

# Opportunities and Challenges in integrating Cloud Computing and Big Data Analytics to E-governance

Bhushan Jadhav  
Research Scholar,

Thadomal Shahani Engineering College, University  
of Mumbai, Mumbai-400050

Archana B. Patankar, PhD  
Professor,

Thadomal Shahani Engineering College, University  
of Mumbai, Mumbai-400050

## ABSTRACT

The rapid advancement in web and computing technologies have made significant improvement in the computing architectures. Today most of the businesses runs on internet and related technologies. The E-governance uses Information and Communication technologies (ICT) to provide essential government services and information exchange with the citizens and other stakeholders through internet. However, the traditional E-governance solutions are incapable to fulfill the current need because of its increasing demand, application complexity, infrastructure management, cost overhead and other technical challenges.

Therefore the emerging technologies like Cloud computing and big data analytics can welcome these challenges and overcome them using the modern approach for computing, storage and data processing. They provide unique features to E-governance like lower cost, scalability, easy management, disaster recovery, accountability, resource provisioning, distributed storage, data analytics, mobility etc. The objective of this paper is to give insights about cloud computing and big data analytics for E-governance. It also proposes a model for cloud and big data enabled E-governance along with their opportunities and challenges.

## Keywords

E-governance, Cloud Computing, Big Data Analytics, Hadoop, Map-reduce

## 1. INTRODUCTION

Upon tremendous success of internet, today every organization is trying to expand their businesses through the web because of its popularity and high usage. In government sector, traditionally accessing the services manually was a quiet difficult task as it has involved so many procedures and formalities with lot of paperwork [2]. Hence governments across the world have come together and started delivering their services through web under the roof of “E-Governance”.

Initially the use of E-governance was very limited but today it becomes very high. Therefore the existing technologies are not capable to meet the current demand of E-governance. So the government needs to think for technological improvement to meet the demand at reduced costs with increased efficiency. The technologies like cloud computing and big data analytics have huge capability to overcome the current challenges of E-governance and can be easily opted for implementation.

The initial part of this paper explain the basics of E-governance along with Cloud computing and Big data analytics. The later part proposes a model for cloud and big data enabled E-governance along with their opportunities and challenges.

## 2. E-GOVERNANCE

The E-Governance refers to the use of information and communication technologies (ICT) to improve the efficiency, effectiveness, transparency and accountability of government by means of exchanging information, services within government, citizens and businesses through the web [1], [2], [6]. The resulting benefits can be increased transparency, less corruption, cost reductions, and convenience of access” [4].

The main objective of E-governance are

- Seamless and improved delivery of services
- Minimize the corruption by making government services transparent
- Cost reduction and improving revenue
- Promoting the economic development
- Ease of access to public information and related services
- Improved public administration.
- Eliminate paperwork and formalities involved in government offerings
- Government can interact with citizens and other stakeholders easily and rapidly

### 2.1 Models of E-Governance

The applications in E-Governance are categorized based on its interaction with stakeholders. The E-Governance is categorized in to three models (See Figure 1). [1], [2],[3],[6]

#### 2.1.1 Government to Citizens (G2C)

In G2C model the interaction happens between Government and citizens. It include information and services specifically designed for citizens by the government delivered through a web. Some of the G2C Services are agricultural service, vehicle licensing service, passport service, revenue service etc.

#### 2.1.2 Government to Business (G2B)

The G2B model provides interaction between Government and Business organizations. It include various services designed for businesses like policy enforcement, Contract management, auditing, security, renewing and obtaining licenses, registration of companies, and payment of taxes etc.

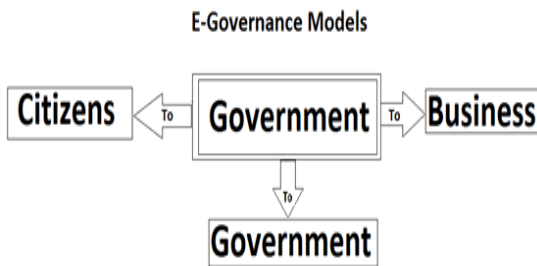


Fig. 1 Models of E-governance

### 2.1.3 Government to Government (G2G)

The G2G model handles the interaction between the government with the other government departments and agencies. This model facilitates the exchange of information, sharing of resources, decision making, fund transfer, shared services and law enforcement between the different Government departments.

