IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-8, Aug 2018 E-ISSN: 2347-2693

A Comparative And Statistical Approach To Leverage Cloud Computing And Big Data Analytics In E-Governance

Bhushan Jadhav[1*], Archana B. Patankar[2]

[1]Thadomal Shahani engineering college, University of Mumbai, Mumbai, India
[2]Computer Engineering department, Thadomal Shahani engineering college, University of Mumbai, Mumbai, India

[*]*Corresponding Author: bhushan.jadhav@tsec.edu, Tel.: 9702868662*

Available online at: www.ijcseonline.org

Accepted: 14/Aug/2018, Published: 31/Aug/2018

Abstract— Nowadays, the governments across the world uses internet for providing their maximum number of services to the different stakeholders with the help of ICT. But due to the legacy computing architecture and data processing technologies, it become difficult for the governments to fulfill the current requirement of E-governance which makes impact on its usage. Therefore the emerging technologies like cloud computing and big data analytics can be used together in E-governance to overcome the existing challenges related to computing and data processing. The Cloud computing and big data analytics can be used together to introduce the many features which are not there in existing system. But before opting them their relevance needs to be studied. Therefore the main objective of this paper is to perform the in-depth study of Cloud and Bigdata enabled E-governance along with the statistical and comparative analysis. This research paper also proposes a model for integrating cloud computing and big data analytics in to E-governance along with advantages and disadvantages.

Keywords—E-governance, Cloud Computing, Big data analytics.

I. INTRODUCTION

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 main objectives of E-governance are
• Efficient and automated delivery of services
• Minimize the corruption by providing transparent services
• Reducing the cost of delivery and improve the revenue
• Promote the economic development
• Ease of access with the better public administration.
• Eliminate paperwork and minimize the formalities involved in government offerings

The E- governance comprises three models namely Government to Citizens (G2C), Government to Business (G2B) and Government to Government (G2G). In G2C model provides the information and services specifically designed for citizens by the government delivered through a web. The G2B model has various services designed for businesses that include policy enforcement, Contract management licenses, payment of taxes etc. While G2G model facilitates the exchange of information, sharing of resources, decision making or fund transfer between the different Government departments [10].

But due to the limited infrastructure, lack of fund, lack of skilled manpower, use of legacy technologies etc. the E-governance poses many challenges. Therefore the modern technologies like cloud computing and big data analytics can be used together to welcome these challenges and overcome them using modern approach [12].

Rest of the paper is organized as follows, Section I contains the introduction of Cloud computing and Big data analytics, Section II contains the related work of cloud and big data enabled E-governance, Section III contains the proposed architecture of cloud and big data enabled E-governance, Section IV contains the Research methodology which performs the hypothetical testing and statistical analysis of cloud and big data enabled E-governance, Section V describes results and discussion and Section VI concludes research work with future directions.

1. Cloud Computing

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" [4], [15].

1.1 Cloud Deployment Models

The cloud computing is classified based in to four deployment models based on its functionality and accessibility [11].

The Public Cloud basically runs over the internet which is managed by Cloud service providers (CSP), so users do not have to concern about resources and their maintenance. This services are available on subscription or pay-per-use basis. The disadvantage of public cloud is lack of security. Some of the public cloud platforms are Amazon web services, Microsoft azure, Google cloud platform, Salesforce etc.

The private cloud is hosted on internal servers for getting the high security. The infrastructure of private cloud is completely managed and maintained by the organization itself. It is highly secure than public cloud as the local administrators write their security policies for users access. The disadvantage of private cloud is it incurs huge capital expenditure. Some of the popular private cloud platforms are Openstack, Eucalyptus, and Open Nebula etc.

The Hybrid cloud is made up of two or more private and public clouds which provide benefits of both the clouds. Hybrid Cloud architecture has both on-premises and off-premise servers. The Hybrid cloud provides the flexibility for in-house applications along with fault tolerance and scalability.

The community cloud is a mixture of one or more public, private or hybrid clouds, shared by multiple organizations whose objective is same The Infrastructure of community cloud is shared by several organizations within the community. The popular Community cloud platform is Salesforce community cloud Google cloud GSuite etc.

1.2 Cloud Service Models

The cloud services are classified in to three service models which are explained as follows [11].

