International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 17 (2017) pp. 6970-6982

Cloud Computing and Big Data is there a Relation between the Two: A Study

Nabeel Zanoon[1], Abdullah Al-Haj[2], Sufian M Khwaldeh[3]

[1] Department of Applied science, Al- Balqa Applied University/Aqaba, Jordan.

[2] Faculty of Information Technology, University of Jordan/Aqaba, Jordan.
[3]Business Information Technology (BIT) Department, The University of Jordan/Aqaba, Jordan.

[1]Orcid ID: 0000-0003-0581-206X

Abstract

Communicating by using information technology in various ways produces big amounts of data. Such data requires processing and storage. The cloud is an online storage model where data is stored on multiple virtual servers. Big data processing represents a new challenge in computing, especially in cloud computing. Data processing involves data acquisition, storage and analysis. In this respect, there are many questions including, what is the relationship between big data and cloud computing? And how is big data processed in cloud computing? The answer to these questions will be discussed in this paper, where the big data and cloud computing will be studied, in addition to getting acquainted with the relationship between them in terms of safety and challenges. We have suggested a term for big data, and a model that illustrates the relationship between big data and cloud computing.

Keywords: big data, Hadoop, Cloud, MapReduce, resources, Five (Vs).

INTRODUCTION

Data is the raw material for information before sorting, arranging and processing. It cannot be used in its primary form prior to processing. Information represents data after processing and analysis [1]. The technology has been developed and used in all aspects of life, increasing the demand for storing and processing more data. As a result, several systems have been developed including cloud computing that support big data. While big data is responsible for data storage and processing, the cloud provides a reliable, accessible, and scalable environment for big data systems to function [2]. Big data is defined as the quantity of digital data produced from different sources of technology for example, sensors, digitizers, scanners, numerical modeling, mobile phones, Internet, videos, e-mails and social networks. The data types include texts, geometries, images, videos, sounds and combinations of each. Such data can be directly or indirectly related to geospatial information [3].

Cloud computing refers to on-demand computer resources and systems available across the network that can provide a number of integrated computing services without local resources to facilitate user access. These resources include data storage capacity, backup and self-synchronization [4]. Most IT Infrastructure computing consist of services that are provided and delivered through public centers and servers based on them. Here, clouds appear as individual access points for the computing needs of the consumer. It is generally expected for commercial offers to meet the QoS requirements of customers or consumers, and typically include service level agreements (SLAs) [5]. They are an online storage model where data are stored on multiple virtual servers, rather than being hosted on a specific server, and are usually provided by a third party. The hosting companies, which have advanced data centers, rent spaces that are stored in a cloud to their customers in line with their needs [6].

The expert Erik Brynjolfsson likened big data to a microscope which was invented in old times, and by which scientists were able to identify and measure things they had never imagined before at the cell level. This is similar to big data which is a modern day microscope by which you are able to see things and measure data that you never have expected. [7] The statistics shown in [8] show that data growth in cloud environments is increasing exponentially and rapidly with the increasing number of internet users around the world. With this rapid growth, the question that comes to mind is how can these vast amounts of data be stored in cloud environments? We need storage technology that meets the needs of rapid data growth on the cloud and we need storage technology with low cost, high reliability and high capability.

The relationship between big data and the cloud computing is based on integration in that the cloud represents the storehouse and the big data represents the product that will be stored in the storehouse, since it is not possible to create storehouses without storing any product in them. The

traditional databases known as 'relational' are no longer sufficient to process multiple-source data. For example, how can these traditional methods deal with data such as record of transactions, customer behavior, mobile phone and GPS navigation, and others. Here comes the role of cloud computing. At this point, a relationship between big data and the cloud will arise. In this paper, the relationship between them will be discussed, in addition to the obstacles and challenges that this relationship may encounter.

BIG DATA

Big data comes and is composed through electronics operations from multiple sources. It requires proper processing power and high capabilities for analysis [9]. The importance of big data lies in the analytical use which can help generate an informed decision to provide better and faster services [10].

The term big data is called on the huge amount of high-speed big data of different types; this data cannot be processed and stored in regular computers. The main characteristics of big data, called V's 5 As in Figure 1 , can be summed up in the fact that the issue is not only about the volume of data, other dimensions of big data, known as 'five Vs', are as follows:

1. Volume: It represents the amount of data produced from multiple sources which show the huge data in numbers by zeta bytes. The volume is most evident dimension in what concerns to big data.
2. Variety: It represents data types, with, increasing the number of Internet users everywhere, smart phones and social networks users, the familiar form of data has changed from structured data in databases to unstructured data that includes a large number of formats such as images, audio and video clips, SMS, and GPS data [11].
3. Velocity: It represents the speed of data frequency from different sources, that is, the speed of data production such as Twitter and Facebook. The huge increase in data volume and their frequency dictates the need for a system that ensures super-speed data analysis.
4. Veracity: It represents the quality of the data, it shows the accuracy of the data and the confidence in the data content. The quality of the data captured can vary greatly, which affects the accuracy of analysis. Although there is wide agreement on the potential value of big data, the data is almost worthless if it is not accurate [12].
5. Value: It represents the value of big data, i.e. it shows the importance of data after analysis. This is due to the fact that the data on its own is almost worthless. The value lies in careful analysis of the exact data, the information and ideas it provides. The value is the final stage that comes after processing volume, velocity, variety, contrast, validity and visualization [13]

Figure 1 . Characteristics Of Big Data

There have been numerous revisions to the big data until they reached (7 v) [14]. In this paper, based on the relationship between cloud computing and big data, will suggest a new term, virtualization, which virtually represents The data structure is by default. The virtualization of big data is a process that focuses on creating virtual structures for big data systems. Virtualization technology is the key technology used to help cloud computing handle large amounts of data flexibly and facilitate the process of managing big data. The virtual storage technology will be studied in section (6.2).

The type and nature of the data

Data in general is a set of values that are in the form of numbers, letters, symbols and other forms where they are concerned with a particular idea and subject .The data does not make sense without analysis, and is, therefore, compiled for use. It represents input, while information is output after processing, i.e. data is entered into the system first, then processed until it comes out in the form of useful information that has a clear meaning and against which decisions are made.

