iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.43407

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

Impact and Implications of Big Data Analytics in Cloud Computing Platforms

Shahana P N

Abstract: ***Big Data and Cloud Computing as two mainstream technologies, are at the center of concern in the IT field. Cloud Computing refers to the processing of anything, including Big Data Analytics, on the "cloud". The "cloud" is just a set of high-powered servers from one of many providers. They can often view and query large data sets much more quickly than a standard computer could. Essentially, "Big Data" refers to the large sets of data collected, while "Cloud Computing" refers to the mechanism that remotely takes this data in and performs any operations specified on that data. Cloud Computing services largely exist because of Big Data. Likewise, the only reason that we collect Big Data is because we have services that are capable of taking it in and deciphering it, often in a matter of seconds. The two are a perfect match, since neither would exist without the other. The combination of both yields beneficial outcome for the organizations. Not to mention, both the technologies are in the stage of evolution but their combination leverages scalable and cost-effective solution in big data analytics. Big data and Cloud computing are perfect combination. Besides that, there are also some real-time challenges to deal with. In this paper, discribes both the aspects. This paper introduces the characteristics, trends and challenges of big data. In addition to that, it investigates the benefits and the risks that may rise out of the integration between big data and cloud computing. KEYWORDS: Cloud Computing, Big Data, Scalable, Analytics.***

I. INTRODUCTION

Big data and Cloud Computing are one of the most used technologies in today's Information Technology world. With these two technologies, business, education, healthcare, research & development, etc are growing rapidly and will provide various advantages to expand their areas with tricks and techniques [1]. Big data deals with massive structured, semi-structured or unstructured data to store and process it for data analysis purpose. There are five aspects of Big Data which are described through 5Vs

1) Volume – the amount of data
2) Variety – different types of data
3) Velocity – data flow rate in the system
4) Value – the value of data based on the information contained within
5) Veracity – data confidentiality and availability

In cloud computing, all data is gathered in data canters and then distributed to the end-users. Further, automatic backups and recovery of data is also ensured for business continuity, all such resources are available in the cloud. We do not know exact physical location of these resources provided to us [2]. Cloud computing offers services to the users on a pay-as-you-go model. Cloud providers offer three primary services, these services are outlined below:

a) *Infrastructure as a Service (IAAS):* Here the service provider offers entire infrastructure along with the maintenance related tasks.
b) *Platform as a Service (PAAS):* In this service, the Cloud provider offers resources like object storage, runtime, queuing, databases, etc. However, the responsibility of configuration and implementation related tasks depend on the consumer.
c) *Software as a Service (SAAS):* This service is the most facilitated one which provides all the necessary settings and infrastructure provides IaaS for the platform and infrastructure are in place.

Big data refers to huge volume of data, its management, and useful information extraction. The two go hand-in-hand, with many public cloud services performing big data analytics. With Software as a Service (SaaS) becoming increasingly popular, keeping up-to-date with cloud infrastructure best practices and the types of data that can be stored in large quantities is crucial. Data storage using cloud computing is a viable option for small to medium sized businesses considering the use of Big Data analytic techniques. Cloud computing is on-demand network access to computing resources which are often provided by an outside entity and require little management effort by the business. A number of architectures and deployment models exist for cloud computing, and these architectures and models are able to be used with other technologies and design approaches.

©IJRASET: All Rights are Reserved

Owners of small to medium sized businesses who are unable to afford adoption of clustered NAS technology can consider a number of cloud computing models to meet their big data needs. Small to medium sized business owners need to consider the correct cloud computing in order to remain both competitive and profitable [3].

A. Relationship Between big Data & Cloud Computing

Cloud Computing providers often utilize a "software as a service" model to allow customers to easily process data. Typically, a console that can take in specialized commands and parameters is available, but everything can also be done from the site's user interface. Some products that are usually part of this package include database management systems, cloud-based virtual machines and containers, identity management systems, machine learning capabilities, and more. The combination of both yields beneficial outcome for the organizations [4].