The Infrastructure-as-a-Service (IaaS) can be defined as the use of computing power, storage, network, servers, OS and virtualization to form utility like services for users. The IaaS also allow users to create, destroy, copy, migrate, scale and manage the virtual machine instances and storages running on servers.

The Platform as a Service provide a web platform to create and run web applications easily. The PaaS provide built-in development tools, runtime engines middleware, databases, programming languages and APIs for developers to develop the applications.

The Software-as-a-Service is specifically designed for on demand applications or software delivery to the cloud users. It gives remote access to Softwares that resides on cloud server not on the user's device.

2. Big data analytics

The development in web and related technologies have increased the use of internet at massive level. Today most of the people in the world access Social Media websites like YouTube, Facebook, and Instagram etc. every day. The access to such sites involves uploading and downloading of huge amount of data. As per social media usage report 2017, there are 400 hours of new videos uploads on YouTube with access by 4.1 million users every day, Facebook has 6.5 million posts per minute include uploading of images, videos, audios and texts while Instagram generates 4.67 million posts every minute [14]. Therefore there is a need to process and analyze such data which is impossible with the traditional databases and tools. Such data is called Big data.

According to Gartner the definition of "Big data is a high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation" [16]. While the Big data analytics is the process of acquiring, analyzing and processing of big data. In big data, the captured data has many formats like structured, semi-structured or unstructured [13]. The structured data has fixed pattern or schema which can be stored and managed using tables in RDBMS, The semi-structured data does not have pre-defined structure or pattern as it involves scientific or bibliographic data which can be represented using Graph data structures while unstructured data also do not have a standard structure, pattern or schema. The examples of unstructured data is videos, audios, images, pdfs, compressed, JSON files. The big data is difficult to process and analyze because of its four characteristics. The Volume refers to huge amount of data being generated, Variety means different types of data generated including structured, unstructured or semi-structured data, Velocity refers to speed of data generated p and Veracity refers to uncertainty of data [11].

II. RELATED WORK

There are many researchers who have proposed their views on cloud and big data based E-governance.

A. In [1] Satyabrata Dasha, et al. have discussed the National e-Governance Plan in India along with benefits and challenges of Cloud based E-governance.

B. In [2] Smitha K.K et al. have explained the conceptual framework for cloud integrated E-governance along with E-governance challenges and cloud benefits.
C. In [3] Rajkumar Sharma et al. have proposed a datacenter oriented cloud architecture along with Suitability of distributed data-centers in E-Governance.
D. In [4] Yannis Charalabidis et al. have explained the open innovation process for co-creation of public services.
E. In [5] Rajiv Ranjan, Saurabh Garg et al. have explained the big data analytics workflow for smart cities.
F. In [6] Deka Ganesh Chandra et al. explains about cloud computing model for Indian national E-governace.
G. In [7] Rajagopalan M.R et al. proposed a big data framework for Indian national E-governace.
H. In [8] Saeid Abolfazli et al. discussed a case study of Malaysia with perspective to cloud computing adopted in telecom, healthcare, agriculture and travel sectors.

The above literature explains the theoretical models of cloud and big data enabled E-governance. So this paper proposes a proposed model along with the statistical and comparative analysis for use of cloud computing and big data analytics for E-governance.

III. Proposed Architecture

The proposed architecture for E-governance has four layers that incorporate the existing components of E-governance along with the integration of cloud computing and Big data analytics as shown in Figure 1. It has four layer namely Infrastructure layer, Cloud computing layer, Big data analytics layer and Access layer.

The bottom layer of proposed architecture is and infrastructure layer which manages and provide all the infrastructural components to the cloud computing layer for virtualization. It has many physical servers with compute, storage and network components.

The second layer of architecture is Cloud computing layer which is meant to provide all the cloud computing features to the various stakeholders. It provide cloud services like IaaS, PaaS or SaaS by means of running and executing applications over the virtual machines (VMs). For E-governance these VMs can run different e-governance servers along with the hosting and storage services [11].

The Third layer of proposed architecture is Big data analytics layer which has two sublayers namely Data access layer and Business logic layer. The data access layer provide different databases, data warehouse and servers for storing and accessing the data by the users while business logic layer uses specialized middleware solutions like SOA, Web services, EJB to run business logic over the stored data.