Big data comes from multiple sources including sensors and free texts such as social media, unstructured data, metadata and other geospatial data collected from web logs, GPS, medical devices, etc. [15]. The big data is gathered from different sources ,so it is in several forms, including:

1. Structured data: It is the organized data in the form of tables or databases to be processed.
2. Unstructured data: It represents the biggest proportion of data; it is the data that people generate daily as texts, images, videos, messages, log records, click-streams,etc.

3. Semi-structured data: or multi-structured ,It is regarded a kind of structured data but not designed in tables or databases, for example XML documents or JSON [16].

Difference between traditional data and big data

In general, the data in the world of technology is a set of letters, words, numbers, symbols or images, but with the evolution of multitasking technology tools the data has become different in content and source[17]. In light of this, big data emerged which differs from traditional data. Differences between traditional data and big data are shown in Table1:

Table 1 Comparison between traditional and big data[18]

	Traditional Data	**Big Data**
Volume	MB and GB	PBs And ZBs
Data Generation Rate	Long periods of time	More rapid
Data Type	Structure	Sim-Structure , Unstructured
Data sources	Centralized	multiple sources, and distributed
Data Store	RDBMS	HDFS, No SQL

CLOUD COMPUTING

it is a term that refers to on-demand computer resources and systems that can provide a number of integrated computer services without being bound by local resources to facilitate user access. These resources include data storage, backup and self-synchronization, as well as software processing and scheduling tasks [19]. Cloud computing is a shared resource system that can offer a variety of online services such as virtual server storage, and applications and licensing for desktop applications. By leveraging common resources, cloud computing is able to achieve expansion and provide volume [20].

Characteristics of cloud computing.

That cloud computing is one of the distributed systems that represents a sophisticated model. NIST has identified important aspects of the cloud, as it shortened the concept of cloud computing in five characteristics as follows:

- **On-demand self-service**: Cloud services provide computer resources such as storage and processing as needed and without any human intervention.
- **Broad network access**: cloud computing resources are accessible over the network , mobile and smart devices even sensors can access computing resources on the cloud.
- **Resource Pooling**: Cloud platform users share a vast array of computing resources; users can determine the nature of resources and the geographic location they prefer but cannot determine the exact physical location of these resources.
- **Rapid Elasticity**: Resources from storage media, network, processing units and applications are always available and can be increased or decreased in an almost instantaneous fashion, allowing for high scalability to ensure optimal use of resources.
- **Measured service**: Cloud systems can measure the processes and consumption of resources as well as surveillance, control and reporting in a completely transparent manner [21] [22] [23].

Cloud computing service models.

Cloud computing types are classified on the basis of two models: cloud computing service models and cloud computing deployment models as in Figure 2:

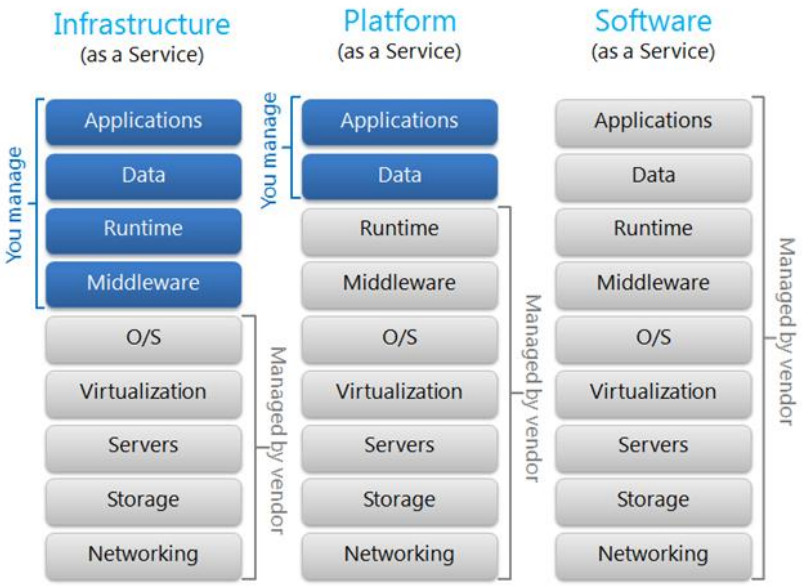

Figure 2 . Cloud computing Models[24]

- **Software as a service (SAAS)**: Cloud service providers provide various software applications to users who can use them without installing them on their computer. The user is not responsible for anything other than adjusting the settings and customizing the service as appropriate to his needs. SAAS helps big-data clients to perform data.
- **Platform as a service (PAAS)**: Cloud service providers provide platforms, tools and other services to users, where the cloud service provider manages everything else, including the operating system and middleware., with resources that enable you to deliver everything from simple cloud-based apps to sophisticated.
- **Infrastructure as a service (IAAS)**: Cloud service providers provide infrastructure such as storage, computing capacity, etc. is a form of cloud computing that provides virtualized computing resources over the Internet , In an IaaS model, a third-party provider hosts hardware, software, servers, storage and other infrastructure components on behalf of its users [25][26].
- **DaaS** : It is the alternative cloud computing model, as it differs from traditional models like (SAAS, IAAS, PAAS) in providing data to users through the network, as data is considered the value of this model [27] in conjunction with cloud computing based on solving some of the challenges in managing a huge amount of data. For these reasons, DaaS is closely related to big data whose technologies must be utilized [28]. DaaS provides highly efficient methods of data distribution and processing. DaaS is closely related to SaaS (storage as a service) and SaaS (software as a service) which can be combined with one of these models or both of them [29].

Cloud Storage

The concept of cloud storage is the same as that of storing files on a remote server to retrieve them from multiple devices at any time we need. Cloud storage is basically a system that allows storing data on the internet. Examples of this system are Google Drive, Dropbox, etc. [30]. Cloud storage , it is stored data while cloud computing is used to complete the specified digital tasks. In most cloud computing applications, data is sent to remote processors over the internet for complete operation, and the resulting data is sent back [31] where you can use the program interface but the bulk of the program activity is remote instead of the computer. Cloud computing is usually more useful for companies than individuals in most cloud computing applications [32]. It is a set of technologies hosting a cloud, and giving resources to hire and consume on demand over the internet on the basis of pay-per-user. Among the best known cloud computing providers are Amazon, Google, and Microsoft [33].

