International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IMPACT FACTOR: 7.056

IJCSMC, Vol. 9, Issue. 11, November 2020, pg.38 – 48

Usage of Cloud Computing and Big data for Internet of Things

Erdal Özbay^{1*}; Feyza Altunbey Özbay²

^{1*}Computer Engineering Department, Faculty of Engineering, Firat University, Turkey, 0000-0002-9004-4802

²Software Engineering Department, Faculty of Engineering, Firat University, Turkey, 0000-0003-0629-6888

^{1*}erdalozbay@firat.edu.tr; ²faltunbey@firat.edu.tr

DOI: 10.47760/ijcsmc.2020.v09i11.004

Abstract— *The Internet of Things (IoT) has been very popular among researchers' study topics, especially for the past few years. Cloud Computing and Big data are two current issues that are increasingly attracted by almost everyone and that we frequently encounter in our daily lives. With these technologies, large amounts of data collected today can be stored in large-sized storage systems using information infrastructures and high-level calculations can be made. There are various requirements for objects to interact with stored data and communicate with other devices, which is an important opportunity in today's competitive environment. In this study, the relationship between the internet of things, Cloud computing, and Big data are examined. The aim of this study is to provide general information to researchers who will conduct data-based studies on Cloud systems in order to use the Internet of Things and related methods and technologies effectively.*

Keywords— *Cloud computing, Big data, Internet of Things*

I. INTRODUCTION

Information technologies are defined as technologies that allow information to be collected, processed, stored, transmitted, or accessed from anywhere. On the other hand, the concept of Information Technology, it is meant that all of the rapidly developing computers, networks, mobile, and similar tools used to provide information to organizations and individuals, as well as information and communication tools such as the internet, intranet, extranet, e-mail and web [1].

There is no consensus on a global definition for Cloud computing yet. From a broad perspective, the service provider is the hardware and software resources in the data center [2]. Cloud computing is a model that provides widespread, practical, and instant network access to a shared pool of adjustable computing resources (e.g. networks, servers, storage, applications, and services) that can be quickly charged and released with minimal management effort or service provider interaction [3]. Cloud computing is a parallel and distributed computing system that includes dynamically charged and connected and virtualized computers with one or more combined computing resources, with a service agreement established between the service provider and its customer [4]. It serves the same purpose as distributed and parallel programming models before, but it contains some functional differences. Although it has the same common denominators in terms of providing the required computing power through remote servers, Cloud computing differs from other models in terms of providing this computing

power. The most important feature of Cloud computing is that it is a flexible server model that provides storage, infrastructure, and software that can be added and removed whenever required.

Details and volumes of various data such as social media, multimedia, and the Internet of Things (IoT) stored by companies and users are increasing day by day. Due to this increase, a huge data flow is produced in structured or unstructured formats. Today, some data used in various sectors in both public and social fields are created in record amounts and as a result of this, a widely known trend, Big data has emerged [5].

Big data can be characterized in three ways:

- Data are countless,
- Data cannot be classified in classical relational databases,
- Data are generated, stored, and processed very quickly.

Also Big data; changes science, engineering, health, finance, business, and therefore society [6].

Big data; the information technology industry has become an attractive topic in finance, business, academia, and scientific research. The amount of data produced in the digital world has increased enormously in recent years. According to the results of the research conducted by the International Data Corporation (IDC), 1.8 zettabytes of data were produced in 2011 and it is estimated that this amount of data will double every two years. In addition, it is predicted that the size of the data created will increase 300 times from 2004 to 2021. With the understanding of the importance of Big data, companies in the advertising industry, biomedical companies, healthcare industry, government agencies, and private organizations can take personal data that has reached enormous dimensions; has started to make many investments in collecting, bringing together, and sharing [7-10].

There are three key features to consider in Big data such as volume, velocity, and variety.

1. **Volume:** It shows the amount of data that has been moved and stored. Today, the amount of data is unpredictably large and constantly increasing. For example, 10 billion messages are sent daily on Facebook alone. Sensors, machines, cameras, etc. Mobil Electronic System Integration cameras, which are recording at any time, continuously generate data and expand the data volume.
2. **Velocity:** This concept expresses the speed in the production of data and the propagation of the produced data. Speed also includes the instant analysis and evaluation of the produced data without being stored. Nowadays, data is produced very fast, it spreads very fast, it must be analyzed very quickly.
3. **Variety:** Indicates the type of data. The diversity feature includes different types of data from photos, click counts, e-mails, voice recordings, videos to ECG data.

