Cloud Storage Architecture: Research Challenges and Opportunities

Anwar Ghani 1, Afzal Badshah 1, Saeedullah Jan 2, Abdulrahman A. Alshdadi 3, and Ali Daud 4

1 Department of Computer Science & Software Engineering, International Islamic University Islamabad, 44000, Pakistan (e-mail: anwar.ghani@iiu.edu.pk, afzal.phdcs120@iiu.edu.pk)
2 Department of Computer Science & IT, University of Malakand, Chakdara, 18800, Pakistan (e-mail:saeedullah@uom.edu.pk)
3 Department of Information Systems and Technology, College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia (email: alshdadi@uj.edu.sa)
4 Department of Information Systems and Technology, College of Computer Science and Engineering, University of Jeddah, Saudi Arabia (e-mail: ali_msdb@hotmail.com)

Corresponding author: Anwar Ghani (e-mail: anwar.ghani@iiu.edu.pk).

ABSTRACT

From home appliances to industrial enterprises, the Information and Communication Technology (ICT) industry is revolutionizing the world. We are witnessing the emergence of new technologies (e.g., Cloud computing, Fog computing, Internet of Things (IoT), Artificial Intelligence (AI) and Block-chain) which proves the growing use of ICT (e.g., business, education, health and home appliances), resulting in massive data generation. It is expected that more than 175 ZB data will be processed annually by 75 billion devices by 2025. The 5G technology (i.e. mobile communication technology) dramatically increases network speed, enabling users to upload ultra high definition videos in real-time, will generate a massive stream of big data. Furthermore, smart devices, having artificial intelligence, will act like a human being (e.g. a self-driving vehicle etc) on the network, will also generate big data. This sudden shift and massive data generation created serious challenges in storing and managing heterogeneous data at such a large scale. This article presents a state-of-the-art review of the issues and challenges involved in storing heterogeneous big data, their countermeasures (i.e., from security and management perspectives), and future opportunities of cloud storage. These challenges are reviewed in detail and new dynamics for researchers in the field of cloud storage are discovered.

INDEX TERMS Internet of Things, Cloud Computing, Storage Architecture, Cloud Security, Cloud Data Management

I. INTRODUCTION

The recent advances and development in smart technology is getting more attention and attraction, resulting in a massive data generation. The 75 billion devices forecast is a big number; even ten times greater than the whole world population [1]. These devices will increase the annual size of the global data-sphere up to 175 ZB [2], [3]. Another report states, as shown in Fig. 1 that more than 331 billion dollars will be invested in cloud up to 2023 [2]. This development not only requires special infrastructural improvements but also special techniques to process and store the incoming data [4]. Furthermore, integrating Artificial Intelligence (AI) in smart devices makes the data storage process more complicated. Fig. 1 shows the devices and revenue forecast of cloud computing.
Cloud storage provides virtual storage resources without buying the physical resources. Remote accessibility (i.e., access from everywhere and anytime) is the core of cloud storage. Leading cloud provider (i.e., Apple iCloud, Microsoft OneDrive, and Google Drive etc) are providing remote services to their users.

Today more embedded devices joined the Internet to monitor and connect everything (e.g., traffic facilities, buildings, environment, and lakes), enlarging the size of the data generation [5], [6], [7], [8], [9]. As the data on the Internet is increasing day by day, therefore, analyzing and storing it through traditional data management method is a great challenge [10], [11]. However, researchers are struggling to design new kinds of databases based on NoSQL for handling unstructured data at such a large scale [12], [13], [14]. There are many proposals in the literature for a universal storage architecture which supports multiple data models at the same time and can store big heterogeneous data in the cloud environment [15], [16], [17], [18].

Fig. 2 shows the structure of cloud storage.

With the advent of technology, computing requirements of organizations grew exponentially prompting the organization to incorporate more computing and storage resources [19]. Setting up systems at such large scale require more efforts and heavy investments prompting the enterprise customers to outsource their computing and storage resources [20], [21], [22], [23]. The users have no full control over the computing resources available through cloud over the Internet [24], [25]. Storage in the cloud is becoming a hot research venue today because new applications are data intensive which doubles storage capacity requirement as well as data usage every year. It prompted some commercial organizations to work for another cloud service called as “on demand storage”. Currently, the storage providers are fixated towards other aspects related to cloud storage like cost issues, performance issues and incorporating multiple storage [26]. Fig. 2 shows the structure of master and data node in cloud storage architecture.

The models of data centers in cloud computing are based on “design-for-failure” principle. Provisioning of global storage services require, cloud storage must use scalable, cheaper and purposed built solutions. Such solutions may include different hardware like, servers, networking equipment, and storage systems. It should use standard delivery models on massive economies of scale. “off-the-shelf” products designed for the traditional IT market may not be suitable to use in cloud data centers since they are not only expensive but also they do not meet the specific requirements of cloud data center environment.

This study explores the cloud storage architecture its challenges and possible solutions. Additionally the cloud storage future and opportunities. Cloud storage issues include but not
limited to Security, Confidentiality, Data Dynamics, Integrity, Data Access, Data Segregation, Authentication and Authorization, Data Breaches, Backup Problem and vulnerabilities in Virtualization.

Rest of this article is structured as follows: Section II provides an insight into the issues related to cloud storage and their countermeasures. Section II discusses the future opportunities of cloud storage. Finally, section IV concludes the article with the key findings and future directions.

II. CLOUD STORAGE CHALLENGES AND POSSIBLE SOLUTIONS

Storage in a cloud is a crucial part of the Infrastructure as a Service (IaaS). The lack of proper storage management in cloud environment, may lead to severe consequences [27]. Cloud storage related issues have been categorized as data security and data management issues [28], [29]. This paper focuses on issues related to these two categories and a review of possible solutions to such issues. Some of the points may overlap both categories, however, this distinction may help in understanding the challenges faced by cloud storage providers and tenants. The following subsections elaborate these issues and their counter measures.