The big data analytics layer use the data stored by data access layer and perform different operations like data acquisition, data processing and data analysis by using different big data analytics tools like hadoop, stroam, spark etc.

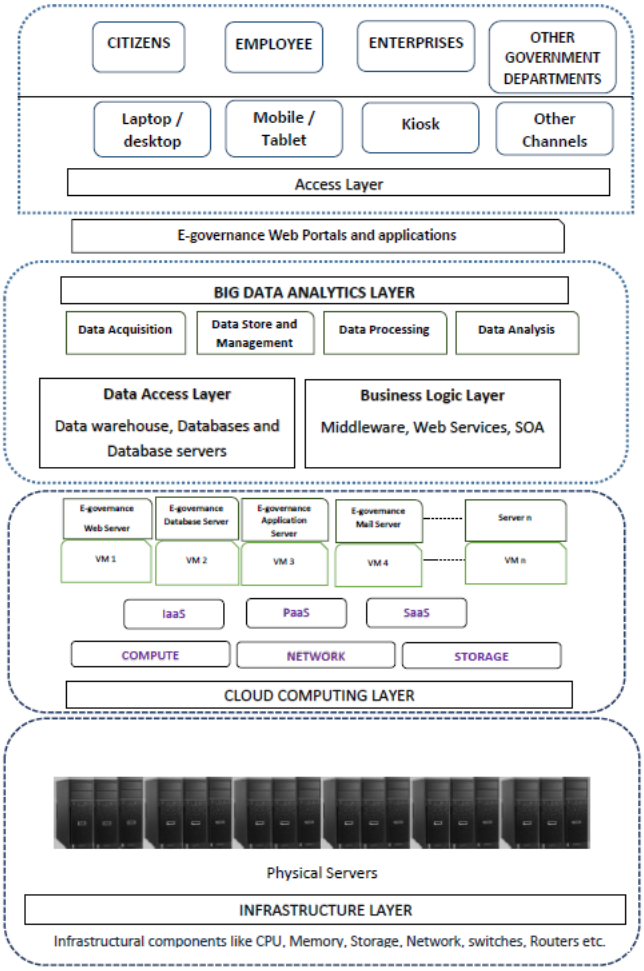

Figure 1. Proposed Cloud and big data analytics enabled architecture for E-governance

Finally E-governance access layer is responsible for providing access to E-government portals over the different channels like laptop, desktop, mobile, kiosk etc. to their various stakeholders like citizens, businesses and other government departments.

IV. RESEARCH METHODOLOGY

The research methodology involves performing the statistical analysis to find out efficiency and effectiveness of cloud computing and big data analytics for E-governance. In the Statistical analysis, the online survey has been carried out to understand the efficiency and effectiveness of E-governance with integrated cloud computing and big data analytics. The

survey is carried out by collecting the data form various stakeholders through an online survey form by means of questionnaire [9]. The questions for survey have been formulated such a way that it can evaluate the importance of cloud computing and big data analytics for current E-governance system. After collecting the responses the Statistical test like chi-square were used to calculate the difference between two population means. To perform Chi-square test the IBMs statistical package for the social science (SPSS) is used for testing the hypothesis and its analysis. The IBM SPSS tool is used for analysis with crosstab method of descriptive analysis to find the chi-square value. The Chi-square test is performed over the collected data based on key indicators.

In the survey, different key indicators are in used to satisfy the research objective and will be used as a decision factors for the acceptability of the system by the various stakeholders.

1) Availability –: The availability is measured in terms of downtime such that minimum the downtime maximum the availability. This indicator measures the time by which the E-government portal was actually available with zero downtime using. Cloud computing.
2) Elasticity -: It checks whether the portal allows to elastically increase or decrease the resource capacity in case of increasing demand by the users so that portal never stopped during the runtime. It's a feature of cloud computing.
3) Response Time and latency-: It checks weather the portal makes timely response to the customer or not along with the delay in getting the response called latency used over the cloud computing infrastructure.
4) Throughput -: It measures how fast system generates the output. The throughput is used to check the performance of the system by means of calculating the request execution time on cloud.
5) Privacy -: privacy refers to the protecting the confidential access to the services in cloud computing and big data analytics.
6) Security -: The security measure test the vulnerabilities in the system based on various security parameters like authentication and authorization mechanisms in cloud computing and big data analytics.
7) Supports Data capture, storage, process and analyze -: It justifies whether the system allows to capture, store, process and analyze the data easily using big data analytics.
8) Cost efficiency -: It calculates the total cost including Capex and Opex incurred in accessing and delivering the services over the cloud platform and big data analytics.