The increasing amount of data requires equipment to store them. The cloud provides storage units, making it easier to navigate without having to carry physical storage equipment while on the move. Limited storage space is a real concern for both consumers and businesses [34]. The storage of data in the cloud is done through a cloud service provider (CSP) in a set of cloud servers where the user interacts with the cloud servers via CSP to access or retrieve its data. Since they no longer have their data locally, it is important to assure users that their data is properly stored and maintained. This means that users should be provided with security means so that they can ensure that their stored data is consistently maintained even without local copies [35].

DATABASE MANAGEMENT SYSTEM.

Data is collected in the form of an organized structure called the database which is the food of any information system. Data huge amount is the major component of the cloud infrastructure. Data can be shared among many tenants. As a result, data management in particular is a key aspect of storage in the cloud [36]. Data in the cloud is distributed across multiple sites and may contain certain privileges and authentic information. It is therefore very important to ensure that data consistency, scalability and security are maintained. In order to address these issues and many other important data issues, there is a need for a database management system for cloud data [37].The database management system shows the mechanism of storage and retrieval of user data with maximum efficiency, taking into consideration the appropriate security policies [38]. The database management system always provides data independence. No change is made to the storage mechanism and shapes without modifying the entire application. There are several types of database organization, relational database, flat database, object oriented database, hierarchical database [39].

Structured data work with relational databases while non-relational databases work with semi-structured data [40]. The non-relational database is known as (No-SQL), which is a non-relational database. This category of databases has been steadily adopted in recent years with the emergence of big data applications, since the purpose of designing non-relational databases is to overcome the limitations of relational databases in dealing with big data demands. Big data refers to data that is growing and moving very rapidly and is very diverse in the structure of traditional technologies to deal with [41] .The difference between relational data and (No-SQL) is that the relational data model consists of a set of interconnected tables through keys, while (No-SQL) is increasingly considered a viable alternative to relational databases, especially for big data applications [42].There are several database management systems in the computed cloud that provide storage and analysis for both relational (SQL) and non-relational (No-SQL) [43]. But No- SQL Big data systems are designed to take advantage of new cloud computing structures, which makes big operational data much easier to manage, cheaper and faster to implement [44].

THE RELATIONSHIP BETWEEN THE CLOUD AND BIG DATA

Cloud computing is a trend in the development of technology, as the development of technology has led to the rapid development of electronic information society. This leads to the phenomenon of big data and the rapid increase in big data is a problem that may face the development of electronic information society [45]. Cloud computing and big data go together, as big data is concerned with storage capacity in the cloud system, cloud computing uses huge computing and storage resources. Thus, by providing big data application with computing capability, big data stimulate and accelerate the development of cloud computing. The distributed storage technology in environmental computing helps to manage big data [46].

Cloud computing and big data are complementary to each other. Rapid growth in big data is regarded a problem. Clouds are evolving and providing solutions for the appropriate environment of big data [47] while traditional storage cannot meet the requirements for dealing with big data, in addition to the need for data exchange between various distributed storage locations. Cloud computing provides solutions and addresses problems with big data [48]. The cloud computing environment is expanding to be able to absorb big amounts of data as it follows the policy of data splitting, that is, to store data in more than one location or availability area. Cloud computing environments are built for general purpose workloads and resource pooling is used to provide flexibility on demand. Therefore, the cloud computing environment seems to be well suited for big data [49].

Big data processing and storage require expansion as the cloud provides expansion through virtual machines and helps big data evolve and become accessible. This is a consistent relationship between them. Google, IBM, Amazon and Microsoft are examples of the success in using big data in the cloud environment [50].In order for the cloud environment to fit with big data the cloud computing environment must be modified to suit data and cloud work together. Many changes are needed to be made on the cloud: CPUs to handle big data and others [51].

The Models between the cloud and big data

The most common models for providing big data analytics solution on clouds are PaaS and SaaS. IaaS is usually not used for high-level data analytics applications but mainly to handle the storage and computing needs of data, Cloud computing models can help accelerate the potential for scalable analytics solutions [52]Cloud computing is a member of distributed computing family that provides resources in the form of user services such as (SaaS), infrastructure like (IaaS) and a platform as service like (PaaS), but with the advent of big data, the cloud computing model is gradually moving to big database service including (AaaS,BDaaS) known as (DaaS) database as a service which means that database services are available for applications that are deployed in any implementation environment [53]. BDaaS is a form of service similar to software as a service or infrastructure as a service. Huge data as a service often relies on cloud storage to maintain continuous data access to the enterprise that owns the information and the provider it works with [54] and is considered to be hosted in the cloud. Similar types of services include (SaaS) or service-based infrastructure, (IaaS) where big, specific data is used as service options to help businesses

handle big data. It provides a lot of value for companies today [55], where a combination of all of these has been made to create the ultimate solution for companies moving forward , DBaaS is still a relatively hazy term, but it mostly refers to a host of outsourced services and functions related to Big Data handling in a cloud-based environment[56] models for cloud-based big data analytics , envision two types of services for Cloud analytics , Analytics as a Service (AaaS), where analytics is provided to clients on demand and they can pick the solutions required for their purposes; and Model as a Service (MaaS) where models are offered as building blocks for analytics solutions ,More recently, terms such as Analytics as a Service (AaaS) and Big Data as a Service (BDaaS) are becoming popular. They comprise services for data analysis similarly as IaaS offers computing resources. However, these analytics services still lack well defined contracts since it may be difficult to measure quality and reliability of results and input data, provide promises on execution times[57]

Virtual Machine (VM) between the cloud and big data

Virtual Machine (VM) is a software application that simulates a virtual computing environment that can run the operating system (OS) and its associated applications with multiple virtual machines installed on a single machine. Distributed systems, network computing and parallel programming are not new as one of the key enabling factors of the cloud is virtual technology. By using virtualization technology, one virtual machine can often host multiple virtual machines [58]. Virtualization technology provides the ability to reduce workload in virtual metering devices and unify them into one physical server. Consolidation has become particularly effective after the adoption of multi-core CPUs in computing environments, where many virtual machines can be allocated to a single physical node that improves resource utilization and reduces power consumption compared to multi-node setup [59].

Virtualization technology is the best platform for big data as well as traditional applications. Assuming big data applications simplifies managing your big data infrastructure, providing faster results and is more cost-effective [60]. The role of infrastructure, whether real or virtual, is to support applications. This includes important traditional business applications, modern cloud, and mobile and big data applications. Virtualized big data applications, such as (Hadoop), provide many benefits that cannot be accessed on physical infrastructure but helps simplify big data management [61]. Today's virtual data constitute a wide range of sources including multidimensional stores, web and data services, XML documents, analytical devices, and indoor and outdoor applications. Data stores (NoSQL) are a modern source type where they support virtual data [62].

