

# Challenges and Opportunities in Big Data and Cloud Computing

Hassan Sohail, Zeenia Zameer, Hafiz Farhan Ahmed, Usama Iqbal,  
and Pir Amad Ali Shah<sup>(✉)</sup>

Computer Science Department, University of South Asia, Lahore, Pakistan  
hassansohail@live.com, zeenia.zameer@yahoo.com,  
farhanahmad1525@gmail.com, uiqbal1970@gmail.com,  
amad.ali@usa.edu.pk

**Abstract.** In this paper we have discussed the approaches and complexity of big data analytics issues in the perspective of cloud computing. Big data and cloud computing are novel methods of distributing computational resources. Currently big data and cloud computing has engrossed romantic behavior, which needs handling huge quantity of data quickly and securely. In the modern era the size of data is dramatically increasing due to cloud computing technology, which poses various challenges with the variety, security, and size of data. Cloud computing is the newest and the main kind of distributed computing systems and it wraps most of their features. It has already been broadly employed for its huge merits and its capability to handle large amount of data like workflows and big data applications. We have identified some gaps in technology, its challenges, limitation, and applications of big data and cloud computing from theoretical perspective. Moreover, this paper give recommendations to the researchers on future direction sand suggests different solutions to the challenges and limitations.

**Keywords:** Big data · Cloud computing · Cloud data · Security

## 1 Introduction

The advancement in science and technology has made the globe as a small town and as a result a huge amount of data is produced and stored daily. Extracting precise and relevant information from this data is very challenging and helpful in business competition. Various data mining solutions that extract structured and unstructured data is main key for organizations to gain insights from company private data as well as huge amounts of publically available data. The capability of validating customer private information such as product preferences, likes and dislikes etc. with the data obtained from tweets, blogs, feedback, product evaluations, and information available on social networks opens a broad variety of potentials for organizations to know the requirements of their clientele, forecast their needs, demands, and efficiently utilize the resources. This standard is being popularly known as Big Data. In other words, the big data is described as a dataset whose volume is afar the processing capability of classical databases or systems. Four essentials are highlighted in the description of big data that

are capture, store, manage, and analyses [2]. The focal point of the four essentials is the final phase, the big data analytics that is repeatedly pulling out information from a huge quantity of data. It can be seen as the drawing out or handing out of the enormous data and helpful facts can be regained from the large dataset [3]. The conventional technique for analyzing data is laid on the arithmetical models of the problems first and then looks if data vigorous the models. With the increase of the diversity of sensible data, these arithmetical models might be unproductive in resolving problems. The pattern must move from the model-driven to the data-driven techniques. The data-driven technique spotting on forecasting what is happening and also ponders on what is occurring right now and getting prepared for the potential events.

Cloud computing has been transfiguring the Information Technology industry by giving liveness to the way Information Technology is utilized, allowing organizations to pay only for the services and resources they employ. In an attempt to lessen IT assets and running cost, from small to large size are using Clouds to offer the wherewithal needed to run their applications. Clouds differ considerably in their precise technologies and implementation, but frequently give infrastructure, platform, and software resources as services. The most frequently stated benefits of Clouds include offering resources in a pay-as-you-go fashion, enhanced accessibility and flexibility, and fee decrease. Clouds can avoid organizations from costs for preserving peak-provisioned IT assets that they are improbable to utilize the majority of the time. Even as at first glimpse the price suggestion of Clouds as a podium to perform analytics is strong, there are numerous issues that require trouncing to craft Clouds a perfect platform for scalable analytics. In cloud computation, usually many computers are being used to lodge all the users' requirements on instantaneous base. So an enormous facts and/or figures have to be transmitted from one place to another place for implementation of some programs based on the necessities of memory, processor, disk space etc. especially in big data Cloud computing network have some attractive manners like servers are clustered as sub-network within the network. The arrangement can be shaped to specific arrangement techniques. Cloud computing usually has two kinds of connection, i.e. direct and indirect. Peer-to-peer data movement is handling by direct connections while observing system holds all the situations in the network using indirect connections. Fetching huge amount of data is hard as there are many barriers to conquer. The first barrier is handling the gigantic quantity of data rapidly. The size of data influences the performance of cloud network. The big data analytics also experiences from problems where the huge amount of data will be executed in a short time with a sensibly good performance.

The rest of this paper is organized as follows. In Sect. 2, challenges and issues is presented followed by the discussion in Sect. 3. The paper in concluded and future research directions and gaps are discussed in Sect. 4.

## 2 Challenges and Opportunities

With any kind of advancement in technology, cloud computing ought to be thoroughly assessed prior to its extensive adoption. A small number of researches have scientifically considered cloud computing affect on information technology by classifying it as

challenges of cloud computing and its opportunities. In this paper we assess the opportunities and challenges from the following perspective; with opportunity and challenges [6–8].