The Samples of data is collected through two well defined questionnaires where Questionnaire I was designed for citizens and Questionnaire II was designed for E-governance service Implementers. The number of responses collected from E-governance Service implementers was 102 and from citizen was 342. Therefore the total Sample size considered is 444. The responsive data generated by primary questionnaire from Citizens and Service implementer for key indicators with options Yes or No are depicted in Figure 2.

A) Hypotheses testing

Based on the response provided by stakeholders on different key indicators, the hypothesis is tested using chi-square test [9].

Test Statistic -: Chi-square test is applied with 1 degree of freedom at 5% level of Significance.

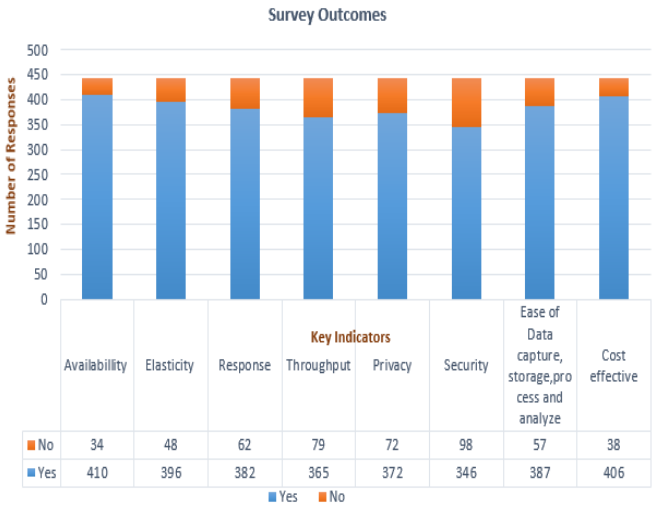

Figure 2. Survey outcomes for Questionnaire

Hypothesis 1-: Cloud computing provides high availability and elasticity for current E-governance solutions.

H0: Null Hypothesis: less than 70% citizens and service implementers think cloud computing is essential in current E-governance solutions to get the high availability and elasticity. (H0: $p < 0.70$)

H1: Alternate Hypothesis: 70% or more citizens and implementers think cloud computing is essential in current E-governance solutions to get the high availability and elasticity. (H1: $p>=0.70$)

The chi-square ($\chi 2$) value can be calculated as is

$\chi 2= \Sigma$ (fi-ei)^2/ ei

Where fi is observed count and ei is expected count

Table 1. Cross tabulation and Chi Square test for high availability and elasticity

Availability * Elasticity Crosstabulation

			Elasticity		
			NO	YES	Total
Availability	NO	Count	0	34	34
		Expected Count	3.7	30.3	34.0
		% within Availability	0.0%	100.0%	100.0%
		% within Elasticity	0.0%	8.6%	7.7%
		% of Total	0.0%	7.7%	7.7%
		Standardized Residual	-1.9	.7	
	YES	Count	48	362	410
		Expected Count	44.3	365.7	410.0
		% within Availability	11.7%	88.3%	100.0%
		% within Elasticity	100.0%	91.4%	92.3%
		% of Total	10.8%	81.5%	92.3%
		Standardized Residual	.6	-.2	
Total		Count	48	396	444
		Expected Count	48.0	396.0	444.0
		% within Availability	10.8%	89.2%	100.0%
		% within Elasticity	100.0%	100.0%	100.0%
		% of Total	10.8%	89.2%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4.463[a]	1	.035		
Continuity Correction[b]	3.331	1	.068		
Likelihood Ratio	8.114	1	.004		
Fisher's Exact Test				.038	.017
Linear-by-Linear Association	4.453	1	.035		
N of Valid Cases	444				

a. 1 cells (25.0%) have expected count less than 5. The minimum expected count is 3.68.

b. Computed only for a 2x2 table

From above table it is been observed that $\chi 2$ value is found out to be 4.463 which is greater than the table value 3.84 at 5% level of significance. Hence, Null hypothesis is rejected and alternate hypothesis is accepted. From survey outcome it is seen that 92 % of citizens and implementers agreed that cloud computing can provide high availability for E-governance while 89% agreed that cloud computing can provides elasticity for E-governance.