A. DATA SECURITY ISSUES

Data security is an important requirement from tenant as a right. Secure services attract users to store their data in a cloud. Companies providing the cloud storage services are searching for techniques that can control access to cloud data and improve security. With increase in size of the data, there is also an increase in data attacks and interceptions. The cloud computing provides storage services as a vitalized environment where a user has no control over the data [30]. In such situation, a user may ask questions like “where exactly is my data located?”, “what happen if I delete my data?” and “is the deleted data really deleted?”.

Many solutions to data security in cloud can be found in literature. Authors in [30] divided the security solutions into four layers (i.e. availability, authentication, confidentiality and integrity). They argued that if confidentiality is achieved, it automatically ensures integrity. However, this sub section is dedicated to a more elaborate study of the issues related to data security. A recent study exploring data security and privacy in cloud storage [27] pointed out the three main reasons based on the features of cloud computing independent of the technology being used on the server. It includes outsourcing and multi-tenancy.

A Time Stamp Authority (TSA) and Public Key Infrastructure (PKI) technologies are introduced into the cloud storage in cloud computing. Further, the study also concluded that the current data security solutions are not sufficient to protect the data and therefore, new approaches and solutions are required.

FIGURE 3. Master and data node in cloud storage architecture.

FIGURE 4. Overall flow of this study.
storage system for authentication and security with minimum cost and less system overhead. Trusted time stamp helps in audit and recording [31, 19, 32, 33]. The three points considered are User Identification, Time Stamping and User Verification through cloud storage system. The use of PKI improves security whereas authentication is done through directory services. The use of a time stamp provides security services like audit and evidences with a very minimum overhead. TSA also performs data management and optimization in cloud storage system. The workload is increased by TSA and client communication and the verification of users’ operations. As during the communication process no certificate is used so extra overhead is not involved. The operation commands are converted into time stamp and sent to TSA server, which communicates with directory server and verify certificate. On validating the certificate, a time stamp is issued. The corresponding time stamp is then sent to the cloud and further operations may be performed. The cloud system stores the time stamps and operations record. The operations may be queries, downloads and uploads etc. The basic approaches used in designing data security techniques are shown in Table 1. Furthermore, AI, 5G, IoT and block chain are improving the privacy and security [34].

1) Confidentiality Issues
Cloud storage is a collection of storage servers on which multiple customers’ data is stored, which makes privacy a major concern. The fundamental requirement for confidentiality of the information stored or processed in the cloud is the guaranteed protection of confidential or sensitive information. Based on the requirements of a specific scenario, this may relate to all or part of the externally stored data, the identity of the users who have access to the data or the actions that the users take on the data cite t05. Encryption techniques are used to achieve confidentiality in such systems. Cloud computing is a technology that uses the internet and servers to maintain and manage data and applications. Cloud computing has improved computing capabilities without large investments.

In the existing situation in order to avoid confidentiality issues, the system may want to implement encryption and decryption techniques [35] which lead to limited system operations and the user must know encryption decryption Keys. Some systems may implement both encryption and obfuscation depending on the type of data to be stored [30].

A system based on proxy encryption, which supports various functions during the distributed storage system, is proposed in cite p38, p55, which consists of four stages: 1) system configuration, 2) data storage, 3) data transfer and 4) data recovery. An RSA-based algorithm is used to generate keys. The solution is when a sender “A” wants to send a message to recipient “B”, “A” signs the message with his secret key and then encrypts it with the public key of “B” and downloads the encoded text. After retrieving the message, “B” decrypts it with its public key and then checks the public key sign “A”. The whole process involves two communication stages; a download from the sender “A” and download by the recipient “B”. This is why the proxy recoding scheme is used to reduce the overhead of the data transfer function in the secure storage system. Here are some crucial points regarding data privacy in a cloud storage environment.

1) In a cloud computing paradigm confidentiality of governmental and business information as well as privacy of personal information has the highest insinuations.
2) The level of confidentiality and privacy of a user depends upon the privacy policies and terms of service provided by a cloud provider.
3) Disclosure of information to a cloud provider by a user may change information of some specific types as well as certain user categories, rights and obligations of privacy and confidentiality.
4) Personal and business information may be adversely affected in terms of legal status protection.
5) Protecting confidentiality and privacy and the privacy rights of those processing and storing this information in a cloud environment may be highly affected by the location of information.
6) A cloud may store information at different venues with different legal implications leading to different legal consequences at the same time.
7) Different laws against criminal activities and other matters can oblige/force a provider to disclose or examine user records for the sack of evidence.
8) In addition to the legal protection for protecting a user’s privacy and confidentiality, various legal qualms resist against gauging an information in a cloud for its status.

2) Integrity Issues
Data integrity is one of the most crucial elements of any system. Integrity requires that the authenticity of the parties (i.e. users and vendors) communicating in the cloud guarantee the data stored with third-party vendors and the responses resulting from the calculation of requests cite t05. In a standalone system, data integrity may be achieved with a single database using constraints and transactions. To insure integrity of the data transactions must adhere the mostly used property in databases known as the ACID (atomicity, consistency, isolation and durability) property. But distributed systems are entirely different in complexity where multiple databases and multiple applications execution is a normal trait. In a distributed environment, data may be maintained at different sites. Therefore, any transaction involving data shared by multiple sites must be handled carefully in a way to avoid transaction failure and allow various distributed applications through a resource manager to be a part of the global transaction.

With the entrance to the world of Service Oriented Architecture (SOA) and Cloud computing, issues of data integrity grow exponentially because a mixture of local and SaaS (Software as a Service) applications are displayed as a service. SaaS model supports multi tenancy in applications.
which usually hosted by third party and their functionality is exposed through XML based APIs (Application Programming Interface). Similarly in other environments like SOA various applications uses web services for example SOAP and REST to expose their functionality. However, managing transactions using web services is a serious challenge. Since guaranteed delivery or transactions are not supported by HTTP protocol level giving the only way out of implementing these SOA at the API level.