The purpose of Big data management is to discover the data value hidden in Big data. The above-mentioned features of Big data are used to reach value. Value is seeing patterns, insights, relationships in data, discovering information from data, and predicting the future. In order to provide these, data analysis should be done effectively.

The large amount of data stored without access and use does not provide any benefit on its own. Data should be processed and transformed into meaningful information in order to trigger defined actions. In some cases, the data obtained from various sensors are analyzed, the Big data system is fed and final reports are produced. At this point, the IoT has provided an opportunity for interaction between machines and people. Its operations in various sectors are provided by the interaction between machines and people (M2H) and machines and machines (M2M). The interrelation between these two technologies is shown in figure 1.

Fig. 1 The intersection of IoT and Big data

The use of the internet in the field of large data objects, data storage, has brought new challenges, such as integration and analysis. However, it has benefited technology far more from these challenges. In the next decade, IoT is predicted to cover the Internet industry's budget of approximately \$ 19 trillion. This situation leads to more research and development of both IoT and Big data fields.

When analyzing Big data, the following questions are sought:

1. How much data is produced,
2. How data is transformed into usable information,
3. How these data are used in making decisions,
4. How data is defined and managed.

Cost and complexity are increasing in the Big data model. Prominent factors for Big data are access, storage, and analytics. In this context, the purpose of Big data is to collect data and turn it into important information. Today, institutions and organizations organize data models to meet their Big data needs. Cloud computing technologies are used to meet the needs of Big data.

There is a need to design and develop effective and efficient systems to process large amounts of data that reach very high speeds and come from different sources. Since the various types of data used today are distributed, new technologies and techniques are being developed to storage, store, and process Big data. For example, Cloud computing-based technologies such as Hadoop MapReduce are being researched for the purpose of storing and processing Big data [11].

IoT enables people to share the burden of life with machines by transferring the right information to people on time. IoT establishes a feedback link in correcting the differences between the actual output and the desired output in order to make the right decision. This process is known as a feedback loop. This cycle contains information to be decided on instantaneous and future behaviors. In this respect, it provides an important gain in planning the ever-changing outputs and determining new behaviors.

In this study, the relationship between Big data and Cloud computing, and IoT is examined. In the second part of the study, Big data in Cloud computing is introduced. It also addresses the challenges for secure Big data storage and processing in the Cloud. In the third part, the use of Big data in Cloud computing is mentioned. The fourth chapter is about the Big data issue in IoT. In the fifth chapter, the study is concluded by giving results.

II. CLOUD COMPUTING AND BIG DATA

Nowadays, the usage area of the latest technological stationary vehicles has expanded and started to be no longer fixed. The reason for this can be explained by the fact that it becomes more important to access information independently from time and place and on the move. Cloud computing technology is one of the most innovative technologies that make it fast and easy to work in different locations, with different devices, and to access, transfer, and share data. Many definitions have been made for Cloud computing. However, Cloud computing, in its most common form, is a model of providing network connectivity to the pool where computer resources are shared, ready to use, and accessible at any time. Apart from this definition of the US National Institute of Technology and Standards (NIST), a broader definition has been introduced by the same institute as follows; Cloud computing can be transferred to a shared pool of adjustable computing resources (e.g. networks, servers, storage, applications, and services) that can be swiftly picked up and dropped with low management effort or service provider interaction, on-demand and with an appropriate It is a model that provides network access [12].

One of the factors that make Cloud computing important is that it allows more than one person to edit a file or document at the same time. For example, with the Google Documents Application, multiple users can have the opportunity to work on the same file or document (files with extensions such as .ppt, .xls, and .doc). In addition, since the storage process is not performed on a fixed computer, users can access the document from any place with an internet network using a computer or a mobile device and can also monitor the changes made on the document retrospectively. In this way, many obstacles are removed with Cloud computing, and the opportunity to work in groups for any text, visual, or project emerges. Another factor that makes Cloud technology important is low costs. A fee is paid to the service provider depending on the amount of service utilized, without the need for an excessive budget for software licenses, hardware infrastructure, experts, and managers. Microsoft 365, Google Apps, Zoho, Dropbox, etc. It is seen that services such as transportation, health, banking, education, and many other areas can be used. Usage types of Cloud computing services; public Cloud, private Cloud, community Cloud, and hybrid Cloud. However, it is also stated that there are some security problems and concerns regarding Cloud computing systems [13].