## 2.2 Technical Challenges in E-governance

Traditionally E-governance services were hosted on shared webservers which had many flaws. The technology used for implementation now becomes technically obsolete as it cannot fulfill the current requirement of E-governance. The Challenges with traditional E-governance implementation are given as follows [3].

- 1) Scalability issue:- In traditional setup the scope of scalability is very limited as in-house servers has limitation of running the number of instances on shared infrastructure therefore scalability factor is very low. The traditional E-governance applications doesn't supports horizontal scalability. Therefore the servers can be scaled by adding the resources physically.
- 2) Disaster Recovery:- The recovery of traditional E-governance application in case of disaster is impossible task as many of the E-governance applications are hosted in-house within the same geographical boundaries and the cost incurred for implementation of disaster recovery is also very high.
- 3) Application Complexity:- The traditional E-governance applications are complex in nature therefore one need to keep the track of development, testing and debugging, managing the structured data used by the applications and avoiding duplication of same resource across different departments.
- 4) Software licensing:- The Software licensing is the another concern in E-governance as applications hosted on server requires a separate bundle of application and system licenses for each physical servers and needs to be renewed periodically that makes a huge financial burden on government agencies.
- 5) Accountability:- The traditional E-governance doesn't have accountability to track the record of usage as well as account management.

## 3. CLOUD COMPUTING

Most of the E-governance services are provided through web applications hosted on multiple web servers. It is accessed by millions of users at a time which requires a huge computing power as well as lots of data is being generated that requires a limitless storage. The cloud computing can fulfill both the requirements of E-governance.

The Cloud Computing is a distributed computing model through which user can gain access to shared resources, applications and services dynamically on any connected devices. According to National Institute of Standards and Technology (NIST), one of the most accepted definition of cloud computing is "It is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [7].

## 3.1 Characteristics of Cloud Computing

Cloud computing exhibits the following key characteristics:

- 1) *On-demand self-service* :- A user can access the different cloud services dynamically under a single self-service portal that doesn't require a human interaction with service provider.
- 2) *Broad network access* :- The Cloud computing services are easily available for everyone through a network and standard mechanisms used by cloud clients like thin or thick clients, mobile phones, tablets, laptops, and workstations etc.
- 3) *Resource pooling* :- To fulfill the multiple users request at the same time the provider's computing resources are pooled together in to one or more resource pools which allocate or de-allocate the resources as per users demand.
- 4) *Rapid elasticity* :- The resources in cloud are elastically provisioned and released as per demand.
- 5) *Measured service* :- The Cloud computing comprises measured services which meters the resources as well as monitor, control and provision them for accounting and optimum utilization.
- 6) *Location independence* :- Most of the cloud services are location independent where user can get access to their devices, applications or services across the globe through internet irrespective of their geographical location.

## 3.2 Cloud Deployment Models

Based on the functionality and accessibility of cloud Services [1],[2]. The cloud computing is classified in to four deployment models explained as follows

### 3.2.1 Public Cloud

The public cloud basically runs over the internet. It is designed for the general public where resources, applications and services are provided over the internet by cloud service providers (CSPs) based on subscription or pay per use model.

### 3.2.2 Private Cloud

This type of cloud is called in-house cloud as the services are hosted on in-house infrastructure and provided through intranet. It is tightly secure than public cloud as its local administrators write their own security policies for user's access.

### 3.2.3 Hybrid Cloud

Hybrid cloud composed of two or more private and/or public clouds that offers the benefits of multiple deployment models.

### 3.2.4 Community Cloud

The community cloud is a mixture of one or more public, private or hybrid clouds, which is shared by many organizations for a same cause (mostly security).

### 3.3 Cloud Service Models

The cloud service models define the core functionality of services implemented at different level of cloud. The cloud service models are categorized in to three service models explained as follows [5].