Not to mention, both the technologies are in the stage of evolution but their combination leverages scalable and cost-effective solution in big data analytics. So, can we say big data and Cloud computing a perfect combination? Well, there are data points in support of it. Besides that, there are also some real-time challenges to deal with. In this blog, we will discuss both the aspects. We assume you have some idea and knowledge on Big data and Cloud computing. In turn, Big Data is often generated by large, network-based systems. It can be in either a standard or non-standard format. If the data is in a non-standard format, artificial intelligence from the Cloud Computing provider may be used in addition to machine learning to standardize the data.

From there, the data can be harnessed through the Cloud Computing platform and utilized in a variety of ways. For example, it can be searched, edited, and used for future insights.

This cloud infrastructure allows for real-time processing of Big Data. It can take huge "blasts" of data from intensive systems and interpret it in real-time. Another common relationship between Big Data and Cloud Computing is that the power of the cloud allows Big Data analytics to occur in a fraction of the time it used to [5].

However, Cloud Computing allows us to use state-of-the-art infrastructure and only pay for the time and power that we use! Cloud application development is also fueled by Big Data. Without Big Data, there would be far fewer cloud-based applications, since there wouldn't be any real necessity for them. Remember, Big Data is often collected by cloud-based applications, as well!

In short, Cloud Computing services largely exist because of Big Data. Likewise, the only reason that we collect Big Data is because we have services that are capable of taking it in and deciphering it, often in a matter of seconds. The two are a perfect match, since neither would exist without the other!

II. CLOUD COMPUTING ROLE FOR BIG DATA

Cloud Computing is the delivery of computing services such as servers, storage, databases, networking, software, analytics etc., over the Internet (" the cloud") with the aim of providing flexible resources, faster innovation and economies of scale [6]. Cloud computing has revolutionized the way computing infrastructure is abstracted and used. Cloud paradigms have been extended to include anything that can be considered as a service (hence x a service). The many benefits of cloud computing such as elasticity, pay-as-you-go or pay-per-use model, low upfront investment etc., have made it a viable and desirable choice for big data storage, management and analytics [7]. Because big data is now considered vital for many organizations and fields, service providers such as Amazon, Google and Microsoft are offering their own big data systems in a cost-efficient manner. These systems offer scalability for business of all sizes. This had led to the prominence of the
term Analytics as a Service (AaaS) as a faster and efficient way to integrate, transform and visualize different types of data.

Big data and Cloud computing relationship can be categorized based on service types:

1) *IAAS in Public Cloud:* IaaS is a cost-effective solution and utilizing this Cloud service, Big Data services enable people to access unlimited storage and compute power. It is a very cost-effective solution for enterprises where the Cloud provider bears all the expenses of managing underlying hardware.
2) *PAAS in Private Cloud:* PaaS vendors incorporate Big Data technologies into their offered service. Hence, they eliminate the need for dealing with the complexities of managing single software and hardware elements which is a real concern while dealing with terabytes of data.
3) *SAAS in Hybrid Cloud:* Analysing social media data is nowadays an essential parameter for companies for business analysis. In this context, SaaS vendors provide an excellent platform for conducting the analysis.

A. *Big Data Analytics Cycle*

According to processing big data for analytics differs from processing traditional transactional data. In traditional environments, data is first explored then a model design as well as a database structure is created. In order to provide a methodology to organize the work and deliver clear insights from Big Data, there is a cycle with different stages [8]. All the stages of the Big Data life cycle are related to each other. A data analytics architecture maps out such steps for data science professionals. It is a cyclic structure that encompasses all the data life cycle phases, where each stage has its significance and characteristics. Figure 1. depicts the flow of big data analysis. As can be seen, it starts by gathering data from multiple sources, such as multiple files, systems, sensors and the Web [6].