1. Management as a challenge in Big Data and Cloud Computing
  - a. Less expenses on Information Technology infrastructure
    - i. Lack of trust by health care professionals
  - b. Computing resources available on demand
    - i. Organizational inertia
  - c. Payment of use on a short-term basis as needed
    - i. Loss of governance
2. Technology as a challenge in Big Data and Cloud Computing
  - a. Reduction of IT maintenance burdens
    - i. Uncertain provider's compliance
  - b. Scalability and flexibility of infrastructure
    - i. Resource exhaustion issues
    - ii. Unpredictable performance
  - c. Benefits for green computation
    - i. Secure Data
    - ii. Restricted access on data transfer
    - iii. Faults in hefty distributed cloud computing
3. Security as a challenge in Big Data and Cloud Computing
  - a. No resource constraints on protecting data
    - i. Non-centralized collapse
  - b. Secure data by placing its various copies at different locations
    - i. Public administration problems
  - c. Vigorously leveled protective assets intensification pliability
    - i. Meager key encryption
4. Legal as a challenge in Big Data and Cloud Computing
  - a. Supplier's promises to guard consumer's information/privacy
    - i. Privilege abuse
  - b. Mature strategies/technologies for enabling the erection of reliance policies by nonprofit groups
    - i. Data jurisdiction issues
  - c. Development of rules by government regarding information/privacy security
    - i. Privacy issues
5. Scalability as a challenge in Big Data and Cloud Computing
  - a. Distributed data storage systems
    - i. Relational Database Management Systems are not supported by cloud technologies
6. Availability as a challenge in Big Data and Cloud Computing
  - a. Data is accessible from everywhere
    - i. deliver high-quality services
    - ii. data integrity and security

7. Transformation as a challenge in Big Data and Cloud Computing
  - a. Structured and Unstructured data
    - i. In unstructured data, data need to saved in a distributed databases prior to processing
8. Heterogeneity is a challenge in Big Data and Cloud Computing
  - a. Data from multiple sources
    - i. users can save data three formats, i.e. structured format, semi-structured format, or unstructured format
9. Privacy as a challenge in Big Data and Cloud Computing
  - a. Personal details are one click away
    - i. Personal details are exposed to scrutiny,
    - ii. Profiling, stealing, and loss of control issues
10. Governance is a Challenge in Big Data and Cloud Computing
  - a. Use of standards
    - i. Applications that use huge amounts of data obtained from outside sources.

### 3 Discussion

The key limitations and challenges in big data and cloud computing are the amount of data and the storage capacity, accessing and fetching the data. Cloud computing usually uses direct and indirect connections. Small amount of data on direct connections have no issues however when the amount of data increases then fetching large dataset has different hurdles to overcome. The first hurdle is the amount of data which ultimately affect processing on data. The big data analytics experiences from problems where the huge amount of data will be executed in a short time with a sensibly good performance. Another hurdle is the frequency of changes in the data content. As the data is constantly growing so efficient and intelligent algorithms are need to be modified. Likewise, another hurdle is the diversity of data, i.e. various types' data coming from various sources. Sometimes various types of formless data require to be pre-processed to semi-formed and/or well formed data before final processing. In some situations, several goals need to be attained concurrently in large datasets. Mainstream conventional techniques can only produce good result with constant and various algorithms, and need to do a sequence of disconnect runs to assure different goals [7, 8]. With the advancements in various communication fields such as intelligent transportation systems [9], Internet of Things (IOT), Software defined networks, there is requirement for more study to break the multi-goals problems with less constraint.

The concept of cloud computing shift the computing power and data storage away from computer into the cloud, here are the potential benefits of cloud computing using swarm intelligence for big data. It helps to overcome processing limitations of normal systems with high amount of data storage, and power consumption. It also increases security level for data through a centralized monitoring and maintenance of software. Similarly, the power of the cloud computing can be seen in E-commerce and E-marketing where huge amount of data is produced. For example, use of the cloud technology in business allows us to efficiently utilize resources which reduce business

cost. Thus it will charm enterprise's kernel competitive control and finally complete goods and services trading.

Further, we give detail discussion of the challenges and opportunity discussed in previous section, i.e. management, technology, security, legality, scalability, availability, transformation, heterogeneity, privacy, and governance.

### 1. **Management**

The principle advantage of cloud computing is its low cost due to its quick deployment and maintenance. This ability shows that due to change in demand organizations are not required to change infrastructures. Similarly a cultural confrontation; to share information and changing conventional ways of working is a general administration challenge to accept cloud technology.

### 2. **Technology**

Smaller organizations typically do not have internal IT staff. Hence, removing the new information technology infrastructure and maintenance burdens can eliminate lots of barriers. Similarly, cloud computing has advantages for so-called *green computing*—the more efficient use of computer resources to help the environment and promote energy saving. However, the main issues are resource exhaustion, unpredictability of performance, data lock-in, data transfer bottlenecks, and bugs in large-scale distributed cloud systems.