Big data and cloud computing point to the convergence of technologies and trends that make IT infrastructure and their applications more dynamic, more modular and more expendable. Currently, the virtual platform building technology is only in the primary stage, which is mainly based on cloud data center integration technology [63]. Cloud computing and big data projects rely heavily on virtualization. Virtual data is the only way to access and improve heterogeneous environments, such as environments used in big data projects. The cloud computing model allows users to have a default data center that can access data sets that were not previously available by using a shared (API) for disparate data sets [64].

Big data Security in cloud computing

Big data and cloud are among the most important stages of IT development. Information privacy and security are one of the most important issues for the cloud because of its open environment with very limited user control [65]. Security and privacy affect big data storage and processing because there is a huge use of third party services and the infrastructure used to host important data or to perform operations as growing data and application growth bring challenges [66].

A solution is provided for the security services and the level of confidence needed through the third party services within the cloud. The data is stored in a central location known as the cloud storage server, where the data is processed somewhere on the servers, so the client has confidence in the service provider as well as data security. The service level agreement must be standardized to gain trust between service providers and customer [67]. The security of cloud client data varies in protection requirements. Customers require protection of their data only through basic logical access controls, while intellectual property, structured or classified data are confidential and require advanced security controls including encryption, data hiding, login, logging, etc. [68].

The Service Level Agreement (SLA) reflects a service level contract between the user and the service provider. It is one of the ways to enhance the security level, where different levels and complexities of security are determined depending on services to better understand security policies for a cloud consumer, and to protect data [69]. There are rules with service level agreements to protect the data, capacity, scalability, security, privacy, and availability of issues such as data storage and data growth [70]. The technologies available to secure big data, such as registry entry, encryption, and trap detection are essential. In many organizations, big data analytics can be used to detect and prevent malicious hackers and advanced threats. The security of big data in cloud computing is necessary because of the following issues:

1. Protection of big data from malicious intruders and advanced threats.
2. Knowledge about how cloud service providers securely maintain huge disk space and erase existing big data.

3. Lack of standards for checking and reporting big data in the public cloud [71].

Challenges in big data and cloud computing

The security challenges in cloud computing environments fall under several levels: the network level which includes dealing with network protocols and network security such as distributed nodes, distributed data, and communications between the nodes; authentication level where the user handles encryption / decryption techniques, authentication methods such as contract administrative rights, authentication of applications and nodes, and logging entry; the data level which is concerned with data integrity and availability as well as data protection and data distribution [72]. Cloud computing follows the policy of shared resources, where the privacy of data is very important because it faces some challenges like integrity, authorized access, and availability of (backup / replication). Data integrity ensures that data is not corrupted or tampered with during communication. Authorized access prevents data from infiltration attacks while backups and replicas allow access to data efficiently even in case of technical error or disaster in some cloud location [73].

Big data face some challenges as they can be classified into groups: data sets, processing and management challenges. When dealing with big amounts of data we face challenges such as volume, variety, velocity and verification which are also known as 5V of big data [74]. Also, in the field of computer networks the cost of communications is a major concern compared to the cost of processing the same data, as the challenge is to reduce the cost of communications to the minimum while meeting the requirements of storage and additional data from the general cloud to handle big data [75]. Among the factors and challenges that affect the processing of big data in a timely manner is The bandwidth and latency [76]. where several challenges can be summarized in the relationship between big data and cloud computing.

- Data Storage: The storage of big data through traditional storage is problematic because hard drives often fail, data protection mechanisms are not effective, and the speed of big data requires storage systems in order to expand rapidly, which is difficult to achieve with conventional storage systems. Cloud storage services offer almost unlimited storage with a great deal of error tolerance, which offers potential solutions to address the challenges of big data storage.
- Variety of data: Big data naturally grow, increase and vary, which is the result of the growth of almost unlimited sources of data. This growth leads to the heterogeneous nature of big data. Generally speaking, data from multiple sources of different types and representations are highly interrelated. They have incompatible shapes and are inconsistent. A user can store data in structured, semi-structured, or unstructured formats. Structured data format is suitable for today's database systems, while semi-structured data formats are only fairly suitable. Unstructured data is inappropriate because it contains a complex format that is difficult to represent in rows and columns.
- Data transfer: The data goes through several stages: data collection, input, processing, and output. Big data transfer is a challenge, so data compression techniques need to be reduced to reduce the volume, where data volume is a hindrance to transfer speed. It also affects the cost, while cloud computing provides distributed storage resources and data transfer on high-speed lines, reducing costs through virtual resources and resource use at user's request.
- Privacy and data ownership: The cloud environment is an open environment and the user's role in monitoring is limited. Privacy and security are an important challenge for big data. Big data and cloud computing come together in practice. According to (IDC) estimates, by 2020, around 40% of global data will be accessed by cloud computing. Cloud computing provides strong storage, calculation and distribution capability to support big data processing. As such, there is a strong demand to investigate the privacy of information and security challenges in both cloud computing and big data.

What Is Big Data's Relationship To The Cloud?

How does the cloud computing environment correspond to big data? The answer to this question reflects the relationship between them. This is done through the cloud computing features to handle big data, the resources provided by cloud computing, the resource service to provide service to many users where the various physical and virtual resources are automatically set and reset upon request. Cloud computing has access from anywhere to data resources that are spread all over the world by using a (public) cloud to allow those sources faster access to storage. The nature of big data is generated by technologies and locations worldwide, so the cloud resource service provides and helps in the collection and storage of big amounts of data resulting from the use of technologies.

The cloud computing structure can expand the solid equipment to accommodate small and big data volumes. The cloud can expand to handle big amounts of data by dividing the data into parts, automatically done in IAAS. Expanding the environment is a big data requirement. Cloud computing has the advantage of helping to reduce costs by paying for the value of the resources used, which helps to develop big data. Flexibility is also regarded a requirement for big data. When we need more storage for data the cloud platform can dynamically expand to meet proper storage needs when we would like to handle a large number of virtual machines in a single time period. For error tolerance, the cloud helps to handle big data in the extraction and storage process. Error tolerance helps SLAs, as well as QOS levels. Service level agreements specify different rules for regulating availability of cloud service.