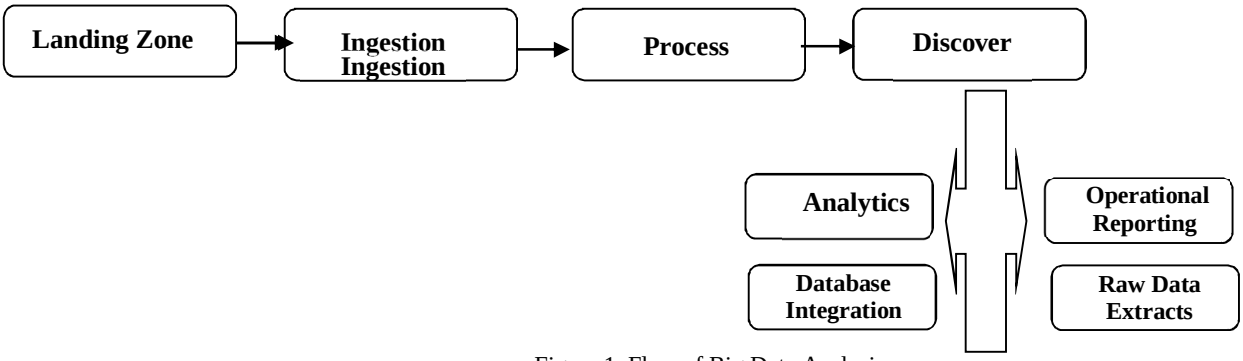


Figure 1: Flow of Big Data Analysis

This data is then stored in the so called" landing zone" which is a medium capable of handling the volume, variety and velocity of data. This is usually a distributed file system. After data is stored, different transformations occur in this data to preserve its efficiency and scalability. After that, they are integrated into particular analytical tasks, operational reporting, databases or raw data extracts [9].

A scientific method that helps give the data analysis process a structured framework is divided into six phases of data analytics architecture.

1) *Data Discovery and Formation*

Data discovery involves the collection and evaluation of data from various sources and is often used to understand trends and patterns in the data. It requires a progression of steps that organizations can use as a framework to understand their data. In this phase, you'll define your data's purpose and how to achieve it by the time you reach the end of the data analytics lifecycle. The initial stage consists of mapping out the potential use and requirement of data, such as where the information is coming from, what story you want your data to convey, and how your organization benefits from the incoming data. Basically, as a data analysis expert, you'll need to focus on enterprise requirements related to data, rather than data itself. Additionally, your work also includes assessing the tools and systems that are necessary to read, organize, and process all the incoming data [6].

2) *Data Preparation and Processing*

Data preparation is the process of gathering, combining, structuring and organizing data so it can be used in business intelligence, analytics and data visualization applications. The components of data preparation include data pre-processing, profiling, cleansing, validation and transformation; it often also involves pulling together data from different internal systems and external sources. This stage consists of everything that has anything to do with data. In this stage attention of experts moves from business requirements to information requirements. The data preparation and processing step involve collecting, processing, and cleansing the accumulated data. One of the essential parts of this phase is to make sure that the data you need is actually available to you for processing [10]. The earliest step of the data preparation phase is to collect valuable information and proceed with the data analytics lifecycle in a business ecosystem. Data is collected using the below methods:

Data Acquisition: Accumulating information from external sources.

Data Entry: Formulating recent data points using digital systems or manual data entry techniques within the enterprise.

Signal Reception: Capturing information from digital devices, such as control systems and the Internet of Things.

3) Design a Model

After mapping out your business goals and collecting a glut of data (structured, unstructured, or semi-structured), it is time to build a model that utilizes the data to achieve the goal. Good data models make data statistics more consistent and reduce the possibility of computing errors. There are several techniques available to load data into the system and start studying it:

ETL (Extract, Transform, and Load) transforms the data first using a set of business rules, before loading it into a sandbox.

ELT (Extract, Load, and Transform) first loads raw data into the sandbox and then transform it.

ETLT (Extract, Transform, Load, Transform) is a mixture; it has two transformation levels.

This step also includes the teamwork to determine the methods, techniques, and workflow to build the model in the subsequent phase. The model's building initiates with identifying the relation between data points to select the key variables and eventually find a suitable model [4].