#### 3.3.1 "Infrastructure as a Service (IaaS)"

Infrastructure-as-a-Service (IaaS) is liable for providing servers, storage, Computing power, network and virtualization to form utility like services for users. The infrastructure consists of communication networks, physical compute nodes, Storage solutions and the pool of virtualized computing resources managed by a service provider.

#### 3.3.2 "Platform as a Service (PaaS)"

The Platform as a Service provides a readymade platform for running web applications or web services. The PaaS composed of readily available runtime execution environment, development and deployment tools, middleware solutions, databases and APIs

#### 3.3.3 "Software as a Service (SaaS)"

Software-as-a-Service is specifically designed for on-demand applications delivery to the cloud users. It gives access to remotely installed Softwares reside on cloud server through the internet. Therefore user do not have to install required software in their local pc instead they are delivered by remote server through a network.

## 4. BIG DATA ANALYTICS

In E-governance lots of data is being generated in the form of texts, images, videos, xml, JSONs etc. such data can't be processed using traditional databases known as Big data. The Big data generated by E-governance has four characteristics (See Figure 2).

The characteristics of E-governance big data are nothing but its concerns. The Volume represents the unpredictable quantity of data is being generated that requires a huge storage volumes, Variety refers to data generated having different formats, Velocity refers to rate of data is generated (per second/per minute) and Veracity means generation of uncleaned and untrusted data.

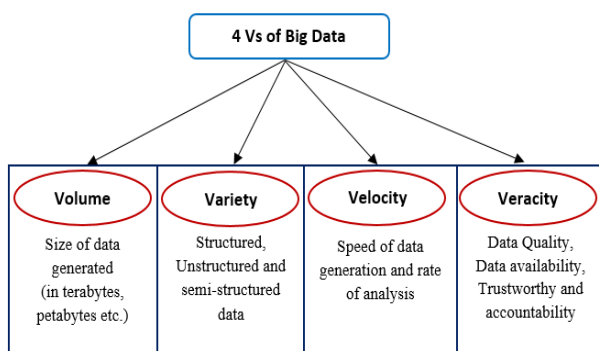


Fig. 2 Characteristics of Big data

The Traditional databases are incapable to acquire, process, manage and analyze such kind of big data because it's too big for storing or too fast it's generated, or too hard for processing. The Big data analytics is the solution for processing, managing and analyzing the vast amount of data over the network or internet.

The Big data analytics has many tools for acquiring, managing, processing and analyzing the big data. But one of the most popular framework is Apache Hadoop.

The Apache Hadoop is an open source framework used for storing and processing a large datasets over the multiple clusters [8]. The Hadoop is designed to scale up from single servers to thousands of machines, where each machine offers a local computation and storage. The Hadoop framework has two essential components mainly Hadoop distributed file system (HDFS) and map-reduce. The HDFS is used to store the distributed data in structured or unstructured form while Map Reduce is a programming paradigm used for analyzing and processing large distributed data sets over the heterogeneous environment that overcome the challenges of traditional data analysis and management tools.

The Hadoop framework has many built-in supported tools called Ecosystem components, they are categorized in to data storage, data processing, data access and data management tools [8]. Each components has predefined functions (See Figure 3).