3) Data Access Issues

Issues in access to data in a cloud storage are mostly due to security policies. For example, a small level business organization may use services of a cloud provider for executing its business processes [36, 37]. Such organizations allow their employees to access a specific organizational data according to its own organizational security policies. These policies may prevent some employees from accessing a specific set of data and allow them to access certain data. To stop intruders from gaining unauthorized access to cloud resources, a cloud must adhere these security policies [38]. The SaaS model must have the ability to allow organizations to integrate their security policies as well as keep organizational data within its boundary in case when multiple organizations use the same cloud environment. The requirement of availability is; there must be a mechanism for verification of Service Level Agreements (SLA) between a user and providers which verifies that the user’s requirements are fulfilled [39].

Many counter measures proposed in the literature can be found to mitigate the problems related to data access in cloud storage. In literature three categories of secure access control can be found (i.e. Role Based Access Control (RBAC), User Based Access Control (UBAC) and Attribute Based Access Control (ABAC)) [40, 41]. Due to the attachment of access control list (ACL) to user data, UBAC is usually not considered as a suitable candidate for cloud storage. Additionally, the involvement of Big Data the computational and communication overhead required for handling ACL is high [42]. Then there is role based classification of users to control access to data. A user matching a specific role is granted access to data. Such approaches are considered suitable for business organizations at enterprise level for example hospitals [43]. The third and often used category in cloud storage is Attribute Based Access Control (ABAC) where a data owner assigns attributes and policies to users and data respectively [44]. In this case access to the data is granted to users having attributes that satisfy a specific access policy. For a confidential fine-grained access to data in cloud, this category is further divided into two approaches i.e. KP-ABE [45, 31] and CP-ABE [46] [13]. In case of KP-ABE the key of a user is linked with an access policy whereas attributes are linked with ciphertext. In contrast to KP-ABE, in CP-ABE the key of a user is linked with an attribute whereas the ciphertext is linked with an access policy.

However, the complexity of attribute based access control techniques grows linearly, as the number of attributes used in decryption raises, incorporating tremendous overhead in computation specially for devices with limited resources like mobile devices [37].

4) Authentication and Authorization Issues

Authentication, in any system that needs a foolproof security, plays a crucial role like an entrance door that allows only trusted individuals, to the premises of a cloud. Access to important information depends on authentication, therefore, due to its sensitive nature, authentication process must be robust to ensure availability to authentic users. In combination with cryptography, not only data confidentiality, but also its integrity can be ensured by granting access only to authenticated individuals. Most of the security concerns can be mitigated through a sophisticated authentication mechanism [47, 48, 35].

A Lightweight Directory Access Protocol (LDAP) server is used by various companies to store information about their employees [49]. Managing users in small and medium size businesses is mostly achieved through Active Directory in the portion of business where the adoption of SaaS model is high (Microsoft White Paper, 2000). This model allows software to be hosted outside the organizational firewall. Many organizations separate user credential database from their IT infrastructure therefore, a customer must keep track of all the employees joining or leaving the organization and must enable or remove their accounts accordingly from the system. This may result in extra management overhead on the customer organization if it uses multiple SaaS products. In such cases different powers can be delegated to the customer by the provider, authentication for example enabling customer organizations internal LDAP/AD server to control their user management.

5) Data Breaches

A cloud environment is usually shared among many customers to store their data. Therefore, a compromise of the cloud environment means a potential threat to the data of all users making cloud an attractive target for attackers [36]. R. Cooper in his report rated external criminals as the highest threat contributing 73% but with least impact compro-
mising 30,000 records producing 67,500 Pseudo Risk Score (PRS). Similarly, insider threats received the minimum rating of (18%) but with greatest impact compromising 375,000 records with a PRS of 67,500. The middle rating has been received by partners with 73.39% compromising 187,500 with a PRS of 73,125. The security provided by SaaS is argued to be better in comparison to conventional means, however insiders may not have direct database access but it still raises a risk with huge impact on data security. Employees of SaaS providers can cause exposure of customers private information since they have access to a lot of information. In order to avoid such complications, standards like PCI-DSS (Payment Card Industry-Data Security Standards) must be followed by SaaS providers.

B. DATA MANAGEMENT ISSUES

The management issues related to data has been explored in this sub sections. The data management issues has been categorized and briefly explained as follows.

1) Data Dynamics Issues

Data management in cloud is considered to be untrustworthy due to the fact that it shifts databases as well as application software to large centralized data centers. This new paradigm introduces various security issues yet to be understood. Data dynamics support through operations in cloud for example insertion, block modification, and deletion is a huge step in the direction of practicality as cloud services are not restricted only to backup and archiving. The following different methods are used for the assurance of data dynamics in cloud storage [51], [52].

- On a large scale the data centers are being transformed into computing pools by “Software as a Service” (SaaS) computing architecture. In addition the fast growth in network resources like bandwidth and reliability enables customers to subscribe services with high quality from the remote data and software applications in data centers.

- A cloud service provider for his own benefits may conceal errors in data or software used by the clients. For example a provider may deliberately delete data of an ordinary client which is accessed less often without the client’s knowledge in order to increase his savings in money and storage space [51].

- For data dynamics various schemes have been designed with the efforts to combine efficiency, unlimited use of queries and information retrievability in these schemes.

One possible solution in this case could be to motivate public auditing system of data storage security in Cloud computing [53]. In addition, fully dynamic protocols for data operations specially for block insertion, must be designed which is a lacking feature in most of the existing approaches. To support public auditing which is efficient and scalable, the existing schemes must be extended. Such extension should achieve batch auditing enabling a third party auditor (TPA) to perform auditing tasks delegated from multiple users simultaneously.

2) Data Segregation Issues

Cloud computing architecture became popular because of its multi-tenancy nature [49], [53]. Multi-tenancy in cloud through SaaS applications allow storage of data from multiple users simultaneously. This may create an opportunity for a user’s data to intrude into another user’s data since data of different users reside at single location. This intrusion may exploit application’s loopholes or by injecting SaaS system with malicious client code. If an application injected with a masked code executes it without verification shows that there are high possibilities of intrusion into others data. Therefore, a SaaS model must ensure that the data of each user is bounded both at physical and application levels. Data from different users must be ghettosize intelligently by the SaaS service [54].