Especially two of the most important requirements for Cloud computing are massive computation and storage. Similarly, Big data needs these requirements just like Cloud Computing. However, Cloud computing has some advantages such as scalability and cost savings. Various businesses and some companies can easily adapt to Cloud technology thanks to these advantages. It also provides companies with enormous data processing power and storage capacity. Difficult jobs to do in traditional systems; Thanks to technologies used in Cloud computing such as virtualization, distributed storage, and processing, it can be realized more easily. However,

Cloud computing has its privacy problems. Some users are reluctant to transfer their sensitive or private data to the Cloud because they are not sure that it will be safe on the Cloud. In this respect, the Cloud needs secure and reliable storage of Big data within its structure, and it should allow outsourcing, multi-tenancy, and massive computation to create the processing system [14, 15].

Outsourcing: In recent years, firms and organizations keep their data in the Cloud by outsourcing to reduce capital and operational expenses. But this process of keeping in the Cloud using outsourcing; causes companies to lose physical control of customer data. Loss of control over data is one of the main causes of insecurity in Cloud usage. Insecurity can seriously damage the privacy of Cloud users and customers. These problems can be overcome with a secure computing environment and secure data storage. In addition, data kept in the Cloud must be verifiable for customers in terms of integrity and confidentiality [14].

Multi-tenancy: By using virtualization, the same Cloud platform can be shared by more than one user. By using some resource allocation principles, data belonging to different Cloud users can be stored in the same physical storage area. A malicious person can more easily access data that does not belong to him/her through illegal means through such an environment. In addition, some problems such as data and calculation violations may occur. For this reason, it is very important to develop new methods to overcome possible security and privacy risks.

Massive computation: Cloud computing has massive storage and compute-intensive capabilities. Therefore, classical methods are not enough to protect the privacy of individuals.

Distributed and parallel programming models used to serve the same purpose, but they had some functional differences. Although it has the same common denominators in terms of providing the required computing power through remote servers, Cloud computing differs from other models in terms of providing this computing power. The most important feature of Cloud computing is that it is a flexible server model that provides storage, infrastructure, and software that can be added or removed when desired. In this section, a basic introduction to Cloud computing is made, its features, service models, Cloud types are given, the advantages and disadvantages of Cloud computing are discussed.

Cloud computing has some basic characteristics. It takes a different approach from conventional server systems for computing and storage. These differences and features [16];

- Virtualizability,
- Self-allocation of resources when requested,
- Pay-as-you-go model,
- Virtual storage,
- Public Cloud media interfaces,
- Virtual networks,
- Dynamic sourcing,
- Virtual clusters,
- High level of availability and data recovery.

2.1. Cloud Computing Service Models

Cloud computing is a technology that provides services to the user using the pay-as-you-go idea. Cloud computing offers three main services listed below.

Infrastructure as a Service (IAAS): In this most basic Cloud service model, Cloud providers present their servers as physical or virtual machines. In the model of offering the infrastructure as a Cloud service, the customer can configure the processor, storage, network resource, and other basic information resources themselves and install the operating systems and applications on them. Although the customer does not have full management and control over the network structure, he has full control of the system at the operating system level and can manage some network components (Firewall-like).

Platform as a Service (PAAS): This Cloud provider includes uptime, flexible storage, queuing, databases, etc. Offers different resources such as. Responsibilities regarding configuration and implementation depend on the user. The service provider provides a platform for the customer to develop and run their application. This platform includes the environment in which the application will be developed and operated, as well as complementary services and necessary technological infrastructure. Apart from the application installed by the user, there is no control and management opportunity on the components that make up the platform infrastructure.

Software as a Service (SAAS): This service is the easiest to provide all the necessary settings and infrastructure for the platform, provided that IaaS and infrastructure are in place. It is the ability of users to work by accessing applications on Cloud computing systems via internet browsers without the need to make any installation on their systems to access applications. Customers do not manage or control components in the infrastructure such as network, server, operating system, and storage devices. They can only make adjustments specific to the applications they use.