Hypothesis 2 -: The cloud computing and big data analytics has helped in improving the delivery of Government service with respect to better Response time with low latency and high throughput

H0: Null Hypothesis: less than 70% citizens and service implementers think cloud computing provides better response time with low latency and high throughput for E-governance. (H0: $p < 0.70$)

H1: Alternate Hypothesis: 70% or more citizens and implementers think cloud computing provides better response time with low latency and high throughput for E-governance. (H1: $p >= 0.70$)

Table 2. Cross tabulation and Chi Square test for Response time and throughput

Responsetime * Throughput Crosstabulation

			Throughput		
			NO	YES	Total
Responsetime	NO	Count	7	55	62
		Expected Count	11.0	51.0	62.0
		% within Responsetime	11.3%	88.7%	100.0%
		% within Throughput	8.9%	15.1%	14.0%
		% of Total	1.6%	12.4%	14.0%
		Standardized Residual	-1.2	.6	
	YES	Count	72	310	382
		Expected Count	68.0	314.0	382.0
		% within Responsetime	18.8%	81.2%	100.0%
		% within Throughput	91.1%	84.9%	86.0%
		% of Total	16.2%	69.8%	86.0%
		Standardized Residual	.5	-.2	
Total		Count	79	365	444
		Expected Count	79.0	365.0	444.0
		% within Responsetime	17.8%	82.2%	100.0%
		% within Throughput	100.0%	100.0%	100.0%
		% of Total	17.8%	82.2%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	2.083[a]	1	.149		
Continuity Correction[b]	1.598	1	.206		
Likelihood Ratio	2.293	1	.130		
Fisher's Exact Test				.209	.099
Linear-by-Linear Association	2.078	1	.149		
N of Valid Cases	444				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 11.03.

b. Computed only for a 2x2 table

From above table it is been observed that $\chi 2$ value is found out to be 2.063 which is less than the table value 3.84 at 5% level of significance. Hence, alternate hypothesis is rejected and null hypothesis is accepted. From survey outcome it is seen that 86 % of citizens and implementers says that cloud computing and big data analytics can provide high response time while 82% agreed that it can provides high throughput for E-governance.

Hypothesis 3 -: The cloud computing and big data analytics together can provide better Privacy and security for current E-governance

H0: Null Hypothesis: less than 70% citizens and service implementers think cloud computing and big data analytics has enhanced Privacy and security features for the current E-governance.(H0: $p < 0.70$)

H1: Alternate Hypothesis: 70% or more citizens and implementers think cloud computing is essential in current

E-governance solutions to get the high availability and elasticity. (H1: p>=0.70)

Table 3. Cross tabulation and Chi Square test for Privacy and Security

Privacy * Security Crosstabulation

			Security		
			NO	YES	Total
Privacy	NO	Count	7	65	72
		Expected Count	15.9	56.1	72.0
		% within Privacy	9.7%	90.3%	100.0%
		% within Security	7.1%	18.8%	16.2%
		% of Total	1.6%	14.6%	16.2%
		Standardized Residual	-2.2	1.2	
	YES	Count	91	281	372
		Expected Count	82.1	289.9	372.0
		% within Privacy	24.5%	75.5%	100.0%
		% within Security	92.9%	81.2%	83.8%
		% of Total	20.5%	63.3%	83.8%
		Standardized Residual	1.0	-.5	
Total		Count	98	346	444
		Expected Count	98.0	346.0	444.0
		% within Privacy	22.1%	77.9%	100.0%
		% within Security	100.0%	100.0%	100.0%
		% of Total	22.1%	77.9%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	7.620[a]	1	.006		
Continuity Correction[b]	6.787	1	.009		
Likelihood Ratio	8.851	1	.003		
Fisher's Exact Test				.005	.003
Linear-by-Linear Association	7.603	1	.006		
N of Valid Cases	444				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 15.89.

b. Computed only for a 2x2 table

From above table it is been observed that $\chi 2$ value is found out to be 7.620 which is greater than the table value 3.84 at 5% level of significance. Hence, null hypothesis is rejected and alternate hypothesis is accepted. From survey outcome it is seen that 84 % of citizens and implementers agree that cloud computing and big data analytics has better privacy while 77 % says cloud computing and big data analytics has good security mechanisms for E-governance.