Security checks may be bypassed using vulnerabilities in application by attackers through handcraft parameters. This may lead to the exposure of other tenants sensitive data. Therefore different assessments test must be performed to ensure that data from different users in multi-tenant environment is fully segregated from each other. These tests include; i) Data validation, ii) SQL injection flaws and iii) Storage insecurity. Any possible flaws detected by these tests could be used to illegally access sensitive data of the enterprise or other tenants.

3) Virtualization Issues and Vulnerability

One of the major component of cloud environment which ensures that various instances running over a single machine be ghettosize from each other is known as virtualization. It is the source of major security challenges in a cloud environment which are not fully investigated today [55], [56]. Second issue is the administrative control of the operating systems, operating as guest and host systems and their imperfect provisioning of isolation [57] and scalability issues [58]. Many of the current Virtual Machine Monitors (VMM’s) suffer from bugs allowing escape from VM therefore, “root security” is mandatory in such cases to prevent host operating system from being interfere with by any virtualized guest systems. Some virtualization software has been reported to have vulnerabilities which could allow a local user or an attacker to skip certain security checks and gain illegitimate access [57], [59]. One such example is that of Microsoft Virtual Server and Virtual PC vulnerability where a user of guest operating system could be allowed to execute code on other guest operating system or even the host operating system itself. This could allow a raise in privileges which can lead to unauthorized access of sensitive information. Similarly a validation error in “tools/pygrub/src/GrubConf.py” of Xen which could allow a user with “root” access in a guest operating system through specific crafted contents in grub conf to use domain 0 for running various commands at booting time of guest operating system. Fully functional interposition, inspection
## Table 2. Security solution for cloud storage architecture (Part-I)

| Security Properties | Approaches       | Description                                                                 |
|---------------------|------------------|------------------------------------------------------------------------------|
| Confidentiality     | Cryptography     | Cryptography secure and protect data during communication. It is useful to block unauthorized users from accessing private data. |
|                     | Digital signatures| Digital signature is a symbolic description that can verify the authenticity of messages or digital documents. A true digital signature provides access to the data. |
|                     | Proxy Re-encryption| Proxy encryption is commonly used when a party wants to reveal to a third party the content of messages sent to it that are encrypted with their public key. |
|                     | Obfuscation      | Obfuscation is the intentional creation of source code or machine code that is difficult for a human to understand. Like natural language eclipse, it can use unnecessarily redirected expressions to make statements. |
|                     | Blockchain       | Blockchain is a smart design that offers digital information for sharing, but not for copying. Blockchain technology has created the backbone of a new type of internet. |
| Atomicity           | MC               | Data MC is a world leader in the delivery of highly complex data migrations, specializing in end-to-end delivery of industry-specific, custom and enterprise transformation ERP, CRM projects. |
|                     | Consistency      | The consistency of the database system refers to the fact that the database transaction can only be modified in an authorized manner. |
|                     | Isolation        | Isolation in database systems determines, how the integrity of activities are visible to other users and systems. |
|                     | Durability       | In database systems, sustainability is the ACID property that ensures that closed transactions persist. |
| Data Access         | Role Based Access Control | RBAC is an entrance approach to regulate access to the system to authorized users. It is used by most companies with more than 500 employees and can implement mandatory access control (MAC) or discretionary access control (DAC). |
|                     | User Based Access Control | Role-based access (or role-based permissions) adds another categorization layer in addition to what is provided by user-based access. |
|                     | Attribute Based Access Control | Attribute-based access control (ABAC) is also called as policy-based access control, defines an access control that give access rights to users through the use of policies that combine attributes. |
| Data Breaches       | Directory Services | A Lightweight Directory Access Protocol (LDAP) server is used to provide authentication and authorization services |

and complete isolation are not achieved in VMMM yet and need further investigation.

### 4) Backup Issues
The sensitive data belonging to various business enterprises must be backed up by the SaaS providers to be used for fast recovery in disasters cases. Also, to protect against security threats like accidental leakage of data various encryption schemes be used to protect the back up data. These encryption schemes must be strong enough to resist modern attacks.

Amazon as cloud vendor does not encrypt the data by default at rest in S3. This control is given to the user to secure their back up data separately in order to protect against unauthorized access or tempering. Various tests can be performed to validate that a back up data is secure provided by SaaS model. These tests include; i) Storage insecurity and ii) Configuration insecurity. Any flaws identified by these tests may be potential threats which can lead unauthorized users to access information which is sensitive and stored in cloud backups belonging to different enterprises.

### 5) Availability
The SaaS applications guarantee around the clock services to a client. This involves architectural level changes in SaaS infrastructure and applications to attain availability and scalability. Multitier cloud architecture needs to be adopted, cloud architecture must also support load balancing of application instances, running on different servers. Cloud storage must be resilient to software and hardware failures further, it must be protected from both distributed denial of service attacks (DDOS) as well as denial of service DOS attacks[60], [61], [62], [63].

For any unforeseen disaster, appropriate disaster recovery and operational sustainability action plan should be considered. This is important for certifying organizational data security and organizational nominal downtime. For example, at Amazon, the AWS API endpoints are hosted by the same world-class Internet infrastructure that Amazon supports and use connection throttling. To further reduce the potential impact of a DDOS attack, Amazon internally maintains the bandwidth that surfs on its vendor’s internet bandwidth to validate the SaaS vendor’s availability and evaluation tests.