Fig. 2 Cloud computing service models

Cloud computing service models are models created to authorize and access the Cloud provider's hardware and software resources according to the needs of the user. After determining what level of the model is needed, the service model that will serve the purpose in the most appropriate way can be rented through Cloud providers. As the authority and freedoms increase in the transition from software to infrastructure, pricing will increase in parallel. Although there are different opinions about the number, basically three types of service models can be mentioned. These service models are shown in figure 2.

2.2. Types of Cloud Computing

Cloud computing can be examined under three different types: public, private, and hybrid Cloud.

Public Cloud: In the public Cloud, the service provider makes applications and storage areas accessible over the internet. The most obvious example for the public Cloud is videos uploaded by YouTube users. Video resources and information are organized by the Cloud provider and are accessible to anyone with internet access.

Private Cloud: The user has the authority to provide his security on the storage and applications provided by the service provider in the private Cloud. In this way, the private information and data of the user are not accessible to third parties, including the Cloud provider. Google Drive service can be given as an example of this application. Personal information is stored by Google servers and can only be accessed and edited by the respective user.

Hybrid Cloud: In the hybrid Cloud, the Cloud provider does not secure the entire leased area, but instead only secures the private information of the user. In this system, private information is not stored in publicly accessible areas. For example, publicly shared files on Google Drive. The sharing of files to be hidden can be turned off and only certain files can be shared.

Fig. 3 Public, private, and hybrid Cloud environments

Figure 3 shows the public, private, and hybrid Cloud environments. Private Cloud environment is a type of Cloud that is not integrated with Cloud systems and contains computing and storage infrastructures. Public Cloud is a type of Cloud that provides computing and storage infrastructures from a Cloud provider. In some special cases, it may be desirable to use both private and public Cloud environments together. This approach is called hybrid Cloud. Hybrid Cloud, while having private Cloud infrastructure, provides an environment that can provide some computing and storage needs from public Cloud providers.

Cloud computing systems differ from conventional servers at some points and provide greater benefits. Some advantages of Cloud computing systems are:

- Low software and hardware cost,
- Enhanced performance,
- Instant update,
- Infinite storage capacity,
- Increased data security,
- Improved Integration between Operating Systems,
- Compatible file formats.

Since some of the required hardware and software are provided by the Cloud infrastructure, there is no additional charge for these systems. In addition, more performance can be achieved as the software and hardware support of these systems is provided by professional teams. In addition to these, a more secure environment is provided for personal servers.

Cloud computing systems have some disadvantages as they run on remote servers. Some of the disadvantages of Cloud computing systems are:

- Stable internet connection,
- Slow running at low connection speed,
- Vulnerabilities,
- System updates,
- Inexperienced Cloud operator,

Since Cloud computing systems provide access over the internet, a constant internet connection is required. In this respect, the regular operation of applications depends on high-speed internet access. The smallest security-related mistake can cause great damage. Also, when the Cloud infrastructure is updated, some applications on the infrastructure may not work properly.

In this section, a basic introduction to Cloud computing is made, its features, service models, Cloud types are given, the advantages and disadvantages of Cloud computing are discussed.

III. BIG DATA USAGE IN CLOUD COMPUTING

Today, advanced technologies such as satellite and sensor technologies, internet technologies, and mobile device innovations greatly increase the type and amount of data stored. Data formats can now be stored as images, audio, or video, while previously only text could be stored. In addition, the effective use of computer technologies in different areas is a reason that increases the data size. In most systems, when they need to store new data arises, if the available space is insufficient, either new hardware must be purchased or this hardware must be integrated into the server. However, with the flexible storage model provided by Cloud systems, new storage areas can be purchased within minutes and new data can be easily recorded. Cloud computing infrastructures are capable of storing and processing all kinds of data on it. A fixed storage area was allocated on conventional servers to store data, and when new data had to be stored, extra hardware was needed. Thanks to

Cloud computing systems, new storage areas can be reserved and used within minutes. Digitally stored data are analyzed in three parts according to their characteristics; structured, unstructured, and semi-structured.

Structured Data: Structured data usually refers to data that has a default type, size, or format. It is the data that has a certain structure, usually kept in the database. Structured data is primarily based on the data model. Structured data are easy to read, process, and understand [17]. Structured data can be easily classified, queried, and analyzed by a computer. For example, when a user enters data such as his / her name, address, the contact information on a website, it actually creates structured data. Sensor data, Weblog data, shopping data, financial data, input data, click data, game-related data are some of the structured data types.