Hypothesis 4 -: The big data analytics in E-governance provide cost efficiency with better strategies for data capture, storage, process and analysis.

H0: Null Hypothesis: less than 70% citizens and service implementers think big data analytics integrated E-governance will have efficient and cost effective mechanisms to capture, storage, process and analyze the Big data. (H0: p < 0.70)

H1: Alternate Hypothesis: 70% or more citizens and implementers think big data analytics integrated E-governance will have efficient and cost effective mechanisms to capture, storage, process and analyze the Big data. (H1: p>=0.70)

Table 4. Cross tabulation and Chi Square test for data capture, storage, process, and analysis with Cost efficiency

DataCaptureStorageProcessAnalysis * CostEfficiency Crosstabulation

			CostEfficiency		
			.00	1.00	Total
DataCaptureStorageProcessAnalysis	NO	Count	0	57	57
		Expected Count	4.9	52.1	57.0
		% within DataCaptureStorageProcessAnalysis	0.0%	100.0%	100.0%
		% within CostEfficiency	0.0%	14.0%	12.8%
		% of Total	0.0%	12.8%	12.8%
		Standardized Residual	-2.2	.7	
	YES	Count	38	349	387
		Expected Count	33.1	353.9	387.0
		% within DataCaptureStorageProcessAnalysis	9.8%	90.2%	100.0%
		% within CostEfficiency	100.0%	86.0%	87.2%
		% of Total	8.6%	78.6%	87.2%
		Standardized Residual	.8	-.3	
Total		Count	38	406	444
		Expected Count	38.0	406.0	444.0
		% within DataCaptureStorageProcessAnalysis	8.6%	91.4%	100.0%
		% within CostEfficiency	100.0%	100.0%	100.0%
		% of Total	8.6%	91.4%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	6.121[a]	1	.013		
Continuity Correction[b]	4.930	1	.026		
Likelihood Ratio	10.953	1	.001		
Fisher's Exact Test				.009	.004
Linear-by-Linear Association	6.107	1	.013		
N of Valid Cases	444				

a. 1 cells (25.0%) have expected count less than 5. The minimum expected count is 4.88.

b. Computed only for a 2x2 table

From above table it is been observed that $\chi 2$ value is found out to be 6.121 which is greater than the table value 3.84 at 5% level of significance. Hence, null hypothesis is rejected and alternate hypothesis is accepted. From survey outcome it is seen that 87 % of citizens and implementers agree that big data analytics has better efficient mechanisms for data capture, storage, process and analysis while 91 % agrees that the cloud computing and big data analytics are cost efficient technologies for E-governance.

V. RESULTS AND DISCUSSION

The integration of Cloud computing and big data analytics in E-governances delivers many eminent features which were not there in existing system. The statistical study implies that the cloud computing and big data analytics are essential for E-governance. To prove the effectiveness of Cloud computing and big data analytics, the comparative analysis is done for Cloud and big data enabled E-governance as shown in Table 5 and Table 6.

Table 5. Comparative Analysis of E-governance with and Without Cloud Computing

Sr. No.	Feature	Traditional E-governance Without Cloud Computing	E-governance With Cloud Computing
1	Cost (Including Capital expenses and Operational expenses)	High Capital expenses and High operational expenses	Very low Capital operational expenses
2	Infrastructure available and management	Very Limited and managed by local maintenance teams	Very Huge and managed by third party cloud service providers
3	Resource pooling	Not supported	Highly supported
4	Elasticity	Not supported	Highly supported
5	On demand Self Service	Not available as portals are hosted on local infrastructure	Highly available as portals are hosted on Cloud platforms
6	Disaster recovery	Very limited	Highly provided using on fail-over and failback clusters from geographically diverse datacenters
7	Load Balancing	Very limited	Provided through very efficient cloud load balancing algorithms run at master node
8	Availability	Less compared to Cloud hosting	Very high
9	Throughput	Less	Very high
10	Scalability	Supports only Vertical scaling over limited infrastructure	Supports both Horizontal and vertical scaling
11	Reliability	Less	Very high
12	Security	Medium to high	Medium to high
13	Flexibility	Limited	Good
14	Mobility	Not supported	Supported
15	Backup and replication	Limited over the physical infrastructure	Limitless over the virtual infrastructure
16	Data storage mechanisms	Only object level raw storage	Supports Object level and block level storage
17	Speed of access and	Slow because of shared hosting and	Very high due to cloud hosting
18	Interrupt in services	Interrupt the service if number of users exceeds the access slots	Never interrupt the website because of auto scaling feature
20	Storage support	Limited	Very high
21	Integrated Devops support	Not available	Available through cloud integrated Devops tools like Ansible, Puppet, Chef, Jenkins, Docker, Git,Nagios etc.
22	Software defined data center	Not supported	Highly supported
23	Integrated Software defined network	Not supported	Highly supported
24	Integrated Big data analytics Solutions	Not supported	Highly supported