Many applications automatically provide security locks for user accounts after successive incorrect credentials. Also, improper implementation and configuration of these functions can be vulnerable to malicious users as a result of DDOS attacks.
### Table 3. Security solution for cloud storage architecture (Part-II)

| Security Issue     | Solution          | Description                                                                                                                                 |
|--------------------|-------------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| Data Dynamics      | Public Auditing   | Efficient and scalable public auditing system should be introduced to extend the existing schemes                                             |
| Data validity      | Data Segregation  | Security layers give you the flexibility to consolidate vast amounts of data while controlling who can see what, through a sophisticated system of work groups, organizational rollups, and access levels, combined with field and function level security. |
| Virtualization Vulnerabilities | Root Security      | Root protection enables users (e.g. smartphones, tablets) with the android mobile operating system to get privileged control (known as root access) over different Android subsystems. |
| Backup Issues      | Encryption schemes| Different encryption schemes coded the data before storing. This secure the backup data from unauthorized users.                              |
| Availability       | Multitier architecture | Multitier architecture (often referred to as multi-level architecture) or multi-layer architecture is a client-server structure in which the functions of presentation, application processing and data management are separated. |
| Data Locality      | Regional backup servers | Due to the different, region cyber rules, data should be kept in the same region servers to avoid data locality, cultural and cyber rules issues. |

6) Data Locality

In SaaS cloud model, a client uses the application provided by the SaaS and their own business data, but the client is unaware of storage location of the data in the cloud [38], [64]. This may lead to several issues and many cases. For example, due to data privacy laws in different counties, data locality is of utmost importance in enterprise business architecture. For instance in many Southern American States and several countries in European Union, certain types of data may not be allowed to leave the country premises because of the sensitivity of the information. Similarly local Governmentâ€™s laws and jurisdiction issues may arise in case of any type of investigation [65]. A secure SaaS model may be capable to provide reliability to its clients at the consumer data locality.

III. CLOUD STORAGE FUTURE AND OPPORTUNITIES

The future of the cloud is not less than a dream. AI-enabled objects (e.g, self-driving vehicle), the web of IoT devices, and 5G connectivity (i.e, mobile communication technology) is changing the way of living [3]. The IT industry is rapidly changing everything. Its simple and easy user interface; no cost and capacity constraints; and other numbers of features are attracting the individual and market [66].

The recent advances in smart technology generate a massive data traffic. The 51 billion devices forecast is a big number; even seven times greater than the whole world population [1]. These devices will increase the annual size of the global data-sphere up to 175 ZB [2]. [3]. Another report states, as shown in Figure [1] that more than 331 billion dollars will be invested in cloud up to 2023 [2]. It needs special techniques and infrastructure to process the incoming data [4]. Furthermore, integration of Artificial Intelligence (AI) in smart devices increases the data production value dramatically. With this rapid development in smart technology, cloud storage is getting more and more attention. Along with the AI, block-chain is adding safety and security to the cloud storage. This technology in storage is getting mature and will increase the customer trust on cloud.

This section presents a quick review of the cloud storage future and its opportunities [67], [38], [68], [69], [70], [71], [72], [73], [74], [75].

#### A. REMOTE ACCESSIBILITY

Remote accessibility (i.e, access from everywhere and anytime) is the core of cloud storage. The fast network speed and AI is making it more smarter and faster. Leading cloud provider (i.e, Apple iCloud [76], Microsoft OneDrive [77], and Google Drive [28] etc) are providing fast and reliable remote services to their users. Remote access allows to store and retrieve items from a cloud storage without needing to create a physical connection. Accessibility of storage devices is getting interested after introducing high storage devices and high bandwidth network. Remote access increases the usage of cloud storage and business. In the presence of
internet services, cloud storage can provide seamless access to data files. The coming 5G internet service will make the accessibility very easy and smart as real-time access.

B. 5G CONNECTIVITY
With this high-speed technology, humans will be able to virtually operate any machine at a distance of thousands of KMs. This will reduce the latency of up to 0 ms. Such a big speed will minimize the need for the local hard drive. This technology will able to store and process data on the cloud without facing any jitters or delay. It is making real-time use possible. 5G is a new era of cloud storage.

C. INTERNET OF THINGS (IOT)
With the introduction of the Internet of Things (IoT), the number of devices connected to the internet has increased enormously. By 2025, 75 billion devices are expected to be connected to the Internet processing 75 ZB data annually. This is a great number and will need a high technology to process and store this data. These figures clearly shows that the cloud storage has great worth in coming years. Furthermore, the use smart devices are also dramatically increasing. These devices are small in size and have not enough space to store or process big data therefore, they depend on cloud.

D. ARTIFICIAL INTELLIGENCE (AI)
From facial expressions to self-driving vehicles, AI is progressing very rapidly. AI is making smart decisions in complex situations. The today AI is called the weak AI which performs limited tasks such as recognizing facial expression and driving a car, however, the future will have general AI which will perform a task just like human beings. The AI is making the cloud storage further smarter and attractive.

E. USABILITY
The provider business directly depends on resources utilization. Today technologies massively increase the cloud usage because it provides a very easy and reliable user interface. Usually, cloud storage has a local desktop folder for PCs and mobile devices which allows users to move files back and forth between the cloud and the local system using drag and drop facilities. The integration of smarts technologies (i.e, IoT, AI, fog and 5G), making the cloud storage usability very easy. The 5G will provide a high bandwidth like real-time access. Its cost is very low compared to buy the devices; which is very appealing.

F. DISASTER RECOVERY
In today modern world, data is the most valuable asset. Losing it, cause irreversible damage to the business (including loss of productivity, income, reputation and even customers). Business enterprises use cloud storage as a backup for their important files. In cloud storage, data is stored in three different locations and in case of any disaster, data may easily be recovered. Furthermore, cloud storage provides remote access to files therefore, these files can be used for recovery of their system in case any emergency or disaster. 5G technology made the recovery process very easy and fast. Comparatively to the traditional disaster recovery, cloud storage recovery is very easy, cheaper and fast. High investment, staff and maintenance are required for local disaster recovery site.

G. COST SAVINGS
When we talk about cloud, it means that we are getting the resources of a supercomputer at our home without buying it. We actually, hire these resources on very cheaper rates which save the capital investment of the consumer. Cloud storage is used by various types of business enterprises to reduce their annual database operation expenses. Especially, the medium corporations, which are not able to invest too much on storage infrastructure, hire the cloud storage. This saves their major investment. Storing one gigabyte of data using cloud storage services cost about three whereas a user can achieve further saving in terms of power consumption as remote cloud storage does not need internal power.