Unstructured Data: Unstructured data refers to raw data. Most of the Big data exist in the form of data that has not been put into the classical formats specified in databases. Most of the data stored in today's world consist of unstructured data. Therefore, the processing of unstructured data is important due to the information potential it contains. Until recent years, unstructured data was a data type that was not emphasized and did not make any effort for its analysis. Today, the need to analyze large-scale data has required the compromise of structured data and the unstructured storage of data in order to be able to analyze the data quickly. Satellite images, seismic, atmospheric, and high energy physics data, photo and video data, radar or sonar data, special text data, social media data, mobile data, website content are some of the unstructured data types. Unstructured data has approximately 4 times the usage of structured data. Currently, approximately 85% of the stored data is unstructured data and the rate of increase increases much faster than structured data.

Semi-Structured Data: Semi-structured data arises when structured and unstructured data types are used together in the same record. Certain parameters have structured data properties, certain parameters have unstructured data properties. For example, since IP address or ID number are mandatory fields, they are stored as structured data, while non-mandatory fields such as personal information are stored as unstructured data. Semi-structured data can be enumerated as unstructured data or classified separately.

Today, there are various applications that provide data storage services commercially. Basically, all data storage systems can store Cloud data, but traditional methods are insufficient for many Cloud applications due to reasons such as the distributed operation of Cloud systems and the availability of data on different physical servers. Some storage systems are more advantageous than others with additional system approaches. For successful data management, it is important to improve concepts such as data transport and storage. Data storage is divided into three separate types.

Local Storage: Data that can be found on local devices and can be accessed directly. HDD, USB stick, CD, DVD, etc. The data on it can be shown as examples.

Centralized Storage: It is a storage type where data is stored and shared in a single center. This data can be accessed by many devices over the internet or network. The use of central data storage servers may cause bottlenecks in access to data, congestion, inefficient operation, and problems such as blocking of access by errors originating from a single center.

Distributed Storage: The data center is managed by a DBMS (Database Management System). But data is stored and copied in many different locations, rather than in one center. This allows data sharing to be more efficient and easy. Distributed data can be accessed at the local and global levels. Since there is not a single center in the distributed system, even if a center is not operating, access to the system continues as copies of the data will be found in different locations. This structure is the most uninterrupted structure to access data. Cloud computing is an example of distributed data.

Cloud databases are divided into two parts, relational and NoSQL databases.

Relational Database: Relational Cloud databases are systems in which the classical relational database model is adapted to the Cloud. The data is stored in related tables and has a predefined chart. The most important disadvantage of relational database systems is that queries run slowly due to the need to define the relationships. Relational database systems can be recommended for Cloud systems that do not work heavily with data. Oracle, IBM DB2, Microsoft SQL Server, PostgreSQL, and MySQL are examples of relational Cloud databases [18].

NoSQL Database: NoSQL databases use a different approach to data storage compared to conventional relational databases. It provides file-based data storage via key-value. Therefore, file access, file reading, and writing operations can be performed very quickly. NoSQL databases need to process and use raw data. Although additional operations bring extra costs, it can run faster than relational database approaches due to its working on memory and being highly parallelized [19].

Cloud computing and Big data are two inseparable concepts that complement each other. Often users with multiple data set process Big data with distributed queries and can use commodity computation to return final sets as needed. Hadoop is an open-source project developed to process Big data in a reliable, scalable, and

distributed manner on ordinary server clusters. Cloud computing provides an engine based on the use of Hadoop. The use of Cloud computing in Big data is shown in figure 4. As shown in figure 4, unstructured Big data from the Web and the Cloud is stored in a distributed and fault-tolerant database. This stored data is processed by a programming model with distributed and parallel algorithms. The goal of data visualization in the decision-making phase is to visualize the analytical results evaluated with different graphics [5].

Fig. 4 Cloud computing in Big data

Big data uses Cloud computing-based distributed storage system instead of using the storage space of a computer or a fixed electronic device. Cloud-based applications have increased rapidly with the use of virtualized technologies and have had significant effects on the evaluation of Big data. For this reason, Cloud computing is used as a service model apart from the processing and calculation of Big data. Table 1 shows a comparison of some Cloud providers for Big data.