Table 5. Comparative Analysis of E-governance with and Without Big data Analytics

Sr. No.	Feature	Traditional E-governance Without Big data analytics	E-governance With Integrated Big data analytics
1	Type of data supported	Structured	Structured, Semi-Structured and unstructured
2	Data architecture	Centralized	Fully Distributed
3	Volume of	Gigabyte to	Hundreds of

	data supported	terabyte	terabyte to petabyte
4	Database Schema	Fixed	Dynamic mostly schema less
5	View of data relationship	Easy and Simple	Difficult and Complex
6	Databases in use	Sql databases like Oracle, Mysql, Sibel etc.	NoSql databases like MongoDB, Cassendra, Couchdb, Hbase, memcache, Neo4j etc.
7	Data processing mechanism	Extract transform load (ETL) batch processing Operations	Parallel data processing operations
8	Scaling	Absent	Present
9	Cost per bit	Very high	Low
10	Cost for setup	Very expensive	Cost effective
11	Data Analytics	Can be Performed over Stored data in a RDBMS tables	Can be Performed over Real time accessed through web like websites or social medias as well as stored data in the tables
12	Administration	Difficult	Easy
13	Hardware Support	Physical servers with high cpu, memory and storage are required	Commodity hardware can be used or can be run on virtual machines in cloud
14	Accuracy of result	Low to medium	Very high
15	Input and output format	Tuple in a table	Key-Value pair
16	Data processing tools and frameworks	ETL tools like SQL Server Integration Services (SSIS), Informatica, Cognos, IBM Infospher etc.	Frameworks and tools like Hadoop, Spark, Storm, Flink, Samza, SaS, R, EMR etc.
17	Privacy	Very Good	Limited to moderate
18	Data store	Relational DBMS	Distributed file system or NOSQL databases
19	Read Write operations	Multiple read and Multiple write	Multiple read and Single write
20	Graph analytics	Limited	Supported using graph based NoSql databases

CONCLUSION AND FUTURE SCOPE

Based on the statistical and comparative analysis it is observed that the cloud computing and big data analytics are essentiality for E-governance because of their eminent features. This research paper has carried out the hypothetical study of cloud computing and big data analytics based on key indicators like Availability, Elasticity, Response Time, Throughput, Privacy ,Security ,Data capture, storage, process and analyze mechanisms, Cost efficiency. The study proves that the above factors are essential for evolving the E-governance. The limitations of cloud computing and big data enabled E-governance has complex architecture and needs and professional expertise to work on that.

The future scope of this research will be implementation of cloud computing and big data analytics for E-governance to test the performance parameters like throughput, speed, latency, scalability etc.

ACKNOWLEDGMENT

We gratefully acknowledge the experts who have contributed towards the development of this research paper especially Dr. G.T. Thampi and Prof. Arun Kulkarni.

REFERENCES

[1] Satyabrata Dash and Subhendu Kumar Pani, "E-Governance Paradigm Using Cloud Infrastructure: Benefits and Challenges", International Conference on Computational Modeling and Security, Elsevier, pp:843-855,2016.

[2] Smitha K. K., Dr. Tony Thomas, Chitaranjan K,"Cloud based E-Governance System: A Survey", International Conference on Modeling, Optimization and Computing", Elsevier, pp: 3816-3823, 2012.