Big companies such as Yahoo, Google, Facebook, and others offer web-based services, and the amount of data they routinely collect through online user interactions has overwhelmed traditional IT capabilities. Therefore, the development of basic infrastructure components has to be developed. Apache Hadoop has been introduced as a realistic benchmark for managing big amounts of unstructured data. Apache Hadoop is open platform distributed software for storing and processing data. By using Hadoop, you can reliably store big amounts (pet bytes) on tens of thousands of servers while effectively scaling performance in terms of cost. MapReduce is based on the distribution of a data set between multiple servers, partial results are then reassembled.

Big data are characterized by diversity, i.e. they are of different types and therefore require big data. ETL technology, therefore, deals with data diversity, as ETL represents several functions such as extraction, conversion, and loading. These three functions are combined into one tool to pull data from one database and place it in another database. It helps to convert databases from one form to another.

Big data relies on data integrity to be effective. If you store big data at the local level, it will take a huge amount of work to manually merge all data to manage it. The cloud can do this work for the user, offering one site to store and manage all commercial data. In this way, you can get one source of the truth, without exhausting your time and resources to manually merge the data.

Cloud computing offers features and benefits to big data through ease of use, access to resources, low cost in resource utilization on supply and demand, and reduces the use of solid equipment used to handle big data. Both big data and the cloud aim to increase the value of a company while reducing investment costs. The cloud reduces the cost of managing local software, while big data reduces investment costs by encouraging more prudent business decisions. It seems only natural that these two concepts together provide greater value to companies.

Any system in technology must pass through several main stages. The computer system follows the input, processing and output model. Input is done through devices and then processed through the CPU. Thus, the results of the information are produced. In the relationship between the data and cloud computing, the data is stored on external and remote storage units. On the other hand, in the computer system, the data is stored internally or locally. Therefore, the relationship between the data and cloud computing represents the input, processing and output model as in Figure 3. The big data is entered through devices such as the mouse, cellular devices and other smart devices. Processing is carried out through the tools and techniques used by the cloud computing in providing service, and the outputs are the results, it represents the value of data after processing.

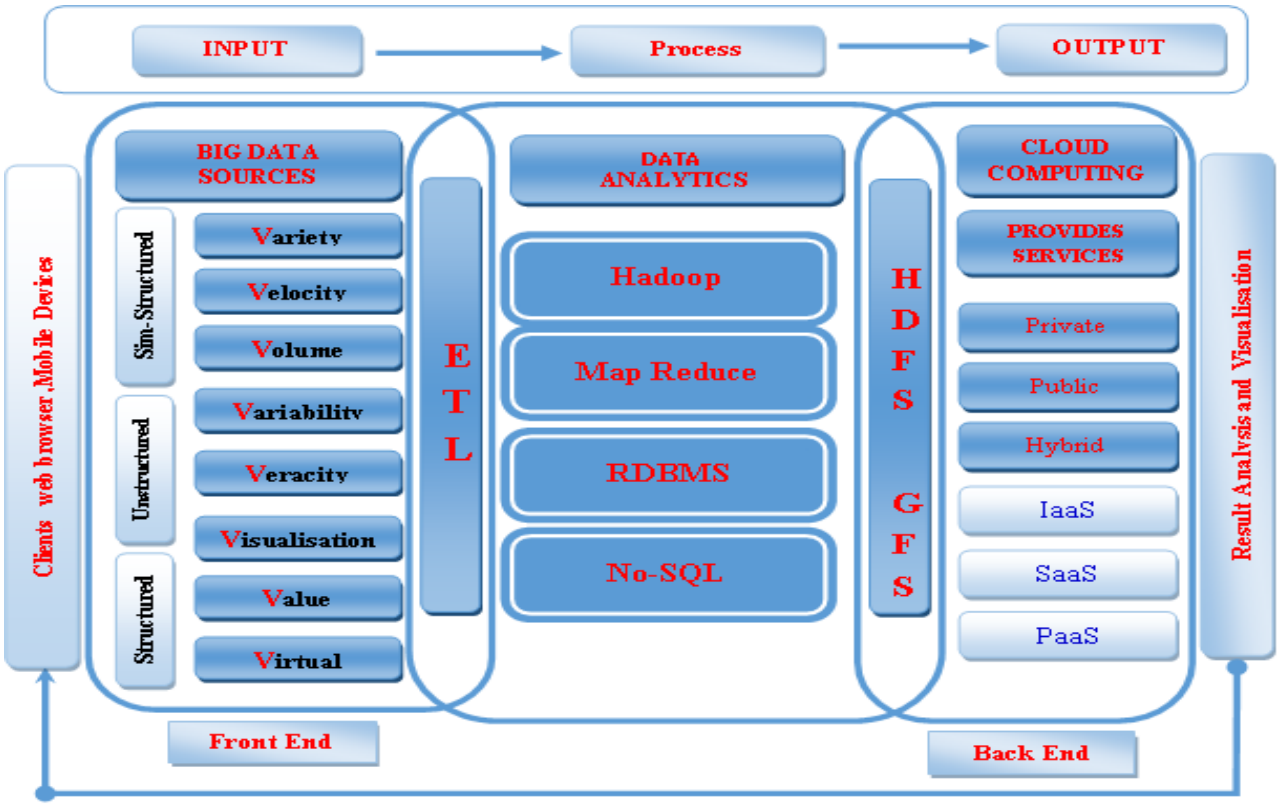

Figure 3 . A Model Showing The Relationship Between Big Data And Cloud Computing

The input and output model defines input, output and processing tasks required to convert input to output. Inputs represent the flow of data and raw materials. The processing step includes all tasks required to transform inputs. The output is data flowing from the transformation process.

Common factor between cloud computing and big data.