### 3. **Security**

Possibly the toughest antagonism to cloud technology implementation in an organization is data security. Most cloud suppliers put users' information in various locations which makes profile redundant and less secure. These includes hacker attacks, network breaks, natural disasters, separation failure, public management interface, poor encryption key management, and privilege abuse.

### 4. **Legality**

Data and privacy protection are necessary for gaining consumer faith required for cloud technology to utilize its complete prospective. If the supplier implements enhanced and efficient rules and practice, consumer will easily evaluate the associated risks they face. However, the use of cloud technology may pose various legal issues like contract law, intellectual property rights, data jurisdiction, and privacy.

### 5. **Scalability**

The capability of memory to grip growing sums of information in a suitable way. Scalable dispersed data storage systems is a vital element of cloud infrastructures. However, due to the lack of cloud computing features to support RDBMS associated with enterprise solutions has made RDBMS less attractive for the deployment of large-scale applications in the cloud.

### 6. **Availability**

It is the resources of a system available to the clients on demand by an authoritative person. In a cloud computing, a major problem regarding cloud service providers is the accessibility of data saved in the cloud. However, the increase in number of cloud users may cause quality of services, and most important issue is the private and secret information on cloud has security threats.

### 7. Transformation

Altering the data into an appropriate format for investigation is a hindrance in the implementation of big data. Due to different types of data formats, big data can be changed into an analysis workflow is different for structured and unstructured data.

### 8. Heterogeneity

Data from multiple sources are generally of different types and representation forms and significantly interconnected; they have incompatible formats and are inconsistently represented, i.e. users can store data in structured, semi-structured, or unstructured format. However, unstructured data are inappropriate because they have a complex format that is difficult to represent in rows and columns.

### 9. Privacy

Privacy apprehensions carry on hampering those clients who outsource personal data into cloud. This apprehension has turned into severe issues with the improvement of big data mining and analytics, which need personal information to produce relevant results, such as personalized and location-based service. However, if information of an individual is leaked which will give rise to concerns on profiling, stealing, and loss of control.

### 10. Governance

The exercise of control and authority over data related rules of law, transparency, and accountabilities of individuals and information systems to achieve business objectives. However, policies, principles, and frameworks that strike stability between risk and value in the face of increasing data size and deliver better and faster data management technology can create huge challenges.

## 4 Conclusion

In this paper we have mentioned a few important challenges and issues in big data and cloud computing. We discussed cloud computing will grow and with the age of big data, and have elaborated some of the key challenges existing in the field of cloud computing. With the existing tool and techniques it is not sufficient to adhere all the challenges relating to big volume of data. It is not feasible to provide better data quality with the existing technology and again privacy is big problem with cloud data. Processing streaming data need some novel algorithms and some efficient tools. Again from privacy point of view hashing is not possible for data with volume of data. Confidentiality challenge for data can be addressed by developing any novel algorithms for key management and key exchange.

## References

1. Wang, Z., Chu, Y., Tan, K.-L., Agrawal, D., Abbadi, A.E., Xu, X.: Scalable data cube analysis over big data. arXiv preprint [arXiv:1311.5663](https://arxiv.org/abs/1311.5663) (2013)
2. Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., Tufano, P.: Analytics: The Real-World Use of Big Data. IBM Global Business Services, Armonk (2012)

3. Akerkar, R.: *Big Data Computing*. CRC Press, Boca Raton (2013)
4. Majhi, S.K., Shial, G.: Challenges in big data cloud computing and future research prospects: a review. *Smart Comput. Rev.* **5**(4) (2015)
5. Assunção, M.D., Calheiros, R.N., Bianchi, S., Netto, M.A.S., Buyyab, R.: Big data computing and clouds: trends and future directions. *J. Parallel Distrib. Comput.* **79–80**, 3–15 (2015)
6. Kuo, A.M.-H.: Opportunities and challenges of cloud computing to improve health care services. *JMIR* **13**(3) (2011)
7. [http://sameekhan.org/pub/H\\_K\\_2015\\_IS.pdf](http://sameekhan.org/pub/H_K_2015_IS.pdf)
8. Targio, H.I.A., Yaqoob, I., Anuar, N.B., Mokhtar, S., Gani, A., Khan, S.: The rise of big data on cloud computing: review and open research issues. *Inf. Syst.* **47**, 98–115 (2015)
9. Alam, M., Ferreira, J., Fonseca, J.: Introduction to intelligent transportation systems. In: Alam, M., Ferreira, J., Fonseca, J. (eds.) *Intelligent Transportation Systems*. SSDC, vol. 52, pp. 1–17. Springer, Heidelberg (2016). doi:[10.1007/978-3-319-28183-4\\_1](https://doi.org/10.1007/978-3-319-28183-4_1)