[3] Rajkumar Sharma and Priyesh Kanungo,"An Intelligent Cloud Computing Architecture Supporting e-Governance", International Conference on Automation & Computing,IEEE, 2011,pp:1-5,2015.

[4] Yannis Charalabidis, Sotirios Koussouris, Antonis Ramfos," A Cloud Infrastructure for Collaborative Digital Public Services", International Conference on Cloud Computing Technology and Science, IEEE, pp:342-347, 2011.

[5] Rajiv Ranjan, Saurabh Garg, Ali Reza Khoskbar, et al.: "Orchestrating Big Data Analysis Workflows".IEEE transaction on Cloud Computing.IEEE Computer Society, pp:20-28,2017

[6] Deka Ganesh Chandra and Robin Singh Bhadoria, "Cloud Computing Model for National e-Governance Plan(NeGP)", International Conference on Computational Intelligence and Communication Networks,IEEE, pp:520-524,2012.

[7] Rajagopalan M.R and Solaimurugan vellaipandiyan,"Big Data Framework for National e-Governance Plan", International Conference on ICT and Knowledge Engineering,IEEE,2013.
[8] Saeid Abolfazli, Zohreh Sanaei et al.,"Cloud Adoption in Malaysia: Trends, Opportunities, and Challenges", IEEE CLOUD COMPUTING SOCIETY,2015.
[9] Prajakta N. Warale, Hemalatha Diwakar," A Study of Citizen Satisfaction for e-Governance Initiative SETU in Maharashtra (INDIA)", International Journal of Computer Applications (0975 – 8887),Volume 124 – No.17, August 2015.
[10] Bhushan Jadhav, Archana Patankar, "Opportunities and Challenges in integrating Cloud computing and Big data analytics to E-governance", International Journal of Computer Applications, (0975 – 8887), Volume No.182, January 2018.
[11] Bhushan Jadhav, Archana Patankar, "A Novel Solution for Cloud enabled E-governance using Openstack: Opportunities and Challenges", International conference on computers network and communications, proceedings in Communications in computer and information science, Springer CCIS, 2018.
[12] Ibrahim Abaker, Targio Hashem, Ibrar Yaqoob et al.:" The rise of "big data" on cloud computing : Review and open research issues",International Journal of Information Sciences, Elsevier,(2014)
[13] Ethiranjan D.,S. Purushothaman et al.,"Adoption of E-governance applications towards Big data approach", International Journal of Applied Engineering Research, ISSN 0973-4562,pp: 11336-11340, Volume 12, (2017)
[14] https://blog.microfocus.com/how-much-data-is-created-on-the-internet-each-day/
[15] Fang Liu, Jin Tong, Jian Mao, Robert Bohn, John Messina, Lee Badger and Dawn Leaf,"NIST Cloud Computing Reference Architecture", Recommendations of the National Institute of Standards and Technology, Special Publication 500-292, 2011.
[16] https://www.gartner.com/it-glossary/big-data

Authors Profile

Bhushan Jadhav pursed Bachelor of Computer Engineering from University of Mumbai, in 2008 and Master of Engineeering from University of Mumbai in year 2015. He is currently pursuing Ph.D. and currently working as Assistant Professor in Department of Information technology, Thadomal shahani engineering college, University of Mumbai,. He is a life member of ISTE since 2015, He has published many research papers in reputed international journals and conferensces including IEEE and Springer. His main research work focuses on Virtualization, Cloud Computing, Big Data Analytics, machine learning, data science and IoT.He has conducted more than 80+ hands on workshops across the india on Python programming,Cloud computing, big data analytics, Web programming, Linux administration etc. He has total 10 years of teaching experience and 4 years of Research Experience.

Archana B. Patankar pursued Bachelor of Computer Engineering from Walchand College of Engineering Sangli in 1996 Master of Engineering from VJTI in 1999, Mumbai.She has pursued Ph.D in Computer Engineering from NMIMS, Mumbai in 2011. She is working as Professor in Department of Computer Engineering, Thadomal Shahani Engineering College, Mumbai. She is a member of CSI & Life member of ISTE since 2004. She has published about 80 research papers in reputed international journals and conferences including IEEE, Springer. Her main research work focuses on Image processing, Cloud computing and Security. She has more than 18 years of teaching experience. She has received Distinguished Professor Award from CSI During CSI Technext in 2017 as well as 2018.