H. INVISIBILITY
The word storage create the imagination of a big physical device to store big data. Big data and storage mean a big physical device, which will need operation and maintenance. However, cloud storage does not need physical space and user access it remotely. Cloud storage services, use virtualization techniques to provide resources to the customers. Customers do not know the complexities and working of the back end. Cloud storage is invisible and provides storage transparency, with no physical presence on the user side. It does not take up valuable space in the office or at home. It does not need to spare a huge space for rocks and storage. Customers only hire the services and use them on the go.

I. PRIVACY AND SECURITY
Security of cloud storage for sensitive and confidential information is usually higher than that for the locally stored data, especially for enterprises. It uses advance security (i.e, advanced firewalls, event logging, internal firewalls, intrusion detection, data replication, encryption, and physical security) to protect the data from outside attacks. Different type of security layers is used to protect the data houses. Concerning individual storage, enterprises invest more in security. Storage services in the cloud used encrypted data both in transmission as well as at rest ensuring no unauthorized access to data files. AI, 5G, IoT and block chain are improving the privacy and security.
J. AUTOMATION AND SYNCHRONIZATION

Cloud automation is a term for the processes and tools that are used to reduce the manual effort involved in provisioning and managing cloud workloads. The cloud storage is self-managed and does not need any human efforts [87]. The main issue most businesses and customers have, the proper follow up of data backup. Cloud storage provides an automated data backup service to ease this tedious process. A user simply needs to tell the system what and when to back up, and the cloud service takes care of it by itself.

Another attraction with the cloud storage is automatic synchronization. Synchronization process ensures that user data files are automatically updated across all of the user devices. In this way the latest versions of the user’s data files are saved on his/her local device and available on all of other user devices like user Smartphone etc. 5G made the syncing more easy and now the devices works on real time [87].

K. SHARING AND COLLABORATION

Cloud storage makes the sharing easy. Either it is a photo or a file or even a folder containing hundreds of information files, storage service in cloud make it convenient for a user to share it with a few clicks. Furthermore, it makes the files availability everywhere and every time [88]. Online cloud storage services are also ideal for collaboration purposes. It allows multiple users to collaborate and edit on a single document or data file. User do not have to concern about tracking the up-to-date version or who has made what changes [88].

L. MASSIVE DEVICES AND DATA

As mentioned earlier, up to 2025, approximately 75 billion devices will connect to the internet and this will process more than 175 ZB of data per year. This is a very large figure and requires a lot of cloud storage. These predictions will drastically change the need for cloud storage. This shows that cloud storage has a very bright future ahead.

IV. CONCLUSION AND FUTURE DIRECTIONS

The recent advances in IT industry (e.g. Cloud computing, Internet of Things (IoT), Fog computing, Artificial Intelligence (AI) and Block-chain) is rapidly revolutionizing the cloud storage. Especially, the 5G facilitation (i.e, minimum access delay and ultra high speed) boost the use of cloud storage dramatically. This article presented different challenges, their counter measures, opportunities and future of cloud storage, it seems that cloud storage is designed to be highly scalable and conveniently manageable storage system rather than an efficient file system.

Further, it is revealed that despite the ease of use and economic benefits, cloud storage technology still suffers from numerous problems. The cloud storage architecture is mostly clouded by security (e.g, confidentiality, integrity, access, authentication, authorization and data breaches) and data management issues (e.g. dynamics, data segregation, backup, and virtualization). To counter these threats, various measures are proposed in the literature. For example, for the security of data, digital certificates are used along with a trusted timestamp approach. Similarly, confidentiality is ensured through cryptography solutions while for data access issues, attribute-based encryption is mostly used. Access control is achieved through authentication and authorization.

To maintain the integrity of the data, a global transaction manager is used to ensure fail-safe management of transaction across multiple databases.

Finally, it can be concluded that cloud computing (along with the integrated technologies) is a fast-growing technology which rapidly changing traditional computing. However, still, a lot of research efforts are needed to attract customers, especially business and enterprise customers to store their sensitive data, using cloud storage.

REFERENCES

[1] G. Davis, “2020: Life with 50 billion connected devices,” in IEEE International Conference on Consumer Electronics (ICCE). Las Vegas, NV, USA: IEEE, Jan 2018, pp. 1–1.

[2] C. Louis, “Forbes: Cloud computing forecast,” 2017. [Online]. Available: https://www.forbes.com/sites/louiscolumbus/2017/04/29/roundup-of-cloud-computing-forecasts/2017#5c4232c31e8f

[3] S. Kinza, K. Bilal, S. Farah, Q. Sameer, and M. Muhammad, “Internet of things (iot) for next-generation smart systems: A review of current challenges, future trends and prospects for emerging 5g-iot scenarios,” Antenna and Propagation for 5G and Beyond, vol. 8, no. 1, pp. 2169–3536, 2020.

[4] X. Wang, B. Wang, and J. Huang, “Cloud computing and its key techniques computer science and automation engineering (csea),” in International Conference on Cloud Computing. Shanghai, China: IEEE, 2011, pp. 404–410.

[5] W. Songyun, Y. Jiabin, L. Xin, Q. Zhuzhong, A. Fabio, and Y. Ilsun, “Active data replica recovery for quality-assurance big data analysis in ic-iot,” IEEE Access, vol. 7, pp. 106997 – 107005, 2019.

[6] S.Karen, “Iot big data security and privacy versus innovation,” IEEE Internet of Things Journal, vol. 6, no. 2, pp. 1628 – 1635, 2019.

[7] S. Syed Attique, Z. S. Dursun, H. Sufian, and D. Dirk, “The rising role of big data analytics and iot in disaster management: Recent advances, taxonony and prospects,” IEEE Access, vol. 7, pp. 54 595 – 54 614, 2019.

[8] Q. Bashir, A.-F. Ala, A. Gupta, B. Driss, A. Safas, and Q. Junaid, “Leveraging machine learning and big data for smart buildings: A comprehensive survey,” IEEE Access, vol. 7, pp. 90 316 – 90 356, 2019.