The internet of things represents the new concept of the Internet network, which enables communication between several parties to communicate together, and these parties include smart devices, mobile devices, sensors and other [77] where it is considered effective communication between all elements of architecture so that it can Rapidly deploy applications, process and analyze data quickly to make decisions as quickly as possible. The architecture represents several systems: objects, gates, network infrastructure, cloud infrastructure. [78] Internet objects can benefit from the scalability and performance of cloud computing infrastructure. In fact, Internet applications produce large amounts of data and consist of multiple computer components upon request. [79]

The Internet of Things (IOT) is going to generate a massive amount of data and this in turn puts a huge strain on Internet Infrastructure. As a result, this forces companies to find solutions to minimize the pressure and solve their problem of transferring large amounts of data. . [80] But cloud computing has played a major role in IT, by migrating its data operations to the cloud. Many cloud providers can allow your data to either be transmitted over your traditional Internet connection or via a dedicated direct link. [81] That the real purpose of cloud computing and Internet of things increase efficiency in daily tasks and both have a complementary relationship. The Internet of things generates huge amounts of data, and cloud computing provides a pathway for these data to navigate [82]. By storing data in the cloud, most companies find that it is possible to access large amounts of big data through the cloud. [83] And internet of things are all parts of a continuum. Difficult to think of Internet things without thinking about the cloud, it is difficult to think of the cloud without thinking about the Big data analyzes. Which generates a lot of data, this data is stored in the cloud computing, cloud computing is the only technology suitable for filtering, analysis, storage and access to IoT and other data in ways that are useful, as these data constitute large quantities must be analyzed, Objects is a common factor between the erased cloud and big data.

Common points between big data and the cloud

- The cloud computing environment consists of several user terminals and service provider. The big data comes on both sides, as the user collects the data and, in dealing with the technology tools, the big data is produced. The role of the service provider is to save, store and process the big data at the user's request, so cloud computing represents the big data infrastructure. The service provider must ensure that users have on-demand resources or otherwise access their data, systems and applications on a regular basis and is available throughout the service without any interruption.
- Data, whether small or big, require storage, processing and security, but the volume and capacity of data requirements differ in accordance with the volume of the data, so cloud computing must provide storage, processing and security requirements for big data in its environment. The cloud environment is scalable and uses sophisticated high-end data management techniques and security policies as the service provider protects and manages data.
- Cloud computing provides security, depending not on data volume but the availability of security and protection for small and big data. The service provider guarantees complete confidentiality of user data of all kinds and only allows access to authorized users. Therefore, identity management and access control must be provided for information resources and service resources, according to user needs. The user can connect to the network in these resources through a simple software interface that simplifies and ignores many internal details and processes.
- Cloud computing saves the cost of storing and processing data to the user through the availability of geographically dispersed servers and the availability of virtual server technology. The service provider must ensure that the devices and equipment are sufficiently available, and restricted by an integrated and documented entry system for reference when needed. Cloud computing offers the use of high-level applications and software, regardless of the efficiency of the devices the user uses, because it depends on the strength of the network servers and not on the personal resources of your device, regardless of the efficiency of the user's device he can benefit from the cloud service.
- Cloud computing is considered as a distributed system; it is distributed over a geographical distance. An example is the general cloud, where resources are distributed everywhere. This makes it easier for the user to speed up access to the data. Thus, cloud computing is based on solving the problem of geographical divergence between devices and resources. It also enables multiple users to share a single database and share resources such as web pages, files and other physical resources.
- Cloud computing is characterized by continuity, i.e. the ability to withstand failure by providing resources even in the absence of defect in the components. The nature of the cloud is that it is geographically distributed, so there is a high probability of errors. These events increase the need for failure tolerance techniques to achieve reliability.

All these points represent the relationship between big data and cloud computing, as it shows the important requirements for the continuous increase in the growth of big data and provides the appropriate environment to deal with big data.

Table 1: Compatibility between big data and cloud computing in terms of characteristics.

Characteristics Big Data	Concept	Characteristics Cloud Computing
Velocity **Visualisation**	Data Rates Data Representation	• Network Bandwidth • Gigabit rates today • Broad network access • Anywhere access - public cloud • Resource pooling:
Variety **Veracity**	Data Type Data Sources Trustworthiness Of The Data	• Cloud data management ,No-SQL Databases • Anywhere Access - Public Cloud • Mapreduce/Hadoop Is A Data Processing And Analytics Technology • SLA , QoS • ETL technology
Volume	Size Data	• Scalability - Elasticity According To Demand • Cost : Pay-As-You-Go Based On Usage. Reduced cost Reduced cost • Resource Pooling: • On-Demand Self-Service
Virtual	Physical infrastructure data	• Virtual Machine (VM) Is A Software Application • Resource Pooling: Physical Infrastructure
Value	Data Analysis Results, Reports	• OLAP • OLTP

CONCLUSION

Big data and cloud computing have been studied from several important aspects, and we have concluded that the relationship between them is complementary. Big data and cloud computing constitute an integrated model in the world of distributed network technology. The development of big data and their requirements is a factor that motivates service providers in the cloud for continuous development, because the relationship between them is based on the product, the storage and processing as a common factor. Big data represents the product and the cloud represents the container. The big data is concerned with the capacities of cloud computing. On the other hand, cloud computing is interested in the type and source of big data. Depending on the relationship between them, a model was prepared to show the relationship between them as in Figure 3. Compatibility between them is summarized in Table 2. Cloud computing represents an environment of flexible distributed resources that uses high techniques in the processing and management of data and yet reduces the cost. All these characteristics show that cloud computing has an integrated relationship with big data. Both are moving towards rapid progress to keep pace with progress in technology requirements and users.

REFERENCES

[1] Charmaz, K., and A. Bryant. "The SAGE Handbook of Grounded Theory: Paperback Edition." (2010).

[2] Neves, Pedro Caldeira, Bradley Schmerl, Jorge Bernardino, and Javier Cámara. "Big Data in Cloud Computing: features and issues."

[3] Lopez, Xavier. "Big data and advanced spatial analytics." In Proceedings of the 3rd International Conference on Computing for Geospatial Research and Applications, p. 5. ACM, 2012.

[4] Kshetri, Nir. "Cloud computing in developing economies." Computer 43, no. 10 (2010): 47-55.

[5] https://en.wikipedia.org/wiki/Cloud_computing

[6] Klous, Sander, and Nart Wielaard. We are Big Data: The Future of the Information Society. Springer, 2016.

[7] https://www.internetworldstats.com/stats.htm

[8] https://www.ibm.com/big-data/us/en/ Bello-Orgaz G, Jung JJ, Camacho D. Social big data: Recent achievements and new challenges. Information Fusion. 2016 Mar 31;28:45-59.

[9] Boyd, D., & Crawford, K. (2011, September). Six provocations for big data. In A decade in internet time: Symposium on the dynamics of the internet and society (Vol. 21). Oxford: Oxford Internet Institute.