Table. 1 Cloud providers for Big data

Property	Cloud provider for Big data			
	Google	Microsoft	Amazon	Cloudera
Big data storage	Google cloud services	Azure	S3	-
MapReduce	AppEngine	Hadoop on Azure	Flexible MapReduce (Hadoop)	MapReduce YARN
Big data analytics	BigQuery	Hadoop on Azure	Flexible MapReduce (Hadoop)	Flexible MapReduce (Hadoop)
Relational database	Cloud SQL	SQL Azure	MySQL or Oracle	MySQL, Oracle, PostgreSQL
NoSQL database	AppEngine data store	Table storage	DynamoDB	Apache Accumulo
Stream processing	Search API	StreamInsight	Prepackaged nothing	Apache Spark
Machine learning	Prediction API	Hadoop + Mahout	Hadoop + Mahout	Hadoop + Oryx
Data import	Network	Network	Network	Network
Data sources	Several sample datasets	Windows Azure marketplace	Open datasets	Open datasets
Availability	Some services in private beta	Some services in private beta	Open production	Industries

Big data uses Cloud computing-based distributed storage system instead of using the storage space of a computer or a fixed electronic device. Cloud-based applications have increased rapidly with the use of virtualized technologies and have had significant effects on the evaluation of Big data. For this reason, Cloud computing is used as a service model apart from the processing and calculation of Big data. Table 1 shows a comparison of some Cloud providers for Big data.

In a study, the complexity and diversity of data types were examined while calculating the processing power needed to analyze data sets [20]. Cloud computing plays an active role in meeting the need for data storage while analyzing Big data. Cloud-based technologies must be successful in order to keep up with these emerging technologies. The biggest reason for this is that the process becomes more complex when processing Big data simultaneously [21]. MapReduce is one of the best examples of this situation. Moreover, MapReduce is good at

handling large data sets stored in parallel. However, through Cluster computing, good performance can be obtained in some distributed system environments such as storage, network communication, and computing power [22]. In another study, the issue of providing a good environment for data growth in cluster calculation was discussed [23]. However, in a study, they were examined that the lack of data availability would have a costly result. The reasons for this are seen as the inherent weaknesses of incorrectly used analytical methods and methods that produce a wrong and burdensome decision [24].

Another essential part of Cloud computing architecture is Database Management Systems (DBMS). The most important benefit is that it facilitates the transition of the outdated infrastructures of various applications used by some institutions or companies to new architectures using Cloud infrastructure. Various institutions, companies, and businesses must overcome storage and Big data processing problems in order to keep up with technology. However, this often leads to various risks and errors. In this respect, researchers have discussed many studies in which Cloud computing technologies are used with Big data [25].

IV. THE INTERNET OF THINGS AND BIG DATA

The IoT is a system formed by devices that communicate, connect, share information, and form a smart network over the Internet through various communication protocols. “Thing” in IoT represents a person or a device assigned an IP address. “Thing” obtains and transfers data from the internet environment thanks to embedded technologies. Also, “Thing” helps the data to communicate with the external environment and internal situations to make decisions.

IoT, smart devices, mobile devices, sensors, etc. developed the concept of the internet network that enables the communication between objects. It provides effective communication between all elements of the architecture such as the architecture of the IoT, objects, gates, network infrastructure, Cloud infrastructure.

Using IoT and Cloud computing together provides many advantages:

- Different applications are used in a Cloud infrastructure to process and analyze data quickly.
- Cloud computing should be used to minimize the large pressure caused by the increasing amount of data and to transfer data. Cloud computing provides the performance required to store and run Big data.
- While IoT enables large amounts of data to be produced, Cloud providers transfer data. This situation proves that these two concepts are complementary to each other.
- Many IoT developers can store and easily access data remotely using the Cloud platform.
- Analysis and monitoring of IoT devices are made easier thanks to Cloud computing.
- IoT devices using common APIs can make important security updates over the Cloud when any security problem occurs in the infrastructure.

The development of the IoT concept and technology changes the social structure by making life easier, increasing living standards, increasing efficiency, and contributing to economies. As with all good things, it can cause seriously bad results if not taken care of.