[9] S. Rui, L. Shanyun, W. Shuo, X. Ke, and F. Pingyi, “Importance of small probability events in big data: Information measures, applications, and challenges,” IEEE Access, vol. 7, pp. 100 363 – 100 382, 2019.

[10] B. Afzal, G. Anwar, S. Shahahbodhin, A. Giuseppe, and P. Antonio, “Performance based service level agreement (psla) in cloud computing to optimize penalties and revenue,” IET Communications, 2020.

[11] B. Afzal, S. Shahahbodhin, G. Anwar, and C. Anthony, “Optimizing iaaS provider revenue through customer satisfaction and efficient resource provisioning in cloud computing,” IET Communications, Vol. 13, no. 9, pp. 2913–2922, 2019.

[12] A. Ankita, V. Gregory, H. Seghbroecka, F. Morah, T. De, and V. Bruno, “Spech: A scalable framework for data placement of data-intensive services in geo-distributed clouds,” Journal of Network and Computer Applications, vol. 142, pp. 1–14, 2019.

[13] Z. Lei, F. Ammin, Y. Shui, S. Mang, and K. Boyu, “Data integrity verification of the outsourced big data in the cloud environment: A survey,” Journal of Network and Computer Applications, vol. 112, pp. 1–15, 2019.

[14] A. Sidra, U. Saif, K. Abid, A. Mansoor, A. Adnan, and K. K. Muhammad, “Information collection centric techniques for cloud resource management: Taxonomy, analysis and challenges,” Journal of Network and Computer Applications, vol. 100, pp. 80–94, 2019.

[15] W.-T. Tsai, Z. Jin, and X. Bai, “Internetware computing: issues and perspective,” in Proceedings of the First Asia-Pacific Symposium on Internetware, ACM, 2009, p. 1.

[16] H. Raj, R. Nathuji, A. Singh, and P. England, “Resource management for enhanced cloud services,” in Proceedings of the ACM workshop on Cloud computing security. ACM, 2009, pp. 77–84.
[17] V. Chang and G. Wills, “A model to compare cloud and non-cloud storage of big data,” Future Generation Computer Systems, vol. 57, pp. 56–76, 2016.

[18] R. Kumar and A. K. Bose, “Internet of things and opc ua,” ICNS 2015, p. 52, 2015.

[19] S. Kamara and K. Lauter, Cryptographic Cloud Storage, Canary Islands, Spain: Springer Berlin Heidelberg, January 25-28 2010, pp. 136–149. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-14992-4_58, 2013.

[20] P. Paola, C. Roberto, B. Alberto, and P. Lorenzo, “Amazon, google and microsoft solutions for iot: Architectures and a performance comparison,” IEEE Access, vol. 8, pp. 5455 – 5470, 2020.

[21] T. Ye, X. Peng, and J. Hai, “Secure data sharing and search for cloud-edge-collaborative storage,” IEEE Access, vol. 7, pp. 15 963 – 15 972, 2019.

[22] Q. Shuaiqing, Z. Qisheng, Z. Qimao, G. Feng, and L. Wenhao, “Hybrid seismic-electrical-data acquisition station based on cloud technology and green iot,” IEEE Access, vol. 8, pp. 31 026 – 31 033, 2020.

[23] K. P. Chan and J. B. Seung, “Blockchain of finite-lifetime blocks with applications to edge-based iot,” IEEE Internet of Things Journal, vol. 7, no. 9, p. 10.1109/JOIOT.2019.2959599, 2020.

[24] G. Murugaboopeathi, C. Chandravathy, and P. Vinoth Kumar, “Study on cloud computing and security approaches,” International Journal of Soft Computing and Engineering (IJSCSE), vol. 3, pp. 212–215, 2013.

[25] B. A. P. Rajan et al., “Evolution of cloud storage as cloud computing infrastructure service,” arXiv preprint arXiv:1308.1303, no. 1, 2013.

[26] V. Chang, Y.-H. Kuo, and M. Ramachandran, “Cloud computing adoption framework: A multi-case study on business clouds,” Future Generation Computer Systems, vol. 57, pp. 24–41, 2016.

[27] N. Kaaniche and M. Laurent, “Data security and privacy preservation in cloud storage environments based on cryptographic mechanisms,” Computer Communications, vol. 111, pp. 120–141, 2017.

[28] X. Sun, S. Qu, X. Zhu, M. Zhang, Z. Ren, and C. Yang, “Cloud storage architecture achieving privacy protection and sharing,” Appl. Math, vol. 9, no. 1699–1694, 2015.

[29] L. Geng, “The research of digital library mass information storage system architecture,” in International Symposium on Computers & Informatics. Atlantis Press, 2015.

[30] A. L. Arockiam and S. Monikandan, “Efficient cloud storage confidentiality to ensure data security,” in International Conference on Computer Communication and Informatics. IEEE, Jan 2014, pp. 1–5.

[31] J. Han, W. Susilo, Y. Mu, and J. Yan, “Privacy-preserving decentralized key-policy attribute-based encryption,” IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 11, pp. 2150–2162, 2012.

[32] N. H. Hussein, A. Khalid, and K. Khatfan, “A survey of cryptography cloud storage techniques,” International Journal of Computer Science and Mobile Computing, no. 5, 2016.

[33] C. Wang, K. Ren, W. Lou, and J. Li, “Toward publicly auditable secure cloud data storage services,” IEEE network, vol. 24, no. 4, pp. 19–24, 2010.

[34] G. Hittu and D. Mayank, “Securing iot devices and securelyconnecting the dots using rest api and middleware,” in 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU). IEEE, 2019, pp. 1 – 6.

[35] Z. Kartit, A. Azougaghe, H. K. Idrissi, M. El Marraki, M. Hedabou, G. Hittu and D. Mayank, “Securing iot devices and securelyconnecting the dots using rest api and middleware,” in 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU). IEEE, 2019, pp. 1 – 6.