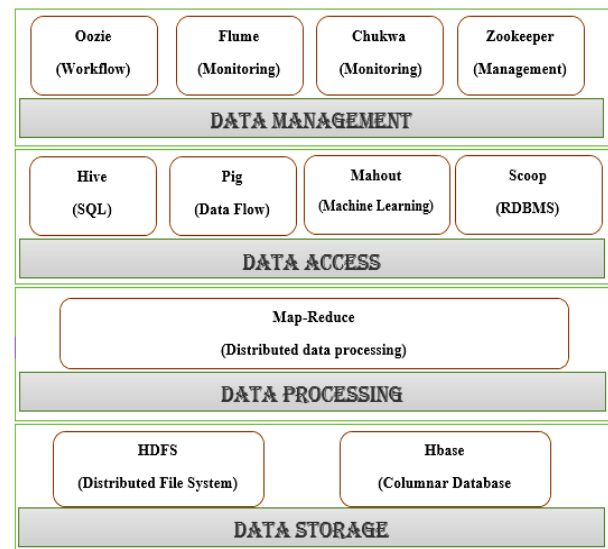


Fig. 3 Hadoop Ecosystem Components

In Hadoop ecosystem, the bottom layer is data storage which provides HDFS distributed file storage and Hbase for NOSQL Columnar data storage. The data processing layer gives Map-reduce framework for distributed data processing while the data to map reduce is being provided through data access tools. The data access tools has four components like Hive, Pig, Mahout and Scoop. The Hive is used to store the data in a data warehouse using SQL, Pig is used for dataflow and analysis, Scoop is used for importing and exporting data from RDBMS in to Hadoop and vice and mahout used for machine learning. The top layer is data management which provides oozie for managing workflow of Hadoop jobs, Flume is used for collecting, transferring and monitoring a large set of data in the form of log files ,chukwa is used for monitoring large distributed data while zookeeper is used for managing the configuration information an distributed Synchronization.

## 5. PROPOSED MODEL FOR CLOUD AND BIG DATA INTEGRATED E-GOVERNANCE

The Big data analytics require massive computing power along with huge distributed storage which are the basic offerings of Cloud infrastructure and big data can works

efficiently over the Cloud computing platforms. Therefore to resolve the challenges associated with traditional E-governance the integrated cloud and big data framework can be used.

There are many cloud service solutions are available in the market. Therefore the proposed model is designed such a way that it can fit in to any IaaS based cloud platforms like Openstack, AWS, Azure, Open nebula or Google cloud platform. The proposed model for Cloud and Big data enabled E-governance comprises the architectural components of cloud computing and big data along with E-governance (See Figure 4).

The proposed model has bottom up architecture. It is made up of five layers, where each layer provides a functionality and services to its upper layer.

The bottom layer of proposed architecture is Infrastructure layer which provides infrastructural components to the cloud computing layer. It includes servers with cpu, memory and storages, networking devices like L3 switches IDS, IPS, and firewalls etc. The cloud computing layer is responsible for making a resource pool using different infrastructural components. It segregate the compute, storage and network components together in to a shared pool of resources. The main function of cloud layer is to run multiple virtual machine (VM) instances over the shared infrastructure and provide unlimited storage space. It accommodate various servers like web servers, application servers, Hadoop clusters, database servers etc. at one place over the VM instances.

The Big data analytics layer is responsible for performing various E-governance big data operations over Hadoop clusters running inside the cloud layer. It facilitates acquisition, processing, analysis and management of E-governance data easily over the cloud infrastructure. This layer also allows to create a data warehouse using sql and nosql databases over the storage provided by cloud layer.

The Application layer is responsible for running different E-governance applications and Websites over cloud infrastructure .Each application and website runs a separate VM instance over a cloud platform. It also gets analytics services through big data layer for processing a large amount of data.

The user access layer is responsible for providing access to different E-governance applications to its stake holders through supported cloud clients and open source APIs.

### 5.1 Opportunities in proposed model

The proposed model exhibit combined advantages of both cloud computing and big data analytics which were not there in traditional E-governance solution. There are many new opportunities introduced by proposed model, some of them are listed in Table 1.

**Table 1. Opportunities introduced by proposed model**

Sr. No.	Opportunities	Description
1	Scalability	It allows to scale up or scale down the compute, storage and network resources to a required extent through resource pool.
2	Disaster recovery	This feature of cloud computing allows to recover the data and services in case of disaster