4) Model Building

This step of data analytics architecture comprises developing data sets for testing, training, and production purposes. The data analytics experts meticulously build and operate the model that they had designed in the previous step. They rely on tools and several techniques like decision trees, regression techniques (logistic regression), and neural networks for building and executing the model. The experts also perform a trial run of the model to observe if the model corresponds to the datasets. It is a theoretical representation of data objects and relationships between them. The process of formulating data in a structured format in an information system is known as data modeling. It facilitates data analysis, which will aid in meeting business requirements [3].

5) Result Communication and Publication

The communication step starts with a collaboration with major stakeholders to determine if the project results are a success or failure. The project team is required to identify the key findings of the analysis, measure the business value associated with the result, and produce a narrative to summarise and convey the results to the stakeholders.

6) Measuring of Effectiveness

As your data analytics lifecycle draws to a conclusion, the final step is to provide a detailed report with key findings, coding, briefings, technical papers/ documents to the stakeholders. It consists of assessing first the quality of Big Data itself, which involve processes such as cleansing, filtering and approximation. Then, assessing the quality of process handling this Big Data, which involve for example processing and analytics process [8].

B. Big Data Cloud

The cloud can help you process and analyze your big data faster, leading to insights that can improve your products and business. Merging big data with cloud computing is a powerful combination that can transform your organization. When big data computing takes place in the clouds it is known as "Big Data Clouds". Their purpose is to build an integrated infrastructure that is suitable for quick analytics and deployment of an elastically scalable infrastructure. Cloud technology is used to derive quantum-leap advantages inherent in big data [10]. Hence, from the above description, we can see that Cloud enables "As-a-Service" pattern by abstracting the challenges and complexity through a scalable and elastic self-service application. Big data requirement is same where distributed processing of massive data is abstracted from the end users.

There are multiple benefits of big data analysis in Cloud.

1) *Improved Analysis:* With the advancement of Cloud technology, big data analysis has become more improved causing better results. Hence, companies prefer to perform big data analysis in the Cloud. Moreover, Cloud helps to integrate data from numerous sources.
2) *Simplified Infrastructure:* Big Data analysis is a tremendous strenuous job on infrastructure as the data comes in large volumes with varying speeds, and types which traditional infrastructures usually cannot keep up with. As the Cloud computing provides flexible infrastructure, which we can scale according to the needs at the time, it is easy to manage workloads.
3) *Lowering the Cost:* Both Big data and Cloud technology delivers value to organizations by reducing the ownership. The Pay-per-user model of Cloud turns CAPEX into OPEX. On the other hand, Apache cut down the licensing cost of Big data which is supposed to be cost millions to build and buy. Cloud enables customers for big data processing without large-scale big data resources. Hence, both Big Data and Cloud technology are driving the cost down for enterprise purposes and bringing value to the enterprise. Mining big data in the cloud has made the analytics process less costly. In addition to the reduction of on-

premise infrastructure, you can also save on costs related to system maintenance and upgrades, energy consumption, facility management, and more.

4) *Security and Privacy:* Data security and privacy are two major concerns when dealing with enterprise data. Moreover, when your application is hosted on a Cloud platform due to its open environment and limited user control security becomes a primary concern. On the other hand, being an open-source application, Big data solution like Hadoop uses a lot of third-party services and infrastructure. Hence, nowadays system integrators bring in Private Cloud Solution that is Elastic and Scalable.

III. CHALLENGES

Big data challenges include the storing, analyzing the extremely large and fast-growing data. Perhaps the most frequent challenge in big data efforts is the inaccessibility of data sets from external sources. Sharing data can cause substantial challenges. Cloud data is stored and processed in a central location commonly known as Cloud storage server. Along with it the service provider and the customer sign a service level agreement (SLA) to gain the trust between them. If require the provider also leverages required advanced level of security control. The technology of the cloud provides nearly unlimited resources essential for big data management because organizations can always purchase more space on the cloud. Some of the challenges of big data are variety of data, data storage and integration, data processing and resource management. Some of the challenges of cloud computing are availability, transformation, security concern, charging model. However, if businesses plan to manage big data in the cloud, they need to be aware of the common vulnerabilities of cloud technology. When merging big data and the cloud, you get the convenience of the cloud, but also one that comes with security risks. In cybersecurity, all the tools and protocols that companies need to protect their cloud is referred to as cloud application security [6].

