in Technology 3.3 (2017). www.ijariit.com.
[7] C. Yang, Q. Huang, Z. Li, K. Liu, and F. Hu, “Big Data and cloud computing: innovation Opportunities and challenges”, International Journal of Digital Earth,10:1, 13-53, DOI: 10.1080/17538947.2016.1239771
[8] M. Bahrami,and M. Singhal, “The Role of Cloud Computing Architecture in Big Data”, Information Granularity, Big Data, and Computational Intelligence, Vol. 8, pp. 275-295, Chapter 13, Pedrycz and S.-M. Chen (eds.), Springer, 2015 http://goo.gl/0LxxlH
[9] C. Ouyang, and M. Moura, “A vision of hybrid cloud for big data and analytics”, IBM Big Data & Analytics Hub.htm, March 17, 2017.
[10] B. M. Purcell, “Big data using cloud computing”, Journal of Technology Research, http://www.aabri.com/manuscripts/131457.pdf
[11] H. T. Dinh, C. Lee, D. Niyato, and P. Wang, “A survey of mobile cloud computing: architecture, applications, and approaches”, Wireless Communication & Mobile Computing, vol. 13, issue 18, 013, pp. 1587-1611, https://doi.org/10.1002/wcm.1203
[12] G. Lackermair, “Hybrid cloud architectures for the online commerce”, Procedia Computer Science 3 (2011) 550–555, doi:10.1016/j.procs.2010.12.091
[13] F. Doelitzscher, "Private cloud for collaboration and e-Learning services: from IaaS to SaaS." Computing 91.1 (2011): 23-42.
[14] M. Bahrami, "Cloud Template, a Big Data Solution", Journal of Soft Computing and Software Engineering , Vol. 3, No. 2, pp.13-17, 2013.
[15] M. Singhal, "A Client-centric Approach to Interoperable Clouds", International Journal of Soft Computing and Software Engineering [JSCSE], Vol. 3, No. 3, pp. 3-4, 2013.
[16] W. J. Rittinghouse, and J. F. Ransome, “Cloud computing: implementation, management, and security”, CRC press, 2009.
[17] A. Sharma, T. Goyal, E. S. Pilli, A. P. Mazumdar, M. C. Govil, and R. C. Joshiz, “A Secure Hybrid Cloud Enabled Architecture for Internet of Things”, 978-1-5090-0366-2/15 ©2015 IEEE
[18] C. R. Senna, L. G. C. Russi, and E. R. M. Madeira, “An Architecture for Orchestrating Hadoop Applications in Hybrid Cloud”, 14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, 2014.
[19] https://www.infinera.com/wp-content/uploads/2015/07/Infonetics-Global_Service_Provider_Survey_ExcerptsOTN_MPLS_Control_Plane_Strategies. pdf

Archna Aggarwal, M.Tech(CSE), Department of Computer Science, Shri Ram College of Engineering and Management,Palwal, Haryana, India, 7734018007.

Ms. Nisha Pandey, Assistant Professor, Department of Computer Science, Shri Ram College of Engineering and Management, Palwal, Haryana, India, 9991646463.