Fig. 5 Interrelationship between IoT, Cloud computing, and Big data technologies

According to the information given on IoT, Cloud computing, and Big data technologies, there is a mutual relationship between these three concepts. Cloud computing plays a common business role for Big data, where IoT is used as a data source and Big data is the analytics platform of the data.

In the upcoming years, it is predicted that almost all of the IoT data will be kept on Cloud Computing platforms. The reasons behind this information are as follows;

- The enormous amount of data generated in IoT feeds Big data systems.
- Reducing the complexity of data blends, one of the criteria for maximizing the benefits of IoT.

According to the two important issues given above, it seems that there is a need to adopt Cloud-based systems for IoT and Big data.

The interrelationship between IoT, Cloud computing, and Big data technologies is shown in figure 5.

Fig. 6 IoT data collection tools

In IoT technology, many network sensors are placed in machines and devices with different functions in the real world. With these sensors placed in the devices, different data such as geographical, environmental, logistic, and astronomical data are obtained [26]. In Figure 6, various vehicles such as transportation vehicles, mobile devices, public facilities, and household appliances used to collect data in IoT are shown.

Big data generated by IoT devices appear to contain different characteristics when compared to general Big data. This is because the data collected are of different types. General characteristics of such data: heterogeneity, diversity, unstructuredness, noise, and high redundancy. Today, the majority of Big data is not IoT data. However, it is predicted that the number of data collection tools used in IoT will reach 1 trillion by 2025. This suggests that a significant portion of Big data in the future may be IoT data. According to the report obtained from Intel, the data produced by IoT devices include three features of the Big data approach.

1. Data chunks created by terminals,
2. Data generated by IoT devices are generally unstructured or semi-structured,
3. IoT data is useful only when analyzed.

Recently, the data capacity that IoT devices can process has lagged behind the obtained data size. It is predicted that IoT will accelerate with the development of new Big data technologies. The success of IoT technology depends on the effective integration of Big data and Cloud computing. The deployment of large-scale IoT devices in different areas and regions will enable many cities to move into the Big data age. The rapid development of Big data requires IoT applications to also adopt Big data. Additionally, the idea that IoT and Big data technologies are interrelated and must be developed together has been accepted around the world. Widely located IoT devices cause the data volume to increase. Thanks to Big data, development, and application possibilities have increased. In addition, with the application of Big data techniques to IoT technology, IoT research, and development of business models have gained speed [26].

V. CONCLUSION

Among the developing technologies, there is no other innovation that grows and changes faster than the internet. It is thought that IoT, which is described as a technology revolution, will accelerate this progress and change many lifestyles in a short time. In this respect, the IoT develops with the principle of connecting with everybody, every object, and always from anywhere and finds an important place in internet technologies.

Cloud computing is one of the popular computing and storage approaches in recent years. It is important for the users to get more professional service that they provide calculations and data storage systems through remote servers and by using professional teams, such as maintenance, system design, backup, etc. that users need to perform within their systems. The management and processing of data on Cloud computing infrastructures have become a necessity due to the increasing demand for Cloud computing. The processing of data on remote servers by computing units on remote servers and delivering the results to the user accurately

and precisely is one of the issues that need to be considered and solved in terms of Cloud computing. Therefore, data-driven operations on Cloud infrastructures are one of the most widely studied topics.

Nowadays, new technologies such as Cloud computing, IoT, social networks, and data analytics enable large-scale data collection. Data are geographically diverse in nature. Researchers working in the field of Big data are interested in the information obtained from the integrated data collection located in these places. The main purpose of this study is to guide the researchers who will work on Big data in Cloud computing and IoT and provide general information on this subject. In the study, Cloud computing and Big data, Big data usage in Cloud computing, and finally IoT and Big data were mentioned. The increasing number of studies on these issues signals that these areas will gain more importance in the future.