To have a working cloud, it's necessary that it's connected to the internet. Any outage will create issues you wouldn't otherwise have if you use big data in the traditional way. We may lose access to your big data if your internet connection fails or our internet connection might even experience lag that will affect your teams' productivity and disrupt workflow. Our cloud network provider also has to be connected. In case something doesn't work on their end, this means that you're locked out of the cloud and can't do the work on big data. The fundamental issue that should be considered is the security of the big data cloud environment. There are some security vulnerabilities that rise because of this integration between both and creating a new unfamiliar platform.

IV. CONCLUSION

Cloud computing seems to be a perfect vehicle for hosting big data workloads. However, working on big data in the cloud brings its own challenge of reconciling two contradictory design principles. The integrating big data with cloud computing technologies, businesses and education institutes can have a better direction to the future. The capability to store large amounts of data in different forms and process it all at very large speeds will result in data that can guide businesses and education institutes in developing fast. Cloud Computing and Big Data Analytics have truly impacted the way organizations function and humans operate. Cloud Computing provides benefits which are applicable to all sizes of businesses and all kinds of individuals. Data is perceived as a resource and organizations are scrambling to implement Hadoop to exploit this resource. It is interesting to know that although these technologies have become mainstream, companies are still investing huge amounts in R&D. We can expect more growth of Cloud Computing and Big Data Analytics in coming years [7].

Finally, it's important to note that both Big Data and Cloud Computing play a huge role in our digital society. The two linked together allow people with great ideas but limited resources a chance at business success. They also allow established businesses to utilize data that they collect but previously had no way of analyzing. More modern components of cloud infrastructure's typical "Software as a Service" model such as artificial intelligence also enable businesses to get insights based on the Big Data they've collected. With a well-planned system, businesses can take advantage of all of this for a nominal fee, leaving competitors who refuse to use these new technologies in the dust.

REFERENCES

[1] S. Yadav and A. Sohal," Review Paper on Big Data Analytics in Cloud Computing," International Journal of Computer Trends and Technology (IJCTT), vol. IX, 2017.