		happens. For that it uses replicated VMs running across diverse locations.
3	High Availability	The fail-over and fail-back clusters in cloud provide high availability of cloud services without disturbance.
4	Cost efficiency	The cloud can accommodate multiple servers in to a single physical server using virtualization which saves the hardware cost as well as reduce the infrastructure management cost.
5	Monitoring & Measuring	The cloud platform provides different monitoring tools to monitor the different cloud instances as well as performance. It also supports different measuring tools to keeps the account of resource usage.
6	Load Balancing	This feature of cloud computing allows to balance the load of data and running services equally across multiple servers to get the faster performance.
7	Mobility	The mobility feature allows to access E-governance applications on the move. It can be accessed on any device which has internet connectivity with supported browser.
8	Ease of access	The proposed model provides ease of access to many E-governance applications concurrently hosted on cloud and big data platforms
9	Stores Structured and unstructured data	The Big data solution in proposed model allows to store different formats of data such as structured, semi-structured and unstructured using different SQL and NOSQL databases.
10	Data processing & data Analytics	The Big data framework in proposed model is liable to process a large amount of data over Hadoop clusters and provides different analytics over that data.

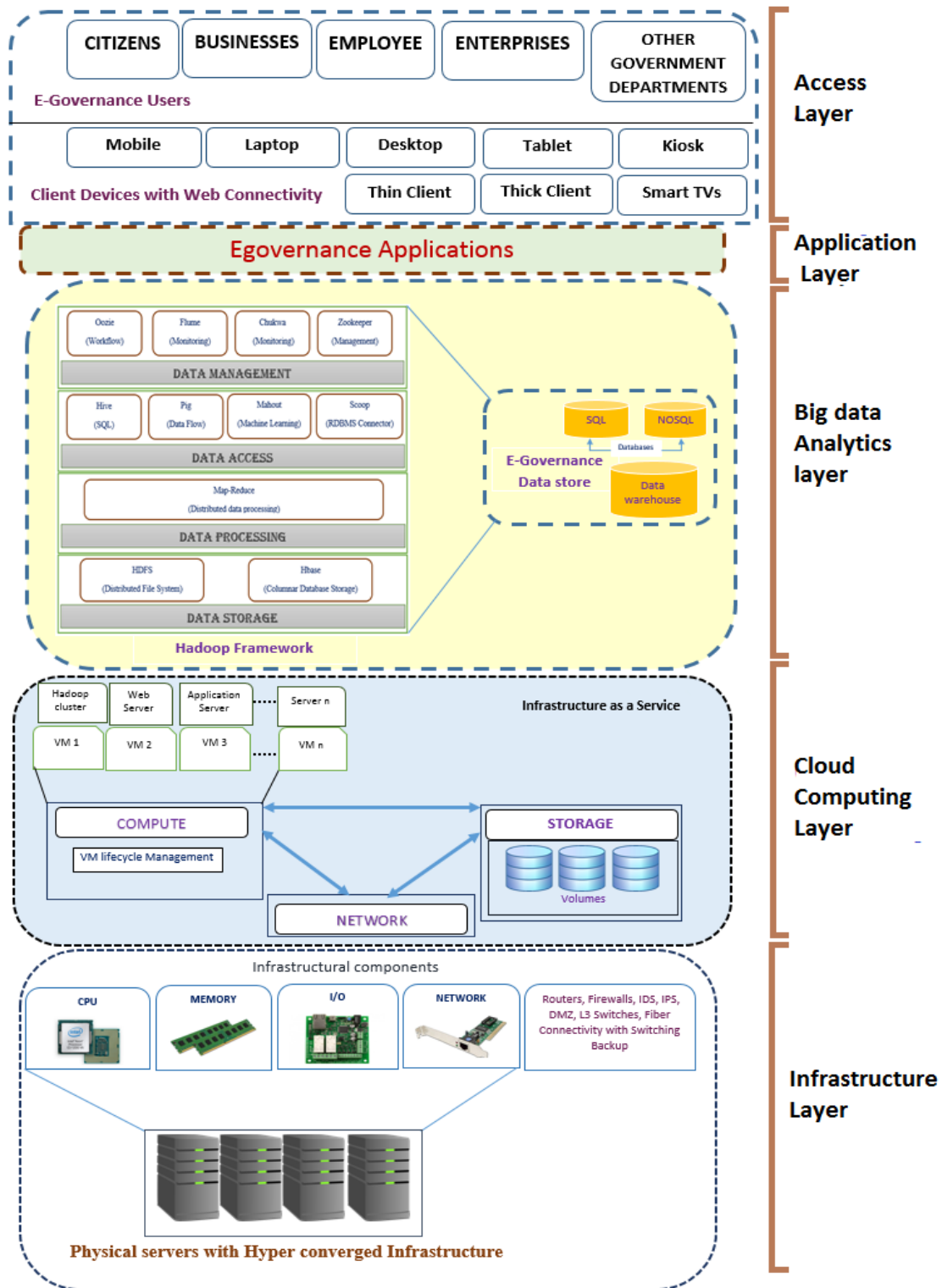


Fig. 4 Proposed model for Cloud and big data integrated E-governance

### 5.2 Challenges in proposed model

As like opportunities, the proposed model has some challenges listed in Table 2.

Table 2. Challenges in proposed model

Sr. No.	Challenges	Description
1	Complex Architecture	Because of complex components in Cloud and Big data frameworks, the proposed model comprises a complex

		architecture that makes difficulty in understanding the functionality of each component.
2	Security of services and Big data in E-governance	The proposed model is prone to lack of security when integrated with public clouds. As public cloud is runs over the public network, so any compromise in data security may leads to huge financial losses.
3	Lack of trust and identity	There are many public clouds which runs over untrusted networks. Therefore integration of such clouds in proposed model may leads to untrusted access of confidential data over public network and lack of identity of authorized users.
4	Lack of interoperability and difficulty in migration	The proposed model supports some public clouds which has lack of interoperability due to the vendor lock-in solution that does not allow to migrate the services on other cloud platform.