[36] Z. Kartit, A. Azougaghe, H. K. Idrissi, M. El Marraki, M. Hedabou, G. Hittu and D. Mayank, “Securing iot devices and securelyconnecting the dots using rest api and middleware,” in 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU). IEEE, 2019, pp. 1 – 6.

[37] Z. Kartit, A. Azougaghe, H. K. Idrissi, M. El Marraki, M. Hedabou, G. Hittu and D. Mayank, “Securing iot devices and securelyconnecting the dots using rest api and middleware,” in 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU). IEEE, 2019, pp. 1 – 6.

[38] Z. Kartit, A. Azougaghe, H. K. Idrissi, M. El Marraki, M. Hedabou, G. Hittu and D. Mayank, “Securing iot devices and securelyconnecting the dots using rest api and middleware,” in 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU). IEEE, 2019, pp. 1 – 6.

[39] Z. Kartit, A. Azougaghe, H. K. Idrissi, M. El Marraki, M. Hedabou, G. Hittu and D. Mayank, “Securing iot devices and securelyconnecting the dots using rest api and middleware,” in 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU). IEEE, 2019, pp. 1 – 6.
Transactions on Parallel and Distributed Systems, vol. 26, no. 11, pp. 3194–3207, Nov 2015.

[65] D. Espley, “Legal sector will embrace cloud, big data and mobility in 2016,” http://www.itproportal.com/2015/12/14/legal-sector-will-embrace-cloud-big-data-and-mobility-in-2016/, 2020.

[66] P. Matthias, “5g is coming around the corner [mobile radio],” IEEE Vehicular Technology Magazine, vol. 4, no. 1, pp. 4 – 10, 2019.

[67] R. L. Grossman, “The case for cloud computing,” IT professional, vol. 11, no. 2, pp. 23–27, 2009.

[68] L. Coles-Kemp, J. Reddington, and P. A. Williams, “Looking at clouds from both sides: The advantages and disadvantages of placing personal narratives in the cloud,” Information Security Technical Report, vol. 16, no. 3, pp. 115–122, 2011.

[69] S. Marston, Z. Li, S. Bandypadhyay, J. Zhang, and A. Ghalsasi, “Cloud computing—the business perspective,” Decision Support Systems, vol. 51, no. 1, pp. 176–189, 2011.

[70] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, “Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility,” Future Generation Computer Systems, vol. 25, no. 6, pp. 599–616, 2009.

[71] S. Mansfield-Devine, “Danger in the clouds,” Network Security, vol. 2008, no. 12, pp. 9–11, 2008.

[72] D. Zissis and D. Lekkas, “Addressing cloud computing security issues,” Future Generation Computer Systems, vol. 28, no. 3, pp. 583–592, 2012.

[73] N. Sultan, “Cloud computing for education: A new dawn?” International Journal of Information Management, vol. 30, no. 2, pp. 109–116, 2010.

[74] P. Hofmann and D. Woods, “Cloud computing: the limits of public clouds for business applications,” IEEE Internet Computing, vol. 14, no. 6, pp. 90–93, 2010.

[75] W. Kim, “Cloud computing: Today and tomorrow.” Journal of Object Technology, vol. 8, no. 1, pp. 65–72, 2009.

[76] “Icloud,” https://www.icloud.com/, 2020.

[77] “Microsoft onedrive,” https://products.office.com/en-us/onedrive/online-cloud-storage, 2020.

[78] “Google drive,” https://gsuite.google.com/, 2020.

[79] C. Lalit and B. Rabindranath, “A comprehensive survey on internet of things (iot) toward 5g wireless systems,” IEEE Internet of Things Journal, vol. 7, no. 1, pp. 16 – 32, 2020.

[80] S. Richa, W. Isaac, C. Glauceio, and A. Alagan, “Mobile cloud storage over 5g: A mechanism design approach,” IEEE Systems Journal, vol. 13, no. 4, pp. 4060 – 4071, 2019.

[81] C. Lalit, I. Majitar, and B. Rabindranath, “A comprehensive survey on internet of things (iot) toward 5g wireless systems,” IEEE Internet of Things Journal, vol. 7, no. 1, pp. 16 – 32, 2020.

[82] P. Samuela, C. Claudia, B. Marina, T. Marco, P. Valeria, G. Andrea, and F. Manuel, “Iot enabling technologies for extreme connectivity smart grid applications,” in CTTE–FITCE: Smart Cities & Information and Communication Technology (CTTE–FITCE). IEEE, 2020, pp. 16–32.

[83] X. Shuming, N. Qiang, W. Liangmin, and W. Qian, “Sem-acsit: Secure and efficient multi-authority access control for iot cloud storage,” IEEE Internet of Things Journal, pp. 1 – 1, 2020.

[84] G. Tudor, C.-C. Andrei, A.-C. Madalina, and Z. Alexandru, “Cloud storage: a comparison between centralized solutions versus decentralized cloud storage solutions using blockchain technology,” in 54th International Universities Power Engineering Conference (UPEC). IEEE, 2019, pp. 16 – 32.

[85] W. Shangping, W. Yuying, and Z. Yaling, “Blockchain-based fair payment protocol for deduplication cloud storage system,” IEEE Access, vol. 7, pp. 127 652 – 127 668, 2019.

[86] L. Mingyu, P. Li, and L. Shijun, “To transfer or not: An online cost optimization algorithm for using two-tier storage-as-a-service clouds,” IEEE Access, vol. 7, pp. 94 263 – 94 275, 2019.

[87] R. N. Ridwan, M. Rohit, and L. Palden, “Towards self-managing cloud storage with reinforcement learning,” in IEEE International Conference on Cloud Engineering (IC2E). IEEE, 2019, pp. 1 – 6.

[88] F. Majid, B. Hamideh, and M. Reza, “An efficient secret sharing-based storage system for cloud-based iots,” in 16th International ISC (Iranian Society of Cryptology) Conference on Information Security and Cryptology (ISCISC), vol. 7, 2019, pp. 1–6.