[10] SHAN, Y. C., Chao, L. V., ZHANG, Q. Y., & TIAN, X. Y. (2017). Research on Mechanism of Early Warning of Health Management Based on Cloud Computing and Big Data. In Proceedings of the 23rd International Conference on Industrial Engineering and Engineering Management 2016 (pp. 291-294). Atlantis Press, Paris.

[11] Parvin Ahmadi Doval Amiri and Mina Rahbari Gavgani, 2016. A Review on Relationship and Challenges of Cloud Computing And Big Data: Methods of Analysis and Data Transfer. *Asian Journal of Information Technology, 15: 2516-2525*

[12] Chen, Min, et al. Big data: related technologies, challenges and future prospects. Heidelberg: Springer, 2014.

[13] Demchenko, Yuri, et al. "Big security for big data: Addressing security challenges for the big data infrastructure." Workshop on Secure Data Management. Springer, Cham, 2013.

[14] McAfee, Andrew, and Erik Brynjolfsson. "Big data: the management revolution." Harvard business review 90.10 (2012): 60-68.

[15] Liebowitz, J. (Ed.). (2014). *Bursting the big data bubble: The case for intuition-based decision making*. CRC Press.

[16] Sremack, Joe. Big Data Forensics–Learning Hadoop Investigations. Packt Publishing Ltd, 2015.

[17] Franks, Bill. Taming the big data tidal wave: Finding opportunities in huge data streams with advanced analytics. Vol. 49. John Wiley & Sons, 2012.

[18] Furht, Borko, and Flavio Villanustre. Big Data Technologies and Applications, Chapter 1, Springer, 2016.

[19] Calheiros, Rodrigo N., et al. "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms." Software: Practice and experience 41.1 (2011): 23-50.

[20] R.Subhulakshmi , S.Suryagandhi , R.Mathubala , P.Sumathi, An evaluation on Cloud Computing Research Challenges and Its Novel Tools, International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST) Volume 2, Special Issue 19, October 2016.

[21] Fonseca, N., & Boutaba, R. (2015). Cloud services, networking, and management. John Wiley & Sons.

[22] https://www.ibm.com/blogs/cloud-computing/2014/01/cloud-computing-defined-characteristics-service-levels/

[23] Zhang, Q., Cheng, L., & Boutaba, R. (2010). Cloud computing: state-of-the-art and research challenges. Journal of internet services and applications, 1(1), 7-18.

[24] Ahmed, F. F. (2015). Comparative Analysis for Cloud Based e-learning. Procedia Computer Science, 65, 368-376.

[25] Vacca, J. R. (Ed.). (2016). Cloud Computing Security: Foundations and Challenges. CRC Press. ch-15.

[26] https://support.rackspace.com/how-to/understanding-the-cloud-computing-stack-saas-paas-iaas/

[27] Terzo, O., Ruiu, P., Bucci, E., & Xhafa, F. (2013, July). Data as a service (DaaS) for sharing and processing of large data collections in the cloud. In Complex, Intelligent, and Software Intensive Systems (CISIS), 2013 Seventh International Conference on (pp. 475-480). IEEE.

[28] Motahari-Nezhad, H. R., Stephenson, B., & Singhal, S. (2009). Outsourcing business to cloud computing services: Opportunities and challenges. *IEEE Internet Computing*, *10*(4), 1-17.

[29] Rajesh Saturi, Data as a Service (Daas) in Cloud Computing [Data-As-A-Service in the Age of Data] Data as a Service Daas in Cloud Computing, Global Journal of Computer Science and Technology Cloud & Distributed Volume 12 Issue 11 ,2012.

[30] http://www.gadgetreview.com/cloud-storage-vs-cloud-computing-which-are-you-using.

[31] http://info.cloudcarib.com/blog/cloud-storage-vs.-cloud-computing-whats-the-difference.

[32] Hamlen, K., Kantarcioglu, M., Khan, L., & Thuraisingham, B. (2012). Security issues for cloud computing. *Optimizing Information Security and*

Advancing Privacy Assurance: New Technologies: New Technologies, *150*.

[33] Al-Roomi, M., Al-Ebrahim, S., Buqrais, S., & Ahmad, I. (2013). Cloud computing pricing models: a survey. *International Journal of Grid and Distributed Computing*, *6*(5), 93-106.

[34] Zhao, L., Sakr, S., Liu, A., & Bouguettaya, A. (2014). Cloud data management(pp. 1-189). Springer.

[35] Rai, R., Sahoo, G., & Mehfuz, S. (2013). Securing software as a service model of cloud computing: Issues and solutions. arXiv preprint arXiv:1309.2426.

[36] Alam, M., & Shakil, K. A. (2013). Cloud database management system architecture. UACEE International Journal of Computer Science and its Applications, 3(1), 27-31.

[37] Malet, B., & Pietzuch, P. (2010, November). Resource allocation across multiple cloud data centre's. In Proceedings of the 8th International Workshop on Middleware for Grids, Clouds and e-Science (p. 5). ACM.

[38] http://www.kciti.edu/wp-content/uploads/2017/07/dbms_tutorial.pdf, 2015 by Tutorials Point.

[39] Oppel, A. (2010). Databases demystified. McGraw-Hill Education Group.

[40] Singh, M. P. (Ed.). (2004). The practical handbook of internet computing. CRC press.

[41] Oh, G., Seo, C., Mayuram, R., Kee, Y. S., & Lee, S. W. (2016, June). SHARE interface in flash storage for relational and NoSQL databases. In Proceedings of the 2016 International Conference on Management of Data (pp. 343-354). ACM.

[42] Raj, P. (Ed.). (2014). Handbook of research on cloud infrastructures for big data analytics. IGI Global.

[43] Moniruzzaman, A. B. M., & Hossain, S. A. (2013). Nosql database: New era of databases for big data analytics-classification, characteristics and comparison. arXiv preprint arXiv:1307.0191.

[44] Jain, V. K. (2017). Big Data and Hadoop. KHANNA PUBLISHING.

[45] Goda, K., & Kitsuregawa, M. (2012). The history of storage systems. *Proceedings of the IEEE*, *100*(Special Centennial Issue), 1433-1440.

[46] MALLICK, Pradeep Kumar (ed.). *Research Advances in the Integration of Big Data and Smart Computing*. IGI Global, 2015.