5	Poor performance	The proposed model supports some public cloud platforms which gives poor performance
6	Difficulty in Managing Hadoop cluster	The complex architecture of Hadoop framework makes difficulties in managing the jobs running on Hadoop clusters

## 6. CONCLUSION

In this paper, the reviewed traditional E-governance solutions with their models and associated challenges. To overcome those challenges the proposed a model is given that supports integration of cloud and big data platforms to traditional E-governance. The proposed model for E-governance has five layers which fulfills all the basic requirements and provide many new features which were not there in traditional E-governance. In the final section of this paper the opportunities and challenges of proposed model are discussed.

## 7. ACKNOWLEDGMENTS

We gratefully acknowledge the experts who have contributed towards the development of this research paper especially Dr. Archana B. Patankar and Prof. Arun Kulkarni.

## 8. REFERENCES

- [1] Satyabrata Dash, Subhendu Kumar Pani, "E-Governance Paradigm Using Cloud Infrastructure: Benefits and Challenges," International Conference on Computational Modeling and Security (CMS 2016), Elsevier, doi 10.1016/j.procs.2016.05.274, pp:843-855.
- [2] Smitha K. K., Dr. Tony Thomas, Chitharanjan K, "Cloud based E-Governance System: A Survey", *International Conference on Modeling, Optimization and Computing (ICMOC2012)*, Elsevier, doi:10.1016/j.procs.2016.05.274, pp:3816-3823.
- [3] Dr. Vasudeva Varma, "Cloud Computing for E-Governance", *A white paper*, IIT, Hyderabad, January 2010.
- [4] Tamara Almarabeh, Yousef Kh. Majdalawi, Hiba Mohammad "Cloud Computing of E-Government", *International journal on Communications and Network*, Scientific Research Publishing, 2016, pp: 1-8.
- [5] Aprna Tripathi, Bhawana Parihar, "E-governance challenges and cloud benefits", *International Conference on Computer Science and Automation Engineering, IEEE*, June 2011 pp:351-354.
- [6] Smitha K K, Chitharanjan K, "Security of Data in Cloud based E-Governance System", *International Journal of Computer Applications on Advanced Computing and Communication Technologies for HPC Applications - ACCTHPCA*, June 2012
- [7] Peter Mell, Timothy Grance, "The NIST Definition of Cloud Computing", *NIST Special Publication 800-145*, September 2011.
- [8] Sruthika.s, Dr .N. Tajunisha, "A Study on evolution of data analytics to big data analytics and its research scope", *International Conference on Innovations in Information Embedded and Communication Systems ICIECS'15, IEEE*, 2015.