REFERENCES

- [1] M. Sağır, İ. Türkeri, Bilişim Teknolojilerinin Konaklama İşletmelerinde İnsan Kaynakları Uygulamalarında Kullanımı: Konya İli Örneği, *Kastamonu University Journal Of Economics & Administrative Sciences Faculty*, 8, 2015.
- [2] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, M. Zaharia, Above The Clouds: A Berkeley View Of Cloud Computing, *Dept. Electrical Eng. and Comput. Sciences, University of California, Berkeley, Rep. UCB/EECS*, 28, 13, 2009.
- [3] P. Mell, T. Grance, The Nist Definition Of Cloud Computing, *Nist Special Publication* 800-1452011, 2011.
- [4] Z. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, I. Brandic, Cloud Computing And Emerging It Platforms: Vision, Hype, And Reality For Delivering Computing As The 5th Utility, *Future Generation Computing Systems*, 25, 599-616, 2009.
- [5] R. L. Villars, C. W. Olofson, M. Eastwood, Big data: What It Is and Why You Should Care, *White Paper, IDC Analyze the Future, MA, USA*, 2011.
- [6] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, S. U. Khan, The rise of “Big data” on Cloud computing: Review and open research issues, *Information Systems*, 47, 98-115, 2015.
- [7] J. C. McCune, Data, data everywhere, *Management Review*, 87, 10, 10-12, 1998.
- [8] C. Tankard, Big data security, *Network Security*, 7, 5-8, 2012.
- [9] J. Gantz, D. Reinsel, The Digital Universe in 2020: Big data, Bigger Digital Shadows, and Biggest Growth in the Far East–United States, *IDC Country Brief, IDC Analyze the Future, Framingham, MA, USA*, 2013.
- [10] J. Bamford, The NSA Is Building the Country’s Biggest Spy Center, *Wired*, 2012.
- [11] A. Mehmood, I. Natgunanathan, Y. Xiang, G. Hua, S. Guo, Protection of Big data Privacy, *IEEE Access*, 4, 1821-1834, 2016.
- [12] M. Saritaş, T. Üner, N. Üner, Eğitimdeki Yenilikçi Teknolojiler: Bulut Teknolojisi. *Eğitim Ve Öğretim Araştırmaları Dergisi* 2, 3, 192-201, 2013.
- [13] B. Ege, Bulut Bilişim. *Bilim Ve Teknik Dergisi*, 541, 12-15, 2012.
- [14] A. Mehmood, I. Natgunanathan, Y. Xiang, G. Hua, S. Guo, Protection of Big data Privacy, *IEEE Access*, 4, 1821-1834, 2016.
- [15] Z. Xiao, Y. Xiao, Security and Privacy in Cloud Computing, *IEEE Communications Surveys & Tutorials*, 15, 2, 843-859, 2013.
- [16] R. Buyya, J. Broberg, A. M. Goscinski, Cloud Computing Principles And Paradigms (1st Ed.), *Wiley Publishing*, 2011.
- [17] J. Hurwitz, A. Nugent, F. Halper, M. Kaufman, Big data For Dummies (1st Ed.), *For Dummies*, 2013.
- [18] M. Grossniklaus, The Case For Object Databases In Cloud Data Management, *Proceedings Of The Third Conference On Objects And Databases*, 25-39, Frankfurt, Germany, 2010.
- [19] A. B. M. Moniruzzaman, S. A. Hossain, Nosql Database: New Era Of Databases For Big data Analytics - Classification, Characteristics And Comparison, *International Journal Of Database Theory And Application*, 6, 1-14, 2013.
- [20] D. Talia, Clouds for Scalable Big Data Analytics, *Computer*, 46, 5, 98-101, 2013.
- [21] C. Ji, Y. Li, W. Qiu, U. Awada, K. Li, Big Data Processing in Cloud Computing Environments, *12th International Symposium on Pervasive Systems, Algorithms and Networks (ISPAN)*, San Marcos, TX, USA, 17-23, 2012.
- [22] J. Dean, S. Ghemawat, MapReduce: Simplified Data Processing on Large Clusters, *Communications of the ACM*, 51, 1, 107-113, 2008.
- [23] D. Bollier, C. M. Firestone, The Promise and Peril of Big Data, *The Aspen Institute, Communications and Society Program*, ISBN: 0-89843-516-1, Washington, DC, USA, 2010.
- [24] H. E. Miller, Big Data in Cloud Computing: A Taxonomy of Risks, *Information Research*, 18, 1, 1-19, 2013.
- [25] S. Aluru, Y. Simmhan, A Special Issue of Journal of Parallel and Distributed Computing: Scalable Systems for Big Data Management and Analytics, *Journal of Parallel and Distributed Computing*, 73, 6, 896, 2015.
- [26] M. Chen, S. Mao, Y. Liu, Big Data: A survey, *Mobile Networks and Applications*, 19, 2, 171-209, 2014.