[47] TIAN, Wenhong Dr; ZHAO, Yong Dr. *Optimized cloud resource management and scheduling: theories and practices*. Morgan Kaufmann, 2014.

[48] HU, Han, et al. Toward scalable systems for big data analytics: A technology tutorial. IEEE access, 2014, 2: 652-687.

[49] Wei-Dong Zhu, Manav Gupta, Ven Kumar, Sujatha Perepa, Arvind Sathi, Craig Statchuk ,Building Big Data and Analytics Solutions in the Cloud, IBM Redbooks - 2014 .

[50] Neves, P. C., Schmerl, B., Bernardino, J., & Cámara, J. Big Data in Cloud Computing: features and issues, Conference: International Conference on Internet of Things and Big Data,2016.

[51] Wei-Dong Zhu, Manav Gupta, Ven Kumar, Sujatha Perepa, Arvind Sathi, Craig Statchuk ,Building Big Data and Analytics Solutions in the Cloud, IBM Redbooks - 2014 .

[52] Zomaya, A. Y., & Sakr, S. (2017). *Handbook of Big Data Technologies*. Springer.

[53] AHSON, Syed A.; ILYAS, Mohammad (ed.). Cloud computing and software services: theory and techniques. CRC Press, 2010.

[54] http://searchcio.techtarget.com/definition/big-data-as-a-service-bdaas,

[55] PACHGHARE, V. K, CLOUD COMPUTING, PHI Learning Pvt. Ltd.,2015

[56] https://www.maximizer.com/blog/entering-the-age-of-big-data-as-a-service/.

[57] Assunção, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A., & Buyya, R. (2015). Big Data computing and clouds: Trends and future directions. *Journal of Parallel and Distributed Computing*, *79*, 3-15.

[58] Wikipedia contributors. "Virtual machine." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 4 Aug. 2017. Web.15 Aug. 2017.

[59] Marvin Zelkowitz, Advances in Computers82, Academic Press,2011.

[60] https://www.vmware.com/asean/solutions/big-data.html

[61] Kumar, Manish, Applied Big Data Analytics in Operations Management, IGI Global,2016.

[62] Mike Ferguson, Data Virtualization – Flexible Technology for the Agile Enterprise, Intelligent Business Strategies www.sas.com,2014.

[63] Ji, C., Li, Y., Qiu, W., Jin, Y., Xu, Y., Awada, U., ... & Qu, W. (2012). Big data processing: Big challenges and opportunities. *Journal of Interconnection Networks*, *13*(03n04), 1250009.

[64] Big Data and the Rise of Cloud Services, http://www.ingrammicroadvisor.com/data-center/big-data-and-the-rise-of-cloud-services

[65] Xiao, Z., & Xiao, Y. (2013). Security and privacy in cloud computing. IEEE Communications Surveys & Tutorials, 15(2), 843-859.

[66] KANNAN, Rajkumar (ed.). Managing and Processing Big Data in Cloud Computing. IGI Global, 2016.

[67] Meetei, M. Z., & Goel, A. (2012, October). Security issues in cloud computing. In Biomedical Engineering and Informatics (BMEI), 2012 5th International Conference on (pp. 1321-1325). IEEE.

[68] Sun, Y., Zhang, J., Xiong, Y., & Zhu, G. (2014). Data security and privacy in cloud computing. International Journal of Distributed Sensor Networks, 10(7), 190903.

[69] Hussain, S. A., Fatima, M., Saeed, A., Raza, I., & Shahzad, R. K. (2017). Multilevel classification of security concerns in cloud computing. Applied Computing and Informatics, 13(1), 57-65.

[70] RADHA, K., et al. Service Level Agreements in Cloud Computing and Big Data. *International Journal of Electrical and Computer Engineering*, 2015, 5.1: 158.

[71] Chaturvedi, A., & Lone, F. A. (2017). Analysis of Big Data Security Schemes for Detection and Prevention from Intruder Attacks in Cloud Computing. International Journal of Computer Applications, 158(5).

[72] Venkatesh, H., Perur, S. D., & Jalihal, N. (2015). A study on use of big data in cloud computing environment. *Int. J. Comput. Sci. Inf. Technol.(IJCSIT)*, *6*(3), 2076-2078.

[73] Olshannikova, E., Ometov, A., Koucheryavy, Y., & Olsson, T. (2015). Visualizing Big Data with augmented and virtual reality: challenges and research agenda. Journal of Big Data, 2(1), 22.

[74] Khan, N., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Mahmoud Ali, W. K., Alam, M., ... & Gani, A. (2014). Big data: survey, technologies, opportunities, and challenges. The Scientific World Journal, 2014.

[75] Suthaharan, S. (2014). Big data classification: Problems and challenges in network intrusion prediction with machine learning. *ACM SIGMETRICS Performance Evaluation Review*, *41*(4), 70-73.

[76] Parvin Ahmadi Doval Amiri and Mina Rahbari Gavgani, 2016. A Review on Relationship and Challenges of Cloud Computing And Big Data: Methods of Analysis and Data Transfer. *Asian Journal of Information Technology, 15: 2516-2525*

[77] Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. Computer networks, 54(15), 2787-2805.

[78] Ahmed Banafa,Securing the Internet of Things (IoT), https://ahmedbanafa.blogspot.com/2015/09/securing-internet-of-things-iot.html

[79] Talia, D. (2014). Towards internet intelligent services based on cloud computing and multi-agents. In Advances onto the Internet of Things (pp. 271-283). Springer International Publishing.

[80] The Role Of Cloud Computing In The IOT Revolution, https://pinaclsolutions.com/blog/2017/cloud-computing-and-iot

[81] Manav Gupta, Mandy Chessell, Gopal Indurkhya, Heather Kreger, Anshu Kak, and Christine Ouyang , How IBM leads in building big data analytics solutions in the cloud, https://www.ibm.com/developerworks/cloud/library/cl-ibm-leads-building-big-data-analytics-solutions-cloud-trs/index.html

[82] The Role Of Cloud Computing In The IOT Revolution, https://pinaclsolutions.com/blog/2017/cloud-computing-and-iot

[83] Hwang, K., Dongarra, J., & Fox, G. C. (2013). *Distributed and cloud computing: from parallel processing to the internet of things*. Morgan Kaufmann..