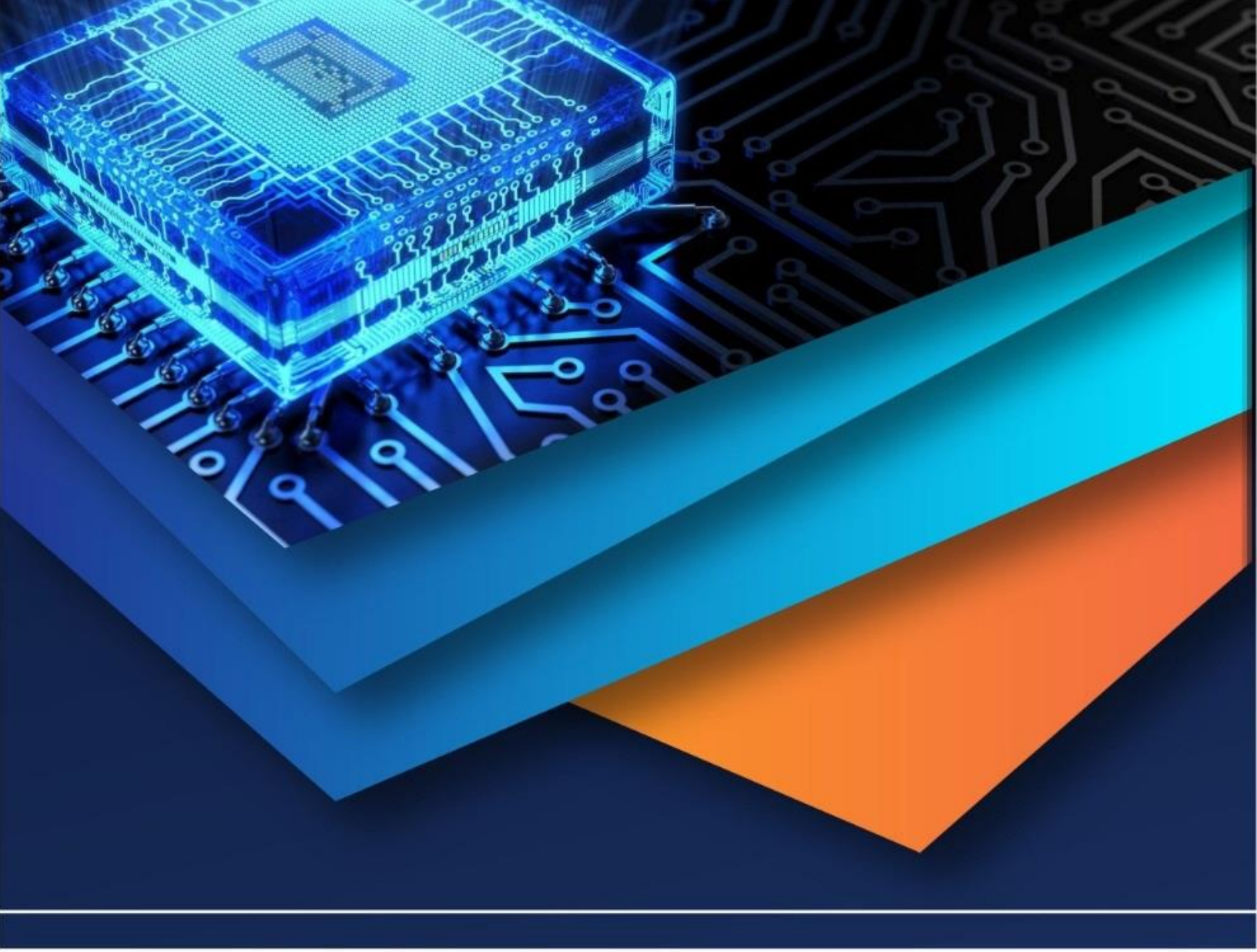
[2] J. Weathington," Big Data Defined.," Tech Republic, 2012.

[3] Rajeev Gupta, Himanshu Gupta, and Mukesh Mohania, "Cloud Computing and Big Data Analytics: What Is New from Databases Perspective?". S. Srinivasa and V. Bhatnagar (Eds.): BDA 2012, LNCS 7678, pp. Springer-Verlag Berlin Heidelberg 42–61, 2012.

[4] Alberto Ferandez, Sara del R, Victoria opez, Abdullah Bawakid, Maria J. del Jesus, Jose M. Benitez, and Francisco Herrera. "Big Data with Cloud Computing: an insight on the computing environment, MapReduce, and programming frameworks". doi: 10.1002/widm.1134. WIREs Data Mining Knowl Discov, 4:380–409, 2014.

[5] Shim K, Cha SK, Chen L, Han W-S, Srivastava D, Tanaka K, Yu H, Zhou X. Data management challenges and opportunities in cloud computing. In: 17th International Conference on Database Systems for Advanced Applications (DASFAA'2012). Berlin/Heidelberg: Springer 323; 2012
[6] Chandrashekar, R., Kala, M., & Mane, D. (2015). Integration of Big Data in Cloud computing environments for enhanced data processing capabilities. International Journal of Engineering Research and General Science, 240-245.
[7] James Kobielus, I., & Bob Marcus, E. S. (2014). Deploying Big Data Analytics Applications to the Cloud: Roadmap for Success. Cloud Standards Customer Council.
[8] Big Data Technologies and Cloud Computing (PDF) | SciTech Connect (elsevier.com)
[9] IOS Press. (2011). Guidelines on security and privacy in public cloud computing. Journal of EGovernance,34 149-151. DOI: 10.3233/GOV-2011-0271.
[10] Managing Data in Motion: Data Integration Best Practice Techniques and Technologies, First Edition (2013) 125- 128. doi:10.1016/B978-0-12-397167-8.00018-2

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)